## 1 Questions for Steve

Steve, below are questions I've compiled whiled reviewing the PC repertoire and your comments document. As you know, I am not a specialist of PC. These questions are all born out of my ignorance of PC, SAC, and related domains, and my general naivete about the scholarly practices of collating and naming such signs. Thank you for your patience with my inquiries.

- 1. There are 88 signs slated for removal from the repertoire due to being out of chronological scope, ie. not from Uruk IV or Uruk III. Do any of these signs occur with signs from these periods? If so, then should they be encoded to support the complete representation of texts, whether in the PC block or in a different one? Also, do you have more details about which chronological periods they are from?
- 2. Sort order: I've fixed the internal sorting of numbers so that 1(N02), 2(N02), 1(N34), etc are grouped by the base N02 and not by the count, ie. 1(N02), 1(N34), 2(N02). I've also moved numbers to the end of the order, after 'Z', instead of keeping ordering after 'M', as in ORACC. Ideo-numeric compounds are ordered similarly, after 'Z'. We had identified the sort order as needing more discussion. How do we want to proceed with the sort order? Ideally, the sort order in Unicode would align with the scholarly order (or vice versa).
- 3. tenu vs non-tenu: Is there a rule or convention for calling a sign a tenu form vs. an allograph? For example, the sign  $\ \Box \$  AD~b is a rotated form of  $\ \Box \$  AD~a, but there is no tenu indication. The sign  $\ \Box \$  ADDA@t is a rotated form of  $\ \Box \$  ADDA, and is indicated as a tenu form, not an allograph.
- 4. Signs in compounds without independent attestation: Following the model for SAC, we're not encoding signs in compounds that do not occur independently. There are 40+ such signs (see section 9.2). While they won't be going into Unicode, I'd like to add them to the scholarly repertoire, similar to the @compoundonly signs you've listed in the PCSL ASL list. I would place them in the PCSL repertoire, and mark their Unicode status as 'compound-only sign, not scoped for inclusion in the Unicode repertoire'. How do you feel about that?
- 5. ZATU signs: Are there other specific names or aliases for ZATU signs? If so, would it make sense to rename the existing names? Are there new interpretations for ZATU642, ZATU645~b, ZATU711, ZATU737, etc.? These seem like they'd have different names, just based on their shapes, which are analogous to other non-ZATU-named signs.
- 6. AN: The sign \* AN~v1 duplicates \* AN, so you've recommended that it be removed. That would leave \* AN and \* AN~v2. I'd recommend renaming \* from AN~v2 to AN~v1 to remove the gap in the variant sequence. Additionally, As \* is a glyphic variant of \*, not an actual alternate form, would it make sense to rename AN~v1 as AN~a, and consider it an allograph? That way it can be referenced in PCSL, even if it's not in the Unicode repertoire.
- 7.  $|(BU\sim a\&BU\sim a).NA_2\sim a|$ : The glyph  $\Rightarrow$  suggests that the two BU $\sim$ a components are 'crossing' instead of 'over', compared to  $\Rightarrow$   $|(BU\sim a\&BU\sim a).NA_2\sim a|$ , in which they are positioned more one 'over' the other. Should  $\Rightarrow$  be renamed to  $|(BU\sim a\&BU\sim a).NA_2\sim a|$ ?
- 8. |(BU~a&BU~a).NA<sub>2</sub>~b|: You state that | |(BU~a&BU~a).NA<sub>2</sub>~b| duplicates | |(BU~a~v1%BU~a~v1).NA<sub>2</sub>~a@n|, and that the former should be deleted. But for the record, should |(BU~a&BU~a).NA<sub>2</sub>~b| be renamed as |(BU~a~v1%BU~a~v1).NA<sub>2</sub>~b@n|? Also which glyph should be the representative form? The | is more

graphically aligned with related signs, such as | (BU~a&BU~a).NA2~a|, than is \( \frac{1}{2} \).

