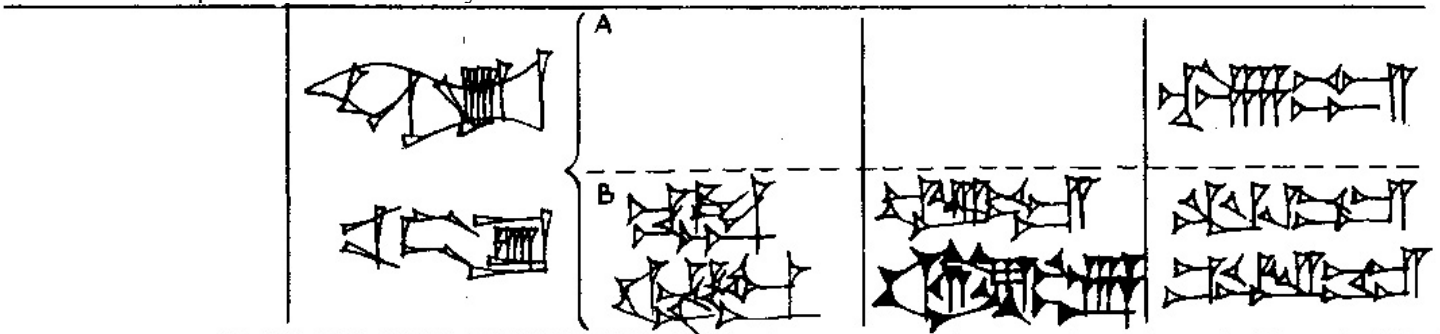


1. Importance of Cuneiform: foundations of western civilization; both the invention of writing, and the paradox (if you will) that it was used to support beaurocracy and complete control of lives of people
2. Optimism on encoding, some movement towards convergence of views:
 - (a) Only a PORTION of older signs need to be encoded now – those which are secure, not the others
 - (b) Some flexibility is now manifest that the time boundary is not as sharp as it was previously
3. Numbers:
 - (a) Almost all archaic numbers are already included, it would be a shame to prevent use for the older period by omitting just a few.
 - (b) Only decision in-principle is diacritic vs. single-sign listing for the "texture-marked" signs
 - (c) The statement at end of N2698 section 3.7 "will need to be revisited" should be "can be revisited ... if warranted"

4. Why consider the full range of cuneiform now, not later, and why encode all securely distinguishable signs?

(a) Because the criteria in N2664(R) and WG2 N2698 for what signs to encode yield **unstable** results depending on inclusion or exclusion of particular stages, even within the more limited time range favored by some. The sign UMBIN 'talon' illustrated below from Labat is historically **not** a sequence of GAD, TAG4, and DU. That analysis is possibly valid only for the NeoAssyrian times (upper right). What Labat calls "Classical Sumerian" shows the sign originated as something like "leg" or "talon" x TAG4 (5 clear examples are cited for Fara under LAK#289), also appearing similar to a combination of two components GAD and DU x TAG4 (4 clear Fara examples). Also in Old and Middle Babylonian the sign does not look like a sequence. Even for NeoAssyrian, the question remains open whether this is a single sign made up of three apparent components, or a sequence of three signs written closely together. Since the single-sign analysis is clearly proven, there is as yet no demonstrated need for a splitting into a sequence of signs. Han CJK characters show examples of similar changes of internal component structure while the identity of the single sign remains constant through time. Fragmenting single signs into sequences which disregard this normal type of development will break many instances of simple linear historical descent of single signs. There is simply no need for such counter-productive false analyses.



