

## Materials for Encoding Proto-Cuneiform

MEPC is divided into three parts:

**Texts:**

An overview of the PC text corpus and how its subset used for PC25.

**Signs:**

Data on the distribution of graphemes in the corpus and subcorpora.

**Lists:**

Top-level sign lists for PCSL and PC25 derived from a combination of the text corpus and extant sign lists.

## Corpus

The PC text Corpus is defined by the CDLI corpus of Proto-Cuneiform texts of Uruk IV and Uruk III data, as adapted for use by PCSL: we call this CDLI-tc; note that CDLI-tc is used here only in the PCSL version of the corpus. This version has some minor modifications to the transliteration and has been converted to use Unicode transliteration conventions rather than the CDLI ASCII ones.

There are 5976 texts in the corpus of which 1752 are attributed to Uruk IV and 4224, or roughly 75% of the corpus, to the subsequent Uruk III period.

### The Digital Corpus

The digital corpus includes the contents of the following major publications:

- Uruk lexical tablets from ATU3 (Uruk IV and Uruk III)
- Uruk administrative tablets from ATU5, ATU6, ATU7 (Uruk IV and Uruk III); these supercede ATU1
- Jemdet Nasr (Uruk III) tablets from MSV01; these supercede Langdon, PI
- Tablets from various locations from MSV04; this includes non-Uruk tablets which were included in ATU1
- Tablets in private collections from various locations in CUSAS1, CUSAS21, and CUSAS31; these post-date the ATU and MSVO volumes
- Tablets from the "Erlenmeyer Collection"; these were to be published in MSV03 which has not yet appeared; many of them have been edited by Englund and others in various publications, however

### PCSL Corpus

PCSL's version of CDLI-tc has the following composition divided by provenience, period, and published/unpublished status:

	IV/pub	IV/unp	IV/all	III/pub	III/unp	III/all
<b>Uruk</b>	1191	599	1790	1575	1522	3097
<b>JN</b>	0	0	0	236	33	269
<b>Umma</b>	0	0	0	90	308	398
<b>Uqair</b>	0	0	0	39	3	42
<b>Misc</b>	29	16	45	369	286	655
<b>total</b>	1220	615	1835	2309	2152	4461

## PC25 Corpus

About 1/3 of the PCSL corpus is available only in transliteration or photograph and has not yet been subjected to the rigorous assessment of a scholarly edition. These texts are removed to create a subcorpus of well-studied text to serve as the basis for the initial (and largest) phase of encoding of Proto-Cuneiform.

The subsetting is based on the CDLI catalogue entries as of February 2025. It is possible that features of the CDLI data may mean that a few texts that should have been omitted are included in the PC25 corpus and vice versa. The impact of this on the final repertoire is negligible, however, because of the cross-checking between corpus and published sign lists.

PC25's subset of the PCSL corpus has the following composition divided by provenience, period, and published/unpublished status:

	IV/pub	IV/unp	IV/all	III/pub	III/unp	III/all
<b>Uruk</b>	1191	88	1279	1574	399	1973
<b>JN</b>	0	0	0	236	1	237
<b>Umma</b>	0	0	0	90	3	93
<b>Uqair</b>	0	0	0	39	3	42
<b>Misc</b>	29	4	33	365	222	587
<b>total</b>	1220	615	1312	2304	2152	2932

## Graphemes

### Grapheme Distribution

A general impression of the amount of graphemic data in the various subcorpora is given in the table below. In each case, the numbers are the count of distinct signs and the total number of instances of signs, with numerical signs and ideograms being given in separate rows.

#### PCSL Corpus Grapheme Distribution

	IV/pub	IV/unp	IV/all	III/pub	III/unp	III/all
<b>Uruk/num</b>	0/0	0/0	0/0	0/0	0/0	0/0
<b>Uruk/idg</b>	1564/10173	906/4710	1925/14883	2152/24011	1714/16361	2794/40372
<b>JN/num</b>	0/0	0/0	0/0	0/0	0/0	0/0
<b>JN/idg</b>	0/0	0/0	0/0	1122/6309	9/11	1123/6320
<b>Umma/num</b>	0/0	0/0	0/0	0/0	0/0	0/0
<b>Umma/idg</b>	0/0	0/0	0/0	725/3242	1478/11482	1724/14724
<b>Uqair/num</b>	0/0	0/0	0/0	0/0	0/0	0/0
<b>Uqair/idg</b>	0/0	0/0	0/0	396/1323	20/24	400/1347
<b>Misc/num</b>	0/0	0/0	0/0	0/0	0/0	0/0
<b>Misc/idg</b>	76/122	36/51	103/173	1040/6738	1100/8278	1566/15016
<b>total</b>	0/0	0/0	0/0	0/0	0/0	0/0
<b>total</b>	1576/10295	917/4761	1941/15056	2983/41623	2756/36156	4055/77779