- 9. |DARA<sub>3</sub>~c×(KAR<sub>2</sub>~c.ŠE~a)|: The component amed DARA<sub>3</sub>~c in this sign differs from the conventional form of DARA<sub>3</sub>~c that is found in |DARA<sub>3</sub>~c×KAR<sub>2</sub>~b|. It also differs from the DARA<sub>3</sub> forms in |DARA<sub>3</sub>~d×KAR<sub>2</sub>~b|, |DARA<sub>3</sub>~d×KAR<sub>2</sub>~b|, |DARA<sub>3</sub>~d2×KAR<sub>2</sub>~a1|, |DARA<sub>3</sub>~d2×(KAR<sub>2</sub>~a1.ŠE~a)|. It also differs from DARA<sub>3</sub>~c~v1. Given that the other DARA<sub>3</sub> forms are distinguished allographically, does it make sense to do the same for the form in , eg DARA<sub>3</sub>~c1, so the compound would be renamed |DARA<sub>3</sub>~c1×(KAR<sub>2</sub>~c.ŠE~a)|?
- 10. |DUG~a×LAM~b|: The sign is given the name |DUG~a×LAM~b|. But, the form of LAM here appears to be LAM~b~v1, not LAM~b. Should the name be changed to |DUG~a×LAM~b~v1|?
- 11. DUG~c@t: The CDLI-gh DUG~c has been separated into  $\bigcirc$  DUG~c~v1 and  $\bigcirc$  DUG~c~v3. Should the name DUG~c@t given to  $\bigcirc$  be changed to DUG~c~v1@t?
- 12. DUG~c2@t : Similar to DUG~c@t above, should the name DUG~c2@t assigned to the new addition be changed to DUG~c~v3@t?

- 15. |EZEN~a×LA~e| and |EZEN~a×LA~e|~v1: These names are assigned to and , respectively. However, the forms and of LA~e in the signs differ from the conventional LA~e (removed due to chronological scope), and appear to occur only in these compounds. Moreover, for the variant mark is applied to the entire compound rather than to the varying component. Should both of these signs be renamed to distinguish the different forms of LA~e using variant marks? For example ||EZEN~a×LA~e~v1| and ||EZEN~a×LA~e~v2|?
- 16. |EZEN~b×6(N57)|: The sign is marked for removal from the Unicode repertoire due to being out of chronological scope. However, the sign is retained as is the variant | EZEN~b×6(N57)|~v1 (see below). Given the inclusion of these signs, would it be suitable to also keep ?
- 17. |EZEN~b×6(N57)|~v1: The variant mark is applied to the entire compound to distinguish it from |EZEN~b×6(N57)| (removed from repertoire due to scope). Should the variant mark be applied to EZEN~b instead, to distinguish the conventional from the compound-only form ? That would result in |EZEN~b×6(N57)| and |EZEN~b~v1×6(N57)|
- 18. |GA~a~v1.ZATU753| and |GA~a~v4.ZATU753|~v1: This names are assigned to and in a respec-

tively. But, there appears to be a mistake in the PCSL names. First, for the ~v1 marker is outside of the compound. Second, the normative form of ZATU753 is . Should these be more appropriately named as [GA~a~v1.ZATU753~v1] (where ZATU753~v1 is compound-only) and [GA~a~v4.ZATU753]?