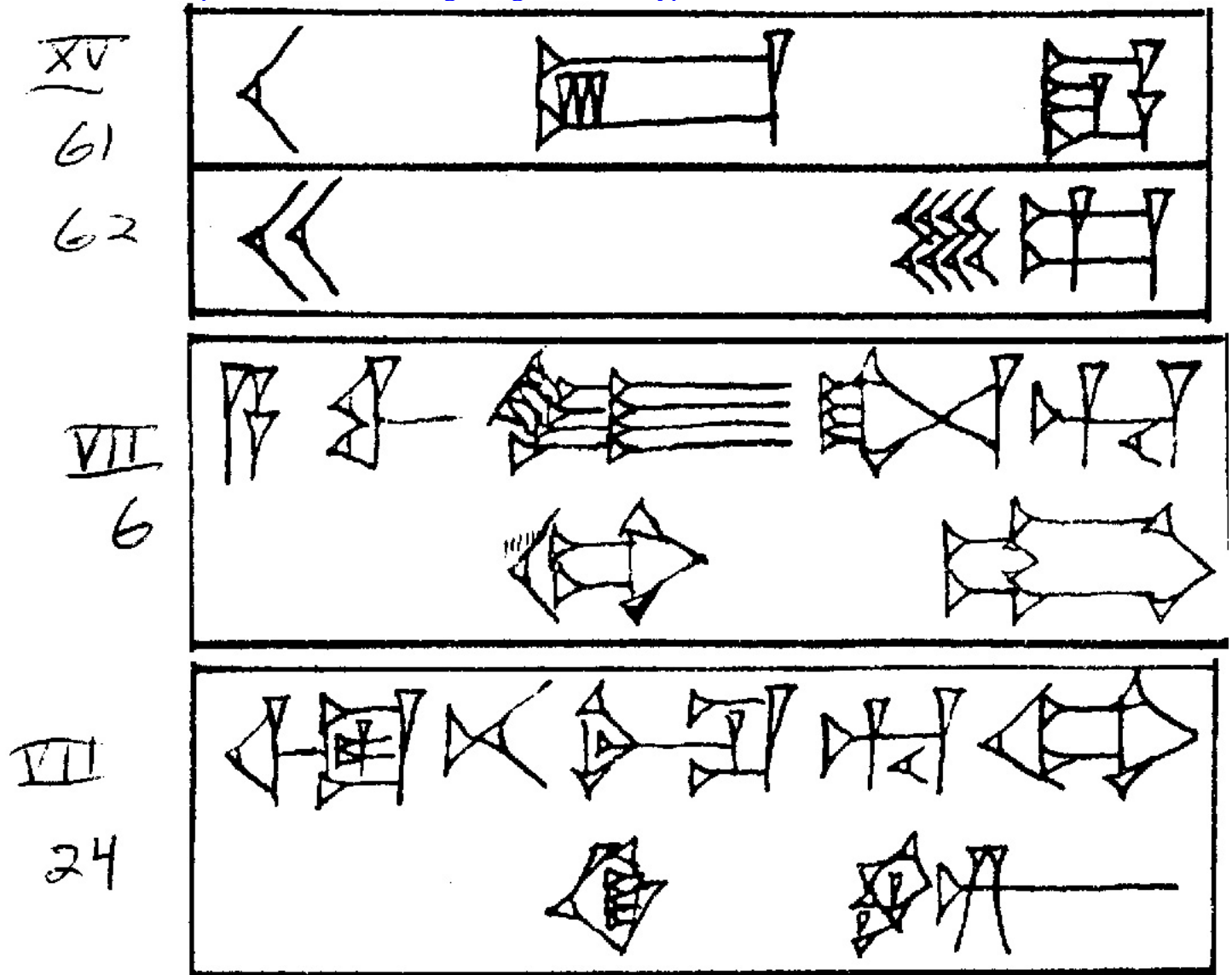
- (b) Because omission of needed signs will lead some users to wrong encodings (example just above)
- (c) Because the entirety is "manageable" now. Triage of ZATU and use of lexical lists and scholarly reviews and articles, and at least unification where possible with secure parts of other older lists (ATU, UET II, KWU, LAK, Recueil) will be complete during February. "Comprehensiveness" is not aimed for in the sense of encoding anything that anyone has thought might be a sign, no matter how insecure. There is no need for the division into stages (N2698 section 3.1), and no need for the statement "Archaic may require an entire separate block." Additional signs, yes, but no principled separation has been justified in any way.
- (d) [Added note after UTC: The claim that the position advocated here implies encoding ENSI as a single sign was careless, negligent and false, both as to the particular example, and as to the generalization intended, that it implies encoding as single signs of what everyone knows are combinations or sequences of separate signs. Negligent in not checking first. False because documents long been in the hands of the person making the statement specifically marked ENSI