#### PC25 Corpus Grapheme Distribution

	IV/pub	IV/unp	IV/all	III/pub	III/unp	III/all
<b>Uruk/num</b>	0/0	0/0	0/0	0/0	0/0	0/0
<b>Uruk/idg</b>	1564/10173	403/1391	1695/11564	2152/24011	1091/7009	2476/31020
<b>JN/num</b>	0/0	0/0	0/0	0/0	0/0	0/0
<b>JN/idg</b>	0/0	0/0	0/0	1122/6309	9/11	1123/6320
<b>Umma/num</b>	0/0	0/0	0/0	0/0	0/0	0/0
<b>Umma/idg</b>	0/0	0/0	0/0	725/3242	163/328	826/3570
<b>Uqair/num</b>	0/0	0/0	0/0	0/0	0/0	0/0
<b>Uqair/idg</b>	0/0	0/0	0/0	396/1323	20/24	400/1347
<b>Misc/num</b>	0/0	0/0	0/0	0/0	0/0	0/0
<b>Misc/idg</b>	76/122	17/18	87/140	1033/6690	843/6228	1363/12918
<b>total</b>	0/0	0/0	0/0	0/0	0/0	0/0
<b>total</b>	1576/10295	412/1409	1708/11704	2979/41575	1569/13600	3409/55175

## Sign Lists

Prior PC proposals were centred on CDLI-gh, treating it as the definitive assemblage of PC signs at the same time as recognizing several important considerations: CDLI-gh is not 100% complete with respect to the PC corpus; includes some signs from ED duplicates of PC lexical texts; and includes a handful of signs which are either duplicates or are apparently place-holders from ongoing work on the Schoyen Umma texts that was never completed.

The published sign list of Uruk Lexical Texts from ATU3 (LLATU) was also utilized as a partial control on CDLI-gh. However, three additional lists in a similar format to the LLATU lists were not used in prior proposals, leading to an inadequate understanding of previously published scholarship on the PC repertoire. Together with LLATU these three previously unutilized lists provide a comprehensive new presentation of the material covered in ZATU and need to be included as part of the foundational data of the PC proposal.

The four lists and their coverage are:

### LLATU:

Lexical lists from Uruk, but with some extraneous signs or forms from ED duplicates, replacing ZATU's coverage of lexical lists

### ATU5:

Administrative texts from Uruk, replacing ATU1 signlist and ZATU

### MSV01:

Administrative texts from Jemdet Nasr, replacing PI and ZATU

### MVSVO4:

Administrative texts from various proveniences, replacing ZATU

The sign lists are based on exhaustive scholarly reassessments of individual portions of the PC corpus and make extensive use of the contrastive notations with subscript letters+numbers, e.g., ABa and ABb. At the same time, these lists gather non-contrastive sign variants under their respective parent signs and this is taken into account in PC25.

The four modern sign lists are an invaluable complement to CDLI-gh because they represent the carefully considered subset of signs which were vetted for publication whereas CDLI-gh is a working collection of signs.

These sign lists make it clear that the unmarked variants in CDLI-gh are non-contrastive variants as opposed to the contrastive variants marked with subscript letter+number sequences.

Importantly, Englund is explicit in the introduction that the reference forms of the signs ("graphemes" in Englund's terminology) are only exemplary forms:

After each sign name a grapheme is presented which represents the general form of the sign on the tablets cited. This graphic must be understood as merely an orientation in understanding the form a particular sign could take, since in particular the texts from the earliest stage of writing exhibit, to varying degrees, a tolerance of graphemic variation. (ATU5 p.107)

### **Sign Lists and Concordances**

Several tables covering the different levels of sign list used in PCSL are on the sign list overview page..

### **MEPC Sign List Overview**

MEPC builds up its top-level lists via mid-level aggregations that incorporate low-level data from individual source sign lists.