- 19. GAN~b: This sign relates to four forms: GAN~b, GAN~b~v1, GAN~b~v2 GAN~b~v3. Should the names for some of these be revised, following the pattern for DUG~b? We have DUG~b and DUG~b~v1. It seems should be GAN~b (squared left portion) instead of GAN~b~v3, and should be GAN~b~v1 (triangular left portion) instead of GAN~b. The signs and seem like different allographs of GAN~b altogether, eg. GAN~b2~v1 and GAN~b2~v2. So would it make sense to adopt the following renaming: GAN~b. GAN~b~v1, GAN~b1~v1, GAN~b1~v2?
- 20. |GAN~c×LAGAB~b|: This name is assigned to GAN~c~v1 rather than GAN~c?
- 21. |GAN~c×SIG7|: This name is assigned to Fix . Should the sign be renamed |GAN~c~v1×SIG7| so that it more accurates indicates the base Fix GAN~c~v1 rather than GAN~c?
- 22. |(GIŠ×(DIN.DIN)):  $\Rightarrow$  |(GIŠ×(DIN.DIN))~a| 'and  $\Rightarrow$  |(GIŠ×(DIN.DIN))~b| have names that suggest a conventional form of GIŠ, but the two are entirely distinct graphically. Should the GIŠ forms be treated differently?
- 23.  $|GIŠ\times ŠU_2\sim b|$ : The form of GIŠ in  $\longrightarrow$  differs from the conventional form  $\longrightarrow$  of GIŠ. Should this form be annotated as a variant of GIŠ, ie.  $GIŠ\sim v1$ ? If so, the compound should be renamed  $|GIŠ\sim v1\times ŠU_2\sim b|$ .
- 24. |LAGAB~a×KUŠU₂~a@t|: The ® sign has a rotated form of KUŠU₂, eg. ♥. but is it ➤ KUŠU₂~a or ➤ KUŠU₂~b? It seems to be the latter, based on the bottom edge being tapered instead of fanning out. Or is ♥ a new allograph, ie. KUŠU₂~c?
- 25. |LAGAB~a×NUN~b|: Seems like + has some extra hashes at the end of NUN~b. Can NUN~b take a varying number of slashes and still be NUN~b?... or do we have something like a NUN~c here? Compare to the allograpphic difference in the number of hashes as distinguished in TUR3~a and TUR3~b.
- 27. LAM~b@s: This name is assigned to \$\display\$, which suggests that the base form is LAM~b. However, the base form resembles LAM~b~v1. Should LAM~b@s be renamed to LAM~b~v1@s?
- 28.  $|LU_2\times GEŠTU\sim c3|$ : Name suggests that  $U_2$  and  $SGEŠTU\sim c3$ . But the  $GEŠTU\sim c3$  here is reminiscent of the  $U_1$  in  $U_2$  in  $U_3$  in  $U_4$  |SAGV1V6V1V6V1V6V1V1V1 is well as V2 in V2 |SAGV6V6V1V1 is defined as such?
- 29. MAR~a@t: The name assigned to suggests that the sign is a *tenu* form of MAR~a. The two forms have the same orientation, but the former is hatched. Maybe the @t here should be *gunu* instead, ie. MAR~a@g?