as not a single sign. False because the general position advocated here, taking the long scholarly tradition as the default guide until individual exceptions are made, means that ENSI would by default **not** be regarded as a single sign (neither Labat nor Borger lists ENSI as a single sign).

5. Why use the combination of the long scholarly tradition (sign lists) as the default criterion for what are the signs of cuneiform --

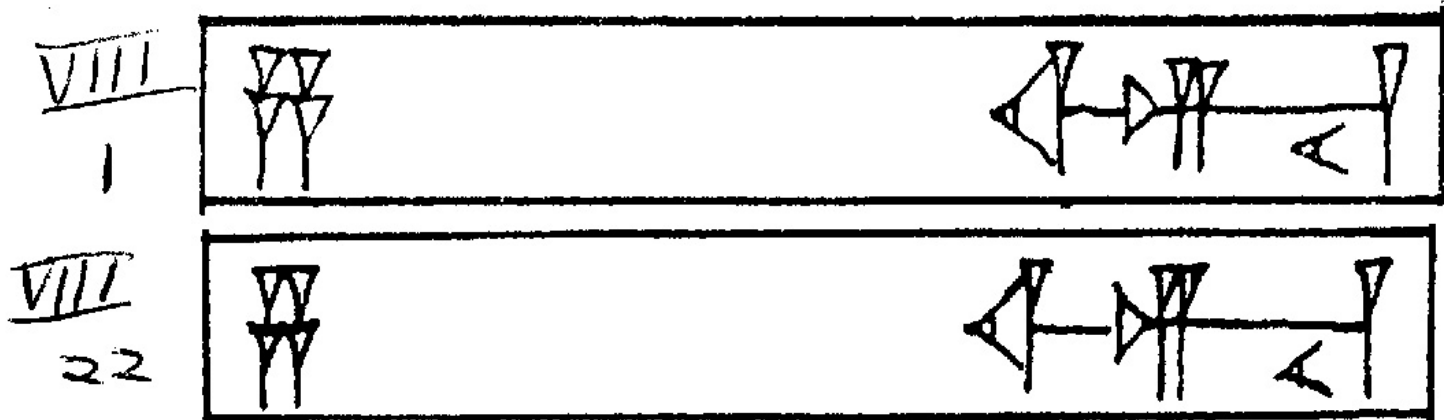
- (a) Because that tradition is much more stable, reflects many kinds of knowledge accumulated over generations. It does not rely on only a single criterion in disregard of many other criteria.
- (b) Because that tradition is confirmed by what happens where typography allows a distinction between spacing between signs and spacing within signs (between components of a single sign). Evidence from typography of Gudea statues and Codex Hammurabi confirm each other. Although the single sign U is written with much space around it as in Hammurabi Prologue xv.61, the component U as part of the sign UL is never so separated even when enormous space is available as in Hammurabi Prologue vii.6, and UL is not broken across end of line ("indent") even when that could yield more spacious typography, as it could have in Hammurabi Prologue vii.24. UL simply is not "U.GU4" even if it may have arisen etymologically that way.

