ORACC IN KORP USER GUIDE (Korp v.9) – July 2024

Heidi Jauhiainen
Centre of Excellence of Ancient Near Eastern Empires, University of Helsinki

Korp is an online concordance tool that contains texts in various languages. It also contains two versions of the Open Richly Annotated Cuneiform Corpus (Oracc) data: Oracc in Korp 2021 is a snapshot from June 2021 and Oracc in Korp 2019 was downloaded in May 2019.

Direct link to **Oracc in Korp 2021**: http://urn.fi/urn:nbn:fi:lb-2022031706
Direct link to **Oracc in Korp 2019**: http://urn.fi/urn:nbn:fi:lb-2019060602

Instead of using the direct links, you can also go to https://www.kielipankki.fi/corpora/oracc/ and click "Open the resource in the concordance service Korp" of either version. The page also has a link to the downloadable version of Oracc in Korp 2019. The downloadable version is a zip file containing .vrt files for each of the projects in that corpus. The files have one word per line with their tab-separated annotations.

You can also go to https://kielipankki.fi/korp 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 "Select corpora" bar next to the Korp logo and click the arrow next to the words **Cuneiform/Nuolenpääkirjoitus**. See below how to select a dataset.

The next snapshot of Oracc data for Korp is planned to be taken at the end of 2023 and to be released in the spring 2024.

Korp has

This guide replaces the one from July 2019. Korp has since then been updated to version 9 and moved to new servers. The Oracc in Korp 2021 been added with some new attributes for both Oracc corpora.

For a general guide to the Language Bank of Finland's Korp tool see https://www.kielipankki.fi/support/korp/. Note that it is for the older version of Korp.

For technical questions or complaints about Korp write to kielipankki@csc.fi.

For any questions on the Oracc data in Korp, contact Aleksi Sahala at firstname.lastname@helsinki.fi.

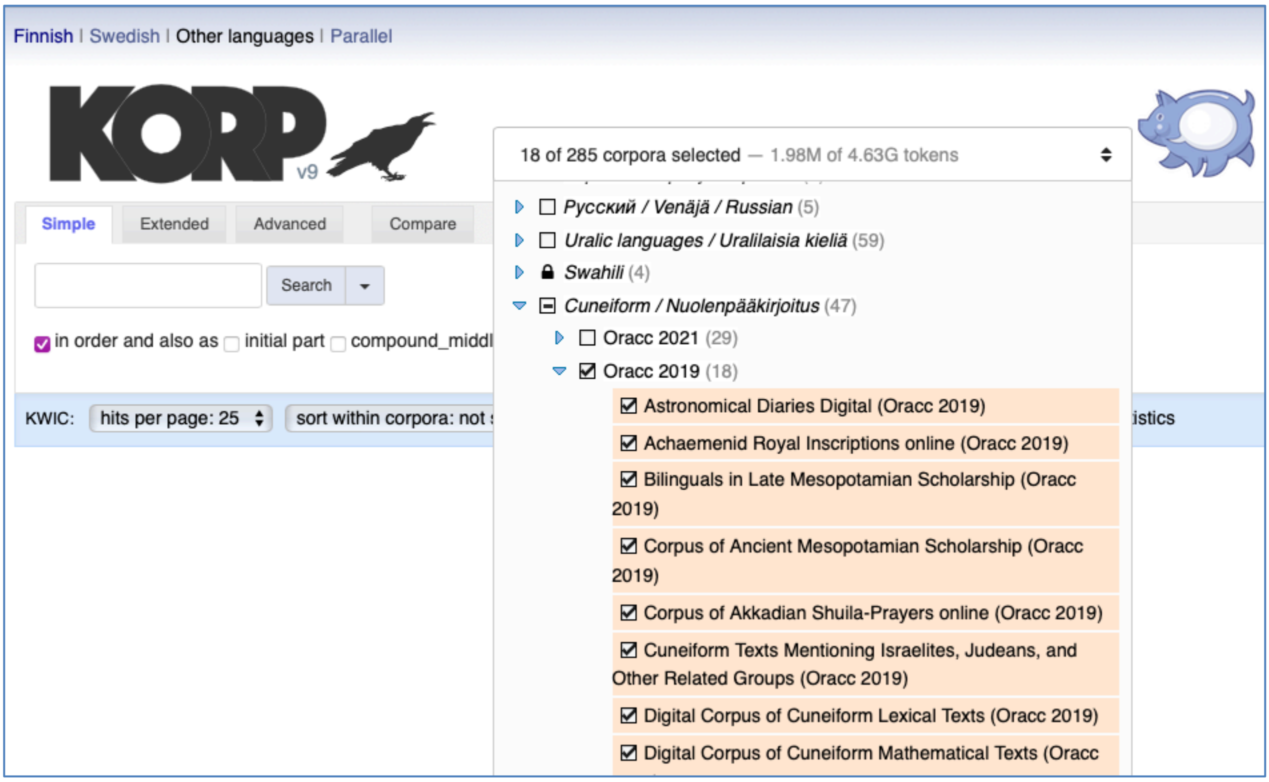
If you have comments/suggestions concerning this user guide, you can send them to Heidi Jauhiainen at firstname.lastname@helsinki.fi.

I would like to thank Tero Alstola and Ellie Bennett for providing me with useful comments on this guide and Tero and Aleksi Sahala for most of the Akkadian examples on how to use Korp.

Table of Contents

1. SELECTING THE DATASETS

You can choose the Oracc version by going to Korp using one of the urn-addresses on previous page or by going to the **Other languages** page in Korp and by ticking the box next to the name of it in the **Select corpora** dropdown. You will then be searching across all the corpora in that version. To select only one or a few corpora click the arrow left of the name of the chosen version to see the different datasets. You can select and deselect all the sets by ticking the box next to the version's name and, when unselected, choose the sets you are interested in.



18 of 285 corpora selected — 1.98M of 4.63G tokens

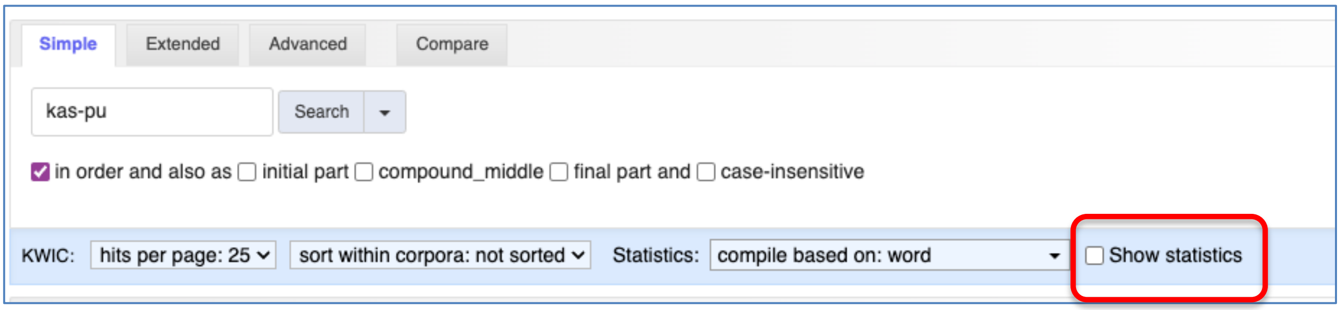
- ▷ □ Русский / Venäjä / Russian (5)
- ▷ □ Uralic languages / Uralilaisia kieliä (59)
- ▷ Swahili (4)
- ▽ ⊟ Cuneiform / Nuolenpääkirjoitus (47)
  - ▷ □ Oracc 2021 (29)
  - ▽ ☑ Oracc 2019 (18)
    - ☑ Astronomical Diaries Digital (Oracc 2019)
    - ☑ Achaemenid Royal Inscriptions online (Oracc 2019)
    - ☑ Bilinguals in Late Mesopotamian Scholarship (Oracc 2019)
    - ☑ Corpus of Ancient Mesopotamian Scholarship (Oracc 2019)
    - ☑ Corpus of Akkadian Shuila-Prayers online (Oracc 2019)
    - ☑ Cuneiform Texts Mentioning Israelites, Judeans, and Other Related Groups (Oracc 2019)
    - ☑ Digital Corpus of Cuneiform Lexical Texts (Oracc 2019)
    - ☑ Digital Corpus of Cuneiform Mathematical Texts (Oracc

Oracc in Korp 2021, contains texts from 29 different Oracc projects.
Oracc in Korp 2019 contains 18 categories, 17 of which correspond to Oracc projects. 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.

Note that the *EPSD2: electronic Pennsylvania Sumerian Dictionary* is a huge corpus that may slow down your search. If you do not specifically need Sumerian texts in your search, you might want to untick both EPSD2 and ETCSRI: Electronic Text Corpus of Sumerian Royal Inscriptions. If you do not need the lexical lists, you can untick the quite large DCCLT: Digital Corpus of Cuneiform Lexical Texts.

2. SIMPLE SEARCH



kas-pu

☑ in order and also as □ initial part □ compound_middle □ final part and □ case-insensitive

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word □ Show statistics

Simple search lets you:

- search for the **transliteration** of a word
    - write the transliteration in the search box and hit Search (*Enter* does not work)
- search for a transliteration that
    - starts with the search query
        - check “✓**initial part**” before clicking Search
    - contains the search query anywhere in the word
        - check “✓**compound_middle**” before clicking Search
    - ends with the search query
        - check “✓**final part**” before clicking Search
- use case-insensitive search (e.g. get *lugal* and *LUGAL* in one search)
    - check “✓**case-insensitive**” before clicking Search

These selections can be combined!

You can add a transliteration of a second word to your search! By default, this searches for words **in order**, that is words that occur together in the order indicated. By unselecting the **in order**… box, the words can be anywhere in the document.

You might want to **deselect the box in front of the words Show statistics** (for statistics see section 6). If you use several subcorpora and statistics are calculated, the query time can get quite long.

Depending on your computer you might be able to write some of the special characters used for the transliteration and transcription of Akkadian and Sumerian words, but probably not all. **Here is a list you can copy and paste from:**

Special characters:

Ā á ē ī í ū ú š ṭ ṣ ʾ ŋ

Ā Á Ē É Ī Ū Ú Ù Š Ṭ Ṣ Ŋ

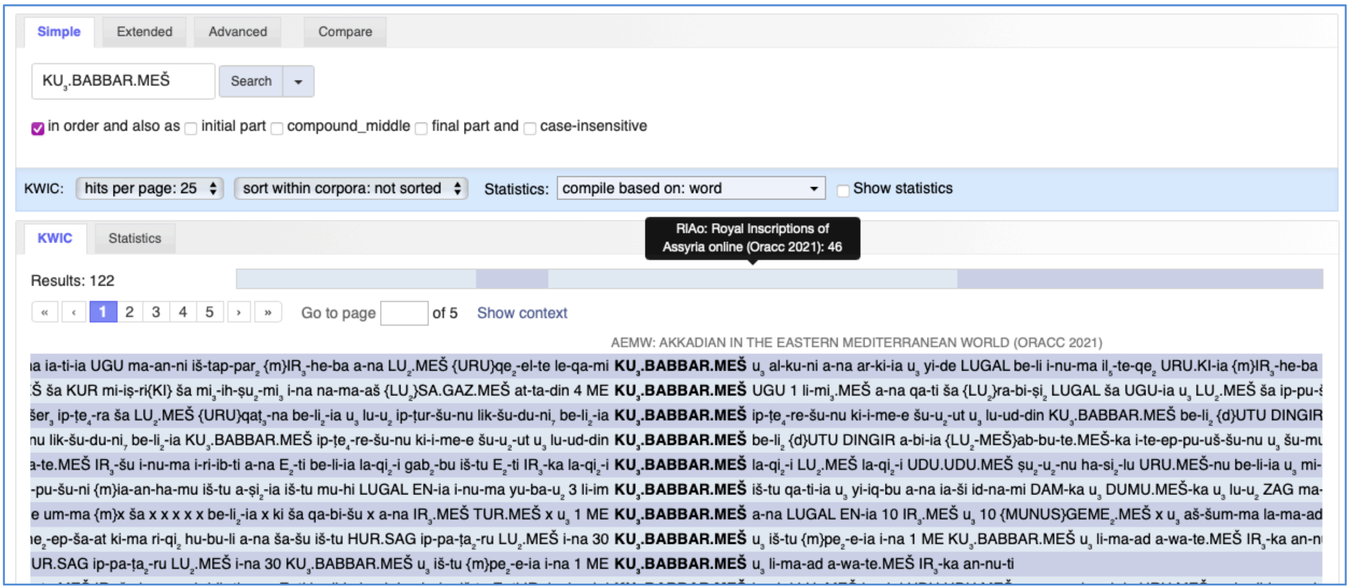
Subscripts:

X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12} X_{13} X_{14}

3. SEARCH RESULTS

The results will show all the instances of the searched word.

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



KU$_3$.BABBAR.MEŠ

☑ in order and also as □ initial part □ compound_middle □ final part and □ case-insensitive

KWIC: hits per page: 25 | sort within corpora: not sorted | Statistics: compile based on: word | □ Show statistics

RIAo: Royal Inscriptions of Assyria online (Oracc 2021): 46

Results: 122

« ‹ 1 2 3 4 5 › » Go to page of 5

AEMW: AKKADIAN IN THE EASTERN MEDITERRANEAN WORLD (ORACC 2021)

ia ia-ti-ia UGU ma-an-ni iš-tap-par$_2$ {m}IR$_3$-he-ba a-na LU$_2$.MEŠ {URU}qe$_2$-el-te le-qa-mi **KU$_3$.BABBAR.MEŠ** u$_3$ al-ku-ni a-na ar-ki-ia u$_3$ yi-de LUGAL be-li i-nu-ma il$_5$-te-qe$_2$ URU.KI-ia {m}IR$_3$-he-ba

Š ša KUR mi-iṣ-ri{KI} ša mi$_3$-ih-ṣu$_2$-mi$_3$ i-na na-ma-aš {LU$_2$}SA.GAZ.MEŠ at-ta-din 4 ME **KU$_3$.BABBAR.MEŠ** UGU 1 li-mi$_3$.MEŠ a-na qa-ti ša {LU$_2$}ra-bi-ṣi$_2$ LUGAL ša UGU-ia u$_3$ LU$_2$.MEŠ ša ip-pu-š

šer$_3$ ip-ṭe$_4$-ra ša LU$_2$.MEŠ {URU}qaṭ$_3$-na be-li$_2$-ia u$_3$ lu-u$_2$ ip-ṭur-šu-nu lik-šu-du-ni$_7$ be-li$_2$-ia **KU$_3$.BABBAR.MEŠ** ip-ṭe$_4$-re-šu-nu ki-i-me-e šu-u$_2$-ut u$_3$ lu-ud-din KU$_3$.BABBAR.MEŠ be-li$_2$ {d}UTU DINGIR

nu lik-šu-du-ni$_7$ be-li$_2$-ia KU$_3$.BABBAR.MEŠ ip-ṭe$_4$-re-šu-nu ki-i-me-e šu-u$_2$-ut u$_3$ lu-ud-din **KU$_3$.BABBAR.MEŠ** be-li$_2$ {d}UTU DINGIR a-bi-ia {LU$_2$-MEŠ}ab-bu-te.MEŠ-ka i-te-ep-pu-uš-šu-nu u$_3$ šu-mu

a-te.MEŠ IR$_3$-šu i-nu-ma i-ri-ib-ti a-na E$_2$-ti be-li-ia la-qi$_2$-i gab$_2$-bu iš-tu E$_2$-ti IR$_3$-ka la-qi$_2$-i **KU$_3$.BABBAR.MEŠ** la-qi$_2$-i LU$_2$.MEŠ la-qi$_2$-i UDU.UDU.MEŠ ṣu$_2$-u$_2$-nu ha-si$_2$-lu URU.MEŠ-nu be-li-ia u$_3$ mi-

-pu-šu-ni {m}ia-an-ha-mu iš-tu a-ṣi$_2$-ia iš-tu mu-hi LUGAL EN-ia i-nu-ma yu-ba-u$_2$ 3 li-im **KU$_3$.BABBAR.MEŠ** iš-tu qa-ti-ia u$_3$ yi-iq-bu a-na ia-ši id-na-mi DAM-ka u$_3$ DUMU.MEŠ-ka u$_3$ lu-u$_2$ ZAG ma-

e um-ma {m}x ša x x x x x be-li$_2$-ia x ki ša qa-bi-šu x a-na IR$_3$.MEŠ TUR.MEŠ x u$_3$ 1 ME **KU$_3$.BABBAR.MEŠ** a-na LUGAL EN-ia 10 IR$_3$.MEŠ u$_3$ 10 {MUNUS}GEME$_2$.MEŠ x u$_3$ aš-šum-ma la-ma-ad

ie$_2$-ep-ša-at ki-ma ri-qi$_2$ hu-bu-li a-na ša-šu iš-tu HUR.SAG ip-pa-ṭa$_2$-ru LU$_2$.MEŠ i-na 30 **KU$_3$.BABBAR.MEŠ** u$_3$ iš-tu {m}pe$_2$-e-ia i-na 1 ME KU$_3$.BABBAR.MEŠ u$_3$ li-ma-ad a-wa-te.MEŠ IR$_3$-ka an-n

UR.SAG ip-pa-ṭa$_2$-ru LU$_2$.MEŠ i-na 30 KU$_3$.BABBAR.MEŠ u$_3$ iš-tu {m}pe$_2$-e-ia i-na 1 ME **KU$_3$.BABBAR.MEŠ** u$_3$ li-ma-ad a-wa-te.MEŠ IR$_3$-ka an-nu-ti

Search results for the transliteration KU$_3$.BABBAR.MEŠ (see previous page for the special characters). The mouse is hovering over the gray bar revealing that the RIAo project texts have 46 instances of the word.

