WRITING IN EARLY MESOPOTAMIA PROJECT

Christopher Woods and Massimo Maiocchi

Overview. *Christopher Woods*

The Writing in Early Mesopotamia (WEM) project endeavors to provide a comprehensive description of how the technology of cuneiform writing represented language. The project will investigate early cuneiform writing from the perspective of both language — how sound and meaning are systematically expressed diachronically and synchronically — and semiotics — the graphic organization and history of the symbols that comprise the system. The scope of the project is the cuneiform written record from the invention of writing in the late fourth millennium BC (ca. 3300 BC) through the Old Babylonian period (ca. 1600 BC). While Sumerian writing is at the center of the project — Sumerian being in all likelihood the language for which writing was invented in Mesopotamia — the adaptation of the script to express Semitic (Akkadian and Eblaite) and the long-term interplay between these writing systems are major concerns.

The following brief report introduces this new Oriental Institute endeavor, initiated the past academic year and supported generously by the Andrew W. Mellon Foundation, the Linda Noe Laine Foundation, and the Oriental Institute's Erica Reiner Fund. Below, we provide an overview of the WEM project and conclude with a description of our database work, which has been our primary occupation for the 2012–13 academic year. The project is currently staffed by Christopher Woods and Massimo Maiocchi (Mellon Postdoctoral Fellow), Howard Farber and Steven George provided part-time assistance, and University of Chicago undergraduate and Metcalf Fellow Abigail Hoskins worked full-time on the project during the summer of 2013. Beginning this upcoming academic year (2013–14), the WEM project will collaborate closely with another new initiative at the University of Chicago concerned with early writing, the Neubauer Collegium's Signs of Writing the Cultural, Social, and Linguistic Contexts of the World's First Writing Systems (described under Christopher Woods's individual research [see separate report]).

1. Rationale

The motivation for this project lies in the fact that to date there has been no rigorous study of cuneiform writing, let alone one that is linguistically oriented, or informed by typological comparisons or interdisciplinary perspectives. In the case of Sumerian, for which our understanding of the writing system is marred by lacunae and misconceptions, this absence is particularly conspicuous. After nearly a century and a half of the decipherment effort, writing remains largely *terra incognita* — a fundamental domain that has yet to be subjected to rigorous investigation. For the philologist, the written record is naturally the point of access for all inquiry, and so the study of the medium that conveys the message stands as a desideratum in its own right. Further, writing and its associated problems are bound up with the higher-level concerns of phonology and morphology. Although not often acknowledged by scholars working on Sumerian grammar, the absence of a coherent theory of Sumerian writing is arguably the greatest obstacle to describing the language. The WEM project will have wide implications for the field of Assyriology, as it will not only provide the basis upon which grammatologically informed theories of Sumerian grammar and lexicography may be built, but will add substantially to our knowledge of early scribal practices and training.

More broadly, writing — often hailed as humankind's greatest technological and cultural achievement — is a fundamental and perennial subject of scholarly inquiry. A comprehensive description of Sumerian writing will make an essential contribution to our understanding of early writing systems and the cognitive processes by which humans first made language visible. Taking its place alongside the nearly contemporaneous Egyptian invention (ca. 3200 BC), as well as the Chinese (ca. 1200 BC) and Mayan (ca. 400 BC) inventions, cuneiform is one of the four "pristine writing systems," the four original writing systems from which all others, directly or indirectly, developed. Cuneiform not only boasts what might very well be the world's oldest writing system, but also stakes claim to the largest corpus of incipient writing and some of the clearest evidence for the cultural and social context out of which writing sprung. As such, cuneiform plays a pivotal role in the story of how humans first made language visible and for the study of writing systems broadly. The WEM project will be of interest not only to cuneiform specialists, but also to those interested in writing systems, literacy, the psycho-linguistics of reading, and cognitive representation more generally.