Recognizing components via a name U x GU4 is more accurate -- please see <http://www.CuneiformSigns.org/ContainerTypes.htm>

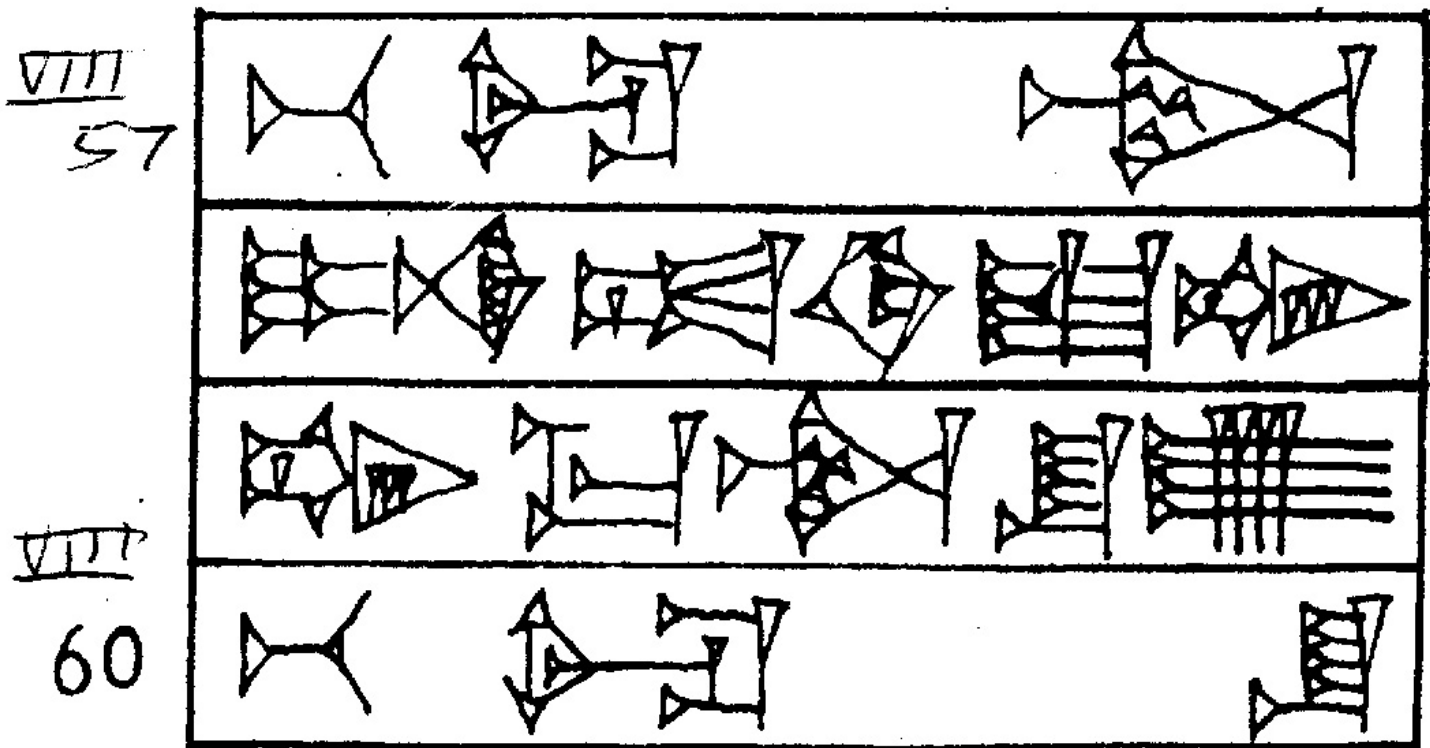


- (c) Because the scholarly tradition confirms the decisions made in correcting N2664 to N2664R (or WG2 N2698), and suggests a limited number of further additions, not additions without any stopping point as claimed in the revised proposal. Typography where there is enough extra space to see distinctions also confirms that scholarly tradition. The signs UL, AR, U3, and a number of others are thus **confirmed** as truly single signs, not merely exceptions noted to keep cuneiform professionals comfortable because they were in the basic school syllabary of Mesopotamia. In a reasonably careful survey of the Gudea statues and the Codex Hammurabi, I have found not a single occurrence of the sign AR as two signs IGI and RI separated by the kind of space which typically does separate independent signs. The tiny amount of space between components IGI and RI in the second example following does not signal a division of signs. In a line with so much extra space overall, only a very wide space could signal that. In addition, the two components, if they were that, were even more fused at the stage of Gudea (one wedge entirely lost).