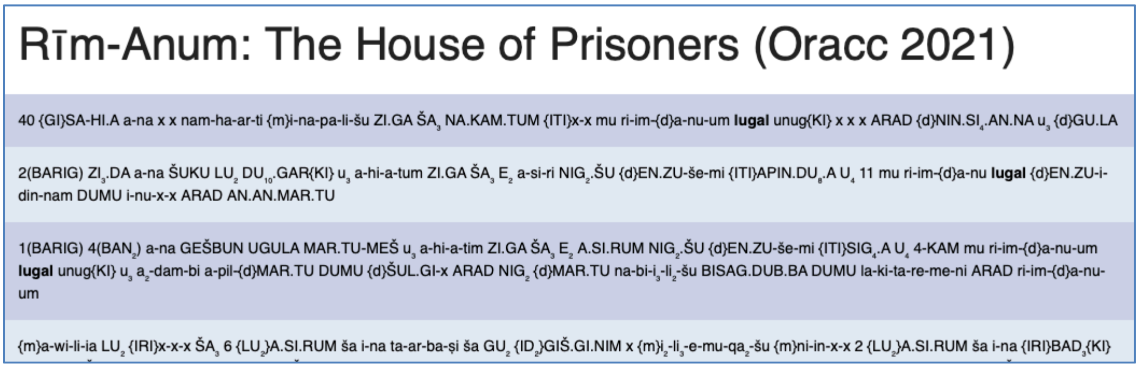
Above the results, there is a list of page numbers with which you can move to another page. Depending on your browser, you might also see a gray bar showing the division of the results according to the subprojects. You can move directly to the results of a subproject by clicking the bar but note that it can take a while to load.

By default, there will be 25 results per page and the results are grouped by the projects chosen. The way the results are shown can be changed from the blue bar (starting with the word KWIC) below the search box:

- Hits per page: choose 50, 75, 100, 500 or 1000 hits on the page (since changing a page can take a while, using a bigger number of hits per page is recommended)
- Sort within corpora:
  - **not sorted** (default) shows the instances in the order they appear in the corpus
  - **matched words(s)** sorts according to the matches
  - **left context** sorts according to the preceding word
  - **right context** sorts according to the following word
  - **random**

If you change any of these settings, you will have to redo your search by clicking the Search button.

The results can also be seen in a larger context by clicking **Show context** to the right of the list of pages (or under the list if your window is narrow). You will then see the whole text of each document in the search results. You can return to the default view with **Show KWIC**.



Rīm-Anum: The House of Prisoners (Oracc 2021)

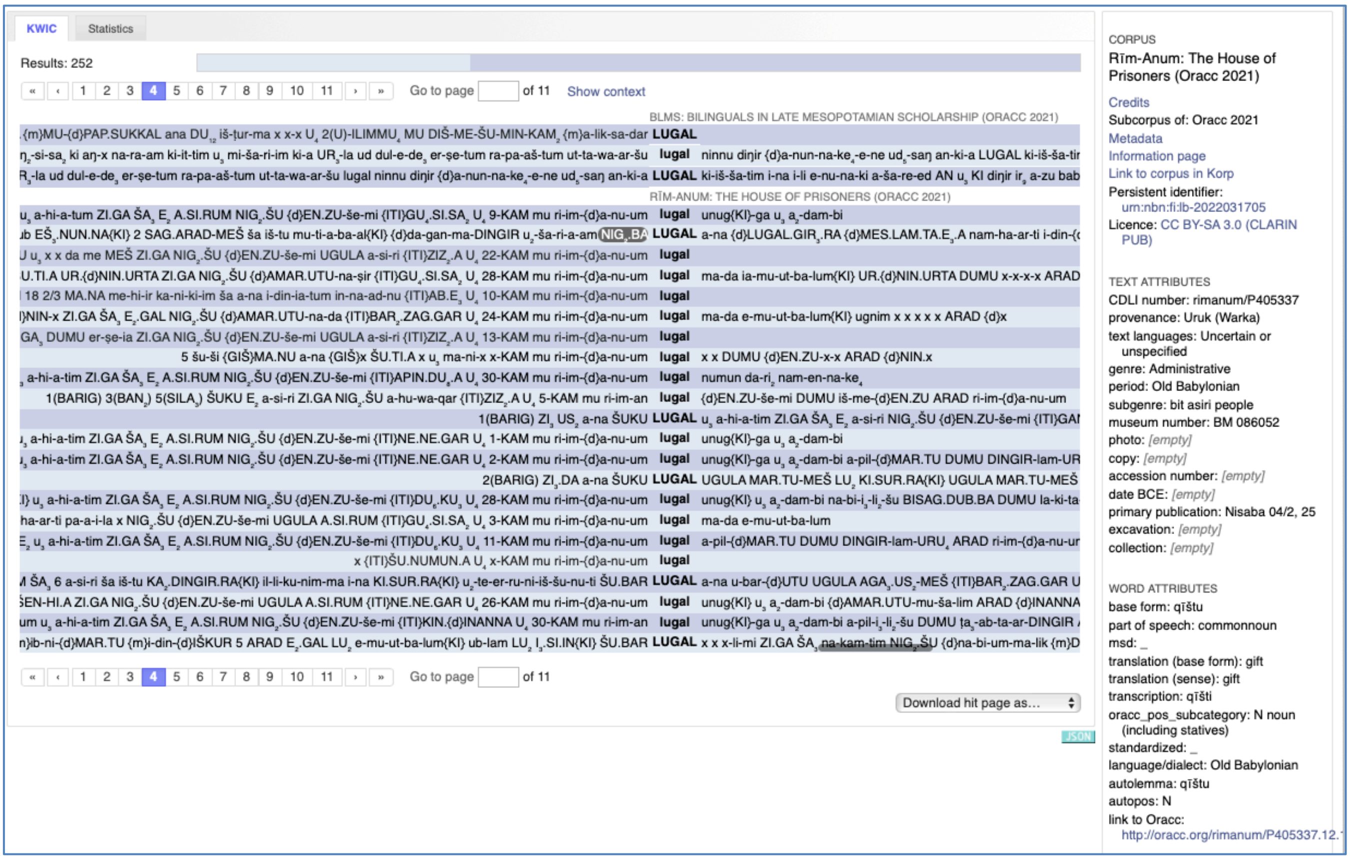
40 {GI}SA-HI.A a-na x x nam-ha-ar-ti {m}i-na-pa-li-šu ZI.GA ŠA$_3$ NA.KAM.TUM {ITI}x-x mu ri-im-{d}a-nu-um **lugal** unug{KI} x x x ARAD {d}NIN.SI$_4$.AN.NA u$_3$ {d}GU.LA

2(BARIG) ZI$_3$.DA a-na ŠUKU LU$_2$ DU$_{10}$.GAR{KI} u$_3$ a-hi-a-tum ZI.GA ŠA$_3$ E$_2$ a-si-ri NIG$_2$.ŠU {d}EN.ZU-še-mi {ITI}APIN.DU$_8$.A U$_4$ 11 mu ri-im-{d}a-nu **lugal** {d}EN.ZU-i-din-nam DUMU i-nu-x-x ARAD AN.AN.MAR.TU

1(BARIG) 4(BAN$_2$) a-na GEŠBUN UGULA MAR.TU-MEŠ u$_3$ a-hi-a-tim ZI.GA ŠA$_3$ E$_2$ A.SI.RUM NIG$_2$.ŠU {d}EN.ZU-še-mi {ITI}SIG$_4$.A U$_4$ 4-KAM mu ri-im-{d}a-nu-um **lugal** unug{KI} u$_3$ a$_2$-dam-bi a-pil-{d}MAR.TU DUMU {d}ŠUL.GI-x ARAD NIG$_2$ {d}MAR.TU na-bi-i$_3$-li$_2$-šu BISAG.DUB.BA DUMU la-ki-ta-re-me-ni ARAD ri-im-{d}a-nu-um

{m}a-wi-li-ia LU$_2$ {IRI}x-x-x ŠA$_3$ 6 {LU$_2$}A.SI.RUM ša i-na ta-ar-ba-ṣi ša GU$_2$ {ID$_2$}GIŠ.GI.NIM x {m}i$_2$-li$_3$-e-mu-qa$_2$-šu {m}ni-in-x-x 2 {LU$_2$}A.SI.RUM ša i-na {IRI}BAD$_3${KI}

When you click any word (i.e. not only the searched word) in the result list, you can see information about the corpus, word, and document in the **sidebar**.



Results: 252

« ‹ 1 2 3 **4** 5 6 7 8 9 10 11 › » Go to page □ of 11 Show context

		BLMS: BILINGUALS IN LATE MESOPOTAMIAN SCHOLARSHIP (ORACC 2021)
.{m}MU-{d}PAP.SUKKAL ana DU$_{12}$ iš-ṭur-ma x x-x U$_4$ 2(U)-ILIMMU$_4$ MU DIŠ-ME-ŠU-MIN-KAM$_2$ {m}a-lik-sa-dar	**LUGAL**	
ŋ$_2$-si-sa$_2$ ki aŋ-x na-ra-am ki-it-tim u$_3$ mi-ša-ri-im ki-a UR$_3$-la ud dul-e-de$_3$ er-ṣe-tum ra-pa-aš-tum ut-ta-wa-ar-šu	**lugal**	ninnu diŋir {d}a-nun-na-ke$_4$-e-ne ud$_5$-saŋ an-ki-a LUGAL ki-iš-ša-tir
R$_3$-la ud dul-e-de$_3$ er-ṣe-tum ra-pa-aš-tum ut-ta-wa-ar-šu lugal ninnu diŋir {d}a-nun-na-ke$_4$-e-ne ud$_5$-saŋ an-ki-a	**LUGAL**	ki-iš-ša-tim i-na i-li e-nu-na-ki a-ša-re-ed AN u$_3$ KI diŋir ir$_9$ a-zu bab
		RĪM-ANUM: THE HOUSE OF PRISONERS (ORACC 2021)
u$_3$ a-hi-a-tum ZI.GA ŠA$_3$ E$_2$ A.SI.RUM NIG$_2$.ŠU {d}EN.ZU-še-mi {ITI}GU$_4$.SI.SA$_2$ U$_4$ 9-KAM mu ri-im-{d}a-nu-um	**lugal**	unug{KI}-ga u$_3$ a$_2$-dam-bi
ub EŠ$_3$.NUN.NA{KI} 2 SAG.ARAD-MEŠ ša iš-tu mu-ti-a-ba-al{KI} {d}da-gan-ma-DINGIR u$_2$-ša-ri-a-am NIG$_2$.BA	**LUGAL**	a-na {d}LUGAL.GIR$_3$.RA {d}MES.LAM.TA.E$_3$.A nam-ha-ar-ti i-din-{(
U u$_3$ x x da me MEŠ ZI.GA NIG$_2$.ŠU {d}EN.ZU-še-mi UGULA a-si-ri {ITI}ZIZ$_2$.A U$_4$ 22-KAM mu ri-im-{d}a-nu-um	**lugal**	
U.TI.A UR.{d}NIN.URTA ZI.GA NIG$_2$.ŠU {d}AMAR.UTU-na-ṣir {ITI}GU$_4$.SI.SA$_2$ U$_4$ 28-KAM mu ri-im-{d}a-nu-um	**lugal**	ma-da ia-mu-ut-ba-lum{KI} UR.{d}NIN.URTA DUMU x-x-x-x ARAD
18 2/3 MA.NA me-hi-ir ka-ni-ki-im ša a-na i-din-ia-tum in-na-ad-nu {ITI}AB.E$_3$ U$_4$ 10-KAM mu ri-im-{d}a-nu-um	**lugal**	
}NIN-x ZI.GA ŠA$_3$ E$_2$.GAL NIG$_2$.ŠU {d}AMAR.UTU-na-da {ITI}BAR$_2$.ZAG.GAR U$_4$ 24-KAM mu ri-im-{d}a-nu-um	**lugal**	ma-da e-mu-ut-ba-lum{KI} ugnim x x x x x ARAD {d}x
GA$_3$ DUMU er-ṣe-ia ZI.GA NIG$_2$.ŠU {d}EN.ZU-še-mi UGULA a-si-ri {ITI}ZIZ$_2$.A U$_4$ 13-KAM mu ri-im-{d}a-nu-um	**lugal**	
5 šu-ši {GIŠ}MA.NU a-na {GIŠ}x ŠU.TI.A x u$_3$ ma-ni-x x-KAM mu ri-im-{d}a-nu-um	**lugal**	x x DUMU {d}EN.ZU-x-x ARAD {d}NIN.x
$_3$ a-hi-a-tim ZI.GA ŠA$_3$ E$_2$ A.SI.RUM NIG$_2$.ŠU {d}EN.ZU-še-mi {ITI}APIN.DU$_8$.A U$_4$ 30-KAM mu ri-im-{d}a-nu-um	**lugal**	numun da-ri$_2$ nam-en-na-ke$_4$
1(BARIG) 3(BAN$_2$) 5(SILA$_3$) ŠUKU E$_2$ a-si-ri ZI.GA NIG$_2$.ŠU a-hu-wa-qar {ITI}ZIZ$_2$.A U$_4$ 5-KAM mu ri-im-an	**lugal**	{d}EN.ZU-še-mi DUMU iš-me-{d}EN.ZU ARAD ri-im-{d}a-nu-um
1(BARIG) ZI$_3$ US$_2$ a-na ŠUKU	**LUGAL**	u$_3$ a-hi-a-tim ZI.GA ŠA$_3$ E$_2$ a-si-ri NIG$_2$.ŠU {d}EN.ZU-še-mi {ITI}GAN
$_3$ a-hi-a-tim ZI.GA ŠA$_3$ E$_2$ A.SI.RUM NIG$_2$.ŠU {d}EN.ZU-še-mi {ITI}NE.NE.GAR U$_4$ 1-KAM mu ri-im-{d}a-nu-um	**lugal**	unug{KI}-ga u$_3$ a$_2$-dam-bi
$_3$ a-hi-a-tim ZI.GA ŠA$_3$ E$_2$ A.SI.RUM NIG$_2$.ŠU {d}EN.ZU-še-mi {ITI}NE.NE.GAR U$_4$ 2-KAM mu ri-im-{d}a-nu-um	**lugal**	unug{KI}-ga u$_3$ a$_2$-dam-bi a-pil-{d}MAR.TU DUMU DINGIR-lam-UR
2(BARIG) ZI$_3$.DA a-na ŠUKU	**LUGAL**	UGULA MAR.TU-MEŠ LU$_2$ KI.SUR.RA{KI} UGULA MAR.TU-MEŠ
I} u$_3$ a-hi-a-tim ZI.GA ŠA$_3$ E$_2$ A.SI.RUM NIG$_2$.ŠU {d}EN.ZU-še-mi {ITI}DU$_6$.KU$_3$ U$_4$ 28-KAM mu ri-im-{d}a-nu-um	**lugal**	unug{KI} u$_3$ a$_2$-dam-bi na-bi-i$_3$-li$_2$-šu BISAG.DUB.BA DUMU la-ki-ta
ha-ar-ti pa-a-i-la x NIG$_2$.ŠU {d}EN.ZU-še-mi UGULA A.SI.RUM {ITI}GU$_4$.SI.SA$_2$ U$_4$ 3-KAM mu ri-im-{d}a-nu-um	**lugal**	ma-da e-mu-ut-ba-lum
E$_2$ u$_3$ a-hi-a-tim ZI.GA ŠA$_3$ E$_2$ A.SI.RUM NIG$_2$.ŠU {d}EN.ZU-še-mi {ITI}DU$_6$.KU$_3$ U$_4$ 11-KAM mu ri-im-{d}a-nu-um	**lugal**	a-pil-{d}MAR.TU DUMU DINGIR-lam-URU$_4$ ARAD ri-im-{d}a-nu-ur
x {ITI}ŠU.NUMUN.A U$_4$ x-KAM mu ri-im-{d}a-nu-um	**lugal**	
M ŠA$_3$ 6 a-si-ri ša iš-tu KA$_2$.DINGIR.RA{KI} il-li-ku-nim-ma i-na KI.SUR.RA{KI} u$_2$-te-er-ru-ni-iš-šu-nu-ti ŠU.BAR	**LUGAL**	a-na u-bar-{d}UTU UGULA AGA$_3$.US$_2$-MEŠ {ITI}BAR$_2$.ZAG.GAR U
ŠEN-HI.A ZI.GA NIG$_2$.ŠU {d}EN.ZU-še-mi UGULA A.SI.RUM {ITI}NE.NE.GAR U$_4$ 26-KAM mu ri-im-{d}a-nu-um	**lugal**	unug{KI} u$_3$ a$_2$-dam-bi {d}AMAR.UTU-mu-ša-lim ARAD {d}INANNA
um u$_3$ a-hi-a-tim ZI.GA ŠA$_3$ E$_2$ A.SI.RUM NIG$_2$.ŠU {d}EN.ZU-še-mi {ITI}KIN.{d}INANNA U$_4$ 30-KAM mu ri-im-an	**lugal**	unug{KI}-ga u$_3$ a$_2$-dam-bi a-pil-i$_3$-li$_2$-šu DUMU ṭa$_3$-ab-ta-ar-DINGIR
n}ib-ni-{d}MAR.TU {m}i-din-{d}IŠKUR 5 ARAD E$_2$.GAL LU$_2$ e-mu-ut-ba-lum{KI} ub-lam LU$_2$ I$_3$.SI.IN{KI} ŠU.BAR	**LUGAL**	x x x-li-mi ZI.GA ŠA$_3$ na-kam-tim NIG$_2$.ŠU {d}na-bi-um-ma-lik {m}D

