ORACC IN KORP USER GUIDE - version July 4, 2019

For a general guide to Korp see https://www.kielipankki.fi/support/korp/.

Direct link to Oracc in Korp: http://urn.fi/urn:nbn:fi:lb-2019060602

Alternatively, you can go to https://www.kielipankki.fi/corpora/oracc/ and click "Open the resource in the concordance service Korp".

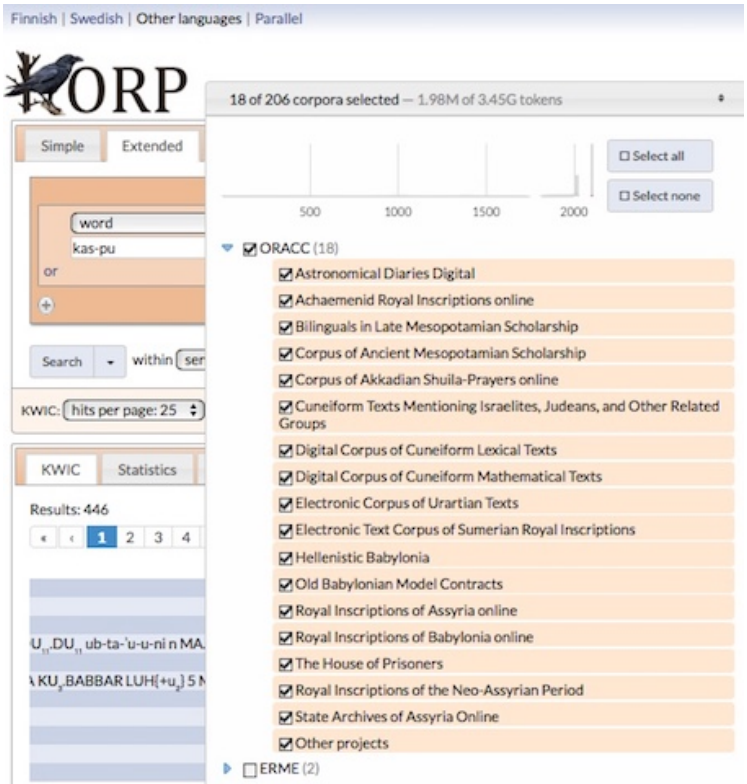
You can also go to https://korp.csc.fi/ and change the language from Finnish to English in the upper right corner of the window. Then select "Other languages" (referring to the languages of the text corpora) in the upper left corner. Finally, click the selection bar next to the Korp logo, click "Select none", and tick the box next to the word "ORACC".

If you do not see Oracc in the list, make sure you are in the "Other languages" section of Korp. Select *Other languages* in the upper left corner of the window. If everything is in Finnish, you can change the language to English from the upper right corner.

SELECTING THE DATASETS

As a default you are searching across all the Oracc corpora in Korp. To select only one or a few corpora click the selection bar to the right of the Korp logo and then click the arrow to the left of ORACC to see the datasets. You can deselect all the sets by unchecking "Oracc" and then choosing the sets you are interested in.

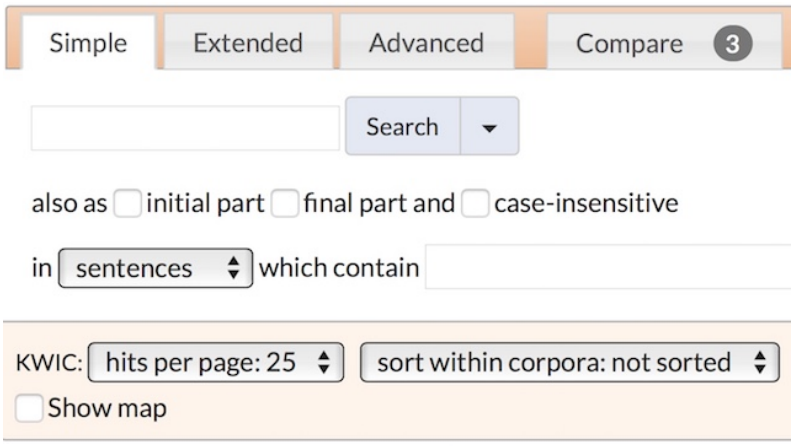
The data has been downloaded from Oracc as JSON-files in May 2019. The data for Korp has been divided into 18 datasets mostly conforming to the projects in Oracc:



The category **Other projects** contains texts from several smaller projects:

- Idrimi: Statue of Idrimi
- akklove: Akkadian Love Literature
- Contributions Amarna
- CKST: Corpus of Kassite Sumerian Texts
- Glass: Corpus of Glass Technological Texts
- LaOCOST: Law and Order: Cuneiform Online Sustainable Tool
- OBTA: Old Babylonian Tabular Accounts
- Suhu: The Inscriptions of Suhu online.