- 30. |MUŠEN.UR₃~b2|: Name suggests the second component of □ is □ UR₃~b2, but should it instead be □, which appears to be a compound-only sign. This □ seems like an allograph of UR₃~b1, if not an actual variant, given the lack of hatches, (or maybe a *nutillu* form?...). I would suggest the sign be renamed, eg. |MUŠEN.UR₃~c|... Also, see note below on |MUŠEN×PAP~a|...
- 31. |MUŠEN×PAP~a|: The sign 🕁 📰 is suggested for deletion due to an incorrect identification in an unpublished Umma text. It seems like this sign should be renamed |MUŠEN.UR<sub>3</sub>~b2| instead of deleted... See note above on |MUŠEN.UR<sub>3</sub>~b2|...
- 32. |NINDA₂×2(N01)|: in the form of 2(N01) does not reflect the conventional ≥ 2(N01). Should this be considered a variant of 2(N01), ie. 2(N01)~v1? Or is this typical of vertical stacking / horizontal sequencing patterns? Could it be a different N-form altogether?...
- 33. |NINDA2×GUDU4|: This CDLI-gh name applies to a sign with three variants: | NINDA2×GUDU4|~v2, | NINDA2×GUDU4|~v3, | NINDA2×(HI~a1.AN.ME~a)|, | NINDA2×(HI~a1.ME~a)|. The first two have been removed and the last two have been retained, but renamed. Wouldn't it make sense to rename the first two for consistency in the record, even though they are not part of the proposed Unicode repertoire? So, change | from |NINDA2×GUDU4|~v2 to |NINDA2×(HI.AN.ME~a)|, and | from |NINDA2×GUDU4|~v3 to |NINDA2×(HI.AN.ME~a)|? If we make these changes, then the original |NINDA2×GUDU4| names should be added as annotations to the PCSL database as well as the Unicode names list.
- 34. Glyphs for former |NINDA2×GUDU4| signs: The glyphs for these signs are identical in terms of composition, but CDLI-gh shows different sizes for them, eg. |NINDA2×GUDU4|~v2 and |NINDA2×(HI.AN.ME~a)|, |NINDA2×GUDU4|~v3 and |NINDA2×(AN.ME~a)|. Which form should be considered representative? (I think the smaller sized components are optimal, optically...)
- 35. NAB: Following on the question of when is a sign tenu and when is it not, we have \* NAB and \*\* NAB~v2. Are there semantic differences between these two? Could \*\* be considered NAB@t?
- 36. |SAG×GEŠTU~a|: Name for 🖾 suggests that SAG should be 🖾, but the form used is the compound-only allograph 💪. Should this SAG be distinguished, ie. |SAG~a×GEŠTU~a|?
- 37. |SAG×SAR~a|: Name for suggests that SAG should be , but the form used is SAG~v1. Should this be renamed to |SAG~v1×SAR~a|?
- 38. SAG@n: In the compound | SAG@n×GEŠTU~b| the SAG@n is intended to refer to an unfinished form of which is named SAG~v2, as in | SAG~v2×GEŠTU~b|. However, the independent form of SAG@n in CDLI is . So, should be considered as SAG~v2@n. It could be added as a new sign. But, in any case | SAG@n×GEŠTU~b| should be changed to | SAG~v2@n×GEŠTU~b| to account for the different forms of SAG@n.
- 39. |SAL.LAM~b|: The sign is given the name |SAL.LAM~b|. But, the form of LAM here appears to be LAM~b~v1, not LAM~b. Should the name be changed to |SAL.LAM~b~v1|? See |DUG~a×LAM~b| above.
- 40. |SILA₃~a×DUG~a|~v1: This name assigned to differentiates it from |SILA₃~a~v1×DUG~a|. The SILA₃~a appropriately refers to →, the sign in →, and SILA₃~a~v1 to the compound-only sign → in →. The two already have distinctive names, so the variant marker in the name for →, which occurs after the compound marker, is not needed. Would it be appropriate to rename |SILA₃~a×DUG~a|~v1 to |SILA₃~a×DUG~a|?

