L2/19-284

PROPOSAL TO ENCODE PROTO-CUNEIFORM IN THE SMP OF THE UCS

UPDATED DRAFT | 1-29-2019

Laura Hawkins

I. BACKGROUND

1. Introduction

Cuneiform is perhaps the oldest writing system attested[1], emerging in a region of Mesopotamia that corresponds with the southern part of modern-day Iraq. Proto-cuneiform refers to the most archaic phases of this writing system and is attested on tablets dating to the end of the fourth millennium B.C., in the latter half of what is known as the Uruk Period (ca. 3200-3000 BC). Unlike later cuneiform, the relationship between the script of these early texts and any spoken language is tenuous. The signs that made up the proto-cuneiform writing system were drawn, and in later cuneiform impressed, with the aid of a stylus into the still soft surface of clay tablets. These tablets naturally dried and hardened in the arid and hot climate, and as a result they survived in great numbers. The most archaic version of this script developed into the later cuneiform that became one of the hallmarks of Mesopotamian history and culture. For more about the excavation history and chronology of the texts, the nature of the writing system and its relationship with the Sumerian language, and the content of these texts, see Englund 1998 (15-81).

The German Archaeological Institute unearthed most of the tablets of this early phase of cuneiform history during their excavations at the ancient city of Uruk in lower Babylonia between 1913 and 1967. During the seasons from 1928 until 1976, nearly 5000 tablets and fragments were discovered, forming the basis for a long-term research project dedicated to the decipherment and edition of these texts. The periods in which proto-cuneiform is attested are often referred to as the Uruk periods (e.g., Uruk V, ca. 3500-3350 BCE; Uruk IV, ca. 3350-3200 BCE; Uruk III, ca. 3200–3000 BCE), after the excavation layers to which the tablets have been assigned (all tablets and fragments were found in trash deposits outside of secure stratigraphy, but within or below Uruk III). The tablets from Uruk, however, are not the only proto-cuneiform documents known from this period. Similar tablets have been found in the northern Babylonian sites of Jemdet Nasr, Khafaje, and Tell Uqair. Although the tablets from these sites are smaller in number compared to the Uruk corpus, they have aided in the advancement of our understanding of this early period in cuneiform's history.

The Berlin-based research project *Archaische Texte aus Uruk* (ATU), led by Hans Nissen, is working to publish and understand the large number of archaic texts found at Uruk. Primary contributors to the decipherment of the archaic proto-cuneiform texts include: H. Nissen, A. Falkenstein, R.M. Boehmer, M.W. Green, K.-H. Deller, J. Friberg, R.K. Englund, P. Damerow, J.-P. Grégoire, A. Cavigneaux, and R. Matthews. The proto-cuneiform texts found at Uruk are published in the ATU series.

2. Evolution of the script

The proto-cuneiform script is traditionally thought to contain approximately 800 separate signs, including nearly 100 numerical signs and sign variants. This count comes from the authoritative list of proto-cuneiform signs published in 1987 by M. W. Green and H. Nissen, *Zeichenliste der Archaischen Texte aus Uruk* (ZATU = ATU 2). While new signs and sign forms have been identified since this volume's original publication, ZATU remains the basis for the sign numbering and naming system for

[1] It remains uncertain if the first Egyptian writings were contemporary with, earlier, or later than the Uruk tablets.

this project since it is the most recent complete work on archaic signs. Other published complete and incomplete sign lists include A. Falkenstein's *Archaiche Texte aus Uruk* (ATU 1; 1936), R. K. Englund and J. Nissen's *Die lexikalischen Listen der archaischen Texte aus Uruk* (ATU 3; 1992), and R. K. Englund and J.-P. Grégoire's *The Proto-Cuneiform Texts from Jemdet Nasr* (*Materialien zu den frühen Schriftzeugnissen des Vorderen Orients* 1; 1991). Recently, an updated list of proto-cuneiform signs with digital graphic renderings has been produced by R. K. Englund and is available on the pages of the *Cuneiform Digital Library Initiative* (http://cdli.ucla.edu/). This updated list consists of 1742 non-numerical signs and sign variants, including compound and complex signs, and 95 numerical signs and sign variants. This vastly adds to the 800 signs originally published in ZATU.

3. Numerical systems

In many ways, the numerical systems are the key to understanding the proto-cuneiform texts since the majority are economic and administrative texts that rely on numerical notations for account-keeping. Multiple numerical systems were used in proto-cuneiform texts to count discrete objects (Sexagesimal System, Sexagesimal System S', Bisexagesimal System, Bisexagesmial System B*, GAN_2 System, EN System), to record capacity measures (ŠE System, ŠE' System, ŠE'' System, ŠE* System, DUG_b System, DUG_c System), and to record time (U_4 System). Each system was used to count particular types or categories of objects. Although the numerical systems themselves were distinct, several signs were utilized in multiple systems. Each sign used in a numerical system is designated as such in sign lists by the letter 'N' followed by a number. The figures below are from Englund (2004, 32-33).