#### *Top-level Lists*

##### **PC25:**

A draft repertoire for a PC encoding. There are also accompanying pages of signs that are removed in PC25 for the following reasons:

- not in corpus;
- sequence not character;
- sequence exceptions encoded as characters;
- broken characters;
- deleted from list;
- numbers not in PC25;
- numbers not in PC25, excluding those in ACN.

##### **PCSL:**

The new version of PCSL on which PC25 is based--includes characters that should not be encoded or which will only be encoded in the future. The most complete sign list of Proto-Cuneiform available today.

#### *Mid-Level Aggregations*

##### **EASL--Englund Archaic Sign List:**

A superset of the ideographic signs in CDLI-gh with alignment columns for ATU3, ATU5, MSVO1, and MSVO4 and category tagging for signs. EASL also includes about 50 signs used in CDLI-tc which do not appear in any of the other source sign lists.

##### **EANM--Englund Archaic Numbers:**

Numerical signs from CDLI-gh with tagging to indicate whether they were included in ACN, should be in PC25, should possibly be encoded in the future, or should not be encoded.

Low-Level Proofing Tables

The proofing tables focus on ideographic signs and aim to align all of the individual glyphs in the various lists with EASL.

EALX--Englund Archaic Lexical:

A proofing table for ATU3

EAAT--Englund Archaic Administrative Texts:

A proofing table for ATU3

EAJN--Englund Archaic Jemdet Nasr:

A proofing table for MSV01

EAVR--Englund Archaic Varia:























A proofing table for MSV04

CSL--CUSAS Sign List:

An EASL-style list containing candidate additional signs from CUSAS 1, 21, and 31.

Glyf Database

## Sequence Database

	ŠA~3ŠE~a~12062 12B3B E0103 2062 12B52 E0101
	ŠE~a~1ŠA~32062 12B52 E0101 2062 12B3B E0103
	AMAR~2.1(N02)~1126CA E0102 200D 125BE E0101
	AN~3.IM~a~1.GI <sub>6</sub> ~2126CD E0103 200D 128FE E0101 200D 1288F E0102
	APIN~a~1.APIN~a~1126D2 E0101 200D 126D2 E0101
	BU~b~1.NA <sub>2</sub> ~a~112716 E0101 200D 12A0F E0101
	BULUG <sub>3</sub> ~1.DU <sub>6</sub> ~a~21271A E0101 200D 12751 E0102
	DA~a~1.LIŠ~112720 E0101 200D 129B5 E0101
	DU~1.ME~a~1.NUN~a~11274D E0101 200D 129E2 E0101 200D 12A74 E0101
	DU~1.U <sub>4</sub> ~11274D E0101 200D 12BE1 E0101
	E <sub>2</sub> ~a~1.LIŠ~1127CA E0101 200D 129B5 E0101
	E <sub>2</sub> ~a~1.NUN~a~1127CA E0101 200D 12A74 E0101
	E <sub>2</sub> ~b~1.LIŠ~1127CE E0101 200D 129B5 E0101
	EN~a~4..ŠE~a@t+EZEN~b~1NUN~a~1.SIG <sub>7</sub> ~2127D6 E0104 200D 200D F3106 E0101 2062 12A74 E0101 200D 12AE3 E0101 200D 12A74 E0101
	EN~a~4.BAHAR <sub>2</sub> ~b~1SIG <sub>7</sub> ~1.ME~a~1.NUN~a~1127D6 E0104 200D 126E6 E0101 2062 200D 12AE3 E0101 200D 12AE3 E0101 200D 12AE3 E0101
	EN~a~4.KID~a~1127D6 E0104 200D 12941 E0101
	EN~a~4.ME~a~1.GI~2127D6 E0104 200D 129E2 E0101 200D 12884 E0102
	EN~a~4.ME~a~1.MU~1127D6 E0104 200D 129E2 E0101 200D 129EC E0101
	EN <sub>2</sub> ~1.E <sub>2</sub> ~a~1127E3 E0101 200D 127CA E0101
	EN <sub>2</sub> ~1.E <sub>2</sub> ~b~1127E3 E0101 200D 127CE E0101
	EŠDA~1.NAM <sub>2</sub> ~1127F0 E0101 200D 12A23 E0101
	EŠDA~1+NAM <sub>2</sub> ~1127F0 E0101 2064 12A23 E0101
	GA~a~1.ZATU753~11280A E0101 200D 12D79 E0101
	GAL~a~1.BUR <sub>2</sub> ~11284E E0101 200D 1271F E0101
	GAL~a~1.NI~a~11284E E0101 200D 12A3A E0101
	GAL~a~1.NIM~a~11284E E0101 200D 12A43 E0101
	GAL~a~1.UKKIN~a~11284E E0101 200D 12C32 E0101
	GAL~a~1.UKKIN~a~21284E E0101 200D 12C32 E0102
	GEŠTU~a~1.NAGA~a~112871 E0101 200D 12A16 E0101
	GI <sub>6</sub> ~2.NUNUZ~c~11288F E0102 200D 12A87 E0101
	GIR~a~1.KU <sub>6</sub> ~a~112893 E0101 200D 12965 E0101
	GIR~b~1.GIR~b~112895 E0101 200D 12895 E0101
	GIR <sub>3</sub> ~c~1.AB~a~21289F E0101 200D 126A2 E0102
	GIR <sub>3</sub> ~c~1.PIRIG~b1~11289F E0101 200D 12A98 E0101
	GIŠ~1.TE~1128A7 E0101 200D 12BB9 E0101
	GIŠ@t~1.E <sub>2</sub> ~a~1128AE E0101 200D 127CA E0101
	GU <sub>4</sub> ~2.ZATU755~b~1128C1 E0102 200D 12D7C E0101