« ‹ 1 2 3 **4** 5 6 7 8 9 10 11 › » Go to page □ of 11

CORPUS
Rīm-Anum: The House of Prisoners (Oracc 2021)
Credits
Subcorpus of: Oracc 2021
Metadata
Information page
Link to corpus in Korp
Persistent identifier: urn:nbn:fi:lb-2022031705
Licence: CC BY-SA 3.0 (CLARIN PUB)

TEXT ATTRIBUTES
CDLI number: rimanum/P405337
provenance: Uruk (Warka)
text languages: Uncertain or unspecified
genre: Administrative
period: Old Babylonian
subgenre: bit asiri people
museum number: BM 086052
photo: *[empty]*
copy: *[empty]*
accession number: *[empty]*
date BCE: *[empty]*
primary publication: Nisaba 04/2, 25
excavation: *[empty]*
collection: *[empty]*

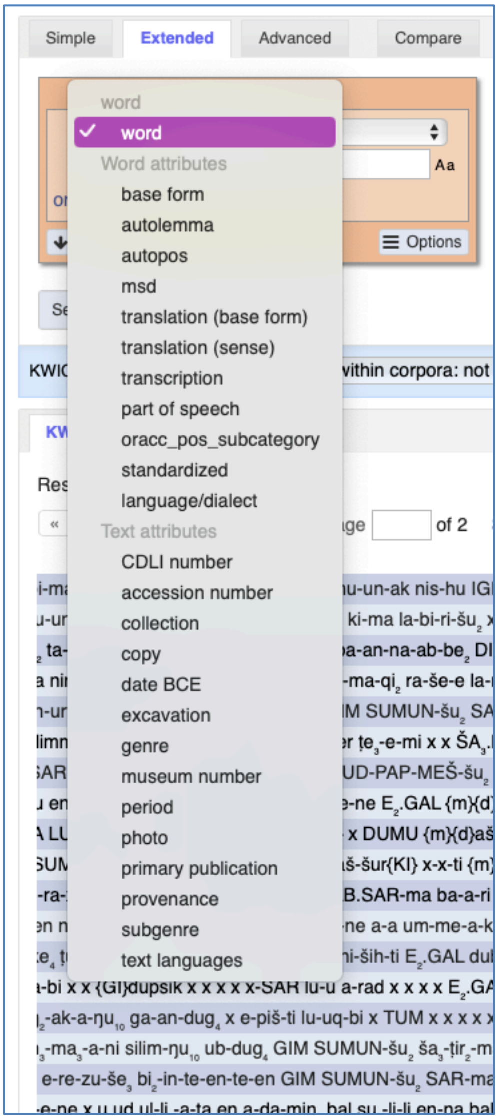
WORD ATTRIBUTES
base form: qīštu
part of speech: commonnoun
msd: _
translation (base form): gift
translation (sense): gift
transcription: qīšti
oracc_pos_subcategory: N noun (including statives)
standardized: _
language/dialect: Old Babylonian
autolemma: qīštu
autopos: N
link to Oracc: http://oracc.org/rimanum/P405337.12.1

CORPUS section has several links:

Credits > the corpus in Oracc
Metadata > MetaShare with information on the whole Oracc in Korp data
information page > Oracc in Language Bank of Finland
Link to corpus in Korp
Persistent identifier of the MetaShare page
Licence

TEXT ATTRIBUTES section has information on the document according to the information available in ORACC (see Section *4. Extended Search* for the attributes and their explanation).
WORD ATTRIBUTES section has information on the word selected (see Section *4. Extended Search* for the attributes and their explanation).

4. EXTENDED SEARCH



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

- word attributes
- text attributes

In the orange box, click the field that says “word” and you get a dropdown list of attributes. For the Oracc explanations of the attributes see here.

The word attributes are:

- word = transliteration
- base form = dictionary headword in *A Concise Dictionary of Akkadian* (The Citation Form in Oracc)
  - 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]
- autolemma = base form produced with a tool called BabyLemmatizer at the University of Helsinki (https://github.com/asahala/BabyLemmatizer)
- autopos = part-of-speech tag produced with BabyLemmatizer
- msd = morphosyntactic descriptors for Sumerian
- translation (base form) = first translation of the dictionary form in CDA (The Guide Word in Oracc)
- translation (sense) = an optional meaning of the word in the context (Sense in Oracc)
- transcription = the normalization for the form (Normalization in Oracc)
- part of speech = the basic grammar of words, i.e. word classes, pronouns, and other tags used in Oracc.
- oracc_pos_subcategory = An optional part-of-speech tied to the current syntactic context (The Effective Part-of-Speech tag in Oracc)
- standardized = normalized form of names of gods and places. See https://github.com/anee-helsinki/OraccInKorp/tree/master/VersionMay2019 for the lists used (the same lists have been used for the Oracc in Korp 2021 version). Not in Oracc.
- language/dialect

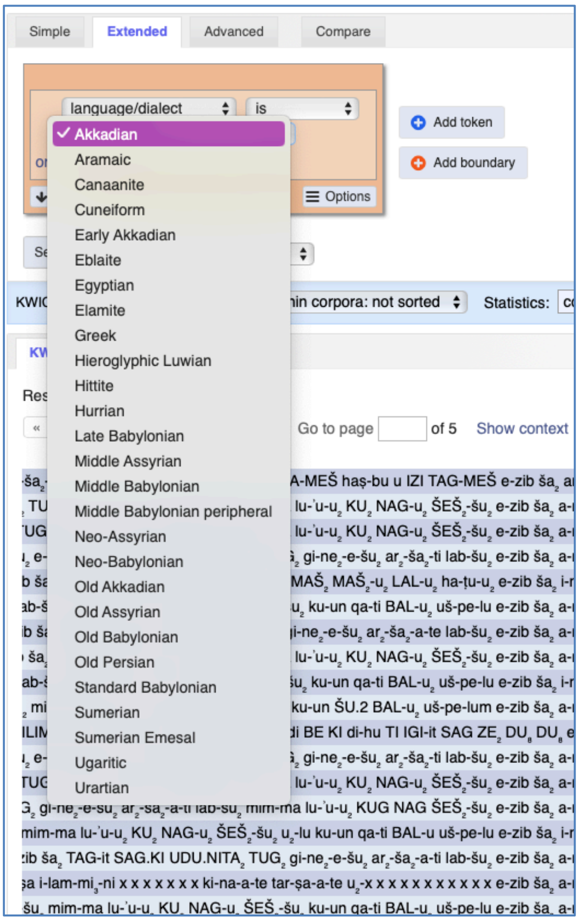
The text attributes are:

- CDLI number: P-numbers and their matching designations have been assigned by the CDLI. Q-numbers refer to Oracc’s central Q-catalogue of composite texts.
- accession number
- collection
- copy (of)
- date BCE
- excavation
- genre
- museum number
- period
- photo
- primary publication
- provenance
- subgenre (as defined in Oracc)
- text languages, i.e. languages assigned to the whole text

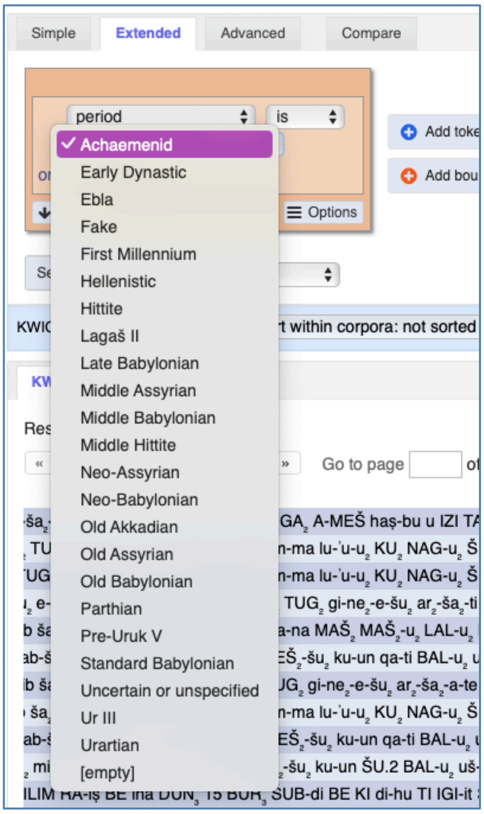
[1] Note that the way compound words are written in different projects in Oracc varies. The joined words which in Oracc have separate translations, word classes, etc. have in Korp been joined with “&&”. Sometimes compound words in Oracc have been defined as one word and the parts have been joined with “-“, 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

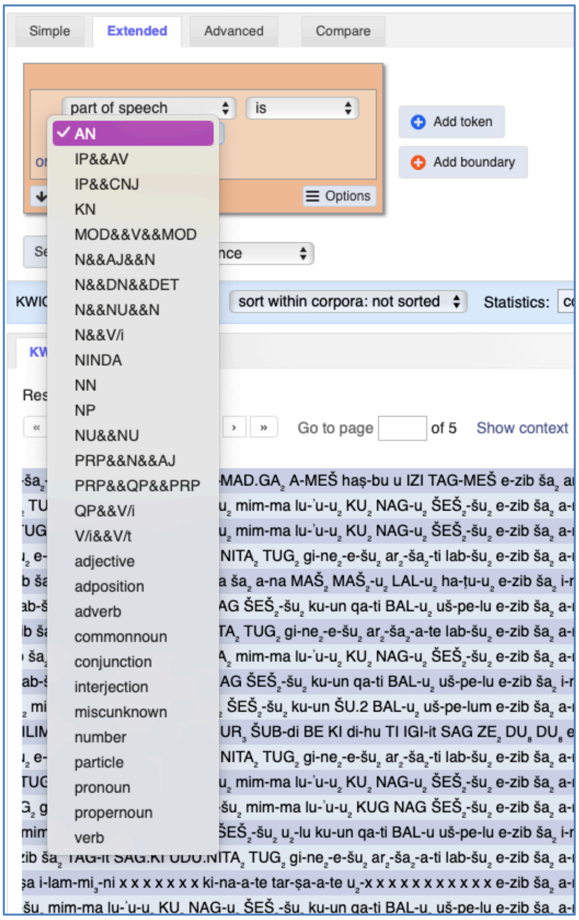
- language/dialect



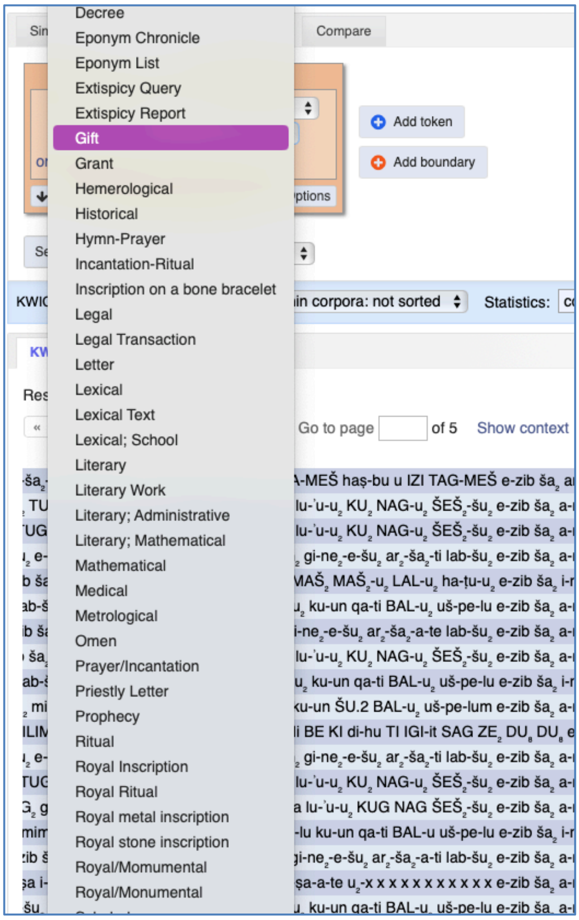
- period



- part-of-speech



- genre



In some of the other cases, you get a list when you click the text box and start writing. In other cases, such as the CDLI-number, you have to write the value you are looking for in the field. In the sidebar, you can find hints on what the different attributes can contain.

In all cases, you can choose whether you want that the search term

- is
- is not

in the results.

When there is no dropdown list, you can also specify if you want that the searched word

- begins with
- contains
- ends with

the given value.

or that it

- is
- is not

the regular expression (regexp) given.

The regular expression is explained in the last section Advanced Search and is only of interest to those who want to do more complicated searches. It is possible to do complicated searches already using the Extended search even without the regular expressions (see next section).

Note that although Korp in Extended search gives you an option (right of the search button) of searching "within sentence", in practice the searches in Oracc corpora are all within the whole text.