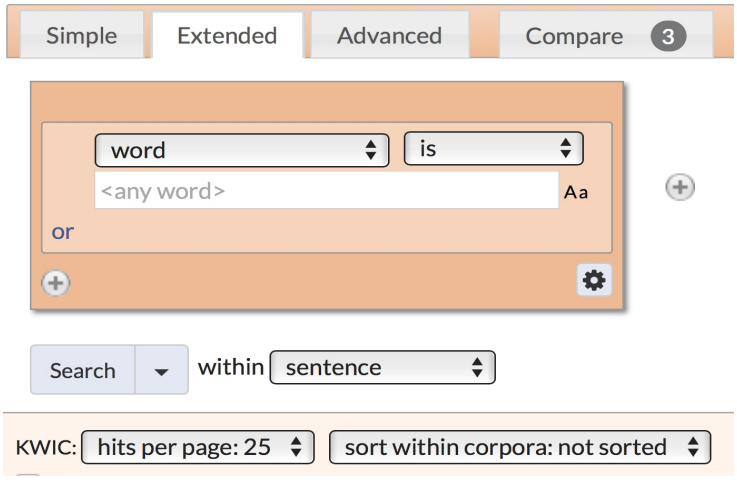
SIMPLE SEARCH



Simple search lets you:

- search for the transliteration of a word
  - write the transliteration in the search box and hit search
- search for the transliteration that
  - starts as you define
    - select "also as ✓**initial part**"
  - ends as you define
    - select "also as initial part ✓**final part**"
- use case-insensitive search (e.g. get *lugal* and *LUGAL* in one search)
  - select "also as initial part final part and ✓**case-insensitive**"
- specify another transliteration that has to be in the same sentence (in the case of Oracc in Korp a "sentence" can mean either a line or several lines connected to an English translation) or document
  - select from the dropdown list "in (sentences) which contain"
    - sentences
    - paragraphs (= document)
    - texts (= document)
  - write the other transliteration in the box after "in (sentences) which contain"
  - hit search
- combine the previous search options

EXTENDED SEARCH



In addition to transliteration (word), extended search lets you search by:

- word attributes
- text attributes

Click the dropdown list that says "word" and you get a list of attributes.
The word attributes are:

- base form = dictionary form in *A Concise Dictionary of Akkadian*
    - if you want to search for a compound word (e.g. *mār bārê*), use "&&" to combine the base forms of the words (māru&&bārû)[1]
- translation (lemma) = the first translation of the dictionary form in CDA
- transcription
- translation (sense) of transcription
- part-of-speech
- part-of-speech (detailed) = subcategory as defined in Oracc
- standardized (form of gods and places), see https://github.com/anee-helsinki/OraccInKorp/tree/master/VersionMay2019
- language/dialect

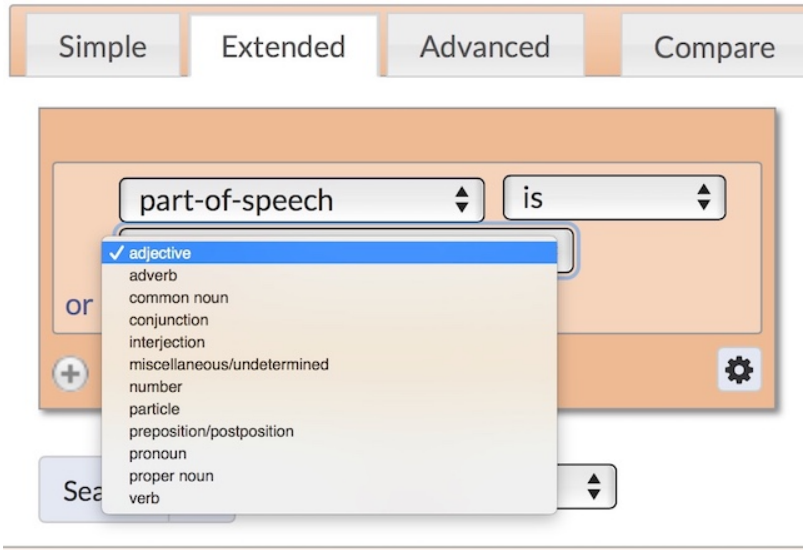
The text attributes are:

- CDLI number
- genre
- period
- provenance
- subgenre (as defined in Oracc)
- text languages

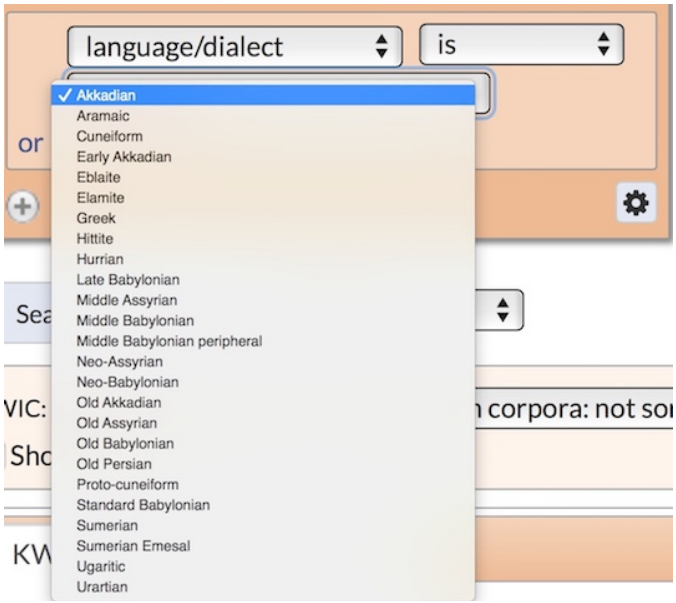
[1] Note that the way compound words are written in different projects in Oracc varies. The joined words which in Oracc have separate translations, word classes, etc. have in Korp been joined with "&&". Sometimes compound words in Oracc have been defined as one word and the parts have been joined with "-", e.g. *EN-MU.MU*, baseform *bēl-zakār-šumi*. Sometimes the parts of a compound word are all defined separately.

