WRITING IN EARLY MESOPOTAMIA

Christopher Woods and Massimo Maiocchi

Introduction

At the root of the Writing in Early Mesopotamia project is a database that endeavors to analyze the development of the cuneiform writing system and its relation to speech. The database provides extended search capabilities, at both word and sign levels, which enables users to search for the presence or absence of grammatical morphemes and their distribution in relation to other morphological elements. The database also facilitates the analysis of the frequency and distribution of textual variants in the epigraphical record through time and space. The database is housed on a FileMaker platform and is manipulated with a series of Perl scripts.

During the last year, the database has been further expanded in terms of both text encoding and additional features. The core of the database remains the encoding of the compositions known as the "Decade." These are ten literary texts that were copied as part of elementary scribal curriculum. On account of this, each composition is typically known from numerous exemplars, each usually providing interesting variants in terms of linguistic analysis. In addition, the database features the composition known as the Instructions of Shuruppak, which offers a substantial array of diachronic variants, from both the third and the second millennia BC. As of September 2015, we finished the complete encoding of the composition Shulgi A, which is now searchable in terms of its morphographemic features, matrixes (see below), and sign distribution (statistics on the frequency of logograms, syllabograms, and determinatives). Two other compositions, namely Gilgamesh and Huwawa A and the above-mentioned Instructions of Shuruppak, are presently being finalized. In its actual state, the database is articulated in several sections, all designed to achieve a better understanding of the strategies adopted by ancient scribes to write down the Sumerian language. Specifically, the database has the following structure: 1) a catalog of all texts (including links where applicable to CDLI, i.e., the largest online repository of cuneiform texts); 2) a transliteration table (cumulative table of all texts in the database); 3) a composite transliteration layout (providing transliterations arranged by compositions); 4) an encoding layout; 5) a search layout; 6) matrixes (used to provide a view at a glance of the variants; see below); 6) a table of signs; 7) a graphemic list (used to display what readings are attached to the individual signs with what frequency); 8) a phonographical variation layout (based on lexical texts, meant to spot variations in phonemes, such as m/n or b/g); 9) a syllabographical variation layout (again based on lexical texts, meant to spot variations in syllabic clusters, for instance a CVC sign, where C = consonant, V = vowel, alternating with a CV-VC sequence). When taken together, the various sections provide great flexibility and diverse access to data, thus promoting research from different complementary perspectives.

Migration to FileMaker Server

During this year, we finalized the migration to the new FileMaker Server platform. The software is now running on a dedicated server hosted at the Oriental Institute. Despite some minor bugs, the new platform has proved to be solid. Most importantly, the database can now accommodate multiple users simultaneously altering records belonging to a given table. This feature has solved the issue of coordinating access for multiple users encoding or adding data within the database. In addition, the database now features different levels of administrative privileges, in order to better deal with different levels of access, from guest access to administrator. The beta version of the database, including a selection of compositions, is expected to be freely available online in the near future.

Color Encoding

In order to increase readability, a system of color encoding was implemented (see fig. 1). Accordingly, strings of text belonging to the composite of the individual compositions are marked in orange. In addition, variant lines are now marked with a yellow background. These small improvements vastly facilitate the navigation within the composite view of the manuscripts. Minor bugs in the calculation field used to automatically apply the color coding have been fixed.

