Archimob Corpus on Sketch Engine and Annis
Getting Started

Noëmi Aepli, Fatima Stadler, Tanja Samardžić

March 23, 2017

1 Searching the Archimob Corpus

This is a brief tutorial describing how to access and search the *Archimob Corpus*[1] with two online search tools namely *Sketch Engine*[2] and *Annis*[3]. The purpose is to show how to get started with these tools and to provide a few examples of possible queries. For further information and advanced search techniques please refer to the respective websites and documentations.

2 Annis

Annis[3] is web browser-based tool to search linguistic corpora. We will explain the most important queries to get started. For a more detailed explanation on corpus queries, we recommend going through the Annis tutorial or reading the Annis user guide: `http://corpus-tools.org/annis/resources/ANNIS_User_Guide_3.4.3.pdf`

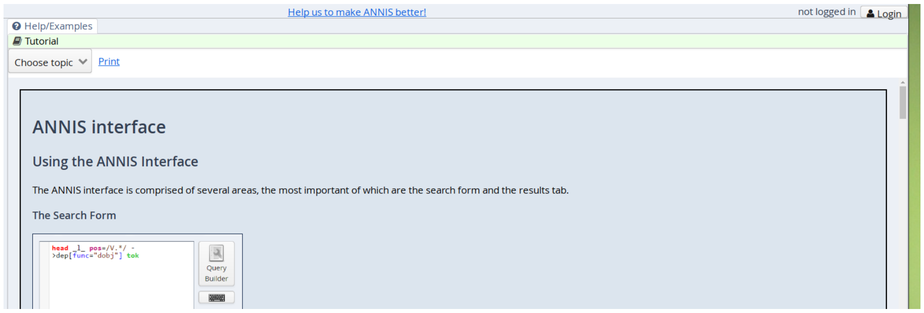
2.1 Access to Annis

To access the Archimob corpus on Annis visit the website `https://annis.linguistik.uzh.ch/annis-gui/archimob`, click the "Login" button in the upper right corner and enter the credentials:

username: *archimob* password: *archimob@UZH4All*

Annis comes with a built-in tutorial that is shown when opening the website (see figure 1). It can always be re-accessed by choosing the tutorial tab. After logging in, you will find the Annis query builder on the left side of the window

[1] `http://www.spur.uzh.ch/en/departments/korpuslab/ArchiMob.html`
[2] `https://www.sketchengine.co.uk`
[3] `http://corpus-tools.org/annis/`



ANNIS interface

Using the ANNIS Interface

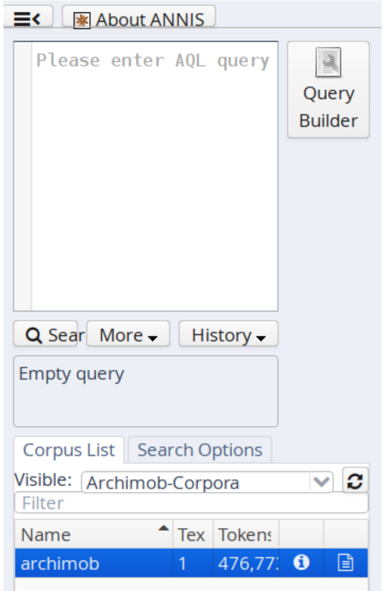
The ANNIS interface is comprised of several areas, the most important of which are the search form and the results tab.

The Search Form

```
head _l_ pos=/V.*/ -
>dep[func="dobj"] tok
```

Figure 1: Annis interface: “Login” (top right), “Help” and “Tutorial” (top left)

and below it the corpus list, as shown in figure 2. In order to perform queries, you first need to activate the Archimob corpus by clicking on its name in the list, which will then be highlighted in blue.



Empty query

Visible: Archimob-Corpora

Name	Tex	Tokens
archimob	1	476,77:

Figure 2: Annis query field and corpus list

2.2 Queries

Within Annis you can perform simple word queries (see section 2.2.1), use regular expressions (see section 2.2.2), search for co-occurrences (see section 2.2.3) and define the context for the search (see section 2.2.4). Moreover, you can for instance search for different dialectal variants of a Standard German word by executing a query on the `normalised` layer or you can use the part of speech information on the `tag` layer for your queries. The following annotation levels are available:

- `id`: identification number of a specific document, utterance or word
- `tok`: transcribed words
- `normalised`: normalisation of the transcribed word[4]
- `tag`: part of speech tags
- `type`: pauses, para- and nonverbal contents, comments

After performing a query, you have the option to do a frequency analysis on the result (see section 3.4) and export (see section 3.5) the query results or the frequency analysis of the query.