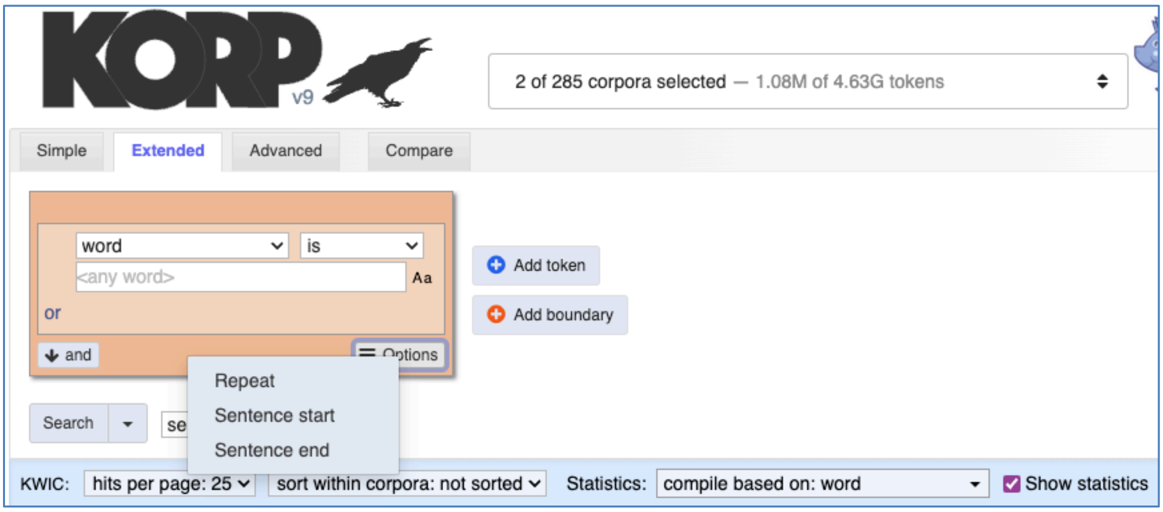
5. MAKING MORE COMPLICATED QUERIES IN THE EXTENDED SEARCH

Click the ☰ Options button on the right lower corner of the orange search box to find more options

- **Repeat** – find words that occur several times right after each other (for how to use this, see Search example 2)
- **Sentence start** – find word instances that are only at the beginning of a text
- **Sentence end** – find word instances that are only at the end of a text

These sentence boundaries will open a Boundary unit box before or after the search box

These can also be added by clicking the ⊕ Add boundary button and choosing ←**first** or **last**→



KWIC: hits per page: 25 | sort within corpora: not sorted | Statistics: compile based on: word | ☑ Show statistics

The orange box is the heart of the Extended search and represents one word.

Click **or** in the left bottom corner of the search box and specify another word to get the attestations of both

- e.g. find all words where word (i.e. transliteration) is either *lugal* or *LUGAL*

Click the ↓ and button in the lower part of the box and specify an attribute that the word must have to narrow down your search (note that searching for the same attribute in both options will not give any results)

- e.g. find words the normalization (i.e. transcription) of which is *kaspa* and the word (i.e. transliteration) is not KU$_3$.BABBAR

Click the ⊕ Add token sign to the right of the box and specify another word that has to follow the first one

- e.g. base form (i.e. dictionary form) is *šarru* + base form is *dannu*

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

You can remove additional words by clicking the x in the top left corner of an orange box. The word attributes added by clicking or-button can be removed by clicking the – on the left side of the box. These options are only available when words or attributes have been added.

Search example 1

Words (i.e. transliterations) that begin with the letter "a" and where the part of speech is verb and the language/dialect is Akkadian.

DO:

word begins with a

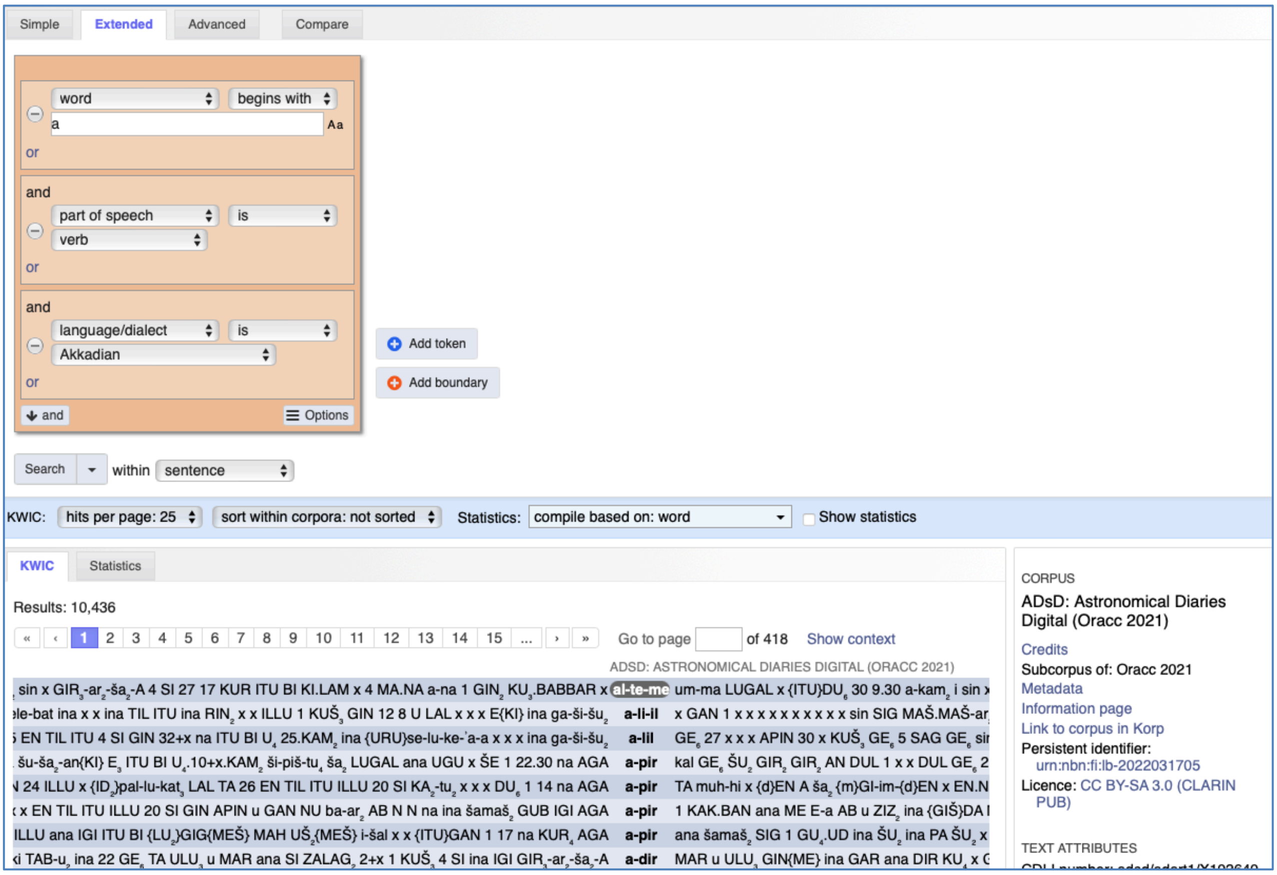
and

part of speech is verb (from the dropdown)

and

language/dialect is Akkadian (from the dropdown)

Search



word | begins with | a

and part of speech | is | verb

and language/dialect | is | Akkadian

Search within sentence

KWIC: hits per page: 25 | sort within corpora: not sorted | Statistics: compile based on: word | □ Show statistics

Results: 10,436

Go to page of 418

ADSD: ASTRONOMICAL DIARIES DIGITAL (ORACC 2021)

sin x GIR$_3$-ar$_2$-ša$_2$-A 4 SI 27 17 KUR ITU BI KI.LAM x 4 MA.NA a-na 1 GIN$_2$ KU$_3$.BABBAR x	al-te-me	um-ma LUGAL x {ITU}DU$_6$ 30 9.30 a-kam$_2$ i sin x
le-bat ina x x ina TIL ITU ina RIN$_2$ x x ILLU 1 KUŠ$_3$ GIN 12 8 U LAL x x x E{KI} ina ga-ši-šu$_2$	a-li-il	x GAN 1 x x x x x x x x x x sin SIG MAŠ.MAŠ-ar
EN TIL ITU 4 SI GIN 32+x na ITU BI U$_4$ 25.KAM$_2$ ina {URU}se-lu-ke-ʾa-a x x x ina ga-ši-šu$_2$	a-lil	GE$_6$ 27 x x x APIN 30 x KUŠ$_3$ GE$_6$ 5 SAG GE$_6$ sir
šu-ša$_2$-an{KI} E$_3$ ITU BI U$_4$.10+x.KAM$_2$ ši-piš-tu$_4$ ša$_2$ LUGAL ana UGU x ŠE 1 22.30 na AGA	a-pir	kal GE$_6$ ŠU$_2$ GIR$_2$ GIR$_2$ AN DUL 1 x x DUL GE$_6$ 2
24 ILLU x {ID$_2$}pal-lu-kat$_3$ LAL TA 26 EN TIL ITU ILLU 20 SI KA$_2$-tu$_2$ x x x DU$_6$ 1 14 na AGA	a-pir	TA muh-hi x {d}EN A ša$_2$ {m}GI-im-{d}EN x EN.N
x EN TIL ITU ILLU 20 SI GIN APIN u GAN NU ba-ar$_2$ AB N N na ina šamaš$_2$ GUB IGI AGA	a-pir	1 KAK.BAN ana ME E-a AB u ZIZ$_2$ ina {GIŠ}DA
ILLU ana IGI ITU BI {LU$_2$}GIG{MEŠ} MAH UŠ$_2${MEŠ} i-šal x x {ITU}GAN 1 17 na KUR$_4$ AGA	a-pir	ana šamaš$_2$ SIG 1 GU$_4$.UD ina ŠU$_2$ ina PA ŠU$_2$ x
i TAB-u$_2$ ina 22 GE$_6$ TA ULU$_3$ u MAR ana SI ZALAG$_2$ 2+x 1 KUŠ$_3$ 4 SI ina IGI GIR$_3$-ar$_2$-ša$_2$-A	a-dir	MAR u ULU$_3$ GIN{ME} ina GAR ana DIR KU$_4$ x G

CORPUS

ADsD: Astronomical Diaries Digital (Oracc 2021)

Subcorpus of: Oracc 2021

Persistent identifier: urn:nbn:fi:lb-2022031705

Licence: CC BY-SA 3.0 (CLARIN PUB)

TEXT ATTRIBUTES

Search example 2

Texts from the Neo-Assyrian period featuring base form (i.e. dictionary form) of the divine name “Aššur” and the translation “king” with no more than 8 words in between.

DO:

base form is Aššur

and

oracc_pos_subcategory is DN Divine Name

and

period is Neo-Assyrian

+Add token

word is <any word>

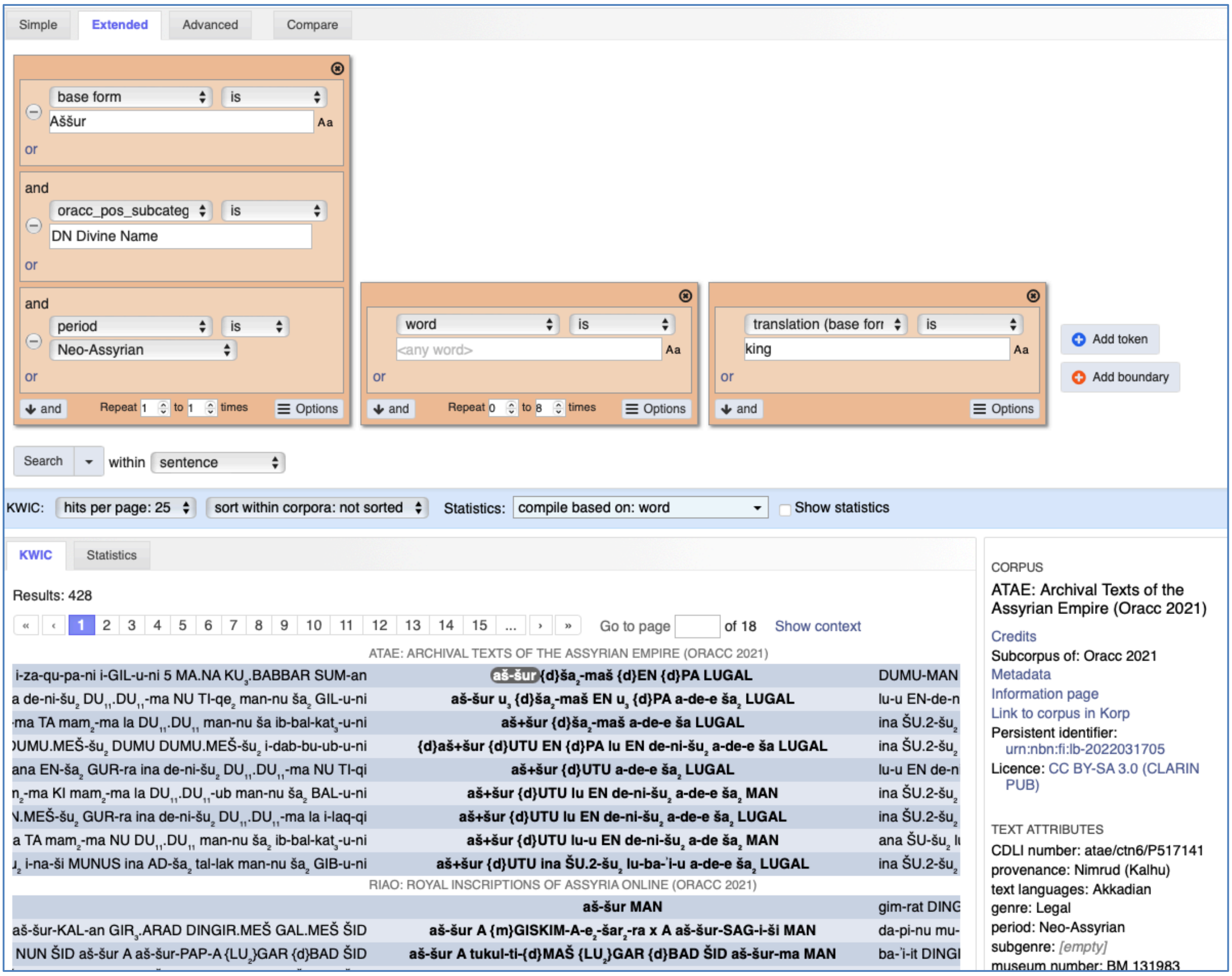
Options: Repeat

Repeat 0 to 8 times

+Add token

translation (base form) is king

Search



base form is Aššur Aa
or
and
oracc_pos_subcateg is DN Divine Name
or
and
period is Neo-Assyrian
or
and Repeat 1 to 1 times Options

word is <any word> Aa
or
and Repeat 0 to 8 times Options

translation (base forı is king Aa
or
and Options

Search within sentence

KWIC: hits per page: 25 | sort within corpora: not sorted | Statistics: compile based on: word | □ Show statistics

Results: 428

« ‹ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... › » Go to page of 18 Show context