## Revised Principles for Encoding Proto-Cuneiform

This page reviews some of the previous assumptions and challenges surrounding a PC encoding and lays out a revised set of principles on which the PC25 repertoire is based.

### *Background*

#### Issues with Sign Lists

- Proposals so far based on lists, especially CDLI-gh and assume list entries are primary source of Unicode characters
- Sign Lists offer one perspective on a repertoire
- Can't assume that every sign list entry should be encoded
- Sign forms are abstractions; two-dimensional sketches of a three-dimensional writing system which tend to offer typical forms
- Sign Lists do not define the use of signs in a corpus
- Sign Lists do not necessarily capture the full range of glyph-variation for any individual character; just because a sign doesn't have unmarked variants in CDLI-gh doesn't mean such variants don't exist (Uruk IV EZEN~c)

#### Contrastive Usage

- Prior assumption that we cannot identify any contrastive/non-contrastive distinctions is not valid
- Historical dimension--Uruk IV and Uruk III forms of same sign (sometimes not a clear distinction)
- Lexical Data -- Uruk IV vs Uruk III manuscripts
- Context -- commodity lists and lexical texts suggest contrastive usage when they have distinct entries for otherwise similar-looking signs
- Transliteration Practice:
  - CDLI-tc can be used as a control on CDLI-gh: if CDLI-tc does not differentiate variants this is an indicator that the variation is non-contrastive
  - SAG example -- CDLI-gh unmarked variants -- SAG has three forms but they are not differentially labelled in CDLI-gh because non-contrastive
  - ŠU<sub>2</sub> example ŠU<sub>2</sub>: ŠU<sub>2</sub>~a and ŠU<sub>2</sub>~b marked in sign lists but not in transliteration because non-contrastive
  - KUŠU<sub>2</sub> example--each of the variants has instances in CDLI-tc

#### Compounds

CDLI-tc and CDLI-gh do not always differentiate compound constituents to the same extent as the independent versions of the constituents. For example, KAR<sub>2</sub> is separated as KAR<sub>2</sub>~a and KAR<sub>2</sub>~b in CDLI-tc and CDLI-gh, but in the DARA<sub>3</sub>×KAR<sub>2</sub> compounds the only notations that occur are DARA<sub>3</sub>~c×KAR<sub>2</sub> and DARA<sub>3</sub>~d×KAR<sub>2</sub>. Where it is clear which version of a compound-constituent is present in the compound, the compound notation should be revised to be specific. Where it is not clear, the compound notation should be left as is, following the CDLI perspective on contrastive/non-contrastive; it may be that variants are considered contrastive when used independently but non-contrastive when used as part of a compound.