2. Approach and Goals

At the core of the project are extensive databases of spellings that are meaningful to the description of the writing system. The collection of this data is an endeavor that requires the systematic review, categorization, and analysis of thousands of texts written between the invention of writing and the end of the Old Babylonian period. Of particular interest are syllabaries compiled by the ancients, the temporal and spatial distribution of logographic and syllabic writings, as well as variant spellings, that is, those writings that do not conform to the standard orthography. Writings of these types are particularly revealing of the mechanics and organization of the system as well as the relationship between the written and spoken languages. Differentiating variants that are phonologically or morphologically significant from those that are merely orthographic is a major concern of inquiries of this kind. Taking a typological and interdisciplinary approach, the WEM project considers individual orthographic phenomena and the cuneiform writing system as a whole within the context of writing systems research more generally and of the other early "pristine" writing systems specifically. The particular interest in the Egyptian, Chinese, and Mesoamerican systems lies in the fact that they exhibit striking similarities to — and some notable differences from — Sumerian writing. As an intrinsically interdisciplinary endeavor, the WEM project also draws upon fields allied to the study of writing systems, including linguistics, semiotics, the philosophy of language, as well as information science and cognitive psychology. Considering cuneiform writing from multiple perspectives allows for insights and a broadly based understanding of the system, which would not be possible adhering only to a philological approach.

The end products of the Writing in Early Mesopotamia project will be two-fold, consisting of synthetic print publications on one hand, and, on the other, electronic research tools upon which the former will rely. The ultimate goal of the project will be a comprehensive description of early cuneiform writing in monograph form. Qiu Xigui's monumental *Chinese Writing* (tr. G. L. Mattos and J. Norman, Berkeley, 2000) provides a suitable model for the publication envisioned. In conjunction with the Neubauer Collegium's Signs of Writing endeavor, the WEM project will host short- and long-term visiting scholars as well as three annual interdisciplinary workshops (successively on the University of Chicago campus in 2014, and the University of Chicago centers in Beijing [2015], and Paris [2016]) on topics seminal to the project and writing systems broadly. Project members, collaborators, and other invited speakers will participate in these workshops; the Oriental Institute will publish a selection of the proceedings at the conclusion of the series. The topics covered by these workshops will range from the more technical, such as the role of logography in pristine writing systems, to the more culturally oriented, such as the interaction between bilingualism and writing in antiquity. As described above, the synthetic, published output of the project will rest upon the collection and organization of large data sets that describe the temporal and spatial attestations of written forms and variant spellings (see further below). At the core of this endeavor are the indigenous syllabaries that preserve native scribal understanding of graph-to-sound correlations, and exemplars of the Sumerian literary corpus of the Old Babylonian period, which represent the largest corpus of textual variants. As these sources have broad Assyriological importance beyond their value for writing, the project will make this material available on-line in the form of open-access, searchable databases.

3. Collaborators

Integral to the project is a network of collaborators and partners with expertise in the fields and writing systems that are of central importance. These members include at this time: Cuneiform — Paul Delnero (Johns Hopkins University), Piotr Steinkeller (Harvard University); Chinese — Wolfgang Behr (University of Zurich), Edward Shaughnessy (University of Chicago); Egyptian — Janet Johnson (University of Chicago), Andréas Stauder (University of Basel); Writing Systems and Linguistics — Richard Sproat (Google Labs).