ATAE: ARCHIVAL TEXTS OF THE ASSYRIAN EMPIRE (ORACC 2021)

i-za-qu-pa-ni i-GIL-u-ni 5 MA.NA KU$_3$.BABBAR SUM-an	**aš-šur {d}ša$_2$-maš {d}EN {d}PA LUGAL**	DUMU-MAN
a de-ni-šu$_2$ DU$_{11}$.DU$_{11}$-ma NU TI-qe$_2$ man-nu ša$_2$ GIL-u-ni	**aš-šur u$_3$ {d}ša$_2$-maš EN u$_3$ {d}PA a-de-e ša$_2$ LUGAL**	lu-u EN-de-n
-ma TA mam$_2$-ma la DU$_{11}$.DU$_{11}$ man-nu ša ib-bal-kat$_3$-u-ni	**aš+šur {d}ša$_2$-maš a-de-e ša LUGAL**	ina ŠU.2-šu$_2$
)UMU.MEŠ-šu$_2$ DUMU DUMU.MEŠ-šu$_2$ i-dab-bu-ub-u-ni	**{d}aš+šur {d}UTU EN {d}PA lu EN de-ni-šu$_2$ a-de-e ša LUGAL**	ina ŠU.2-šu$_2$
ana EN-ša$_2$ GUR-ra ina de-ni-šu$_2$ DU$_{11}$.DU$_{11}$-ma NU TI-qi	**aš+šur {d}UTU a-de-e ša$_2$ LUGAL**	lu-u EN de-ni
n$_2$-ma KI mam$_2$-ma la DU$_{11}$.DU$_{11}$-ub man-nu ša$_2$ BAL-u-ni	**aš+šur {d}UTU lu EN de-ni-šu$_2$ a-de-e ša$_2$ MAN**	ina ŠU.2-šu$_2$
\.MEŠ-šu$_2$ GUR-ra ina de-ni-šu$_2$ DU$_{11}$.DU$_{11}$-ma la i-laq-qi	**aš+šur {d}UTU lu EN de-ni-šu$_2$ a-de-e ša$_2$ LUGAL**	ina ŠU.2-šu$_2$
a TA mam$_2$-ma NU DU$_{11}$.DU$_{11}$ man-nu ša$_2$ ib-bal-kat$_3$-u-ni	**aš+šur {d}UTU lu-u EN de-ni-šu$_2$ a-de ša$_2$ MAN**	ana ŠU-šu$_2$ lı
ı$_2$ i-na-ši MUNUS ina AD-ša$_2$ tal-lak man-nu ša$_2$ GIB-u-ni	**aš+šur {d}UTU ina ŠU.2-šu$_2$ lu-ba-ʾi-u a-de-e ša$_2$ LUGAL**	ina ŠU.2-šu$_2$

RIAO: ROYAL INSCRIPTIONS OF ASSYRIA ONLINE (ORACC 2021)

	aš-šur MAN	gim-rat DING
aš-šur-KAL-an GIR$_3$.ARAD DINGIR.MEŠ GAL.MEŠ ŠID	**aš-šur A {m}GISKIM-A-e$_2$-šar$_2$-ra x A aš-šur-SAG-i-ši MAN**	da-pi-nu mu-
NUN ŠID aš-šur A aš-šur-PAP-A {LU$_2$}GAR {d}BAD ŠID	**aš-šur A tukul-ti-{d}MAŠ {LU$_2$}GAR {d}BAD ŠID aš-šur-ma MAN**	ba-ʾi-it DINGI

CORPUS

ATAE: Archival Texts of the Assyrian Empire (Oracc 2021)

Credits

Subcorpus of: Oracc 2021

Metadata

Information page

Link to corpus in Korp

Persistent identifier: urn:nbn:fi:lb-2022031705

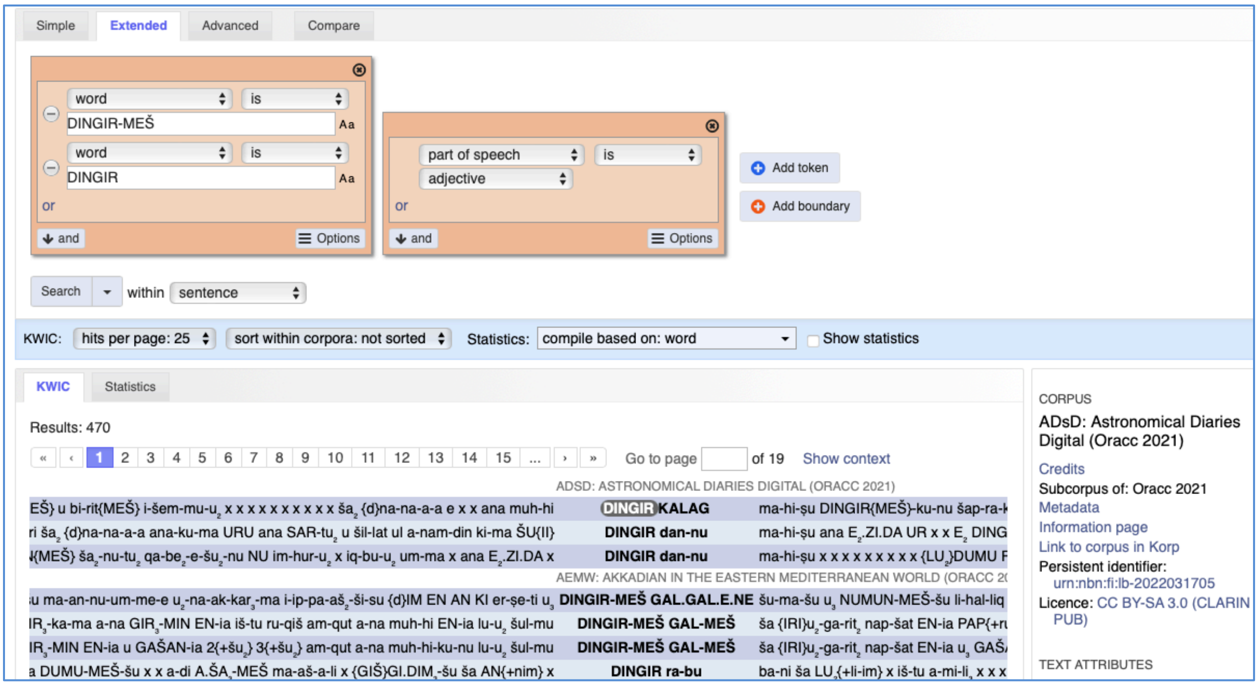
Licence: CC BY-SA 3.0 (CLARIN PUB)

TEXT ATTRIBUTES

CDLI number: atae/ctn6/P517141
provenance: Nimrud (Kalhu)
text languages: Akkadian
genre: Legal
period: Neo-Assyrian
subgenre: *[empty]*
museum number: BM 131983

Search example 3

Word (i.e. transliteration) “DINGIR-MEŠ” **or** “DINGIR” followed by (+Add token) any adjective, i.e. part of speech is adjective (from the dropdown).



word is DINGIR-MEŠ Aa
word is DINGIR Aa
or

part of speech is adjective
or

within sentence

KWIC: hits per page: 25 | sort within corpora: not sorted | Statistics: compile based on: word | □ Show statistics

Results: 470

« ‹ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... › » Go to page of 19 Show context

ADSD: ASTRONOMICAL DIARIES DIGITAL (ORACC 2021)

EŠ} u bi-rit{MEŠ} i-šem-mu-u$_2$ x x x x x x x x x x x ša$_2$ {d}na-na-a-a e x x ana muh-hi	DINGIR KALAG	ma-hi-ṣu DINGIR{MEŠ}-ku-nu šap-ra-k
ri ša$_2$ {d}na-na-a-a ana-ku-ma URU ana SAR-tu$_2$ u šil-lat ul a-nam-din ki-ma ŠU{II}	DINGIR dan-nu	ma-hi-ṣu ana E$_2$.ZI.DA UR x x E$_2$ DING
{MEŠ} ša$_2$-nu-tu$_2$ qa-be$_2$-e-šu$_2$-nu NU im-hur-u$_2$ x iq-bu-u$_2$ um-ma x ana E$_2$.ZI.DA x	DINGIR dan-nu	ma-hi-ṣu x x x x x x x x x {LU$_2$}DUMU F

AEMW: AKKADIAN IN THE EASTERN MEDITERRANEAN WORLD (ORACC 20

u ma-an-nu-um-me-e u$_2$-na-ak-kar$_3$-ma i-ip-pa-aš$_2$-ši-su {d}IM EN AN KI er-ṣe-ti u$_3$	DINGIR-MEŠ GAL.GAL.E.NE	šu-ma-šu u$_3$ NUMUN-MEŠ-šu li-hal-liq
IR$_3$-ka-ma a-na GIR$_3$-MIN EN-ia iš-tu ru-qiš am-qut a-na muh-hi EN-ia lu-u$_2$ šul-mu	DINGIR-MEŠ GAL-MEŠ	ša {IRI}u$_2$-ga-rit$_2$ nap-šat EN-ia PAP{+ru
IR$_3$-MIN EN-ia u GAŠAN-ia 2{+šu$_2$} 3{+šu$_2$} am-qut a-na muh-hi-ku-nu lu-u$_2$ šul-mu	DINGIR-MEŠ GAL-MEŠ	ša {IRI}u$_2$-ga-rit$_2$ nap-šat EN-ia u$_3$ GAŠ
a DUMU-MEŠ-šu x x a-di A.ŠA$_3$-MEŠ ma-aš-a-li x {GIŠ}GI.DIM$_4$-šu ša AN{+nim} x	DINGIR ra-bu	ba-ni ša LU$_2${+li-im} x iš-tu a-mi-li$_3$ x x x

CORPUS

ADsD: Astronomical Diaries Digital (Oracc 2021)

Credits

Subcorpus of: Oracc 2021

Metadata

Information page

Link to corpus in Korp

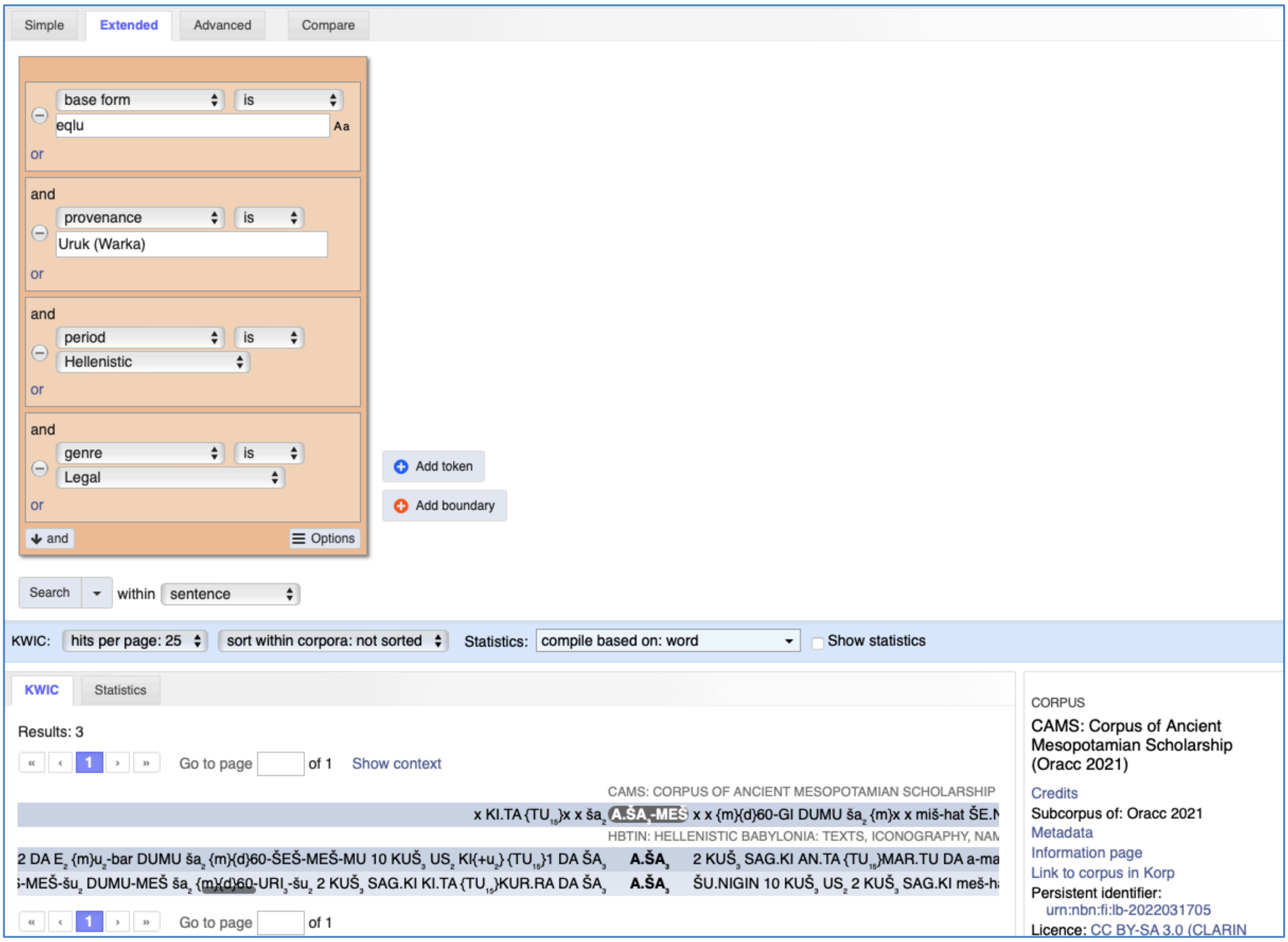
Persistent identifier: urn:nbn:fi:lb-2022031705

Licence: CC BY-SA 3.0 (CLARIN PUB)

TEXT ATTRIBUTES

Search example 4

Base form (i.e. dictionary form) “eqlu” in texts where provenance is Uruk (start writing Uruk), period Hellenistic (from the dropdown), and genre Legal (from the dropdown).



base form is eqlu Aa
or

and
provenance is Uruk (Warka)
or

and
period is Hellenistic
or

and
genre is Legal
or

within sentence

KWIC: hits per page: 25 | sort within corpora: not sorted | Statistics: compile based on: word | □ Show statistics

Results: 3

« ‹ 1 › » Go to page of 1 Show context

CAMS: CORPUS OF ANCIENT MESOPOTAMIAN SCHOLARSHIP

x KI.TA {TU$_{15}$}x x ša$_2$	A.ŠA$_3$-MEŠ	x x {m}{d}60-GI DUMU ša$_2$ {m}x x miš-hat ŠE.N

HBTIN: HELLENISTIC BABYLONIA: TEXTS, ICONOGRAPHY, NAM

2 DA E$_2$ {m}u$_2$-bar DUMU ša$_2$ {m}{d}60-ŠEŠ-MEŠ-MU 10 KUŠ$_3$ US$_2$ KI{+u$_2$} {TU$_{15}$}1 DA ŠA$_3$	A.ŠA$_3$	2 KUŠ$_3$ SAG.KI AN.TA {TU$_{15}$}MAR.TU DA a-ma
-MEŠ-šu$_2$ DUMU-MEŠ ša$_2$ {m}{d}60-URI$_3$-šu$_2$ 2 KUŠ$_3$ SAG.KI KI.TA {TU$_{15}$}KUR.RA DA ŠA$_3$	A.ŠA$_3$	ŠU.NIGIN 10 KUŠ$_3$ US$_2$ 2 KUŠ$_3$ SAG.KI meš-ha

« ‹ 1 › » Go to page of 1

CORPUS

CAMS: Corpus of Ancient Mesopotamian Scholarship (Oracc 2021)

Credits

Subcorpus of: Oracc 2021

Metadata

Information page

Link to corpus in Korp

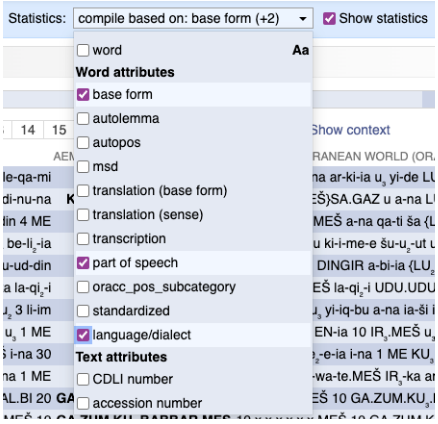
Persistent identifier: urn:nbn:fi:lb-2022031705

Licence: CC BY-SA 3.0 (CLARIN

6. STATISTICS

The Statistics tab to the right of the table called KWIC in the search results gives the number of occurrences for each matched word both in all results (Total) and within individual corpus/dataset. For the statistics to be calculated the box before the words **Show statistics** in the blue bar (starting with the word KWIC) below the search box **must be checked** before performing the search.