2.2.1 Simple Word Query

```
annotation_level="xyz"
```

To find an exact variant of a word, e.g. *miuch* (eng. 'milk' according to a specific pronunciation), you can perform a query on the token level by specifying the annotation level `tok` along with the exact word in quotation marks:

```
tok="miuch"
```

If you are interested in all dialectal variants you can execute the query on the normalisations.[4] In order to do so, set the annotation level to `normalised` and wrap the normalised form in quotation marks.

```
normalised="milch"
```

The above query on the normalisation layer will yield dialectal variants appearing in the "Query Result" as shown in figure 3.

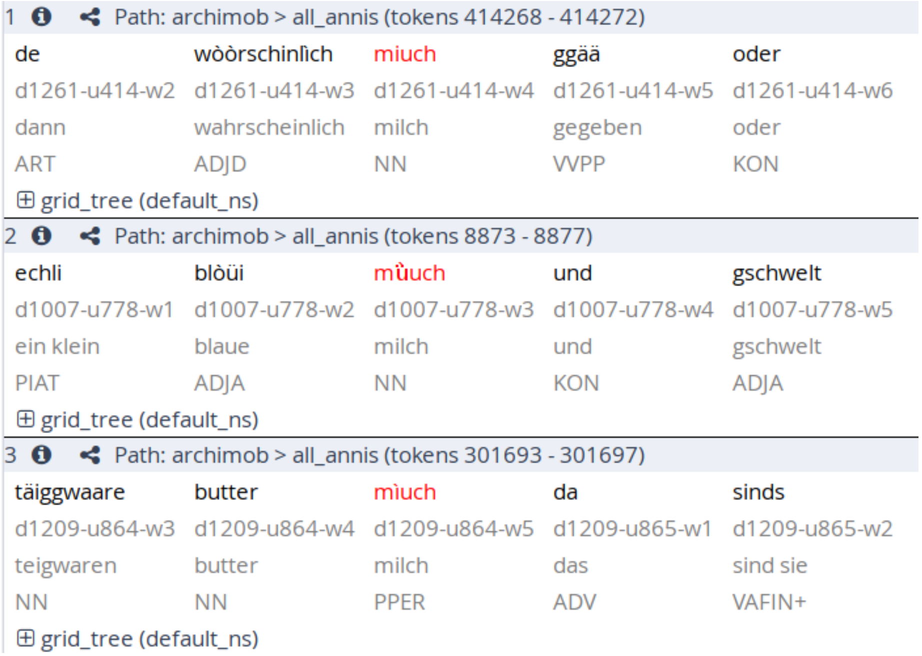
2.2.2 Using Regular Expressions

```
annotation_level=/xyz/
```

With the use of regular expressions,[5] you can specify more flexible and more specific queries. Instead of quotation marks as in the "normal" search, wrap

[4] Normalisations are labels that group dialectal variants together. In most cases they are similar to Standard German forms. More information on normalisation is included in the corpus documentation.

[5] For those who are not familiar with regular expressions, here are some basic operations: `.` matches any character, `*` quantifies the preceding character, meaning it occurs zero or more times whereas with a + it must occur at least once and `?` states that the preceding character is optional.



1 Path: archimob > all_annis (tokens 414268 - 414272)

de	wòòrschinlìch	miuch	ggää	oder
d1261-u414-w2	d1261-u414-w3	d1261-u414-w4	d1261-u414-w5	d1261-u414-w6
dann	wahrscheinlich	milch	gegeben	oder
ART	ADJD	NN	VVPP	KON

⊞ grid_tree (default_ns)

2 Path: archimob > all_annis (tokens 8873 - 8877)

echli	blòüi	mùuch	und	gschwelt
d1007-u778-w1	d1007-u778-w2	d1007-u778-w3	d1007-u778-w4	d1007-u778-w5
ein klein	blaue	milch	und	gschwelt
PIAT	ADJA	NN	KON	ADJA

⊞ grid_tree (default_ns)

3 Path: archimob > all_annis (tokens 301693 - 301697)

täiggwaare	butter	mìuch	da	sinds
d1209-u864-w3	d1209-u864-w4	d1209-u864-w5	d1209-u865-w1	d1209-u865-w2
teigwaren	butter	milch	das	sind sie
NN	NN	PPER	ADV	VAFIN+

⊞ grid_tree (default_ns)

Figure 3: Anni query on the normalisation *milch*

your queries in slashes as shown above. If you are interested in particular pronunciations like for example the merged form of 'have' 1. P. Pl. and the corresponding personal pronoun, you can perform a RegEx search on the `tok`-level: `tok=/h[äe]mm?er/`

2.2.3 Co-occurrences

```
annotation_level="x" . annotation_level="y"
```

Annis can also be used to find co-occurring tokens, the abbreviated Annis syntax for the corresponding query is the dot operator. This makes it possible to find interesting examples of Swiss German syntax like for example word order phenomena. To search for the verb *lassen* preceding any other verb you can search a range of variants on the token level (`tok`) and combine it with any subsequent verb on the part of speech level (`tag`):

```
tok=/l[aoò]+/ . tag=/VV.*/
```

With this expression we find sentences like the following:

(1) *hend mììch deet lò ligge* (Sempach, LU)

If you want to specify a span of tokens within which to search, you can indicate the minimal and the maximal number of positions directly attached to the dot operator, separated by a comma.