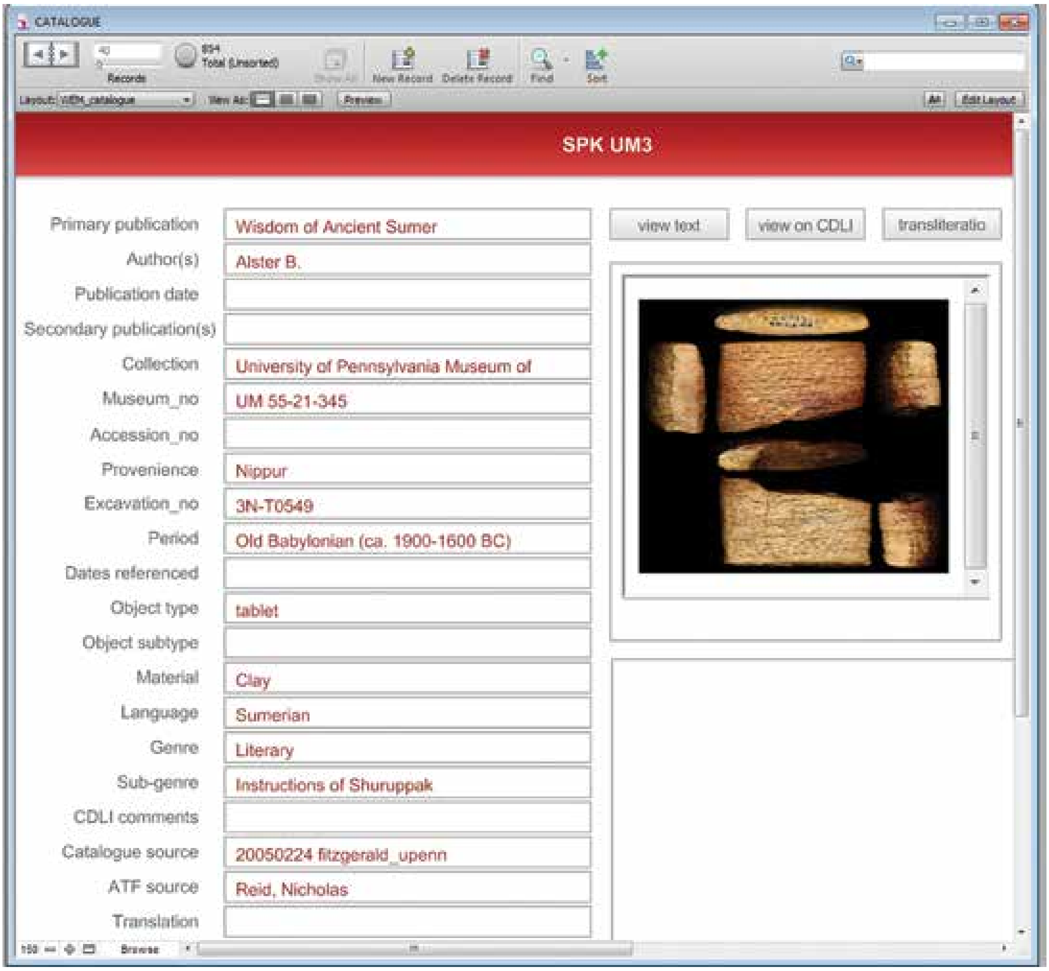
4. 2012–2013 Progress

Our work this year has included compiling a library and bibliography of writing systems research. Our primary focus, however, has been on developing the aforementioned database, which is central to the project, as well as a theoretically suitable and practical method of morphological parsing, which would allow for advanced database queries. In what follows, M. Maiocchi describes our efforts with the database, some of the issues encountered, and some of the protocols developed, with regard to the encoding of text.

4.1. Database Description. *Massimo Maiocchi*

4.1.1. Identification of Textual Corpus

Our first step in creating a database for the analysis of cuneiform writing was to identify an initial test corpus of suitable texts, which is a relatively large collection of tablets that present significant variations in morphology and lexicography. Once the database and search algorithms are sufficiently tested with these initial texts, we will expand the range of documents included in the corpus. The ten Sumerian literary compositions known as the Decad — these were the first ten literary texts apprentice scribes would encounter in the scribal



SPK UM3

Primary publication	Wisdom of Ancient Sumer
Author(s)	Alster B.
Publication date	
Secondary publication(s)	
Collection	University of Pennsylvania Museum of
Museum_no	UM 55-21-345
Accession_no	
Provenience	Nippur
Excavation_no	3N-T0549
Period	Old Babylonian (ca. 1900-1600 BC)
Dates referenced	
Object type	tablet
Object subtype	
Material	Clay
Language	Sumerian
Genre	Literary
Sub-genre	Instructions of Shuruppak
CDLI comments	
Catalogue source	20050224 fitzgerald_upenn
ATF source	Reid, Nicholas
Translation	