## Sequences

- Most sequences are collections of components
- Some sequences are opaque, i.e., the CDLI-tc/CDLI-gh sign name hides the fact that the sign is a sequence (e.g., ENGIZ, ŠAB)
- Several classes of sequences with possibly different handling:
  - Some sequences are reanalysis of originally integral sign forms [esp city names]
  - Animal-ages are indicated by N(N57) followed by signs which in some cases are known to be animals and in others may be assumed to be based on this usage
  - Time measures where years are indicated as N(N57)+U<sub>4</sub>, months as U<sub>4</sub>×N(N14N08), and days as N(N08) following an U<sub>4</sub> notation
- Ordering and placement of sequence components is highly variable and non-contrastive; relative positioning of elements in a sequence occurs because the distribution of signs in cases is not linear and is not an integral part of the structure of the sequence, e.g., GA~a.ZATU753. The arrangement illustrated in CDLI-gh can be effected via ligatures but does not need to be part of the encoding

## PC25 Principles

- shift the basis of the encoding onto the PC corpus and usage, and use the corpus as a control on the lists; use the published lists and corpus as a control on CDLI-gh
- align names with CDLI-tc/CDLI-gh as much as possible, with some exceptions where required to correct names or to improve consistency of naming scheme; if in doubt, retain CDLI names
- take contrastive usage into account to the extent supported by contemporary scholarship
- do not introduce finer-grained allograph notations than CDLI-tc/CDLI-gh is using. The decisions made in the corpus about whether sign variants are contrastive or not are made not only on the basis of form but also context of various kinds; specialists in the corpus should decide if further division is needed in the future
- allograph notations are by default assumed to be contrastive but there are exceptions, e.g., ŠURUP-PAK~a/b/c; evaluate the treatment of ŠE~a and ŠE~a@t (90 vs 45 degrees)
- do not assume that every sign list entry should be encoded as a character
- consider distribution of components when encoding X×Y versus X.Y or even Y.X; sometimes, especially for rare signs, it is not clear whether the juxtaposition of components is part of the sign structure or the distribution of individual elements on the manuscript. In such cases it is preferable to treat the signs as a sequence rather than a complex (e.g., GEŠTU~a×ŠE~a@t treated as ŠE~a@t.GEŠTU~a in PCSL).
- do not generally encode sequences; this includes sequences which are named in CDLI-gh and CDLI-tc as single characters but where the naming is an interpretive mnemonic for a sign group such as ŠAB for PA.IB and the like.
- do make exceptions to sequences rule for items which are:
  - not historically sequences but are later decomposed, i.e., city names and possibly others
  - semantically integral and more convenient to encode as characters, i.e., N(N57)+U<sub>4</sub> year notations
- do not require a minimum number of occurrences for encoding: the corpus consists of mostly fragmentary manuscripts over 5000 years old--if a sign clearly exists and meets the other principles for encoding, it should be encoded
- do not encode signs which occur only in compounds
- do not encode uncertain signs, especially those from unedited texts such as the Schoyen Umma material



- do not encode broken signs; reserve them for the PUA

### **Advantages of the Revised Approach**

- encoding better aligned with transliteration practice
- additional glyph variants can be added without impacting the encoding; encoding every glyph variant would open PC to arbitrary open-ended encoding of slight differences with little basis for distinguishing when a variant should be encoded and when not: adopting the position that scholarly annotation of glyph variance as contrastive is required for encoding would set reasonable boundaries on what can be encoded and what should not be
- new sequences can be added without impacting the encoding; especially important for potentially productive types N57+ANIMAL and U<sub>4</sub>+DAY

### **Mitigations of Issues with Revised Approach**

- variant forms can be managed with font features
- disunifications possible when further research indicates them
- unifications possible if some separately encoded characters are later proved to be non-contrastive

### **Reference Glyphs**

The introduction of 1:several relationships where an encoded character has multiple variants entails the need for a principled selection of reference glyphs.

In order to have some level of consistency it would be preferable to select either Uruk IV glyphs or Uruk III glyphs as the primary choice of reference glyphs. Because the corpus is predominantly Uruk III in date it makes sense to use Uruk III reference glyphs as far as possible.

PC25 reference glyphs are aligned where possible with Uruk III sign forms occurring in published texts originating from Uruk or Jemdet Nasr. The selection of the reference glyph is not necessarily an indication that the other sign forms in EASL do not occur in the same period or place. It means simply that the reference glyph has been confirmed to occur, where possible, in Uruk III Uruk/Jemdet Nasr.

For sequences with multiple forms, the reference form is always the simplest/closest to the sequence description, as long as that form occurs in the corpus. This means that by default the sign looks the way it is described, and ligatures, reorderings, or non-linear dispositions are always accessed by CVNN.