The search field underneath will change into a dropdown list of possible values if you choose

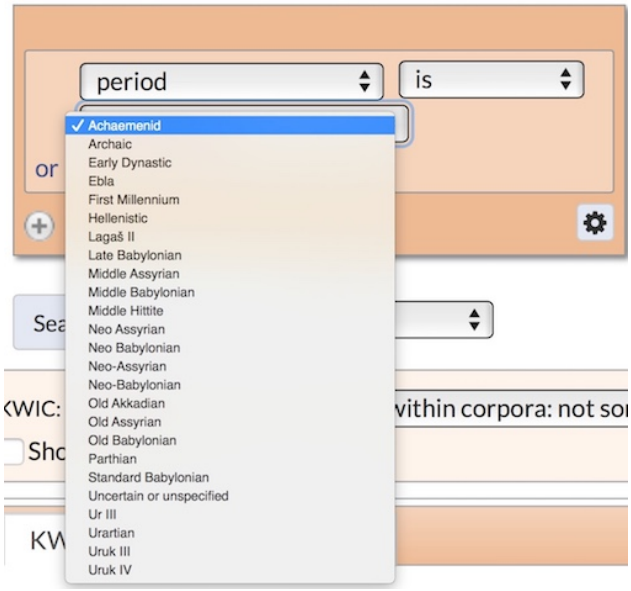
- part-of-speech



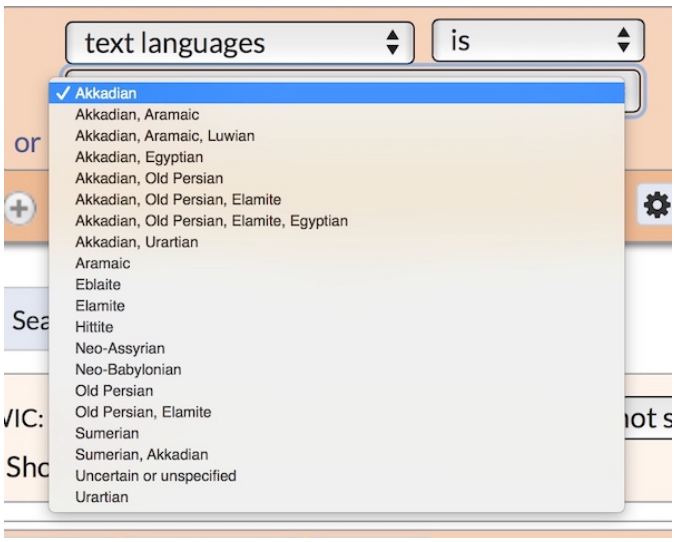
- language/dialect



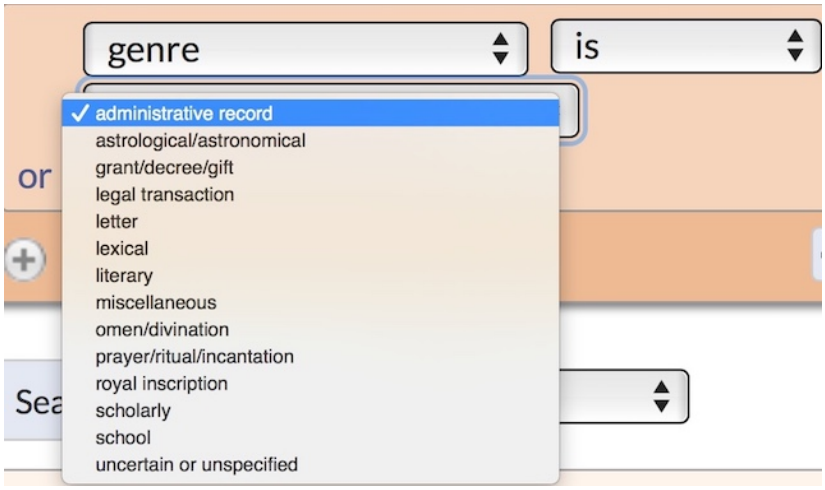
- period



- text languages



- genre



In other cases, you can write the value you are looking for in the field.

In all cases you can choose whether you want that the value specified

- is
- is not

in the resulting sentences.
When there is no dropdown list, you can also specify if you want the word to:

- start with
- contain
- end with

the value you have given or

- is
- is not

the regular expression given.

Hint: By using the default option "Search within (sentence)", you get results that occur either on a line or several lines connected to an English translation. Especially when looking for several words occurring together, it might be beneficial to use "Search within (up to paragraph)" which in the case of Oracc searches within a document, thus over line breaks.