Figure 1. WEM catalog sample

curriculum —seemed an obvious choice, on account of the large number of exemplars preserved and the existence of recent and comprehensive editions compiled by Paul Delnero (Johns Hopkins). That these texts are essentially complete, well attested in multiple copies, and are well understood grammatically are further advantages. Additionally, we included the wisdom text “Instruction of Shuruppak” in this initial group. The long history of this text in the scribal curriculum, spanning from the Early Dynastic III through the Old Babylonian periods (ca. 2600–1600 BC) provides an ideal test case for encoding of diachronic variation.

The relevant information for each text has been entered in a catalog, in which the individual tablets receive an identification number (primary key), as well as bibliographical reference concerning editions, secondary literature, museum number, provenience, dating, physical condition of the text, collection, photos, and copies where available. At present, the catalog of the WEM project includes 853 texts and text fragments, a figure large enough for proper statistical analysis.

From a typological point of view, the test corpus currently contains:

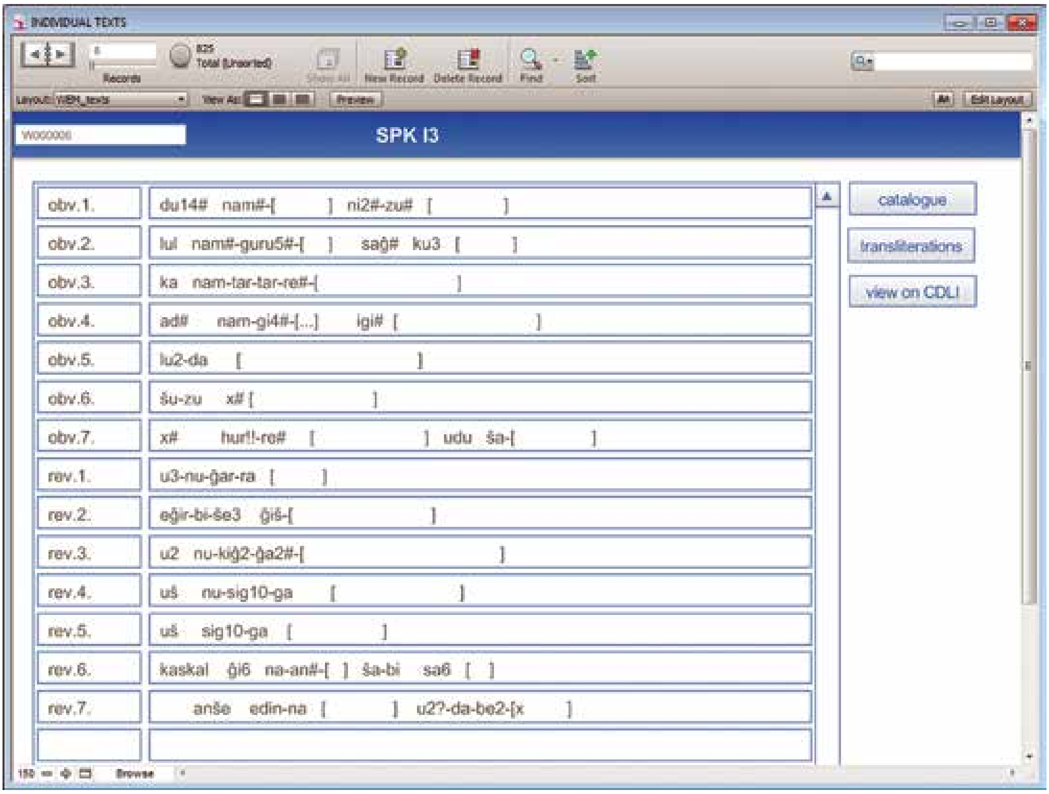
- narrative mythological compositions featuring deities (Enki’s journey to Nippur, Inana and Ebih) and heroes (Gilgamesh and Huwawa A);
- royal praise poetry (Shulgi A, Lipit-Ishtar A);

- wisdom literature (Instruction of Shuruppak);
- hymns to deities (Inana B, Enlil A, Nungal A) and temples (Kesh temple hymn);
- scribal training literature (Song of the Hoe).