- 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 of what attributes were searched for.
- You can change what attribute(s) are considered by selecting the attributes you want from "Statistics: (compile based on:)" in the blue bar and then clicking Search. You can even choose several at once (when you change the selection, hit Search again):
    - e.g. search for all occurrences of words the translation of which contains "love" and base the statistics, for example, on base form, part-of-speech, and language/dialect to see what combinations there are in the results.
- You can sort the statistics according to any column by clicking the heading of that column. The columns with words will be sorted alphabetically and the numbers numerically. Clicking a second time will reverse the sorting.



Statistics: compile based on: base form (+2) ☑ Show statistics

☐ word Aa

Word attributes

☑ base form
☐ autolemma
☐ autopos
☐ msd
☐ translation (base form)
☐ translation (sense)
☐ transcription
☑ part of speech
☐ oracc_pos_subcategory
☐ standardized
☑ language/dialect

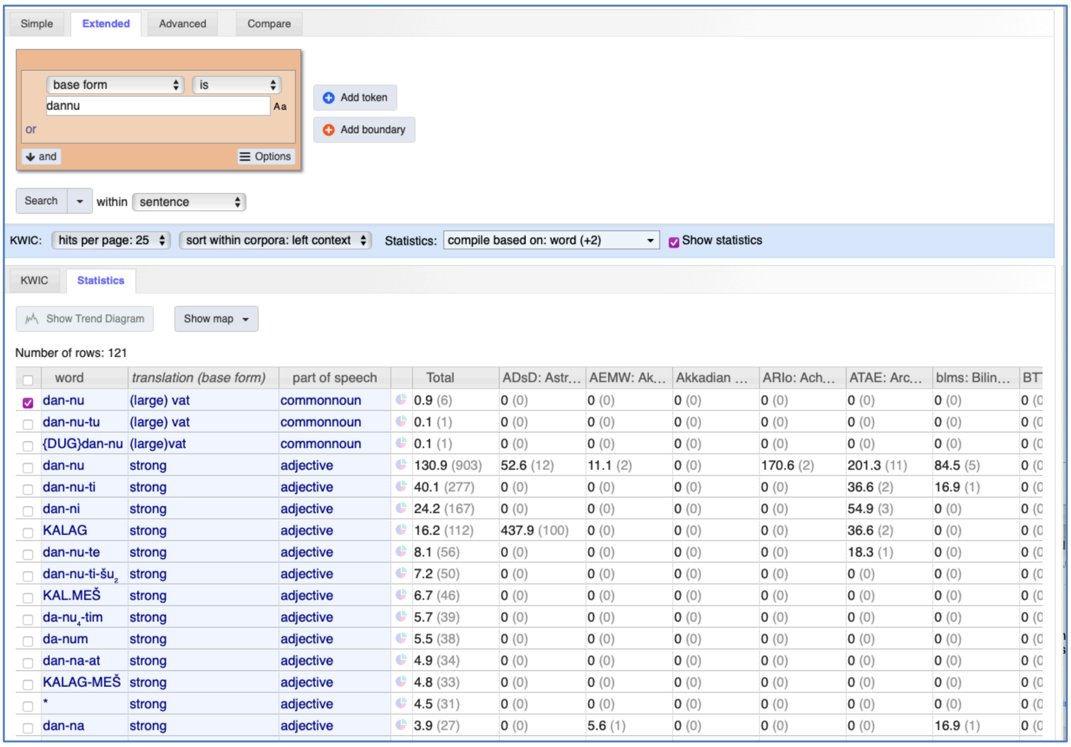
Text attributes

☐ CDLI number
☐ accession number

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

Statistics example

Search for the base form *dannu* and base the statistics on word (= transliteration), translation (of base form), and part of speech. Click Search. Go to Statistics tab and order the data according to translation by clicking the heading *translation*.



base form is dannu Aa

within sentence

KWIC: hits per page: 25 sort within corpora: left context Statistics: compile based on: word (+2) ☑ Show statistics

Number of rows: 121

☐	word	*translation (base form)*	part of speech		Total	ADsD: Astr…	AEMW: Ak…	Akkadian …	ARIo: Ach…	ATAE: Arc…	blms: Bilin…	BT
☑	dan-nu	(large) vat	commonnoun		0.9 (6)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0
☐	dan-nu-tu	(large) vat	commonnoun		0.1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0
☐	{DUG}dan-nu	(large)vat	commonnoun		0.1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0
☐	dan-nu	strong	adjective		130.9 (903)	52.6 (12)	11.1 (2)	0 (0)	170.6 (2)	201.3 (11)	84.5 (5)	0 (0
☐	dan-nu-ti	strong	adjective		40.1 (277)	0 (0)	0 (0)	0 (0)	0 (0)	36.6 (2)	16.9 (1)	0 (0
☐	dan-ni	strong	adjective		24.2 (167)	0 (0)	0 (0)	0 (0)	0 (0)	54.9 (3)	0 (0)	0 (0
☐	KALAG	strong	adjective		16.2 (112)	437.9 (100)	0 (0)	0 (0)	0 (0)	36.6 (2)	0 (0)	0 (0
☐	dan-nu-te	strong	adjective		8.1 (56)	0 (0)	0 (0)	0 (0)	0 (0)	18.3 (1)	0 (0)	0 (0
☐	dan-nu-ti-šu$_2$	strong	adjective		7.2 (50)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0
☐	KAL.MEŠ	strong	adjective		6.7 (46)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0
☐	da-nu$_4$-tim	strong	adjective		5.7 (39)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0
☐	da-num	strong	adjective		5.5 (38)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0
☐	dan-na-at	strong	adjective		4.9 (34)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0
☐	KALAG-MEŠ	strong	adjective		4.8 (33)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0
☐	*	strong	adjective		4.5 (31)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0
☐	dan-na	strong	adjective		3.9 (27)	0 (0)	5.6 (1)	0 (0)	0 (0)	0 (0)	16.9 (1)	0 (0

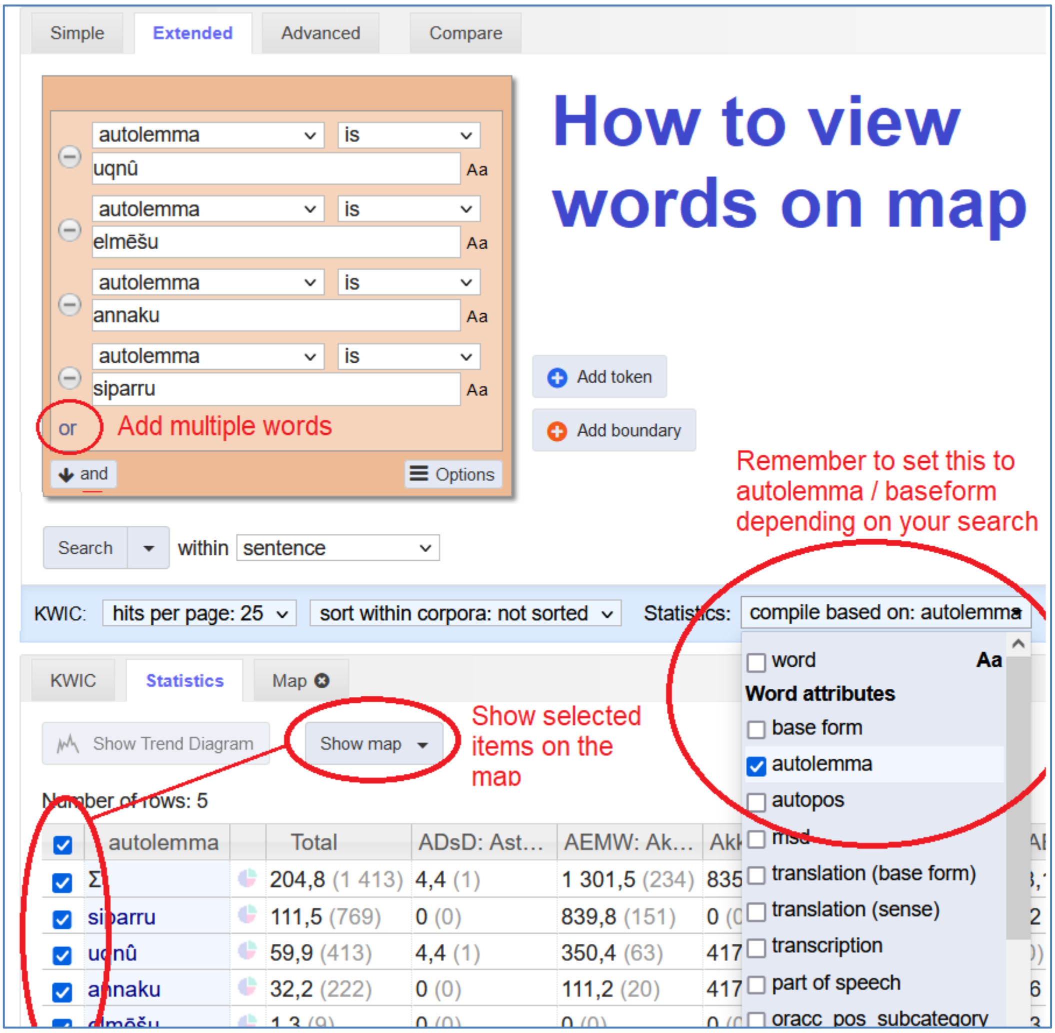
7. MAP

The Statistics tab also allows viewing the provenance of the documents with the searched words in a map. You can use it for all kinds of searches, but it is best suited for comparing words or word forms and their use in different parts of the ancient Near East (but remember that a text was not necessarily written in the place it was found in).

To use the map, make a search with the Show statistics selected. It is best, but not necessary, to compile based on just one attribute such as base form. Then **in the Statistics tab select the words you want to see in the map**. You can easily select all rows by selecting the checkbox in the heading row. If you have searched for a transliteration or based your search on several attributes, you might want to choose only some of the rows to see on the map. **Click the Show map and then select Show map**.

Map example

Search for the base forms of *uqnû (i.e. lapis lazuli)*, *elmēšu* (amber), *annaku* (tin), or *siparru* (bronze)



How to view words on map

autolemma is uqnû
autolemma is elmēšu
autolemma is annaku
autolemma is siparru

Add multiple words

within sentence

Remember to set this to autolemma / baseform depending on your search

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: autolemma

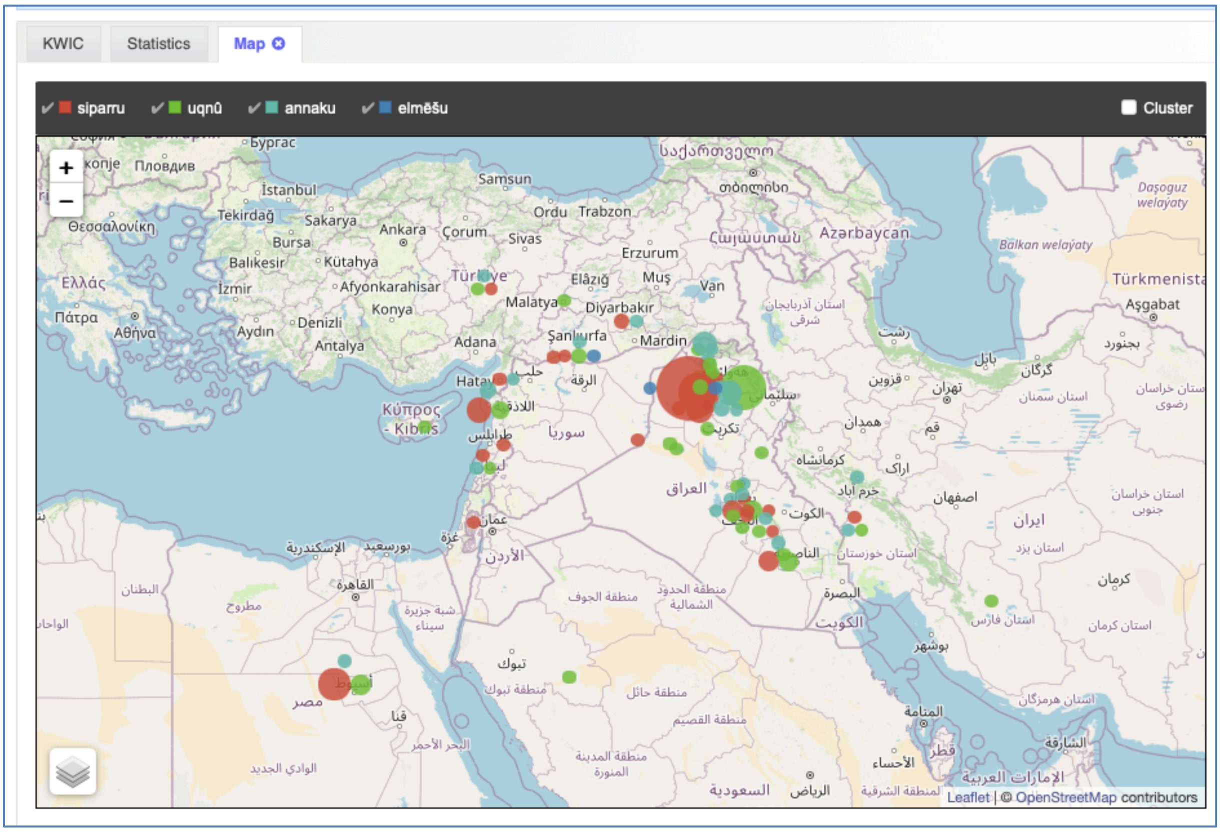
□ word
Word attributes
□ base form
☑ autolemma
□ autopos
□ msd
□ translation (base form)
□ translation (sense)
□ transcription
□ part of speech
□ oracc_pos_subcategory

Show selected items on the map

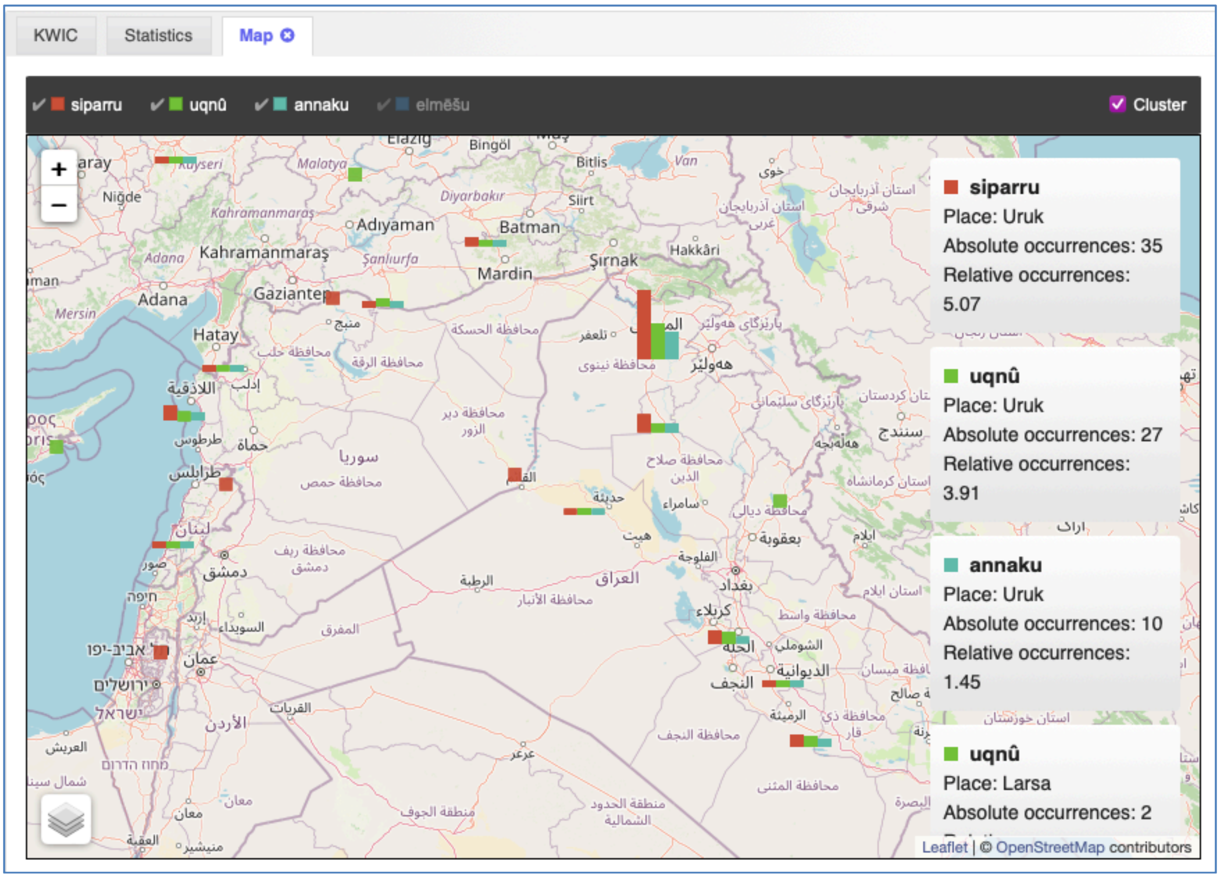
Number of rows: 5

☑	autolemma	Total	ADsD: Ast…	AEMW: Ak…	Akk
☑	Σ	204,8 (1 413)	4,4 (1)	1 301,5 (234)	835
☑	siparru	111,5 (769)	0 (0)	839,8 (151)	0 (0
☑	uqnû	59,9 (413)	4,4 (1)	350,4 (63)	417
☑	annaku	32,2 (222)	0 (0)	111,2 (20)	417
☑	elmēšu	1,3 (9)	0 (0)	0 (0)	0 (0

Example by Aleksi Sahala.