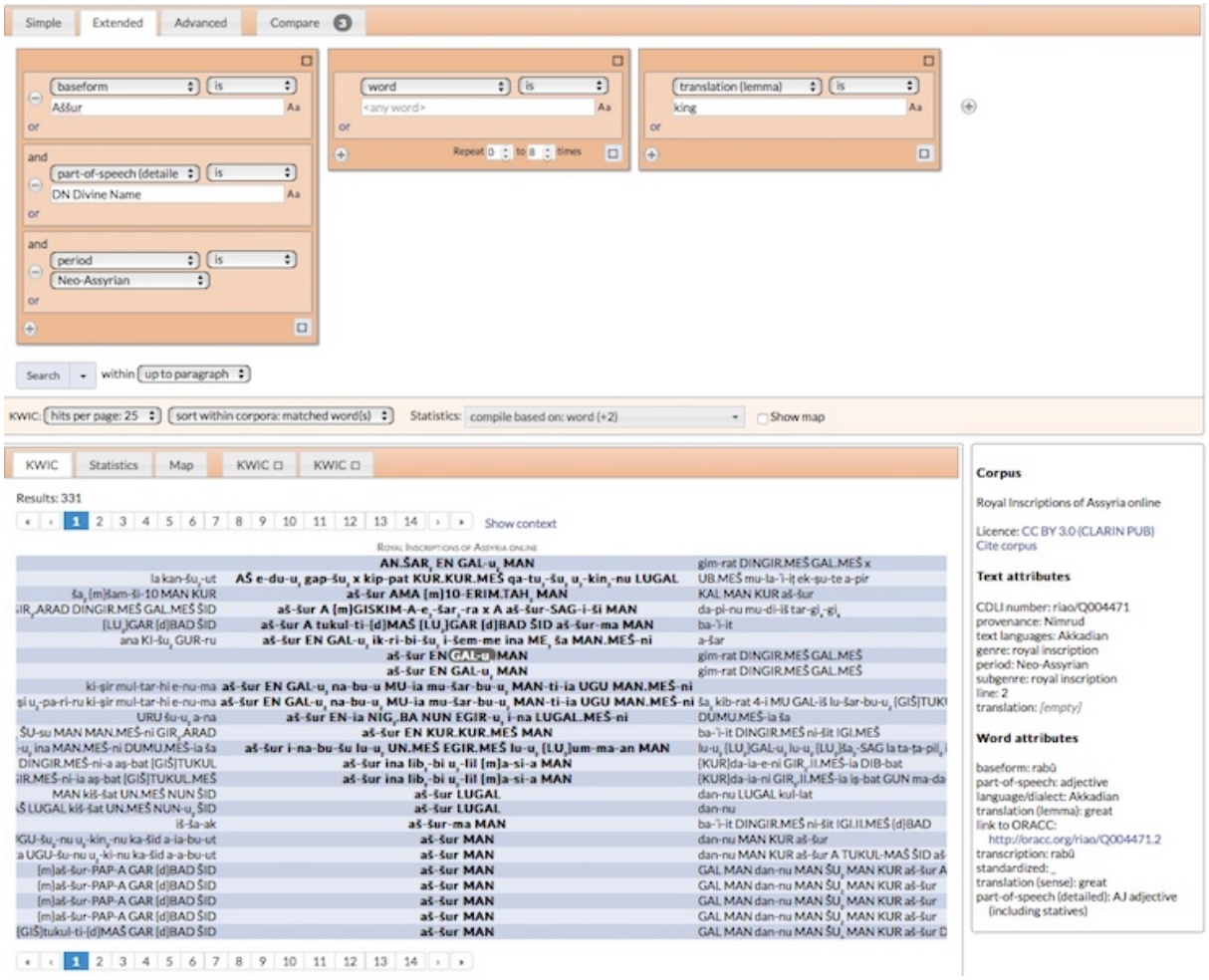
MAKING MORE COMPLICATED QUERIES IN THE EXTENDED SEARCH

- Click the wheel ⚙ on the right lower corner of the box to find options
- Repeat – find words that occur several times right after each other
- Sentence start – find word instances that are only at the beginning of a sentence
- Sentence end – find word instances that are only at the end of a sentence
- Click the word **or** in the box and specify two words/attributes to get the attestations of both
    - e.g. find all words where transliteration is either *lugal* or *LUGAL*
- Click the ⊕ button in the lower part of the box and specify another attribute that the word has to have (note that searching for the same attribute in both options will not give any results)
    - e.g. find words the transcription of which is *kaspa* and transliteration is not KU_3.BABBAR
- Click the ⊕ sign to the right of the box and specify another word that has to follow the first one
    - e.g. base form is *šarru* + base form is *dannu*

All these options can be combined together as many times as you want!

Search example 1:

Texts from the Neo-Assyrian period featuring dictionary form of the divine name “Aššur” and the translation “king” with no more than 8 words in between. Search in documents using “Search within (up to paragraph)”.



SEARCH RESULTS

The results will show all the instances of the searched word(s).

- The word(s) searched for will be highlighted and located in the middle of the result list one below another (Keyword in Context = KWIC).
- You can scroll the screen sideways to see more of the context of the word.