Care has been taken to assure that our database is easily linkable to other major Assyriological projects, such as the Cuneiform Digital Library Initiative (CDLI), the largest repository of cuneiform documents in digital format available on line; the WEM catalog includes references to the CDLI identification numbers for individual texts.

4.1.2. Normalization of Readings and Coherence Check

As many of the transliterations used for our project stem from various and sundry sources, and so exhibit a wide range of incompatible transliteration conventions, it is necessary to normalize all textual witnesses prior to encoding. This process also includes proper Unicode rendering of the texts (some sources use special non-Unicode fonts to represent special characters commonly used in standard Assyriological transliterations). In addition, the readings of individual signs have been altered to conform to WEM conventions, which include broad transcriptions and the use of long values (so for instance the reading šag$_4$ is preferred to ša$_3$ "heart"). Finally, minor collations to the texts are made, and improved descriptions of the textual breaks found in the documenation are given, during the course of encoding.

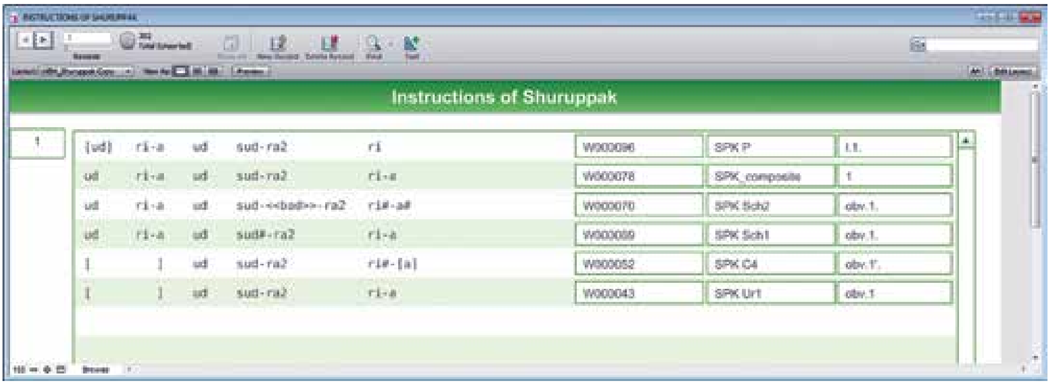


obv.1.	du14# nam#-[ ] ni2#-zu# [ ]
obv.2.	lul nam#-guru5#-[ ] sağ# ku3 [ ]
obv.3.	ka nam-tar-tar-re#-[ ]
obv.4.	ad# nam-gi4#-[...] igi# [ ]
obv.5.	lu2-da [ ]
obv.6.	šu-zu x# [ ]
obv.7.	x# hur!!-re# [ ] udu ša-[ ]
rev.1.	u3-nu-ĝar-ra [ ]
rev.2.	eĝir-bi-še3 ĝiš-[ ]
rev.3.	u2 nu-kiĝ2-ĝa2#-[ ]
rev.4.	uš nu-sig10-ga [ ]
rev.5.	uš sig10-ga [ ]
rev.6.	kaskal ĝi6 na-an#-[ ] ša-bi sa6 [ ]
rev.7.	anše edin-na [ ] u2?-da-be2-[x]

Figure 2. WEM text transliteration

4.1.3. Development of Digital Tools to Process the Texts

To accelerate the process of encoding, as well as to avoid manual mistakes during the delicate step of collecting textual variants, a suite of Perl scripts have been developed. These are small programs, written in a programming language, that makes easy the development of computational methods applied to textual analysis. They take properly formatted transliterations as input, recognize individual words, and search for variants. In order to achieve this goal, the transliterations of the individual exemplars have been formatted and aligned into columns. A side benefit of this process is that the aligned presentation of textual witnesses facilitates the identification of variants at a glance.