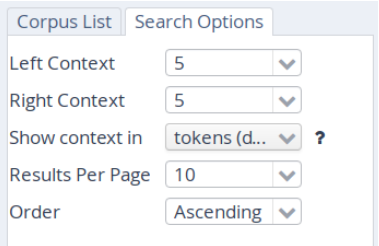
```
tok=/l[aoò]+/ .1,3 tok=/l[aoò]+/
```

The results of the query expression above contains cases of a dialectal reduplication of the verb 'lassen' with the particle *la* and the infinitive *laa* in combination with *z* (dialectal variant for 'zu').

(2) *dä jung pürschtel da la furt z laa* (Bern, BE)

2.2.4 Defining the Range of Word Context

Depending on your research question, the span of the context you are interested in may vary. You can adjust the range around the keyword by selecting the "Search Options" tab below the query window on the left and defining the number of positions around your query match (see figure 4).



Left Context 5
Right Context 5
Show context in tokens (d...
Results Per Page 10
Order Ascending

Figure 4: Options to define the query span

2.3 Frequency Analysis

After performing a query, you can do a frequency analysis on the results and export the search results and rankings. To do so, click the "More" button below the query field and choose the desired action (see figure 5).

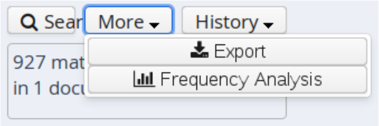
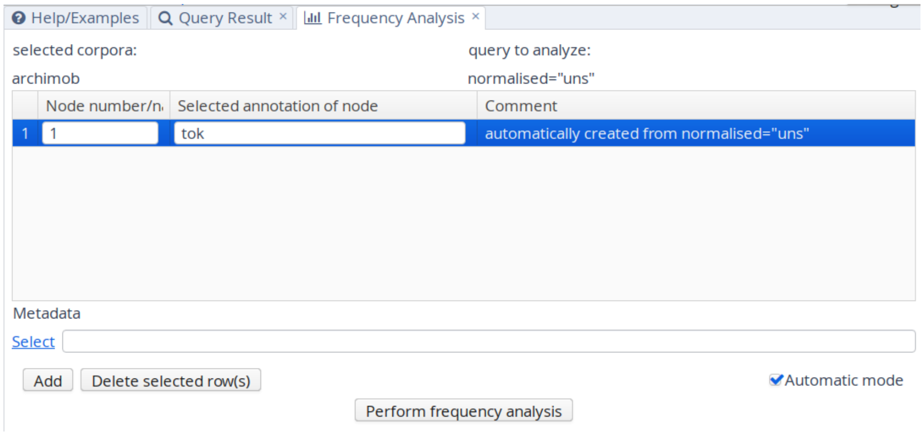


Figure 5: Frequency analysis and export options on the performed query

Choosing the "Frequency Analysis" option will result in a new tab where the analysis can be specified as shown in figure 6. In this step you can choose the level on which the frequency analysis shall be performed by changing “Selected annotation of node” before confirming with “Perform frequency analysis”.



selected corpora: archimob

query to analyze: normalised="uns"

	Node number/n	Selected annotation of node	Comment
1	1	tok	automatically created from normalised="uns"

Metadata

Figure 6: Define options for the frequency analysis

The frequency analysis can take a while. Once finished, the result is displayed as a histogram. Figure 7 shows the frequency analysis result for the query `normalised="uns"`, i.e. a query on the normalised layer to get all the available variants of the first person plural pronoun ’uns’. Furthermore, a table with frequency counts (as shown in figure 8) is generated in which you can sort and rank the result by selecting the corresponding table header. Note that the displayed plot is interactive: by selecting a specific bar in the histogram, the corresponding variant is activated in the table. The analysis can be downloaded by clicking “Download as CSV”.