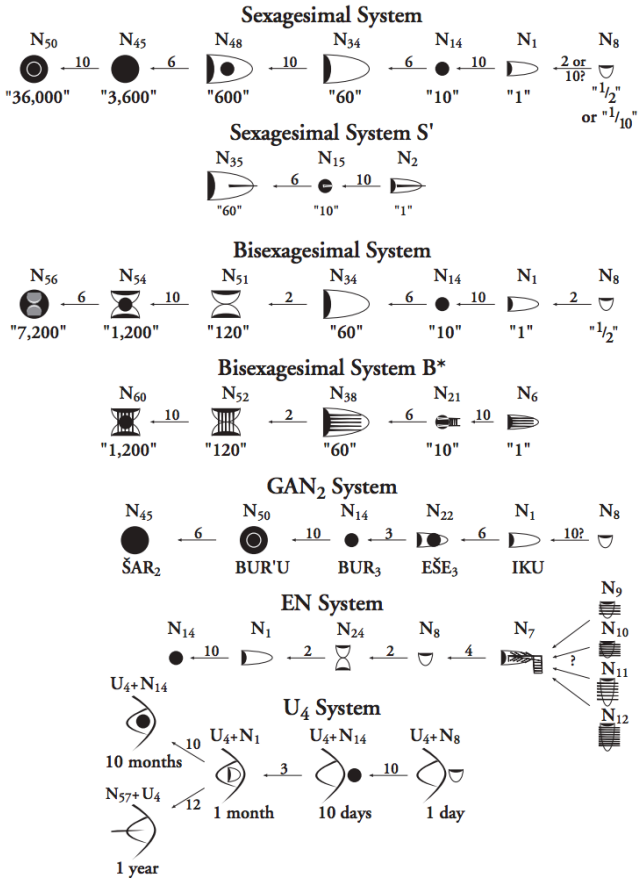

Figure 4a: Proto-cuneiform numerical sign systems.

Several systems of numerical signs served to qualify discrete objects (Fig. 4a), while others qualified measures of grains, (semi-)liquids and time (Fig. 4a and 4b).