Instructions of Shuruppak

1

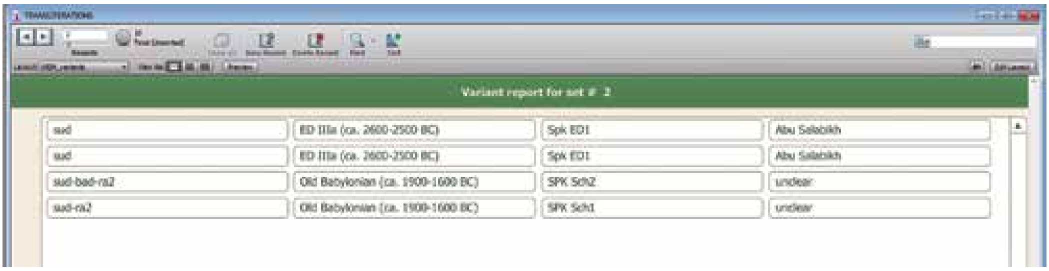
[ud]	ri-a	ud	sud-ra2	ri	W000096	SPK P	I.1.
ud	ri-a	ud	sud-ra2	ri-a	W000078	SPK_composite	1
ud	ri-a	ud	sud-<<bad>>-ra2	ri#-a#	W000070	SPK Sch2	obv.1.
ud	ri-a	ud	sud#-ra2	ri-a	W000059	SPK Sch1	obv.1.
[	]	ud	sud-ra2	ri#-[a]	W000052	SPK C4	obv.1'.
[	]	ud	sud-ra2	ri-a	W000043	SPK Ur1	obv.1

Figure 3. WEM composite view

Specifically, the Perl script parses well-formed transliterations and creates a network of relationships between the individual words of the texts, which are grouped together in bundles of variants. Every lexeme is consequently given a unique identification number, and is linked to all other attestations of the same item. Lexemes can be browsed in the variant report layout, which makes it easy to identify diachronic and regional variations.

4.1.4. Morphological Encoding

Subsequently, individual morphemes are encoded to account for allomorphs. The encoding itself is made on a tabular format related to the one containing the individual lexemes and composed by "unique items" in order to avoid repetitive encoding of identical lexemes in the same context. For instance, in figure 3 one can identify that the word sud-ra₂ occurs five times in the same "column." Instead of encoding each instance separately, it is sufficient to do so only once, the information entered being automatically transferred to the other attes-



Variant report for set # 2

sud	ED IIIa (ca. 2600-2500 BC)	Spk ED1	Abu Salabikh
sud	ED IIIa (ca. 2600-2500 BC)	Spk ED1	Abu Salabikh
sud-bad-ra2	Old Babylonian (ca. 1900-1600 BC)	SPK Sch2	unclear
sud-ra2	Old Babylonian (ca. 1900-1600 BC)	SPK Sch1	unclear