- 41. |SILA<sub>3</sub>~a×GARA<sub>2</sub>~a| : Is truly → SILA<sub>3</sub>~a × GARA<sub>2</sub>~a, or could it be GARA<sub>2</sub>~a × → SILA<sub>3</sub>~a? I would have expected SILA<sub>3</sub>~a to function regularly as the container in compounds, so would have thought the more appropriate representation for |SILA<sub>3</sub>~a×GARA<sub>2</sub>~a| to be → This is the only SILA<sub>3</sub>~a compound where SILA<sub>3</sub>~a is contained instead of the accompanying sign, compare to |SILA<sub>3</sub>~a×DUG~b|, → SILA<sub>3</sub>~a×KAŠ~a|, etc.
- 42. |SILA₃~a×ZATU646|: The compound ♠ has a form of ← ZATU646 that three strokes instead of two: ←. Should this form be catalogued as an allograph of ←? If so, this compound-only form could be named ZATU646~a, and the compound would be renamed as |SILA₃~a×ZATU646~a|.
- 43. |SILA₃~c×NI~a|: This name suggests that ▷ has the component ➤ SILA₃~c, as found in ▷ |SILA₃~c׊U| and ▷ |SILA₃~c×ZATU687|. But, the form > of SILA₃ here differs from the other variants. It also looks like it could be a *nutillu* form of ▷ SILA₃~a, with the bar crossing the two diagonals, but missing the piece to the right. In any case, would it be appropriate to designate > as an allograph similar to the ▷ SILA₃~d in ▷ |SILA₃~d×NI~a|? If so, maybe |SILA₃~e, which would give ▷ the name |SILA₃~e×NI~a|?
- 44. SUKUD@g~b~v1: This sign is named SUKUD@g~b~v1, implying it's a variant of SUKUD@g~b. But, it appears closer to a variant of SUKUD@g~d? Does its name need to be revised? Also, see question about SUKUD@g~d below.
- 45. SUKUD@g~d : Is SUKUD@g~d truly a *gunu* form? Or a separate allograph?
- 46. SUM~a@t: This name is assigned to implying it's a *tenu* form of SUM~a. But, it seems to be a *tenu* form of SUM~a~v1. Should the name for this sign be revised to SUM~a~v1@t?
- 47. ŠAB: There are four ŠAB signs: ŠAB~a, ŠAB~a, ŠAB~a~v1, ŠAB~b, ŠAB~b~v1. The ~v1 forms seem more graphically complex. Should they be considered the representative forms? If so, should the variant annotations be flipped?
- 48. ŠE~a@t~v1: This is a compound-only sign that occurs in |GEŠTU~a׊E~a@t~v1| and |KAŠ~b~v1׊E~a@t~v1|. Its name suggests that it is a form of |ŠE~a@t with a different rotation. However, the form in | differs from the | in | | in | line | in | line | line
- 49. |ŠE~a.NAM₂| and |ŠE~a.NAM₂|~v1: These names are assigned respectively to → and → and → l. However, the independent NAM₂ is □. This is the form in → |ŠE~a.NAM₂|~v1. Should the names be swapped? The name |ŠE~a.NAM₂| is more suitable for → lateral has the variant |, it should be renamed |ŠE~a.NAM₂~a| or |ŠE~a.NAM₂~v1|.
- 50. ŠEN~c and ŠEN~c@t: Should the names for these signs be flipped: SEN~c and SEN~c@t? It seems like the base form is 'horizontal', so should be the regular form and the tenu form. Compare to DUG~c~v1 and DUG~c@t.
- 51. TAR~c: This compound-only sign occurs has four occurrences:  $\langle \neg | UR_2 \times TAR \sim c |$ ,  $\neg \square | UŠ \sim b \times TAR \sim c |$ ,  $\neg \square | UŠ \sim b \times TAR \sim c |$ ,  $\neg \square | UŠ \sim b \times TAR \sim c |$ ,  $\neg \square | UŠ \sim b \times TAR \sim c |$ . Its basic shape appears to be  $\rightarrow$ , as in  $| UŠ \sim b \times TAR \sim c |$  and  $| UŠ \sim b \sim V2 \times TAR \sim c |$ , with a *tenu* form in  $| UR_2 \times TAR \sim c |$ . However, the form in  $| AMAR \sim V2 \times TAR \sim c |$  is different. Is this TAR~c or possibly an allograph TAR~d?