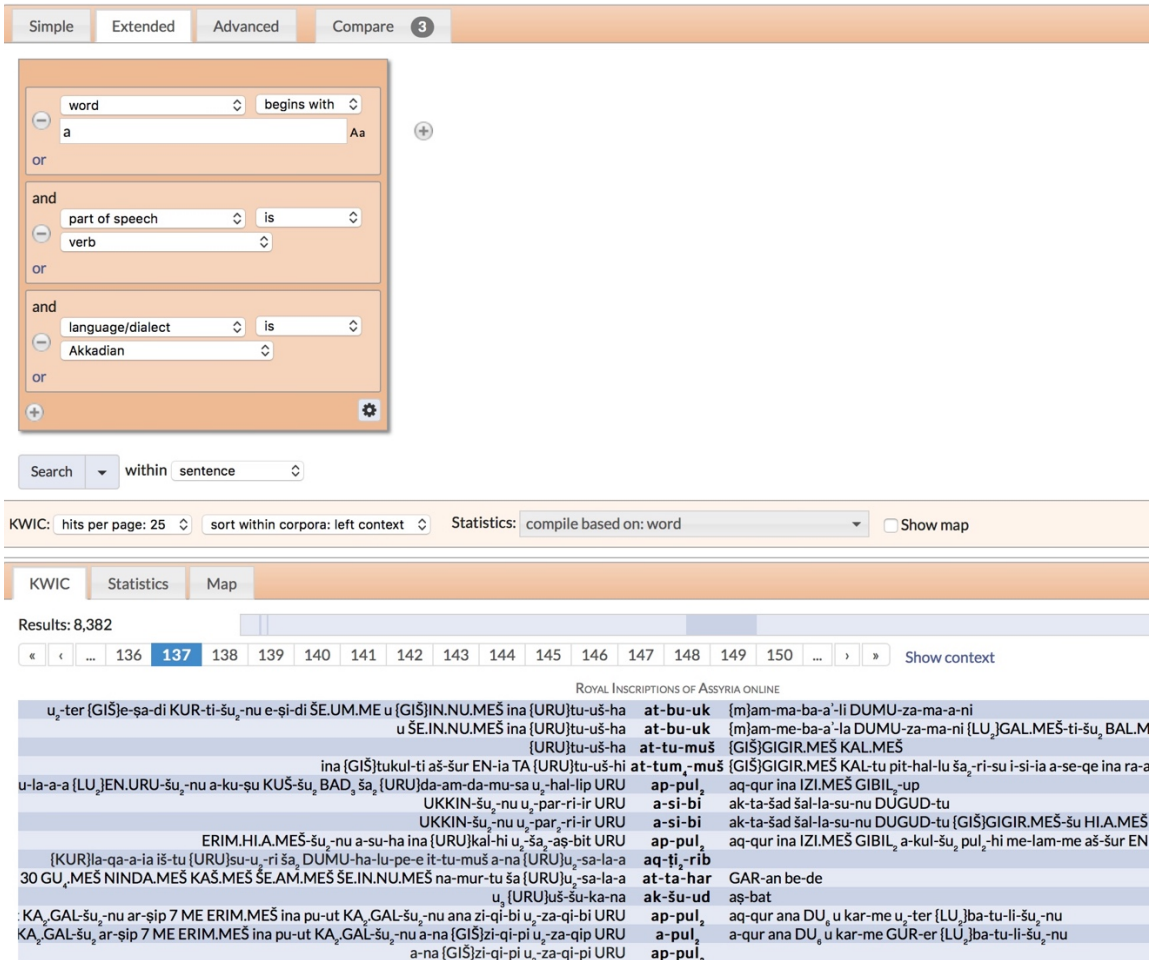
By default, there will be 25 results per page and the results are in the order of the projects chosen. The way the results are shown can be changed from the KWIC bar:

- Hits per page: choose 50, 75, 100, 500 or 1000 hits on the page
- Sort within results:
    - Choose “left context” to sort according to the preceding word
    - Choose “right context” to sort according to the following word
    - Choose “matched words(s)” to sort according to the matches (when they are not all the same word)

The results can also be seen in a larger context by clicking **Show context** just above the KWIC view. When you click any word in the result list, you can see in the **sidebar** information about that word and the document this token appears in. The sidebar also has a link to the page with the *Oracc in Korp* user guide (“Information page”) and a link to the text in Oracc.

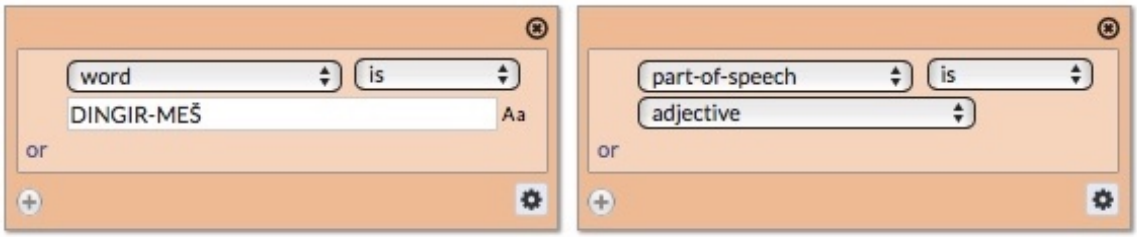
Search example 2

Verb forms starting with the letter “a” in transcription, the language of the word being Akkadian and results ordered by left content, i.e. the preceding word.



