

Oracc in Korp

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Tero Alstola, Tommi Jauhiainen, Krister Lindén (University of Helsinki)

Open Richly Annotated Cuneiform Corpus (Oracc) is an international cooperative undertaking providing free online editions of texts. The text corpora in Oracc have been created by various projects, and it is one of the largest electronic resources of Akkadian and Sumerian texts.

Oracc is a valuable tool for an Assyriologist who works with Akkadian and Sumerian texts, but it has been relatively difficult to do searches across all the different Oracc projects. However, when studying the semantic contexts of a word, it is imperative to have all the attestations of the word at hand. An efficient corpus search tool is therefore a necessity. We have added the texts in Oracc to Korp, an openly available online service provided by the Language Bank of Finland. Korp allows users to make queries in text corpora. The results of the query are presented as *concordances*, i.e. listing every instance of the word matching the query with all its neighbors line by line. Furthermore, the results are, by default, presented in the so-called *KWIC* (keyword in context) view, meaning that each result is displayed with the matching words highlighted in the middle of the display. Additionally, the annotation of each word is presented in the sidebar when a word is selected. Korp is a new and useful tool for studying the contexts in which words appear.

In this poster, we present *Oracc in Korp*. To accommodate the special features of the Oracc data we have added new fields that are not present in the corpora of other languages in Korp. The ways in which words in different Oracc projects have been annotated differ somewhat. Therefore, we have normalized the dictionary forms of certain types of words, such as divine names. Finally, we have added a feature in Korp which allows the user to search for a word using these normalized forms. Each text added to Korp also contains a link to the original text in Oracc.