Important to understand that the selection of an RG does not imply that the form is normative--the corpus is restricted and sign form variation is considerable which means that the concept of a normative form is often inapplicable to any given sign.

### **Repertoire for Encoding Proto-Cuneiform**

The complete unified list of the proposed PC25 repertoire is available [here](#).

The repertoire is created using the CDLI-PCSL text corpus and unified PCSL sign list as starting points. Signs are removed from PCSL for the following reasons:

1. Corpus scope: signs are first removed because they do not occur in the PC25 subcorpus. For a list of these signs see Appendix 2.
2. Non-contrastive variants: XXX glyphs are not removed as such but do not enter into consideration because PC25 does not encode non-contrastive glyph variants. XXX ADD PAGE

3. Ineligible Numbers: signs are removed because they are numbers that are either encoded already in the Archaic Cuneiform Numbers proposal (ACN) or because PC25 does not generally include numbers (exceptions include the N57 and U<sub>4</sub>-time sequences). For a list of these signs, excluding those in ACN, see Appendix 3.
4. Sequences: signs are removed because they are sequences. PC25 makes no distinction between opaque sequences--those whose label hides the fact that the sign is a sequence--and simple sequences. PC25 does propose to encode several exceptional groups of sequences as described in the Sequence Exceptions section. A list of excluded sequences is given in Appendix 4.
5. Broken signs: are removed because they are damaged and partially incomplete. A list of broken signs is given in Appendix 5.
6. Deleted signs: signs are deleted despite surviving the preceding conditions because they are duplicates. A list of deleted signs is given in Appendix 6.

## Numbers

Although PC25 is primarily focussed on ideograms, some numbers, listed here, need to be encoded to support N(N57) sequences. This is a class of additions to AP24, which omitted all numbers.

## PC25: Sequence Exceptions

### *Principle and Practice*

In principle, sign list entries which are sequences of separate signs are not encoded in PC25. In practice, there are several reasons for making exceptions to this policy:

#### **Reanalysis:**

Proto-Cuneiform signs that are originally integral units may be reanalyzed into separate constituents. Where this is demonstrably the case, the original character is encoded and the re-analyzed sequence is considered a glyph variant.

#### **Common Signs:**

Some signs are allowed as exceptions because it would be counter-intuitive not to encode them as characters.

#### **Container Equivalency:**

Some signs are allowed as exceptions because the juxtaposition of elements is the equivalent of a container (TIMES) relationship.

#### **Unencoded Constituents:**

Some sequences contain constituents that are otherwise unattested; in this case the choice is either to encode a sign which may not be attested independently, or to encode the sequence. In general the option adopted is to encode the sequence as a character.

#### **Analogy:**

Some sequences are part of a group and would naturally be considered by users of the encoding to be analogous to each other. Where one or more members of a group fulfils either of the previous conditions for encoding as a character, PC25 encodes the entire group as individual characters by analogy to avoid a possibly confusing mixture of encoded and unencoded group members.

## *List of Exceptions*

### **Reanalysis**

Two city-name signs, ADAB and ARARMA~a, have earlier forms which are distinct from their reanalysis to include an initial U<sub>4</sub> component. Other city names may also have earlier integral forms but without further evidence they are not proposed for encoding as characters at this point.

### **Common Signs**

The signs LUGAL (GAL+LU) and LI (ŠE+ŠA) are encoded as characters as common sign exceptions.

### **Container Equivalency**

The signs ASAR and AZ are the equivalent of containers.

### **Unencoded Constituents**

The following exception signs contain unencoded constituents: ENKUM (EZEN×ŠE), ME<sub>3</sub> (EŠDA-tenu), ŠAGINA (modified UŠ form with additional strokes, unclear with this is an UŠ or not), ZUBI~a (NA<sub>2</sub>-nutillu).

### **Analogy: BAPPIR Group**

The group of signs with the base BAPPIR has one member which is a container (BAPPIR~e) and the entire group is encoded by analogy.

### **Analogy: ŠELU**

Most combinations of ŠE plus another sign are encoded as characters (e.g., TU) so an exception is made for ŠELU.

### **Analogy: Sheep Group**

The groups of signs with the base SILANITA, UDUNITA and UTUA represent various types (ages, genders) of sheep and since some of them have unencoded constituents the members of all groups are encoded as characters.