In the map view, you can select and unselect the words shown in the black bar above the map. By checking the box next to the word Cluster in the right upper corner, you will see provenances for each site as a bar diagram. By hovering with the mouse over a place/circles/bars you will see statistics for the words in that location.



🟥 **siparru**
Place: Uruk
Absolute occurrences: 35
Relative occurrences: 5.07

🟩 **uqnû**
Place: Uruk
Absolute occurrences: 27
Relative occurrences: 3.91

🟩 **annaku**
Place: Uruk
Absolute occurrences: 10
Relative occurrences: 1.45

🟩 **uqnû**
Place: Larsa
Absolute occurrences: 2

Map where the box next to Cluster is ticked and only 3 of the words is selected (in the black bar). The mouse is hovering over the place Uruk revealing the statistics for the selected words in that location.

8. COMPARING RESULTS

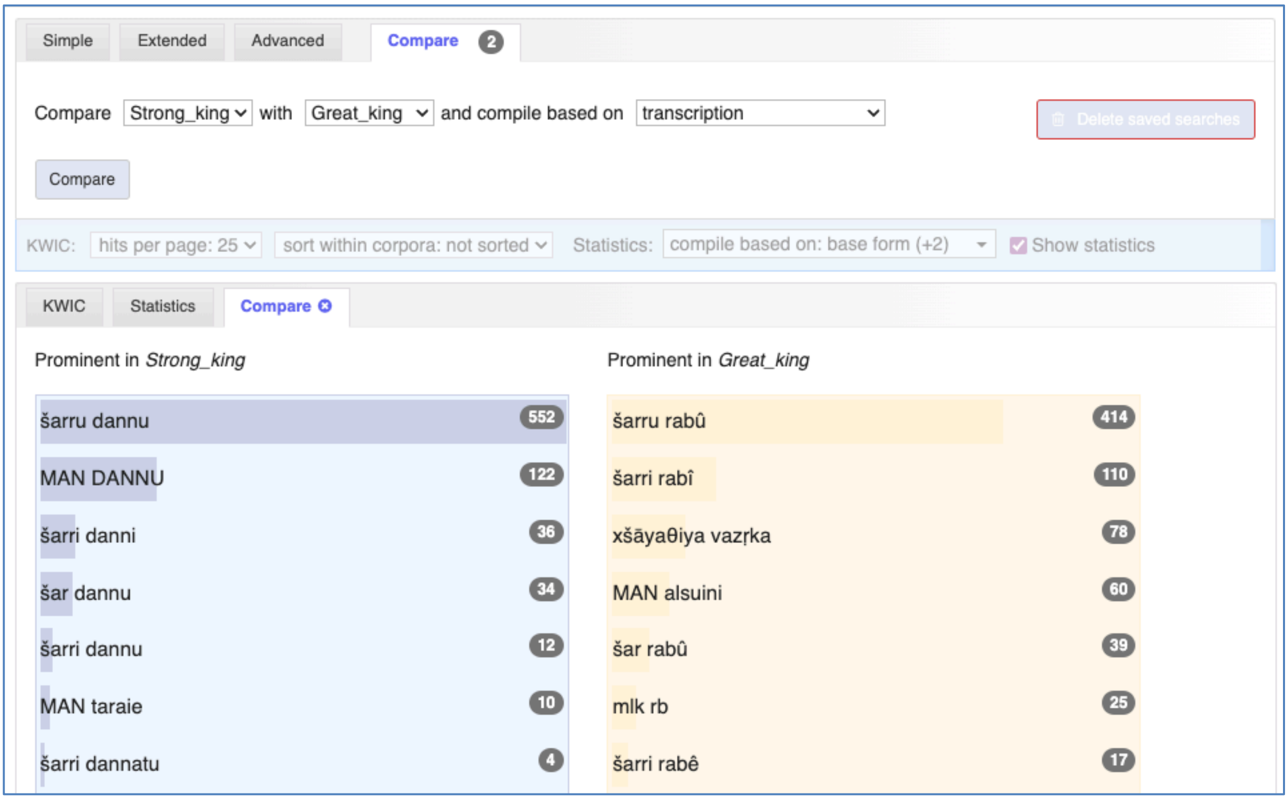
You can save searches for comparison with each other.
When you have typed in your search, click the arrow in the search button and give your search a name.
When you have at least two saved queries, go to the Compare tab (above the search box, to the right of the simple and extended search tabs) and choose from the dropdown lists:

- queries to compare
- what attribute the comparison is based on

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

Comparison example

Search for two consecutive words with the translations (sense) “king” and (+Add token) “strong” (bear in mind that adjectives follow nouns in the Akkadian word order). After the search results have loaded, click the arrow in the search button, name the search Strong_king (in the box that opens) and click Save.
Search for two consecutive words with the translations (sense) “king” and “great”. Name it Great_king and Save.
The Compare tab will now have the number 2 on its title (the number is higher if you have saved more searches). Go to the Compare tab and choose to Compare the saved query Strong_king with Great_king and compile based on transcription. Click Compare.



Compare Strong_king with Great_king and compile based on transcription

KWIC: hits per page: 25 · sort within corpora: not sorted · Statistics: compile based on: base form (+2)

Prominent in *Strong_king*		Prominent in *Great_king*	
šarru dannu	552	šarru rabû	414
MAN DANNU	122	šarri rabî	110
šarri danni	36	xšāyaθiya vazṛka	78
šar dannu	34	MAN alsuini	60
šarri dannu	12	šar rabû	39
MAN taraie	10	mlk rb	25
šarri dannatu	4	šarri rabê	17

Comparison of the number of instances of the different transcriptions of the phrases ‘strong king’ and ‘great king’ in Oracc in korp 2021 (excluding DCCLT, EPSD2, and ETCSRI).

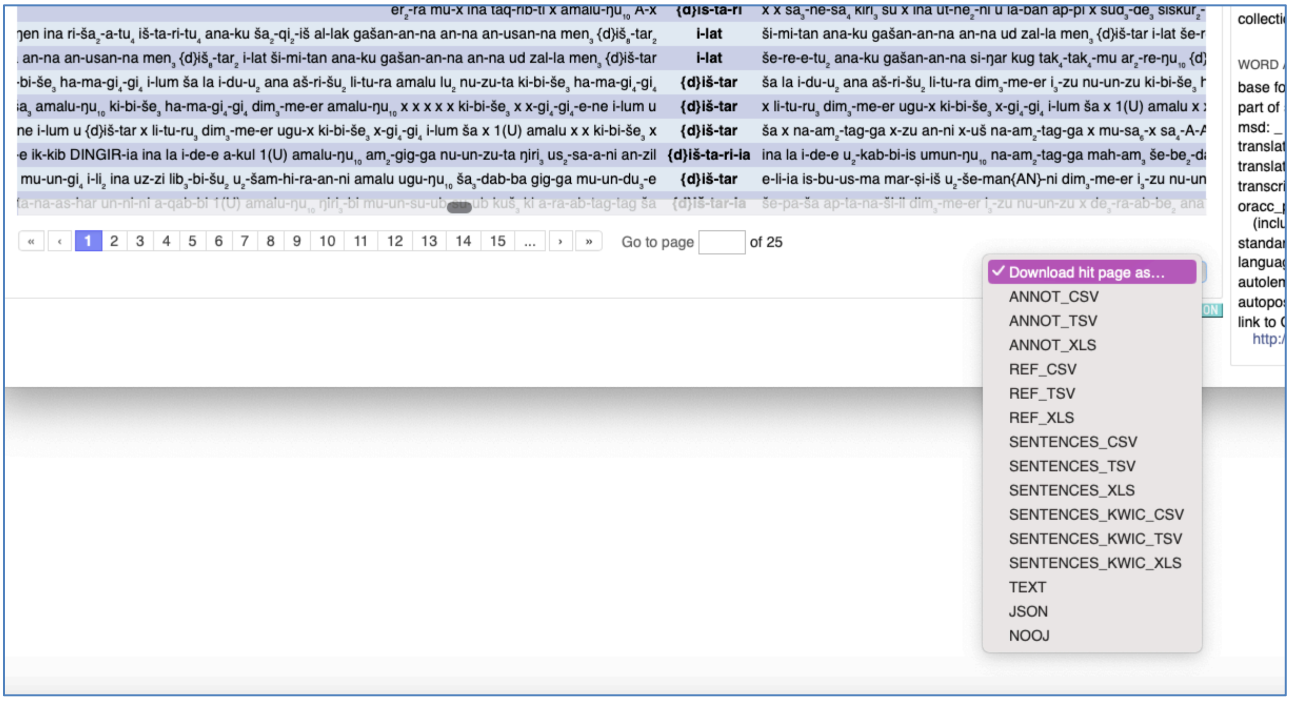
There is a button to the right of the comparison options to Delete saved searches. You do not have to do this, since you can name different searches with different names and use them also later by choosing them from the dropdown once you are in the Compare tab. If you close the window (or even the browser), the cookies will remember your saved searches when you come back to the *Other languages* section of Korp (i.e. not Finnish, Swedish, or Parallel).

9. EXPORTING THE RESULTS

You can export the results of your search in many different formats (the file formats supported in most cases are csv (colon-separated values), tsv (tab-separated values), and xls (excel)):

ANNOT_ = text as a table, one token (with annotations) per row
REF_ = a list with the transliteration of the text and the text attributes
SENTENCE_ = one text per row with transliteration, base forms, and text attributes + more
SENTENCE_KWIC_ = one text per row with transliteration, base forms, and text attributes +
TEXT = transliteration, one text per row
JSON = Structured format with text attributes and annotated words in attribute_name:attribute format
NOOJ = xml format for use in the NooJ-tool

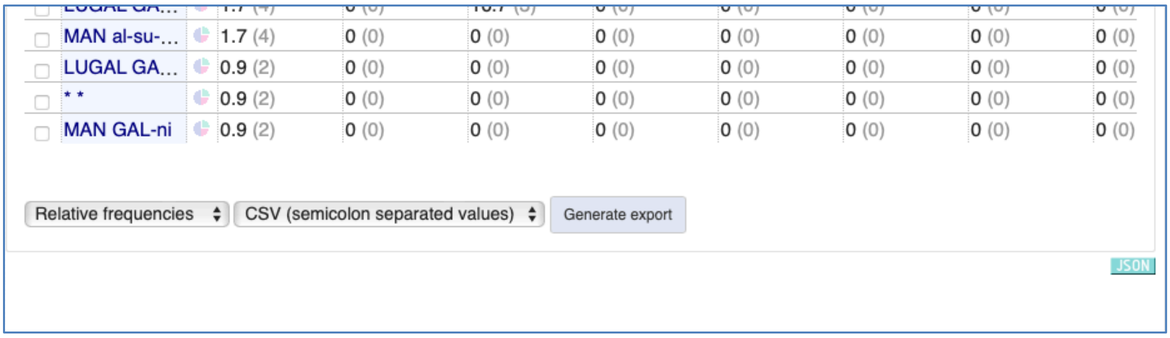
Click the *Download hit page as…* in the lower right corner of the results area and from the dropdown select the desired format and file format by clicking a row. You do not have to do anything else! It may take a while before you see the prompt for where you want to save the file (in Windows) or the file starts downloading to your Downloads folder (in Mac).



er$_2$-ra mu-x ina taq-rib-ti x amalu-ŋu$_{10}$ A-x	**{d}iš-ta-ri**	x x sa$_3$-ne-sa$_4$ kiri$_3$ su x ina ut-ne$_2$-ni u la-ban ap-pi x sud$_3$-de$_3$ siskur$_2$-
ŋen ina ri-ša$_2$-a-tu$_4$ iš-ta-ri-tu$_4$ ana-ku ša$_2$-qi$_2$-iš al-lak gašan-an-na an-na an-usan-na men$_3$ {d}iš$_8$-tar$_2$	**i-lat**	ši-mi-tan ana-ku gašan-an-na an-na ud zal-la men$_3$ {d}iš-tar i-lat še-r
. an-na an-usan-na men$_3$ {d}iš$_8$-tar$_2$ i-lat ši-mi-tan ana-ku gašan-an-na an-na ud zal-la men$_3$ {d}iš-tar	**i-lat**	še-re-e-tu$_2$ ana-ku gašan-an-na si-ŋar kug tak$_4$-tak$_4$-mu ar$_2$-re-ŋu$_{10}$ {d}
-bi-še$_3$ ha-ma-gi$_4$-gi$_4$ i-lum ša la i-du-u$_2$ ana aš-ri-šu$_2$ li-tu-ra amalu lu$_2$ nu-zu-ta ki-bi-še$_3$ ha-ma-gi$_4$-gi$_4$	**{d}iš-tar**	ša la i-du-u$_2$ ana aš-ri-šu$_2$ li-tu-ra dim$_3$-me-er i$_3$-zu nu-un-zu ki-bi-še$_3$ h
ia$_3$ amalu-ŋu$_{10}$ ki-bi-še$_3$ ha-ma-gi$_4$-gi$_4$ dim$_3$-me-er amalu-ŋu$_{10}$ x x x x x ki-bi-še$_3$ x x-gi$_4$-gi$_4$-e-ne i-lum u	**{d}iš-tar**	x li-tu-ru$_3$ dim$_3$-me-er ugu-x ki-bi-še$_3$ x-gi$_4$-gi$_4$ i-lum ša x 1(U) amalu x x
ne i-lum u {d}iš-tar x li-tu-ru$_3$ dim$_3$-me-er ugu-x ki-bi-še$_3$ x-gi$_4$-gi$_4$ i-lum ša x 1(U) amalu x x ki-bi-še$_3$ x	**{d}iš-tar**	ša x na-am$_2$-tag-ga x-zu an-ni x-uš na-am$_2$-tag-ga x mu-sa$_6$-x sa$_4$-A-A
-e ik-kib DINGIR-ia ina la i-de-e a-kul 1(U) amalu-ŋu$_{10}$ am$_2$-gig-ga nu-un-zu-ta ŋiri$_3$ us$_2$-sa-a-ni an-zil	**{d}iš-ta-ri-ia**	ina la i-de-e u$_2$-kab-bi-is umun-ŋu$_{10}$ na-am$_2$-tag-ga mah-am$_3$ še-be$_2$-da
mu-un-gi$_4$ i-li$_2$ ina uz-zi lib$_3$-bi-šu$_2$ u$_2$-šam-hi-ra-an-ni amalu ugu-ŋu$_{10}$ ša$_3$-dab-ba gig-ga mu-un-du$_3$-e	**{d}iš-tar**	e-li-ia is-bu-us-ma mar-ṣi-iš u$_2$-še-man{AN}-ni dim$_3$-me-er i$_3$-zu nu-un
ta-na-as-har un-ni-ni a-qab-bi 1(U) amalu-ŋu$_{10}$ ŋiri$_3$-bi mu-un-su-ub-su-ub kuš$_3$ ki a-ra-ab-tag-tag ša	**{d}iš-tar-ia**	še-pa-ša ap-la-na-ši-il dim$_3$-me-er i$_3$-zu nu-un-zu x de$_3$-ra-ab-be$_2$ ana

The turquoise JSON button underneath the dropdown will open the JSON file in a new tab with all the annotated words of a text on one line.