VIII



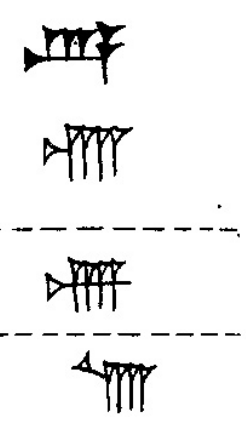
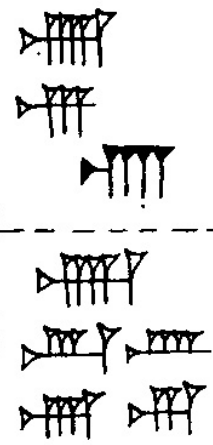
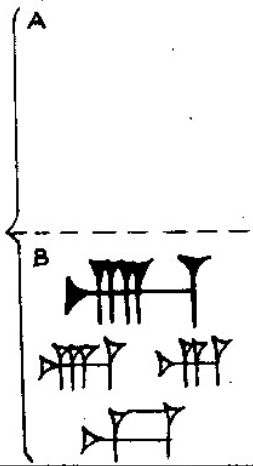
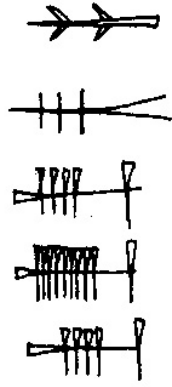
- (d) Justification works quite normally by default in lines of cuneiform text which have enough space to show its effects, **if** we encode as single signs what the scholarly tradition normally treats as single signs. The sign EL, now regarded by all as a single sign, has more space around it and is more stretched in the Codex Hammurabi Prologue viii.60 than it is in viii.57 (note the different width of the single remaining sign at the end of each of these lines), but it does not lose the connection between its component parts. This reflects its acknowledged status as a single sign.



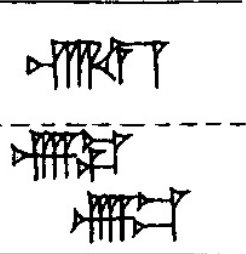
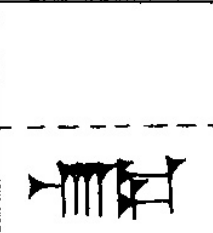
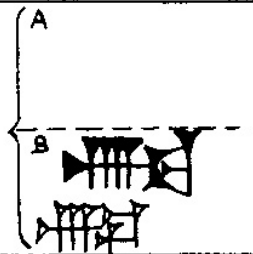
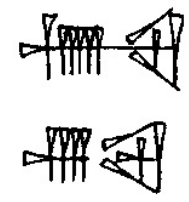
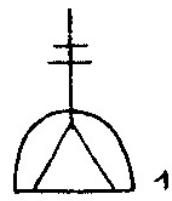
(EL is tymologically perhaps from SAL followed by SI or something which early came to look like SI.)