Figure 4. WEM variant report

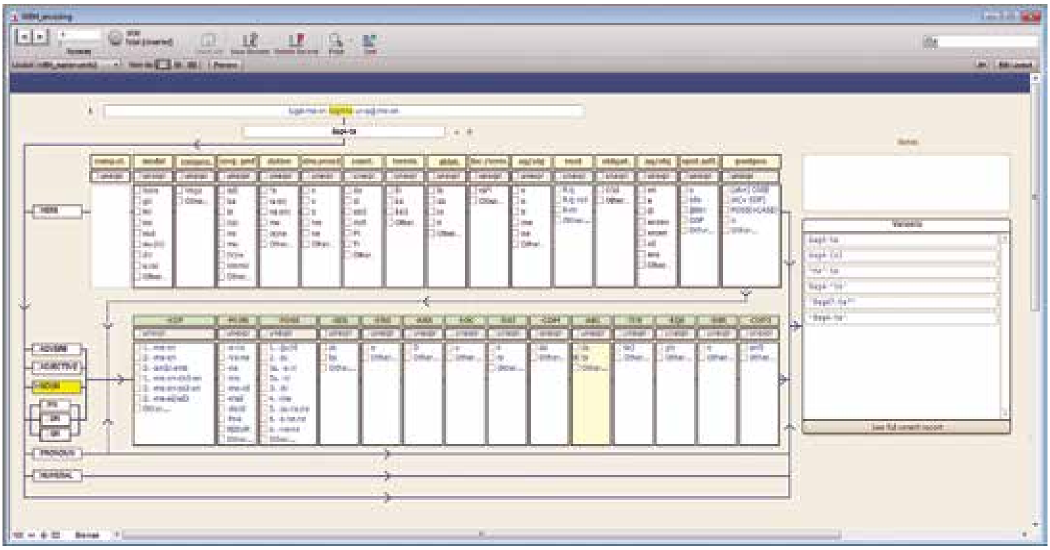


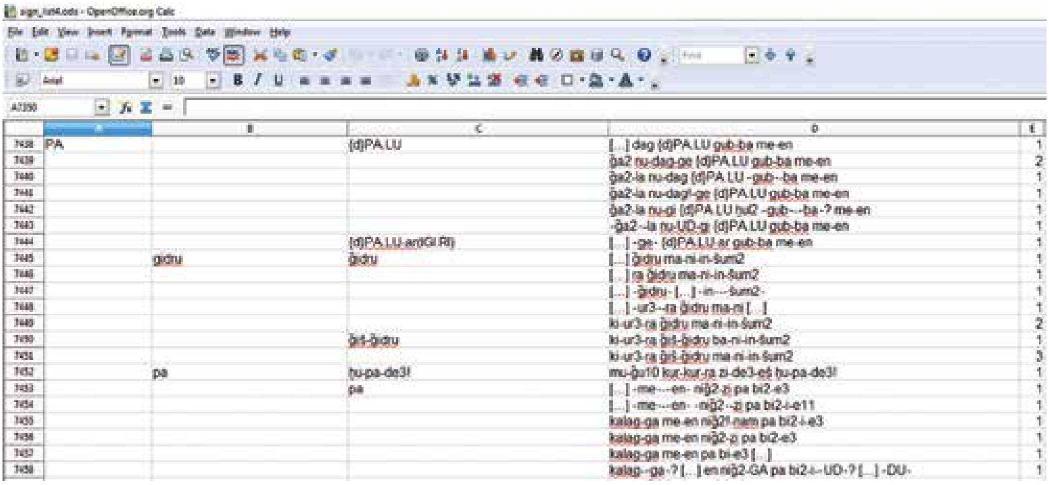
Figure 5. WEM encoding

tations. This reduces by roughly a tenth the amount of work to be done, since the encoding process is manual.

A check-box system reduces the time needed for encoding and minimizes errors introduced by typos. The parsing is broad and avoids controversial aspects of Sumerian verbal morphology. In this way we minimize the potential for forcing interpretation and pre-judging the very grammatical phenomena the database seeks to investigate.

4.1.5. Graphemic Analysis

In the last step of the process every morpheme is analyzed by another Perl script that produces a sign list and syllabary for the individual compositions, and so allows for a statistical analysis based on the distribution of signs and sign values. The list is graphemic in that compound signs are split, when possible, into their individual units. For instance, the word


sign_list4.ods - OpenOffice.org Calc
File Edit View Insert Format Tools Data Window Help
Arial
10
A7350