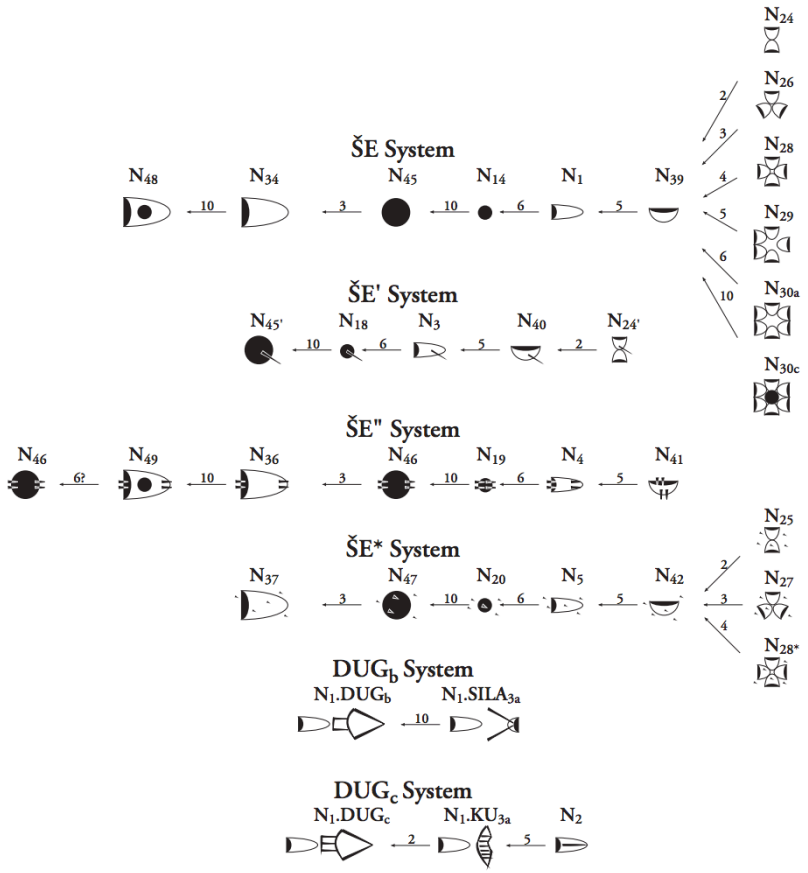

Figure 4b: Proto-cuneiform numerical sign systems

II. METHODOLOGY

1. Basis of the repertoire and methodology

This proposal covers the most commonly occurring 200 Proto-cuneiform lexical and administrative signs that appear within the nearly 4900 texts from the Uruk III period, including Jemdat Nasr. Of these texts, 703 are lexical texts and 4167 are administrative (the genres of the final five texts are catalogued as "legal", "literary", "school", "uncertain", and "votive"; these were not included here). Other characters are not proposed here pending further study; the 200 characters constitute high-frequency characters. This proposal also includes 64 characters with numeric use. There are a total of 1246 signs and sign variants attested in administrative texts and 802 in lexical texts. However, there are many signs that occur only a small number of times, so the top 200 most frequent signs account for 84%[2] of all sign attestations in administrative texts and 84%[3] in lexical texts, meaning a large percentage of texts would be able to be represented using Unicode characters. If the Unicode character set were increased to the top 300 most frequently occurring signs, this would increase the percentage of encodable signs to approximately 91%[4] of sign attestations in administrative and 91%[5] in lexical texts; with the top 400

[2] 27,836 top 200-sign attestations out of a total of 33,177 total sign attestations in administrative texts.
[3] 7771 top 200-sign attestations out of a total of 9260 total sign attestations in lexical texts.
[4] 29,989 out of 33,177 sign attestations.
[5] 8419 out of 9260 sign attestations.

signs encoded, this would increase to 94%[6] and 95%[7] of sign attestations in administrative and lexical texts, respectively.

The signs were pulled from the available texts in the database of the Cuneiform Digital Library Initiative (CDLI). The CDLI is a digital database that catalogues and makes freely available scanned images, hand copies, transliterations, and translations of cuneiform texts, and it currently contains the entire archaic cuneiform corpus of texts in transliteration (and many with images). Although this database represents the most comprehensive corpus of proto-cuneiform texts ever put together, the signs included in these lists were checked against previous published sign lists (Green and Nissen 1987; Englund, Damerow, and Nissen 1993; Englund and Boehmer 1994).

2. Choice of glyphs, names, and ordering

The glyphs encoded here were originally drawn by R. K. Englund.[8] The signs are oriented according to their counterparts' orientation in later cuneiform, as per convention. The proposed names are based on each sign's traditional name, with slight differences in order to abide by the common UCS naming conventions. The sign order is alphabetical. This coincides with the order in the published sign list of Green and Nissen (1987).

Although proto-cuneiform is largely understood, there are some signs that remain undeciphered. Often these signs are those that occur very infrequently within the corpus, and it is sometimes uncertain whether these signs are simply graphical variants of other, better attested signs or semantically unique. For this reason, not all signs were chosen to be initially encoded. However, this could easily be expanded to include all signs and sign variants or those that occur more than once or more than twice.

3. Sign variants, complex and compound signs, and duplicate signs

Proto-cuneiform and later cuneiform scripts used several methods to expand their sign repertoire. This included the use of crosshatchings and other additional marks added to a basic glyph. Because in many of these cases, if not all, the additional markings denote a difference in semantic value, each basic and modified glyph should be encoded separately.

All forms of the proto-cuneiform and later cuneiform scripts combine signs in regular ways to create compound or complex meanings. Compound signs consist of two or more glyphs written adjacently, where the two basic glyphs share one or more strokes, one is written inside another, or one sign is written on top of another. Complex signs consist of two or more glyphs written adjacently that do not share strokes, but that occur so frequently together that they are often thought of as a singular unit and whose meaning is altered by their coinciding. In English, an equivalency could be 'grandmother' and 'grand mother,' which consist of the two same basic words but when combined or separated having very different meanings. Compound signs should be encoded separately, while complex signs should not. In proto-cuneiform, examples of compound signs include NANNA (URI_3+NA, the moon god Nanna), SUSA ($MUŠ_3$+ERIN, the city Susa), and DUR_2 (=LAGABxRUM).

[6] 31,177 out of 33,177 sign attestations.
[7] 8758 out of 9260 sign attestations.
[8] Proto-cuneiform signs in postscript format (CS 4), http://cdli.ucla.edu/?q=downloads.

Numerical glyphs that occur in more than one numerical system should only be encoded once. This keeps with sign naming tradition in the field of Assyriology, where each numerical glyph is given a single name (e.g., N_1 or N_{45}) and that name is used regardless of which system the sign is being used in or its comparative value within each system. For example, the ratio between N_1 and N_{14} in the sexagesimal, bisexagesimal, and EN systems is 1:10, while in the ŠE system the ratio is 1:6. In another example, N_{14} always represents '10', however in the EN system this is an area measure and in the sexagesimal and bisexagesimal systems it's an absolute number.