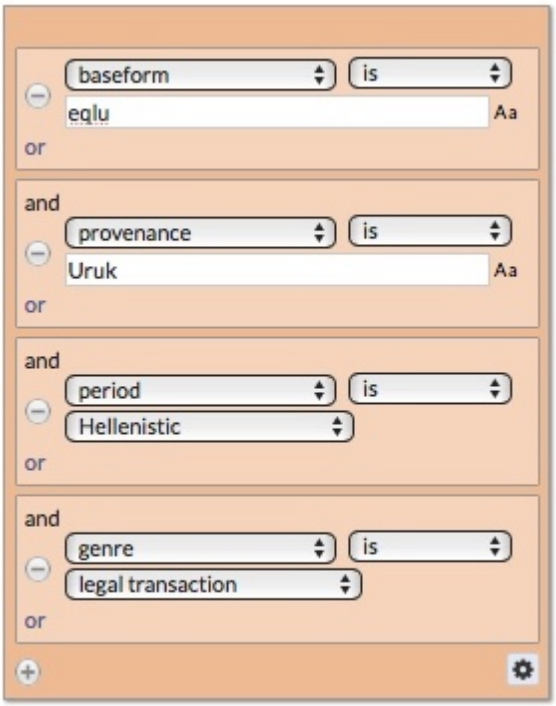
Search example 3

Transliteration “DINGIR-MEŠ” followed by any adjective.



Search example 4

Dictionary form “eqlu” in texts where provenance is Uruk, period Hellenistic, and text genre legal transaction.



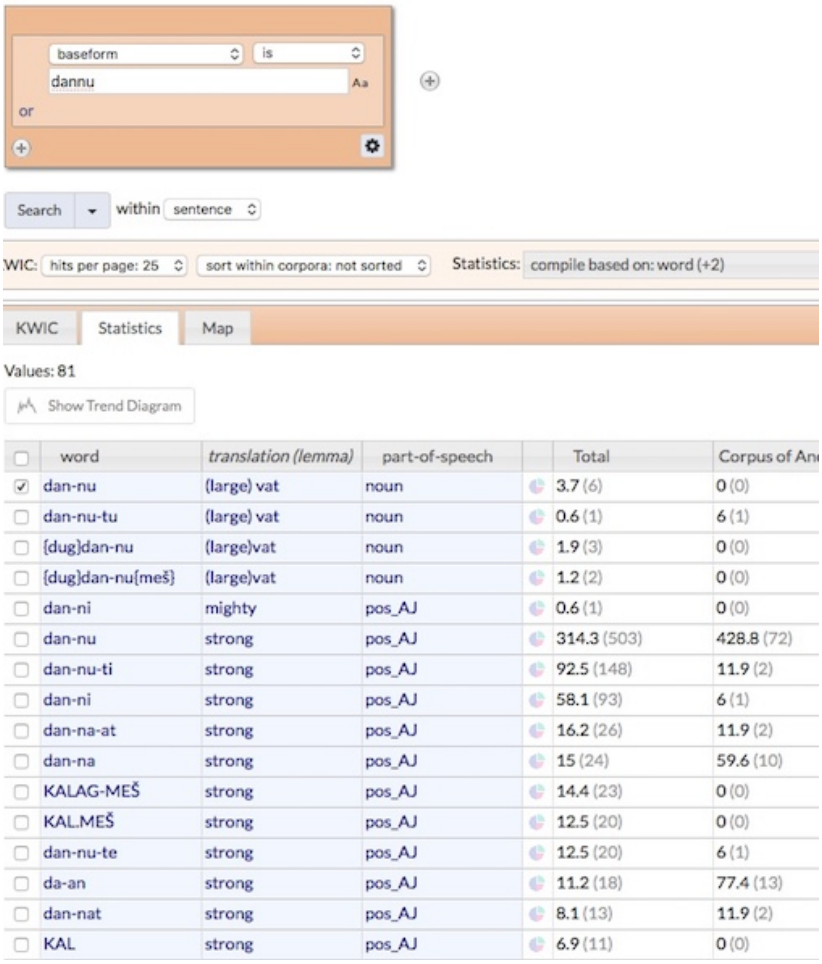
STATISTICS

The Statistics tab gives the number of occurrences for each matched word both in all results and within individual corpus/dataset.

- The number of occurrences is shown as relative frequencies per million tokens, a common measure in corpus linguistics.
- The numbers in parentheses are the absolute frequencies (i.e. the number of occurrences).
- The default view shows the statistics of the transliteration(s) of the word searched regardless what attributes were searched for.
- You can sort the statistics according to any column by clicking the heading of that column.
- You can change what attribute(s) are considered by selecting the attributes you want in “Statistics: (compile based on)” and then clicking “search”. You can even choose several at once:
    - e.g. search for all occurrences of words the translation of which contains “love” and see, for example, what baseform+part-of-speech+word-language combinations there are in the results.

From the statistics you can see the results of an individual line by clicking a word on that line. You will get a new tab with the KWIC view of those results in within your original search.