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**. The JSON button here will open the statistics data in a new tab.



□	MAN al-su-…	1.7 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
□	LUGAL GA…	0.9 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
□	* *	0.9 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
□	MAN GAL-ni	0.9 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Relative frequencies
CSV (semicolon separated values)

The Extended search is sufficient for most searches! The next section introduces the Advanced search function for even more complicated searches, and you might want to take a quick look at it especially if you are familiar with any programming languages. Or you can come back to it later if you feel like it. Anyway, **you should now be ready to use Oracc in Korp**. The best way to learn is by trying things out.
Have fun!

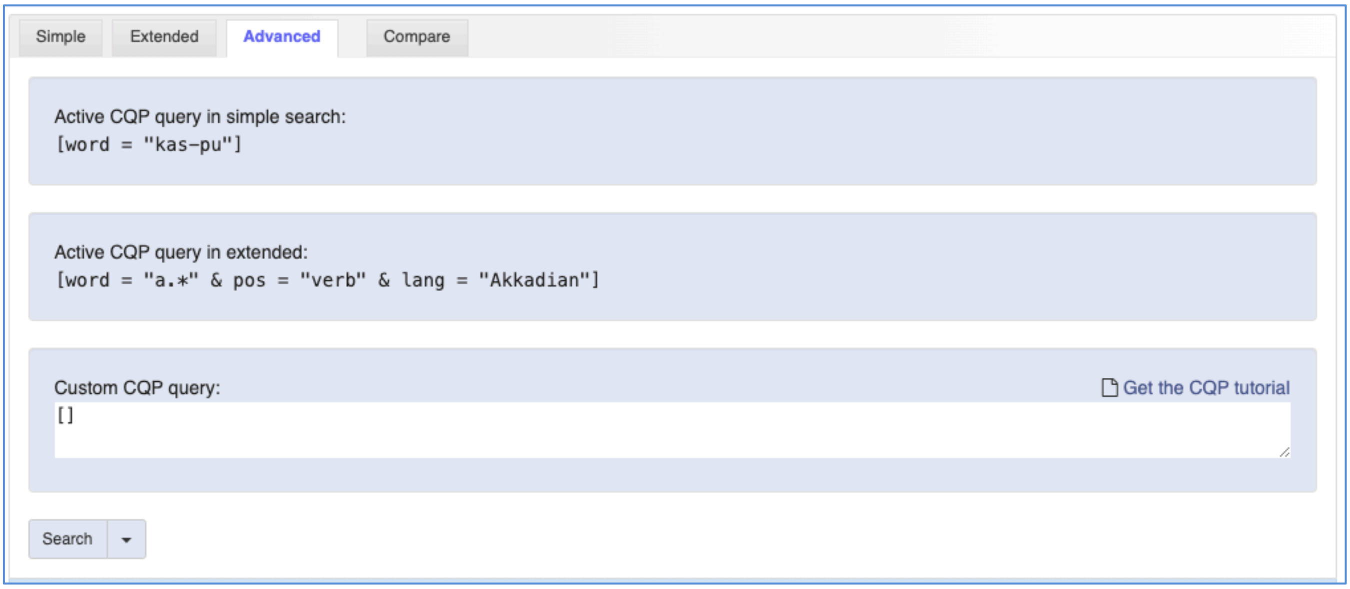
10. ADVANCED SEARCH

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

What you cannot do with Extended search is combine two searches. For example, in Search example 2 above, we performed a search in texts from the Neo-Assyrian period containing the base form of the divine name "Aššur" and the translation "king" with no more than 8 words in between. With the Advanced search, you can add to the same search the cases where the translation "king" is *before* the divine name "Aššur".

The Advanced search is performed by writing the query in the so-called CQP (Corpus Query Protocol) query language. If you want to use the Advanced search, it may take some learning.

The easiest way to start learning CQP is to perform a Simple or Extended search and then check in the Advanced tab what it looks like in CQP. In the Advanced tab, you will see the Active CQP query of the latest Simple and Extended searches.



Active CQP query in simple search:

```
[word = "kas-pu"]
```

Active CQP query in extended:

```
[word = "a.*" & pos = "verb" & lang = "Akkadian"]
```

Custom CQP query:

```
[]
```

Advanced search tab after performing a Simple search "word (i.e. transliteration) is kas-pu" and the Extended search from Search example 1 above (Words (i.e. transliterations) that begin with the letter "a" and where the part of speech is verb and the language/dialect is Akkadian).

Brackets [] always represent an orange box in the search field, empty brackets stand for any word.

The attributes have slightly different names in CQP than in the Extended search:

	CQP	***Extended search***
Word attributes	word	word
	lemma	base form
	translation	translation (base form)
	transcription	transcription
	sense	translation (sense)
	pos	part-of-speech
	oraccpos	oracc_pos_subcategory
	standard	standardized
	lang	language/dialect
Text attributes	cdlinumber	CDLI number
	genre	genre
	provenance	provenance
	period	_.text_period
	subgenre	subgenre
	language	text languages

For example, in Search example 2 above, we performed a search in texts from the Neo-Assyrian period containing the base form of the divine name “Aššur” and the translation “king” with no more than 8 words in between. In the CQP language, the search looks like this:

[lemma = "Aššur" & oraccpos = "DN Divine Name" & _.text_period = "Neo-Assyrian"] []{0,8} [translation = "king"]

The CQP uses mathematical, boolean, and regular expressions used in most programming languages. For example:

[lemma = "Aššur"] words where base form **is** *Aššur*
[lemma **!=** "Aššur"] words where base form **is not** *Aššur*
[lemma = "**.***Aššur**.***"] words where base form **contains** *Aššur* (‘.*’ stands for a letter zero or more times)
[lemma = "Aššur**.***"] words where base form **starts with** *Aššur*
[lemma = "**.***Aššur"] words where base form **ends with** *Aššur*
[lemma = "Aššur" **&** oraccpos = "DN Divine Name"] words where dictionary form is *Aššur* **and** oracc subcategory of the part-of-speech is *DN Divine Name* (must be both to match)
[word = "lugal" **|** word = "LUGAL"] words where transliteration is *lugal* **or** transliteration is *LUGAL* (both are a match)
[]{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* (e.g. LUGAL LUGAL in the text)

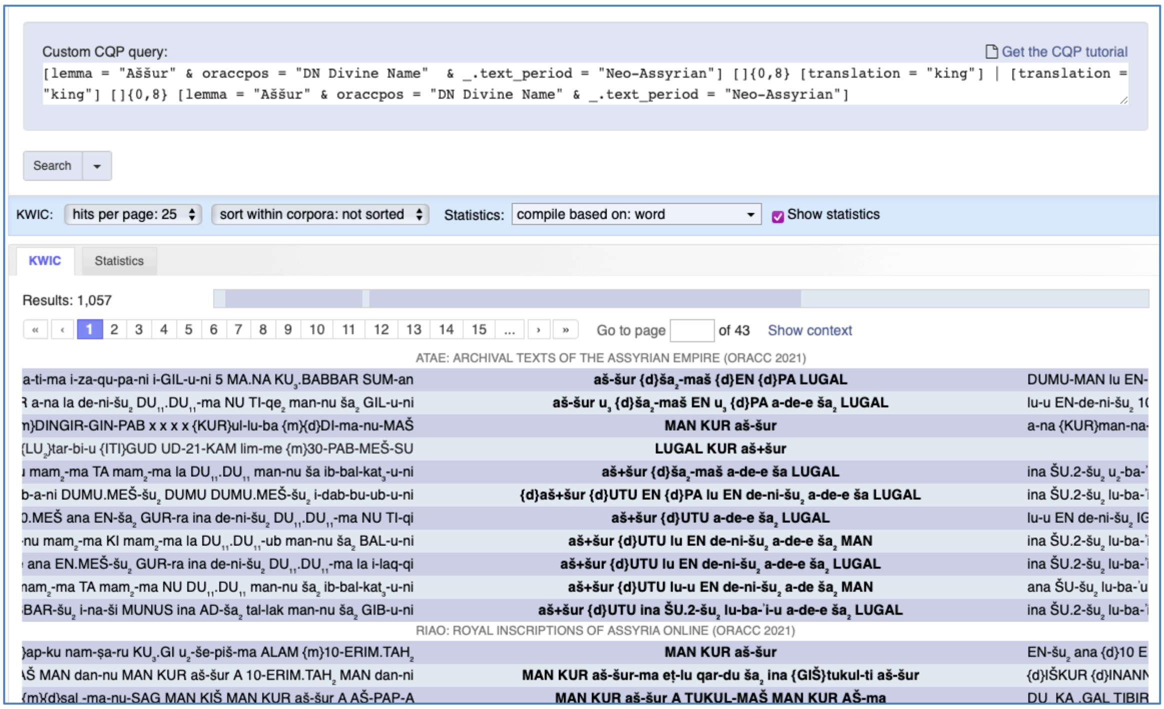
A guide to regular expression used in Korp (from https://www.kielipankki.fi/support/korp-advanced/)

.	any single symbol	
[…]	a set or range of symbols: any of the symbols inside the brackets	[aeiouyäö] matches a single Finnish vowel symbol and [a-hw-z] all the letters from a to h and w to z.
[^…]	the complement of a set or range of symbols, none of the symbols inside the brackets	[^abcw-z] matches any symbol except the letters a, b, c, w, x, y, and z.
RS	concatenation: the substring matched by the expression R if followed by a substring matched by the expression S	[a-z][0-9] matches a lowercase letter followed by a digit
(…)	grouping	
R*	repeat zero or more times; R can be a single character, a set of characters, or parentheses containing a regular expression	a.* matches all strings starting with an a, while a(bc)*matches the strings a, abc, abcbc, abcbcbc etc.
R+	repeat once or more	goo+d matches the strings good, goood, gooood etc.
R{n}	repeat exactly n times	
R{m,n}	repeat m to n times	
R?	optionality (repeat zero or one time)	favou?rite matches favorite and favourite
R|S	alternatives; match R or S	apple|orange matches the strings apple and orange; (read|writ|watch)ing matches reading, writing, watching
\c	escaping; escapes a special character	\. matches a literal full stop

Advanced search example

Texts from the Neo-Assyrian period featuring base form of the divine name "Aššur" and the translation "king" **or the other way round** (using the |) with no more than 8 words (any words) in between.

```
[lemma = "Aššur" & oraccpos = "DN Divine Name"  & _.text_period = "Neo-Assyrian"]
[]{0,8}
[translation = "king"]
|
[translation = "king"]
[]{0,8}
[lemma = "Aššur" & oraccpos = "DN Divine Name" & _.text_period = "Neo-Assyrian"]
```



Custom CQP query:

```
[lemma = "Aššur" & oraccpos = "DN Divine Name"  & _.text_period = "Neo-Assyrian"] []{0,8} [translation = "king"] | [translation =
"king"] []{0,8} [lemma = "Aššur" & oraccpos = "DN Divine Name" & _.text_period = "Neo-Assyrian"]
```

KWIC: hits per page: 25   sort within corpora: not sorted   Statistics: compile based on: word   ☑ Show statistics

Results: 1,057

Go to page ☐ of 43

ATAE: ARCHIVAL TEXTS OF THE ASSYRIAN EMPIRE (ORACC 2021)

a-ti-ma i-za-qu-pa-ni i-GIL-u-ni 5 MA.NA KU$_3$.BABBAR SUM-an	**aš-šur {d}ša$_2$-maš {d}EN {d}PA LUGAL**	DUMU-MAN lu EN-
R a-na la de-ni-šu$_2$ DU$_{11}$.DU$_{11}$-ma NU TI-qe$_2$ man-nu ša$_2$ GIL-u-ni	**aš-šur u$_3$ {d}ša$_2$-maš EN u$_3$ {d}PA a-de-e ša$_2$ LUGAL**	lu-u EN-de-ni-šu$_2$ 1
n}DINGIR-GIN-PAB x x x x {KUR}ul-lu-ba {m}{d}DI-ma-nu-MAŠ	**MAN KUR aš-šur**	a-na {KUR}man-na-
{LU$_2$}tar-bi-u {ITI}GUD UD-21-KAM lim-me {m}30-PAB-MEŠ-SU	**LUGAL KUR aš+šur**	
mam$_2$-ma TA mam$_2$-ma la DU$_{11}$.DU$_{11}$ man-nu ša ib-bal-kat$_3$-u-ni	**aš+šur {d}ša$_2$-maš a-de-e ša LUGAL**	ina ŠU.2-šu$_2$ u$_2$-ba-'
b-a-ni DUMU.MEŠ-šu$_2$ DUMU DUMU.MEŠ-šu$_2$ i-dab-bu-ub-u-ni	**{d}aš+šur {d}UTU EN {d}PA lu EN de-ni-šu$_2$ a-de-e ša LUGAL**	ina ŠU.2-šu$_2$ lu-ba-'
).MEŠ ana EN-ša$_2$ GUR-ra ina de-ni-šu$_2$ DU$_{11}$.DU$_{11}$-ma NU TI-qi	**aš+šur {d}UTU a-de-e ša$_2$ LUGAL**	lu-u EN de-ni-šu$_2$ IG
-nu mam$_2$-ma KI mam$_2$-ma la DU$_{11}$.DU$_{11}$-ub man-nu ša$_2$ BAL-u-ni	**aš+šur {d}UTU lu EN de-ni-šu$_2$ a-de-e ša$_2$ MAN**	ina ŠU.2-šu$_2$ lu-ba-'
ana EN.MEŠ-šu$_2$ GUR-ra ina de-ni-šu$_2$ DU$_{11}$.DU$_{11}$-ma la i-laq-qi	**aš+šur {d}UTU lu EN de-ni-šu$_2$ a-de-e ša$_2$ LUGAL**	ina ŠU.2-šu$_2$ lu-ba-'
nam$_2$-ma TA mam$_2$-ma NU DU$_{11}$.DU$_{11}$ man-nu ša$_2$ ib-bal-kat$_3$-u-ni	**aš+šur {d}UTU lu-u EN de-ni-šu$_2$ a-de ša$_2$ MAN**	ana ŠU-šu$_2$ lu-ba-'u
BAR-šu$_2$ i-na-ši MUNUS ina AD-ša$_2$ tal-lak man-nu ša$_2$ GIB-u-ni	**aš+šur {d}UTU ina ŠU.2-šu$_2$ lu-ba-'i-u a-de-e ša$_2$ LUGAL**	ina ŠU.2-šu$_2$ lu-ba-'
RIAO: ROYAL INSCRIPTIONS OF ASSYRIA ONLINE (ORACC 2021)		
}ap-ku nam-ṣa-ru KU$_3$.GI u$_2$-še-piš-ma ALAM {m}10-ERIM.TAH$_2$	**MAN KUR aš-šur**	EN-šu$_2$ ana {d}10 E
AŠ MAN dan-nu MAN KUR aš-šur A 10-ERIM.TAH$_2$ MAN dan-ni	**MAN KUR aš-šur-ma eṭ-lu qar-du ša$_2$ ina {GIŠ}tukul-ti aš-šur**	{d}IŠKUR {d}INANN
{m}{d}sal-ma-nu-SAG MAN KIŠ MAN KUR aš-šur A AŠ-PAP-A	**MAN KUR aš-šur A TUKUL-MAŠ MAN KUR AŠ-ma**	DU KA.GAL TIBIR