	A	B	C	D	E
7438	PA		(d)PA.LU	[...] dag (d)PA.LU gub-ba me-en	1
7439				ĝa2 nu-dag-ge (d)PA.LU gub-ba me-en	2
7440				ĝa2-la nu-dag (d)PA.LU -gub--ba me-en	1
7441				ĝa2-la nu-dag!-ge (d)PA.LU gub-ba me-en	1
7442				ĝa2-la nu-gi (d)PA.LU hul2 -gub---ba -? me-en	1
7443				-ĝa2--la nu-UD-gi (d)PA.LU gub-ba me-en	1
7444			(d)PA.LU-ar(GI.RI)	[...] -ge- (d)PA.LU-ar gub-ba me-en	1
7445		ĝidru	ĝidru	[...] ĝidru ma-ni-in-šum2	1
7446				[...] ra ĝidru ma-ni-in-šum2	1
7447				[...] -ĝidru- [...] -in---šum2-	1
7448				[...] -ur3--ra ĝidru ma-ni [...]	1
7449				ki-ur3-ra ĝidru ma-ni-in-šum2	2
7450			ĝiš-ĝidru	ki-ur3-ra ĝiš-ĝidru ba-ni-in-šum2	1
7451				ki-ur3-ra ĝiš-ĝidru ma-ni-in-šum2	3
7452		pa	hu-pa-de3!	mu-ĝu10 kur-kur-ra zi-de3-eš hu-pa-de3!	1
7453			pa	[...] -me---en- niĝ2-zi pa bi2-e3	1
7454				[...] -me---en- -niĝ2--zi pa bi2-i-e11	1
7455				kalag-ga me-en niĝ2!-nam pa bi2-i-e3	1
7456				kalag-ga me-en niĝ2-zi pa bi2-e3	1
7457				kalag-ga me-en pa bi-e3 [...]	1
7458				kalag--ga -? [...] en niĝ2-GA pa bi2-i--UD-? [...] -DU-	1

Figure 6. Graphemic sign list and syllabary

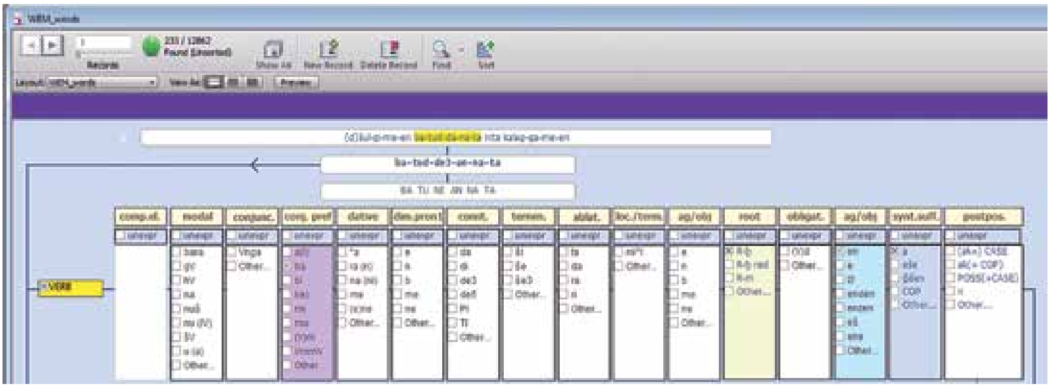


Figure 7. Detail of the encoding of a composite line (first line), the selected word (second line), and the sequence of signs associated in the surrounding context (third line)

for shepherd, sipad in Sumerian, is considered as the union of the PA and UDU signs, and is listed under both. The tool can be fine-tuned to work at the most basic level of sign composition, reaching into the realm of palaeography.

This script also turns standard Assyriological transliterations into sequences of sign names. This is particularly useful for searching the database for signs regardless of their readings in context.

When completed, the database will provide a unique tool for the analysis of writing and grammar, allowing for unprecedented search capabilities on both graphic and morphological levels. It will be possible, for instance, to query the database not only for the presence, but also absence of grammatical morphemes (as for instance in the case of the syntactic cross-referencing between nominal post-positions and verbal infixes). We will also be able to quantitatively describe the relative distribution of morphological elements and generate statistics relating to the frequency of variant spellings, as well as the distribution of signs and sign combinations as a function of period and region.