Statistics example:
Search for the dictionary form *dannu* and base the statistics on transliteration, translation, and the part of speech tag. Order by the translation.



word	*translation (lemma)*	part-of-speech	Total	Corpus of An
dan-nu	(large) vat	noun	3.7 (6)	0 (0)
dan-nu-tu	(large) vat	noun	0.6 (1)	6 (1)
{dug}dan-nu	(large)vat	noun	1.9 (3)	0 (0)
{dug}dan-nu{meš}	(large)vat	noun	1.2 (2)	0 (0)
dan-ni	mighty	pos_AJ	0.6 (1)	0 (0)
dan-nu	strong	pos_AJ	314.3 (503)	428.8 (72)
dan-nu-ti	strong	pos_AJ	92.5 (148)	11.9 (2)
dan-ni	strong	pos_AJ	58.1 (93)	6 (1)
dan-na-at	strong	pos_AJ	16.2 (26)	11.9 (2)
dan-na	strong	pos_AJ	15 (24)	59.6 (10)
KALAG-MEŠ	strong	pos_AJ	14.4 (23)	0 (0)
KAL.MEŠ	strong	pos_AJ	12.5 (20)	0 (0)
dan-nu-te	strong	pos_AJ	12.5 (20)	6 (1)
da-an	strong	pos_AJ	11.2 (18)	77.4 (13)
dan-nat	strong	pos_AJ	8.1 (13)	11.9 (2)
KAL	strong	pos_AJ	6.9 (11)	0 (0)

COMPARING RESULTS

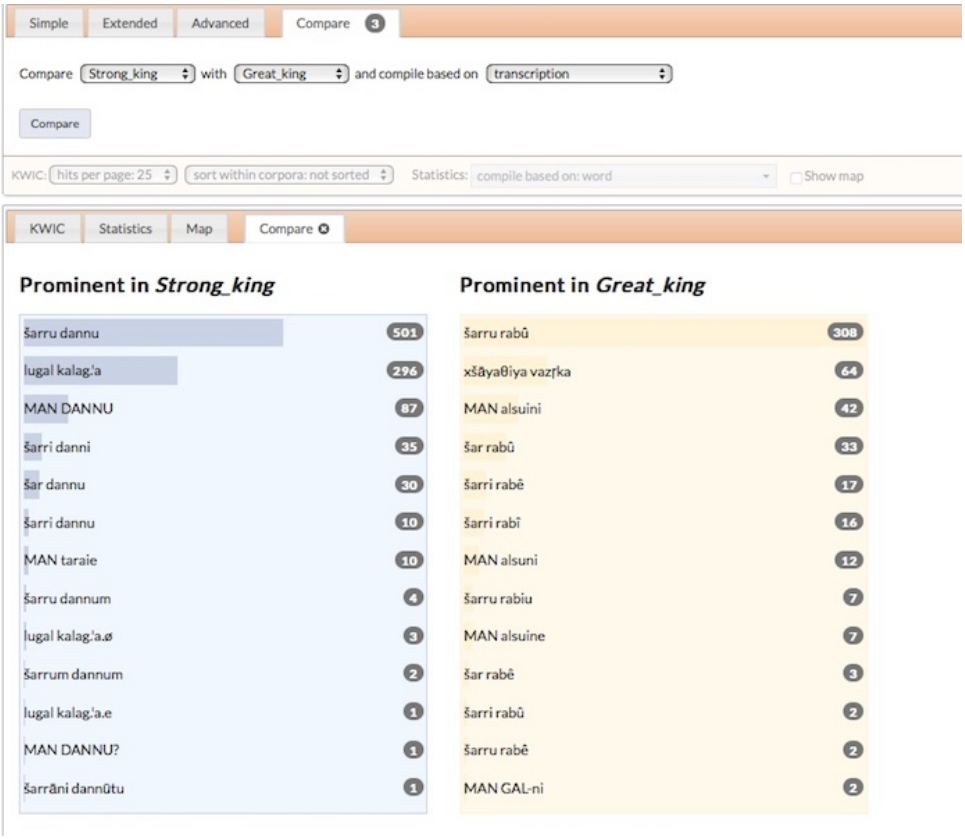
You can save searches for comparison with each other.
When you have typed in your search, click the arrow next to the search button and give your search a name. When you have at least two saved queries, go to the Compare tab and choose:

- from the first two dropdown lists which queries you want to compare
- from the third dropdown which attributes you want to base your comparison on.

The results will be shown in a new tab, thus your latest search is still active.

Comparison example:

Save a search for two consecutive words with the translations (sense) “king” and “strong” (bear in mind that adjectives follow nouns in the Akkadian word order).
Save a search for two consecutive words with the translations (sense) “king” and “great”.
Compare the queries based on transcriptions.



EXPORTING THE RESULTS

You can export the results of your search in many different forms (the formats supported in most cases are excel, csv, tsv and html):

Annotations, i.e. text as a table, token per row
Bibliographical references as a table
Sentence per row
Sentence per row, match and contexts separated
Plain text
JSON
NooJ

Just select your form and format and hit “Download KWIC” at the bottom of the view.

You can also export the statistics data. At the bottom of the Statistics tab select whether you want to export relative or absolute frequencies and whether you want to have the data in the csv or tsv format and hit **Generate export** and then **Export**.

ADVANCED SEARCH

For a general guide for the advanced search in Korp see
https://www.kielipankki.fi/support/korp-advanced/

The advanced search is performed by writing the query in the so-called CQP (Corpus Query Protocol) query language.

