ETCSL
– *The Electronic Text Corpus of Sumerian Literature*

http://etcsl.orinst.ox.ac.uk/

𒂍 𒈨𒌶 𒆥𒁕 𒄑𒐲

Maria de Fátima Rosa

The *Electronic Text Corpus of Sumerian Literature* (ETCSL) is a project of the University of Oxford designed to create a coherent, comprehensive and universally available textual corpus of Sumerian classical literature. The original project team consisted of Jeremy Black (who was also the project director until the time of his death, in 2004), Graham Cunningham, Gábor Zólyomi and Eleanor Robson. The team's goal was to develop an electronic corpus of transliterated and translated compositions that could be searched, browsed and read online. Unfortunately, funding for the project ended in 2006, and the website has not been edited since then.

ETCSL's main page, besides the browsing and searching options, contains four menus with information on how to use the tools available on the website and lots of information about the language and the writing system in which the texts were originally composed. As we know, Sumerian literature is the oldest known to mankind. However, given the fact that the texts were inscribed in clay tablets, often found damaged or broken, it is not always easy to decipher and translate them. Besides, Sumerian is a language for which there is no counterpart. Knowledge of the language has, nevertheless, progressed significantly over the years. Many translations published in the last century are now out-of-date and some can only be found in scattered publications. This was certainly one of the primary motives behind the creation of this project – the need to make Sumerian literature accessible

to scholars, students and also the wider public. I would say the project successfully achieved its original goals.

Making Sumerian literary texts available is no easy task, due to the fact that they often have more than one source. The edition that derives from the compilation of all the sources is called composite text. It should be noted that ETCSL consists of 394 modern editions/composite texts of literary nature. The corpus is divided into seven categories, according to the subjects addressed in the compositions. The criteria is well-suited, comprising narratives featuring deities or heroes (1), court poetry, which includes *balbale*, *adab*, *šir-namšub* and other songs (2), royal correspondence and letter-prayers (3), hymns praising deities and temples (4), scribal training, debates, dialogues and didactic compositions (5) and proverbs (6). There is also one group composed of ancient literary catalogues, for obvious reasons only available in transliteration (0).

To complement the edition of the texts, ETCSL provides extensive information about the writing system, the origins of the literary genre in Sumer and the language of the Sumerians. Since one of the dimensions of the project is interdisciplinary – the possible interest of "those working in comparative literature and history of religion" – this is a very important feature. In this perspective, the "Sumerian" section contains, among other topics, information about the syntactic and grammatical components of the Sumerian language. All the basic structures and key points of Sumerian are mentioned and explained in a simple and understandable way.

Back to the main menu, in the "Browsing and searching" options, we have the possibility to search and analyze the entire corpus by content or by number (i.e., the number assigned to each category). The compositions can therefore be read online one by one. So, if we choose, for instance, the composition "Inana's descent to the nether world" (c. 1.4.1) from the "Narrative and mythological compositions" in the "Corpus content by number" section, we will see that it has two possible reading options: the transliterated version and the translated version. If we then click on the "Transliteration" option we will find

the full text in its original language (naturally transliterated into Latin characters), with its opening line "an gal-ta ki gal-še$_3$ ĝeštug$_2$-ga-ni na-an-gub", followed by a list of cuneiform sources and a short revision history. If we opt instead to follow the "Translation" option, we will be able to read the English version of the text, with matching opening line "From the great heaven she set her mind on the great below", and also its revision history. To facilitate the reader's understanding of the text, proper names are in green and *emesal* (the dialect that often indicates a goddess is speaking) is in purple. This is an important aspect that allows us to distinguish the protagonists and characters of the compositions and also the different forms of speech.

To help with the analysis of the corpus, there is also a "Word lists, sign list and full catalogue" search option. The lists available on the website include the very useful glossaries of Sumerian terms and *emesal*, a list of cuneiform signs and one of proper names. The latter contains the regularized form of each word, its lexeme, the type of name it represents (i.e., royal name, divine name, temple name, month name, etc.) and its frequency (i.e., the number of times it appears in the corpus). Considering the possible importance of this glossary for the analysis and understanding of the themes covered in the compositions, I feel it could be complemented (when possible) with a brief information about the king, deity, temple, etc. in question, as a way to help those who are not well acquainted with the religious and political history of Sumer. For instance, it would be interesting to mention that Utu was the Sumerian god of justice or that Lipit-Eštar was the king of Isin during the 20th century BC.