### **Analogy: UTUL Group**

The UTUL group contains one member which includes an uncoded superposed reduplicated component (UTUL~c), so the entire group is encoded as characters.

## **Numbers**

Most PC numbers have been included in a separate proposal, "Archaic Cuneiform Numbers" (ACN).

Several pages in the PCSL Sign List provide a concordance of ACN and PCSL and notes on numbers not included in ACN; they are all available from ACN Concordance page.

N57 numbers are proposed for inclusion in PC25.

12BFA  
12BFB  
**Character Set**

**Code Charts**  
12BFC

























































































































The code charts are provided in separate HTML pages:

- 12600-126FF
- 12700-127FF
- 12800-128FF
- 12900-129FF
- 12A00-12AFF
- 12B00-12BFF
- 12C00-12CFF


































































































































































**Repertoire List**  
12C00

	PROTO-CUNEIFORM NUMBER FOUR-N58
	PROTO-CUNEIFORM NUMBER FIVE-N58
	PROTO-CUNEIFORM NUMBER EIGHT-N58
	PROTO-CUNEIFORM NUMBER NINE-N58
	PROTO-CUNEIFORM NUMBER ONE-N58
	PROTO-CUNEIFORM NUMBER ONE-N58
	PROTO-CUNEIFORM NUMBER ONE-N58 TENU


































































































































































**Code Chart 1: 12690 - 126FF**

	1269	126A	126B	126C	126D	126E	126F <sup>0</sup>	12690	126A0	126B0	126C0	126D0	126E0	126F0
1														
2														
3														
4														
5														
6														
7														
8														
9														
A														
B														
C														
D														
E														
F														

## Code Chart 2: 12700 - 127FF

274	1275	1276	1277	1278	1279	127A	127B	127C	127D	127E	127F <sup>0</sup>	 12700	 12710	 12720	 12730	 12740	 12750	 12760	 12770	 12780	 12790	 127A0
													 12711	 12721	 12731	 12741	 12751	 12761	 12771	 12781	 12791	 127A1
													 12712	 12722	 12732	 12742	 12752	 12762	 12772	 12782	 12792	 127A2
													 12713	 12723	 12733	 12743	 12753	 12763	 12773	 12783	 12793	 127A3
													 12714	 12724	 12734	 12744	 12754	 12764	 12774	 12784	 12794	 127A4
													 12715	 12725	 12735	 12745	 12755	 12765	 12775	 12785	 12795	 127A5
													 12716	 12726	 12736	 12746	 12756	 12766	 12776	 12786	 12796	 127A6
													 12717	 12727	 12737	 12747	 12757	 12767	 12777	 12787	 12797	 127A7
													 12718	 12728	 12738	 12748	 12758	 12768	 12778	 12788	 12798	 127A8
													 12719	 12729	 12739	 12749	 12759	 12769	 12779	 12789	 12799	 127A9
													 1271A	 1272A	 1273A	 1274A	 1275A	 1276A	 1277A	 1278A	 1279A	 127AA
													 1271B	 1272B	 1273B	 1274B	 1275B	 1276B	 1277B	 1278B	 1279B	 127AB
													 1271C	 1272C	 1273C	 1274C	 1275C	 1276C	 1277C	 1278C	 1279C	 127AC
													 1271D	 1272D	 1273D	 1274D	 1275D	 1276D	 1277D	 1278D	 1279D	 127AD
													 1271E	 1272E	 1273E	 1274E	 1275E	 1276E	 1277E	 1278E	 1279E	 127AE
													 1271F	 1272F	 1273F	 1274F	 1275F	 1276F	 1277F	 1278F	 1279F	 127AF

## Code Chart 3: 12800 - 128FF











1284	1285	1286	1287	1288	1289	128A	128B	128C	128D	128E	128F	12800	12810	12820	12830	12840	12850	12860	12870	12880	12890	128A0
																						
																						
																						
																						
																						
																						
																						
																						
																						
																						
																						
																						
																						
																						
																						
																						

Code Chart 4: 12900 - 129FF

294	1295	1296	1297	1298	1299	129A	129B	129C	129D	129E	129F	0 12900										



## Code Chart 5: 12A00 - 12AFF

12A4	12A5	12A6	12A7	12A8	12A9	12AA	12AB	12AC	12AD	12AE	12AF <sup>0</sup>	 12A00									
												12A10	12A20	12A30	12A40	12A50	12A60	12A70	12A80	12A90	12AA0
												