- (e) Because the scholarly tradition conforms to what linguistic analysis shows to be the productive units of cuneiform script. The most recent revision of N2664R is improved by including a series of the type TUR3 x SAL as single signs, because treating them as a sequence NUN followed by "LAGAR x SAL" would require encoding a list of signs "LAGAR x SAL" with varying infix components, signs which do not exist independently. But the revision does not include the sign TUR3 itself (which is demonstrably the container in this sign of the type "container x infix"). Rather it still would force users into an incorrect encoding of TUR3 as NUN followed by LAGAR instead of the correct NUN x LAGAR proven by what Labat calls "Classical Sumerian" (2nd column in the next illustration) and by the archaic forms (1st column). That is to say, TUR3 x SAL is structurally (NUN x LAGAR) x SAL, not NUN.(LAGAR x SAL). The correct linguistic analysis identifying the combinatorially productive components of such signs thus confirms the conclusion drawn in the traditional scholarship recognizing TUR3 as a single sign. For a survey of some of these sets of signs and which ones have unambiguous linguistic analyses of their components, please see the web page <http://www.CuneiformSigns.org/SignOverSign.htm>. Please also note that the CDLI use of the notation "SIGN x SIGN" using "x" does not refer only to infixation into a completely surrounding container component, even if that is a salient subset of the "SIGN x SIGN" types. Rather, it refers also to components overlapping. I am exploring lineal historical ancestors and descendants of signs in the core of the "SIGN x SIGN" type, to discover the extent of the group, which strong linkages of components are retained across history, which not.

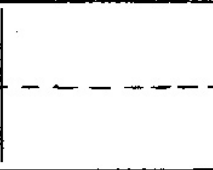
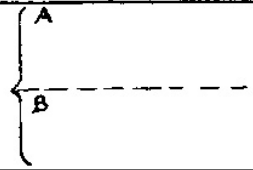
NUN



TUR₃



TUR₃
x
SAL



6. Particular notes on WG2 N2698, on page 7 unless otherwise noted:

[I reaffirm the appropriateness of the comparisons with Han CJK characters, the contrast between Sign and Component of Character being clearer there. Also the appropriateness of the comparisons with Latin "w" which we do **not** encode as the sequence "vv" with ligaturing, and which is **not** a ligature even though it arose that way. And the comparison with the Danish **letter** "æ", which is not a ligature even though it arose that way, and which we do **not** encode as the sequence "ae" with ligaturing. Please see the illustrations of this in paper L2/04-041. There is nothing misleading about these comparisons. This is indeed the issue for Cuneiform as well.]

Section 3.5 on TA* -- this is a special case under 3.3 Mergers and Splits, a late split. The phrase "to be used only when the new interpretation applies" is an implementation question, odd in this context. Fuller statements on implementation are needed, at least pointing to alternatives. In some cases the opposite decision may be made, as plausibly with the signs which later merged as KU, to name complex signs containing them, and to encode them, as their original forms making the distinction at all times, so that historical continuity is preserved. This may or may not be the chosen implementation, but the question does exist. It would be wise to discuss it before a final encoding proposal.

Sentence ending section 3.4 "In general, signs which shift from compound to complex or vice versa have been treated according to their Ur III manifestation." This contradicts section 3.3 on mergers and splits. By that principle, the status as single sign is to be encoded if it occurs at any time, including later, not merely earlier.

Under "complex signs" towards end of section 3.4, the phrase "which present a relative visual unity" applies equally to some signs which are by the proposers regarded as "compound signs". This is one source of the instability of criteria, the subjectivity of these judgements. Notice the slight space in one vs. the other example of AR above (called "IGI.RI"). That tiny difference in white space is irrelevant. Another source of instability of criteria arises through omitting clearly unitary or complex signs which are the lineal descendents or ancestors of ones incorrectly rejected as "compound". Evidence from Gudea and Hammurabi texts in lines whose typography has enough white space to make distinctions applies precisely to this issue of "visual unity". Ancient scribes treated many of these as single signs, by these typographic criteria.

Under "compound signs" middle of section 3.4, the phrase "generally treated by scholars as a single unit" is dangerously ambiguous about the kind of unit -- text unit (sequence of signs) or single sign? This makes it rather explicit that proposal WG2 N3698 is proposing to encode "wedge clusters" (components) as if they were full signs if a full sign exists which is similar. This is at the very least an encoding of components, not of signs. And not a consistent encoding of components, as discussed from various other points of view.