The easiest way to start learning CQP is to perform a simple or extended search and then check in advanced search what it looks like in CQP. For example, we performed above a search in texts from the Neo-Assyrian period containing the dictionary form of the divine name "Aššur" and the translation "king" with no more than 8 words in between. The search looks like this in the CQP language:
[lemma = "Aššur" & possub = "DN Divine Name" & _.text_period = "Neo-Assyrian"] []{0,8} [ltrans = "king"]

Brackets [] always contain one word, empty brackets stand for any word
Thus, the example above contains three words

The attributes have a slightly different names in CQP than in extended search:

CQP	*Extended search*

Word attributes:

- word — transliteration
- lemma — base form/dictionary form
- ltrans — translation (lemma) of the dictionary form
- transcription — transcription
- sense — translation (sense) of the transcription
- pos — part-of-speech (general)
- possub — part-of-speech original in Oracc
- standard — standardized form of divine and place names
- lang — language/dialect of the word

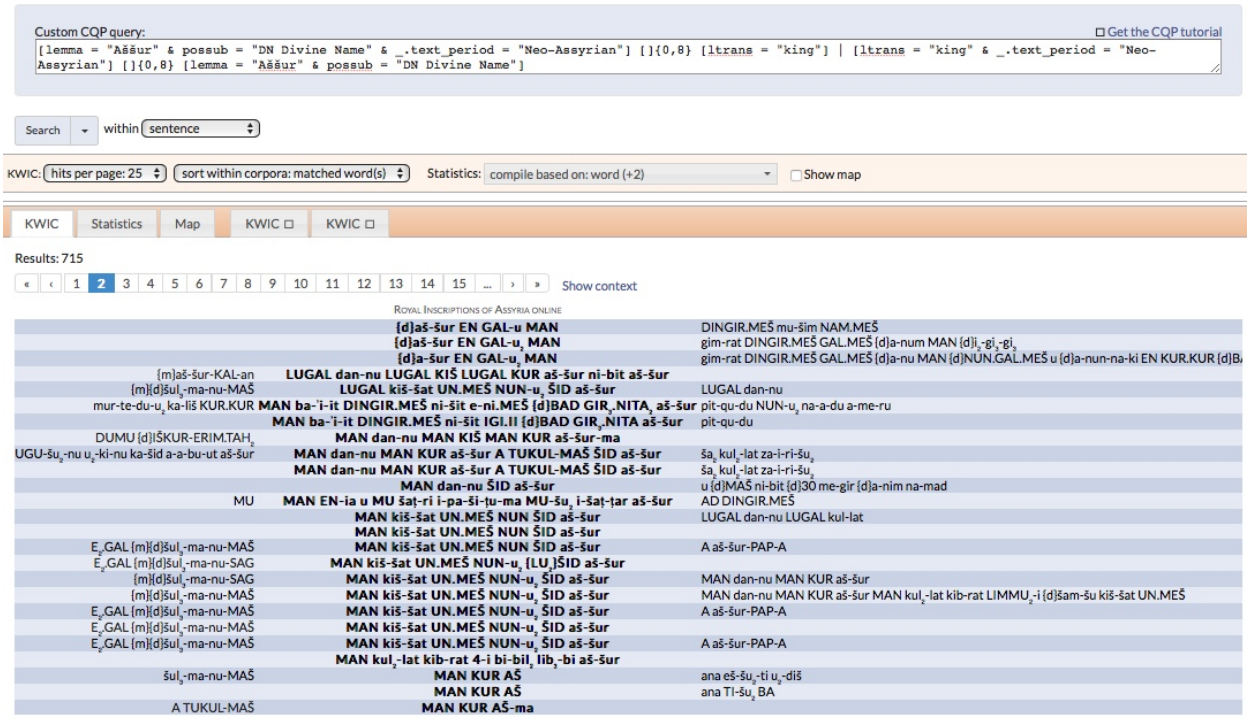
Text attributes:

- cdlinumber — the CDLI number of the text preceded with the abbreviation of the project name
- genre — genre (general)
- provenance — provenance
- period — period
- subgenre — genre as defined in Oracc
- language — language of the text

[lemma = "Aššur"] words where dictionary form **is** Aššur
[lemma **!=** "Aššur"] words where dictionary form **is not** Aššur
[lemma = "**.***Aššur**.***"] words where dictionary form **contains** Aššur
[lemma = "Aššur**.***"] words where dictionary form **starts with** Aššur
[lemma = "**.***Aššur"] words where dictionary form **ends with** Aššur
[lemma = "Aššur" **&** possub = "DN Divine Name"] words where dictionary form is Aššur **and** part-of-speech (detailed) is divine name
[word = "lugal" | word = "LUGAL"] words where transliteration is *lugal* **or** transliteration is *LUGAL*
[]{1,3} one, two, or three words without specifying what word
[word = "LUGAL"]{2,2} words where transliteration is *LUGAL* right after words where transliteration is also *LUGAL* (LUGAL LUGAL in the text)

Advanced search example:

Texts from the Neo-Assyrian period featuring dictionary form of the divine name “Aššur” and the translation “king” **or the other way round** with no more than 8 words in between.
[lemma = "Aššur" & possub = "DN Divine Name" & _.text_period = "Neo-Assyrian"] []{0,8} [ltrans = "king"] | [ltrans = "king" & _.text_period = "Neo-Assyrian"] []{0,8} [lemma = "Aššur" & possub = "DN Divine Name"]



We would like to have comments/suggestions concerning this user guide. You can send them to Heidi Jauhiainen at firstname.surname@helsinki.fi.