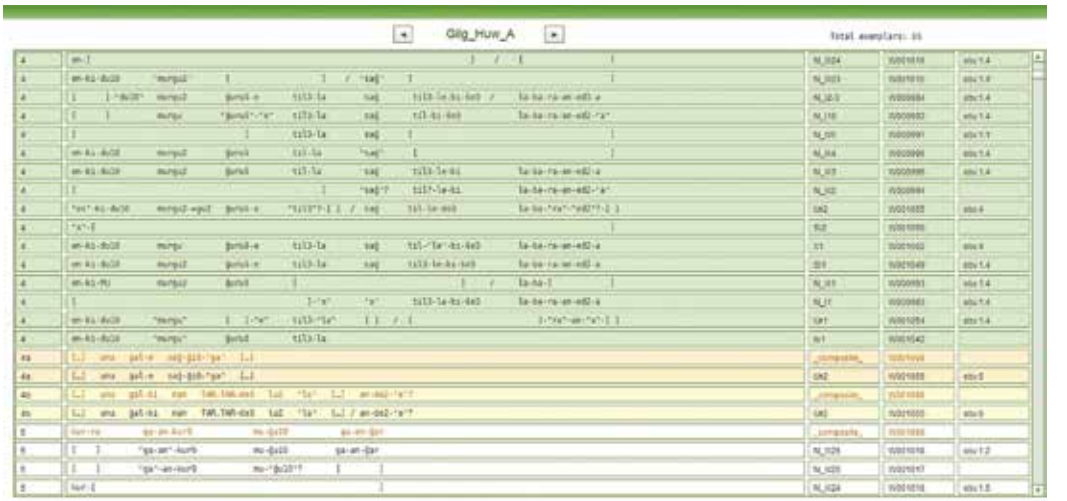

Figure 1. Composite view of the composition Gilgamesh and Huwawa A, with color encoding

Matrixes

The term matrix is used here to refer to a conventional way of representing score transliterations of a given composition in a compact way (see fig. 2). The power of this type of data visualization lies in the fact that the users can achieve an understanding at a glance of the complex variation schemes within the individual compositions. This feature complements the data that the users can extract from the database by means of advanced textual research, at the level of words, or at the level of encoding — or both. Whereas regular queries imply

an interactive exploration of data, producing a list of results that involves active analysis, matrixes are predominantly passive in nature, in the sense that they are meant to be quickly scrolled in order to spot interesting variation patterns. However, users can also perform quick searches on matrixes, to jump to a given segment of special interest. As for the specific encoding used in matrixes, we relied on the conventions introduced by M. Civil to represent variant signs or sequences of signs. Morphographemic writings are especially interesting for the scope of this project, as they may shed some light on both the language and the conventions used by ancient scribes to represent it. These can be easily spotted by simply scrolling through the text, and maintaining the context in which they occur. As of September 2015, the scores of the compositions known as Shulgi A, Gilgamesh and Huwawa A, as well as of the Instructions of Shuruppak are available in the database.

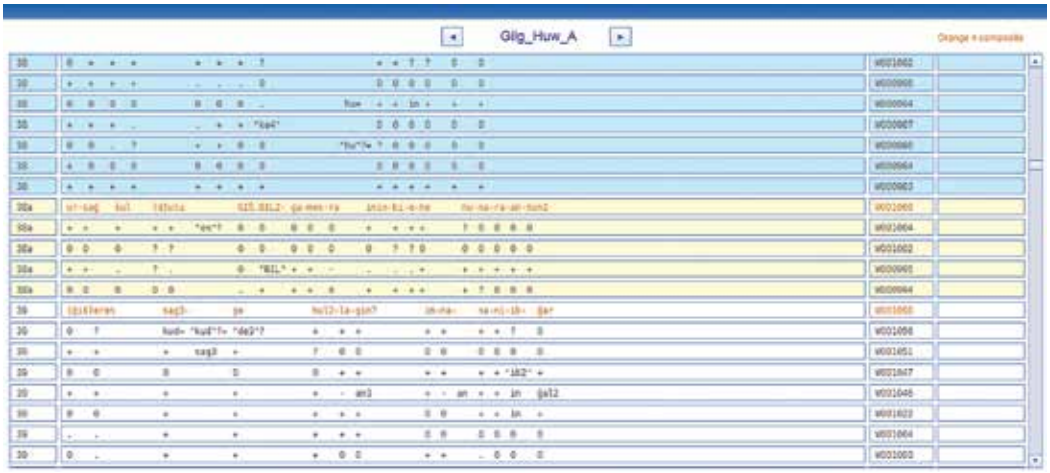

Figure 2. Matrix view of the composition Gilgamesh and Huwawa A, with color encoding