The corpus can also be searched in two different ways: using the "Simple search" or the "Advanced search". The latter is the one that offers more possibilities. We can type a word in the search box and opt to search it as "Lemma" (= lexeme), "form", "emesal", "label" (= translation word) or "pos" (= part of speech). For example, if we enter the lemma "nam" ("destiny"), we will obtain 690 results (i.e., lines of text). The results are not always obvious. As a matter of fact, the first lines of text contain the unit (marked in bold) "na-aĝ$_2$", which corresponds

to the *emesal* form of the word "destiny", written using two distinct cuneiform signs. Further below we have the formula "na-aĝ₂-zu" (*i.e.* the word "destiny" followed by the second-person singular possessive marker).

In fact, to facilitate the search process and to enable consistent search results, each Sumerian word was separated from its grammatical particles and was given a standardized form (a headword = a lexeme) and an English label. This allows for more complete search results as it considers not only the word when it stands alone, but also with all the grammatical morphemes attached to it. Hence, the results obtained for "nam".

This process is called "lemmatisation" (*i.e.*, "grouping all forms of a word, including multiword units, under one lexeme") and constituted a focal point of development of the website. For this process, ETCSL used Steve Tinney's Sumerian lemmatiser. Besides the English label mentioned above, this tool also attached an ePSD (*Electronic Pennsylvania Sumerian Dictionary*) citation form to every word and a part-of-speech tag. As a result, if we read the transliterated form of any composition in the corpus all we need to do in order to know the meaning and the grammatical function of a word is to place the cursor on that same word. For instance, in the opening line of the composition "Inana's descent to the nether world", placing the cursor on the word "an" will display the sequence "an (N) heaven" (*i.e.* the base of the word, its part-of-speech tag in parentheses – a noun – and its English label), placing the cursor on the word "gal-ta" will display the sequence "gal (V) to be big" (*i.e.*, the verb gal which means "to be big") and so on. After this process, there remained some problems arising from the ambiguity of some Sumerian words. Nevertheless, these problems were solved manually by the ETCSL project team.

Going back to our search results page, we can see that each line of text has three links on the left. The third, the "Gl" (Gloss) link, is probably one of the most interesting features of the browse options, as it displays a word-by-word translation of each line.

By clicking on the "Gl" link we will be directed to a new window where a table is presented. The first row displays the word as it appears

in the text, the row below presents its lemma and the last row shows its English label. Every lemma in the second row has an attached link leading to ePSD, where one can find a more complete list of meanings of the word. This is an important aspect because the label attributed by the lemmatiser does not always correspond to the most appropriate translation. Finally, below this table is displayed the paragraph of transliterated text to which that line belongs.

All things considered, we can say that ETCSL provides all the information and resources one needs to correctly interpret and understand the literary legacy the Sumerians have left us.

One last aspect worthwhile pointing out is the series of publications (articles and journals) that arose from ETCSL initiative and editing process. I would like to highlight the work entitled *The Literature of Ancient Sumer* (Black, J. A., Cunningham, G., Ebeling, J., and Zólyomi, G., 2004) and the book *Analyzing Literary Sumerian: Corpus-based Approaches* (Ebeling, J., and Cunningham, G. [eds.], 2007).

In short, ETCSL's purpose – to create a coherent Sumerian literary corpus and make it available to the general public – is entirely fulfilled. The end of funding for ETCSL in 2006 was a rather unfortunate event. One can only imagine how the website would be now if it had developed during the past few years. ETCSL was a breakthrough project in the area of Ancient Near Eastern literature and Sumerian studies. Reading and searching texts composed in the second half of the third millennium BC is now an easy task. A project so exceptional could only be made possible with the collaboration of major names in Sumerian studies, like Graham Cunningham, John Baines, Jacob L. Dahl, Esther Flückiger-Hawker, Jon Taylor, Marc Van De Mieroop and, of course, the master of Sumerian literature Jeremy Black, gently nicknamed by his fellow workers "guardian *me*".