Section 3.9 Character order

This is actually an encoding by components, but it has mixed with it a renaming of many signs away from their traditional names, such as MUL (for constellations etc.) rather than the radically new "AN OVER AN AN" for a cluster of three AN components ('star'). These two issues can be separated. We can support the encoding by components, even by alphabetical order of the most usual naming of components, without any need to discard traditional names.

[Added note after the UTC meeting. There is a frequent assumption that the native tradition of the "DIRI" lists is in fact a list of compounds, or sequences of signs. That may not be correct. It may rather be a listing of "readings" which do not directly reflect components, whether those components are merely components of a sign or are a sequence of components which are themselves independent signs. Or the "DIRI" list may be a mixture of more than one structural category. In any case, native traditions like all meta-discussion **about** language and writing is statistically a less reliable witness to the structural units of language or writing than is study of the actual behavior of units in language or writing. More on these questions later.]

Results of a first-Level Triage on the first 200 signs in ZATU (Uruk period)

The remainder of ZATU will be completed to probably better than this level of validity by mid-February. Additional checking with the Lexical Lists and with reviews of ZATU and the lexical lists and with a number of scholarly articles will be completed before the end of February.

No additional encoded characters are needed (51% of the first 200)

- a) 68 unifiable with later signs
- b) 38 insecure

Additional encoded characters **are** needed (49% of the first 200)

- c) 83 complex, known parts or types of modification, not unifiable
- d) 17 simple (new components = signs)

Balancing of Criteria

This balancing will be both refined and made more explicit as the triage progresses. This is merely a sample to show that balancing has in fact occurred. What we have here is the approximate statistical relation of number of occurrences to presence in lexical lists or not, and to the four categories distinguished above. What this shows is that the recognizability of component parts of complex signs makes their inclusion on average more secure with a smaller number of attestations than for those signs having unique or new single components. Also that among those signs listed as “unidentified” (not unifiable with later signs), the weighting among those proposed for encoding after this first level triage is strongly towards those attested in higher frequencies. No criteria stand alone as determinative, rather a combination of all known factors bear on triage. Especially at the frontiers, many criteria are of course much more subtle than merely whether some experts have chosen to include a form in a list of “signs”. Later updates of this triage will appear by mid-February 2004 at the web site <http://www.CuneiformSigns.org/ZATUSignTriage.htm>

C. Complex with known parts, **not** in lexical lists (note the large proportion with attestation in ATU or elsewhere)

Other ATU	1x	2x	3x	4x	5x	6x	7x	8x	9x	10x
	9	6	8	7	6		1			1

C. Complex with known parts, **in lexical lists** (note the large proportion with only one occurrence)

Other ATU	1x	2x	3x	4x	5x	6x	7x	8x	9x	10x	19x	29x	40x	60x	100x
	27	3	5		3			1	2		1	1	1	1	1

D. Simple signs (single component), **not** in lexical lists (note the larger proportion with many occurrences)

Other ATU	1x	2x	3x	4x	5x	6x	7x	8x	9x	10x	12x	28x	34x
		2			1						1	1	1

D. Simple signs (single component), **in lexical lists** (note the larger proportion with many occurrences)

Other ATU	1x	2x	3x	4x	5x	6x	7x	8x	9x	10x	12x	15x	17x	23x	72x
	1	1							2	1	1	1	1	1	1

“Unidentified” signs (ZATU numbers 620 to 762) (note the especially large proportion with many occurrences)

Other ATU	1x	2x	3x	4x	5x	6x	7x	8x	9x	10x	11x	12x	13x	14x	15x	16x	17x
	4	11	3	2	1	4	1	1	6	1	1	1	1	2	2	2	2

20x	21x	28x	30x	45x	48x	59x
1	1	1	1	1	1	1

A correction note to the version of this partial first-level triage posted on web site (and in paper for UTC):
What had been references to the “Professions” list should instead read “Lu” list. That correction has been made
on the web page.