- 52. TUR@g : This name assigned to  $\square$  implies a *gunu* form of  $\subseteq$  TUR. But, the sign appears to be  $\square$  TUR~v3 instead. Should  $\square$  be renamed TUR~v3@g?
- 53.  $|U_4.3(N08)|$ : The compound  $\nearrow$  has a variant form (different vertical stacking) of  $^{\frac{1}{5}}$  3(N08), which should be identified with a variant marker just as  $^{\frac{1}{5}}*3(N08)\sim v1$  is indicated in  $^{\frac{1}{5}}*100$   $|U_4\times3(N01).3(N08)\sim v1$ .
- 54. |U₄×1(N01).5(N08)|: The sign has a form of 55 5(N08) (different vertical stacking), which should be marked as a variant.
- 55.  $|U_4 \times 5(N57)| ... |U_4 \times 10(N57)|$ : This set of signs have variant positioning of numbers, which differs from the orientation of the normative signs indicated in their names:

```
    ⇒ |U<sub>4</sub>×5(N57)| : ≡ 5(N57)
    ⇒ |U<sub>4</sub>×6(N57)| : ≡ 6(N57)
    ⇒ |U<sub>4</sub>×7(N57)| : ≡ 7(N57)
    ⇒ |U<sub>4</sub>×8(N57)| : ==== 8(N57)
    ⇒ |U<sub>4</sub>×10(N57)| : ==== 10(N57)
```

On the other hand, the following relate set have similar orientations to their normative number signs:

```
• \Rightarrow |U_4 \times 1(N57)| : -1(N57)

• \Rightarrow |U_4 \times 2(N57)| : = 2(N57)

• \Rightarrow |U_4 \times 3(N57)| : = 3(N57)

• \Rightarrow |U_4 \times 4(N57)| : = 4(N57)
```

Should  $|U_4 \times 5(N57)|$  ..  $|U_4 \times 10(N57)|$  be given allograph or variant annotations?

- 56. |UKKIN~b×5(N01)|: This sign has a different orientation of its numerical component than the conventional 5(N01). Is an allograph or variant annotation needed for 5(N01)?
- 57. |URU~a3×KALAM~a|: The sign has a form of KALAM~a that differs from the named form KALAM~a. Should the be renamed to properly identify the component, ie. KALAM~g?
- 58. |UŠ~b×TAR~d|: Even though the sign | UŠ~b×TAR~d| has been removed from the Unicode repertoire, it should be renamed for consistency for record. The form of UŠ~b in the sign is more aligned with the compound-only sign | \*UŠ~b~v2 that occurs in | UŠ~b~v2×TAR~c|, than with the conventional | UŠ~b, eg. | UŠ~b×TAR~c|, | | UŠ~b&UŠ~b|. I suggest this sign | be renamed | UŠ~b~v2×TAR~d|.
- 59. |1(N57).ŠUBUR~v1|: This compound name for w distinguishes the compound-only form of ŠUBUR from the conventional form found in |1(N57).ŠUBUR|. However, the 1(N57) in the former seems to be a *tenu* form, as compared to the 1(N57) in the conventional form. Should |1(N57).ŠUBUR~v1| be renamed as |1(N57@t).ŠUBUR~v1|? Or is the placement of the 1(N57) with regard to naturally expected for 1(N57) given the shape of the variant ŠUBUR?
- 60. |1(N57)×SIG|: This name is assigned to ¬♦, as a revision of CDLI-gh |SIG×1(N57)| due to the ordering of components in the compound. But, is the compound operator truly × 'plus' as opposed to . 'beside'? Compare with ₱♦ |2(N57).AB₂| where there is similar compound behavior, but there is more containment in ₱♦ than in ¬♦. Neither appear to have 'time' behavior, s should |1(N57)×SIG| be renamed as |1(N57).SIG|?

- 62. 8(N57) in compounds: The ==== 8(N57) occurs in ▶ |8(N57).NI~b|, but with a different form. Should this form be marked distinctively?
- 63. 'Ignore' vs 'keep': The data file has recommend = 'keep' for |1(N08@f)×1(N57)| and |1(N34)×1(N58)|. As these are pure numeric compounds, I've changed the recommendation from 'keep' to 'ignore' and added them to the list of numerical signs that should be moved to the archaic numbers proposal. Is this ok?
- 64. |1(N57).1(N57)| is marked for removal, but given that it is a numerical compound, should it be ignored instead, and moved to the archaic numbers proposal?