Encoding of Signs by Relative Typology

The encoding of the Instructions of Shuruppak by the relative typology of signs is currently in progress. When finished, it will provide a minimal contrastive dataset to evaluate the distribution of logograms and syllabograms in compositions, such as Shulgi A, the encoding of which is otherwise complete (see fig. 3). We maintained the conventions introduced last year, namely we distinguish six broad categories of signs: 1) signs that stand for words (logograms); 2) signs that stand for syllables (syllabograms); 3) signs that are not meant to be read, but provide semantic information (determinatives); 4) signs standing for personal names, or part of personal names (this category is motivated by the fact that personal names tend to show peculiar features that are otherwise missing in spelling); 5) unclear (when a sign is physically present on a tablet, but its reading is difficult); 6) unknown (used for broken signs). The purpose of encoding texts according to the relative typology of signs is to detect patterns in the overall use of logography within different compositions or textual corpora, thus improving our understanding of how this structural feature of the writing system evolved through time and space.

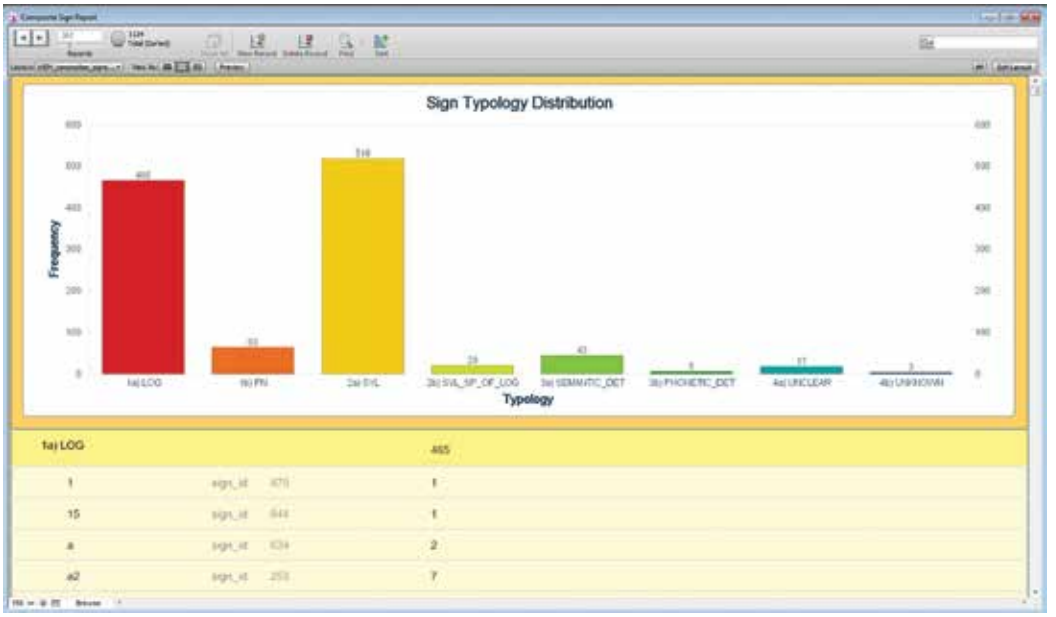

Figure 3. Chart showing the distribution of logograms and syllabograms in the composition Shulgi A

New Texts

As the publication of new cuneiform texts continues apace, each year brings new exemplars of the compositions to be included in our database. This year we added nine new exemplars (including fragments) to our database. One of the new texts, belonging to the composition Gilgamesh and Huwawa A, is particularly interesting for the history of the composition, as well as for the reconstruction of the composite. The latter is either based on the most commonly attested variant, or on the variant that is expected on the basis of our understanding of the underlying grammar. The addition of new texts to the database may fill critical lacunae in a given composition, lead to the reconsideration of difficult passages, or may attest important new variants. In this way, the database may serve as an up-to-date, online score of all the textual witnesses for a given composition.