4. Scholarly use of encoded proto-cuneiform

Electronic renderings of these signs and sign variants have been created to a very high level of detail, which are useful to scholars in the creation of figures and tables for publication. These electronic glyphs are less useful for encoding a single sign or string of signs within running text. For example, see the following text from Nissen, Damerow, and Englund (1994, 32):

> Text 29b is somewhat more explicit. In its first column it has an entry including the sign  (NINDA), again designating cereal rations, plus a second entry, this time with the sign  (DUG_a) pertaining to beer rations. In the second column an individual is

Encoding proto-cuneiform would also make certain computational analyses easier and more replicable. For example, a consistent numbering system would eliminate variation in rendering diacritics that currently is present in electronic versions of sign names: for example, the CDLI uses *sz* to represent *š* while the Electronic Pennsylvania Sumerian Dictionary (ePSD) uses variably *sz* and *c* to represent *š*. This makes it necessary to manually clean electronic transliterations if a researcher wanted to combine electronic transliterations from multiple online databases. Finally, the encoding of these signs would also allow students and scholars to read and reproduce texts more easily in glyph form rather than in transliteration, which is a common practice now when texts are not adequately published.

Bibliography and Further Reading

Bauer, J., Englund, R. K. and Krebernik, M. (1998). *Mesopotamien: Späturuk-Zeit und Frühdynastische Zeit.* Göttingen: Vandenhoeck & Ruprecht.

Cavigneaux, A. (1991). "Die Texte der 33. Kampagne." *BaM* 22:74.

Damerow, P. (2006). "The origins of writing as a problem of historical epistemology." *Cuneiform Digital Library Journal*, 1, 1-10.

Damerow, P. and R. K. Englund (1987). "Die Zahlzeichensysteme der Archaischen Texte aus Uruk," in M. Green and H. Nissen, eds., *Zeichenliste der Archaischen Texte aus Uruk* (ATU 2): 117-66.

Damerow, P. and Meinzer, H.-P. (1995). "Computertomografische Untersuchung ungeöffneter archaischer Tonkugeln aus Uruk: W 20987,9, W 20987,11 und W 20987,12," *Baghdader Mitteilungen* 26, 7-33, pls. 31-34.

Englund, R. K. (1988). "Administrative Timekeeping in Ancient Mesopotamia," *Journal of the Economic and Social History of the Orient* 31, 121-185.

-- (1998). "Texts from the Late Uruk Period," in J. Bauer, R. K. Englund, and M. Krebernik, eds., *Mesopotamien: Späturuk-Zeit und Frühdynastische Zeit* (OBO 160/1): 51-56.

-- (2001). "Grain Accounting Practices in Archaic Mesopotamia," in J. Høyrup and P. Damerow, eds., *Changing Views on Ancient Near Eastern Mathematics* (BBVO 19): 1-35.

-- (2004). "Proto-Cuneiform Account-Books and Journals," in Michael Hudson and Cornelia Wunsch, eds., *Creating Economic Order: Record-keeping, Standardization and the Development of Accounting in the Ancient Near East* (CDL Press: Bethesda, Maryland, USA) pp. 23-46.

Englund, R. K. and Grégoire, J.-P.. (1991). *The Proto-Cuneiform Texts from Jemdet Nasr*. Berlin: Gebrüder Mann.

Englund, R. K. and H. Nissen (1993). *Die lexikalischen Listen der archaischen Texte aus Uruk* (ATU 3).

-- (1994). *Archaic Administrative Texts from Uruk: The Early Campaigns. Archaische Texte aus Uruk 5.* Berlin: Gebrüder Mann.

-- (2001). *Archaische Verwaltungstexte aus Uruk: Die Heidelberger Sammlung* (ATU 7).

Green, M. W. and Nissen, H. J. (1987). Zei*chenliste der Archaischen Texte aus Uruk. Archaische Texte aus Uruk 2.* Berlin: Gebrüder Mann.

Friberg, J. (1978-79). *The Early Roots of Babylonian Mathematics: II. Metrological Relations in a Group of Semi-Pictographic Tablets of the Jemdet Nasr Type, Probably from Uruk-Warka* (= Research Report, 1979-15; University of Göteberg. Department of Mathematics, Chalmers).

Nissen, H., P. Damerow, and R. K. Englund (1993). *Archaic Bookkeeping* (Chicago: University of Chicago Press).

Schmandt-Besserat, D. (1992). *Before Writing*. Austin: University of Texas Press.