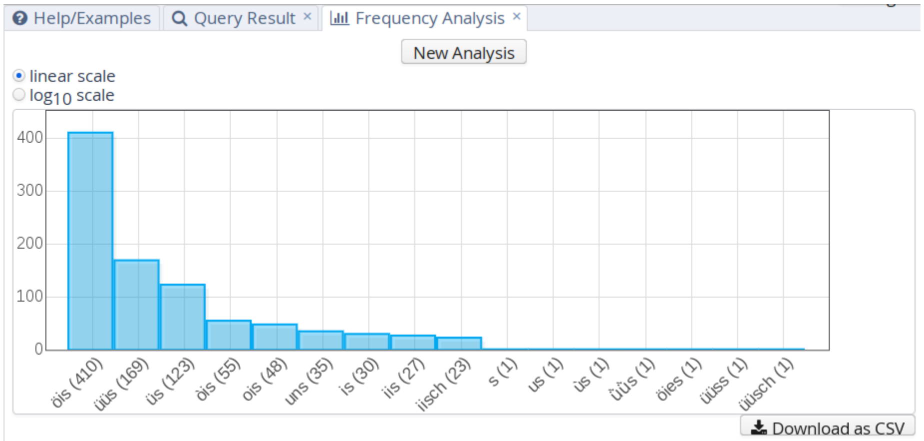
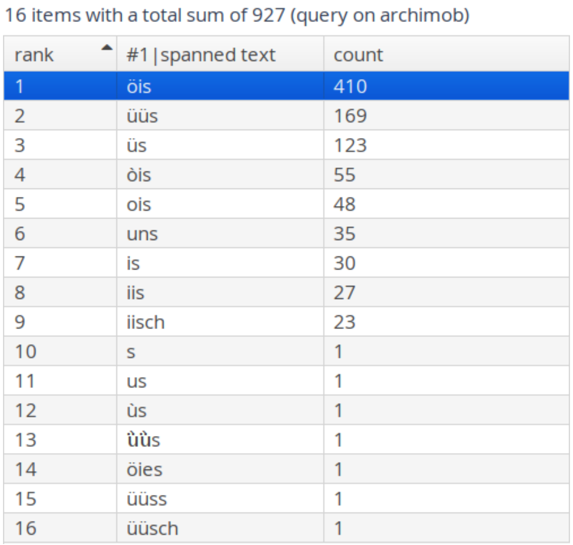


Figure 7: Histogram for the query `normalised="uns"`



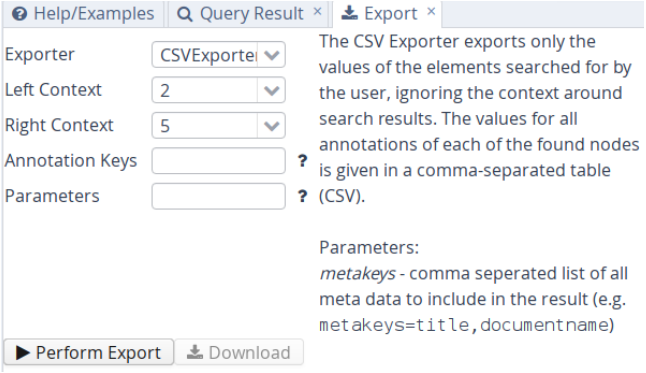
16 items with a total sum of 927 (query on archimob)

rank	#1 | spanned text	count
1	öis	410
2	üüs	169
3	üs	123
4	òis	55
5	ois	48
6	uns	35
7	is	30
8	iis	27
9	iisch	23
10	s	1
11	us	1
12	ùs	1
13	ǜǜs	1
14	öies	1
15	üüss	1
16	üüsch	1

Figure 8: Interactive table with variants and corresponding counts

2.4 Export

Choosing “Export” in the menu “More” (see figure 5), the results of queries can be exported in different formats. According to your needs you can choose between a WEKA, CSV, Token, Grid or SimpleText format (see figure 9). Exports from Annis are time consuming, therefore it can be advantageous to narrow the query scope to sub-queries and then perform several exports. A description of the formats in Annis is visible on the right side when you choose a format in the menu. The following extracts give you a quick impression of the export formats.



Exporter: CSVExporte
Left Context: 2
Right Context: 5
Annotation Keys:
Parameters:

The CSV Exporter exports only the values of the elements searched for by the user, ignoring the context around search results. The values for all annotations of each of the found nodes is given in a comma-separated table (CSV).

Parameters:
metakeys - comma seperated list of all meta data to include in the result (e.g. `metakeys=title,documentname`)

Figure 9: Export function in Annis

2.4.1 SimpleText

```
0. echli blòüi [m`üuch] und gschwelt härdepfu das hemmer
1. bi der [milch] det isch eh dän äu
2. pro litter [milch] und dä händ sich d

...
```

2.4.2 WEKA

```
@relation name
@attribute #1_id string
@attribute #1_span string
@attribute #1_anno_default_ns:id string
@attribute #1_anno_default_ns:normalised string
@attribute #1_anno_default_ns:tag string

@data

’9825126’,’m`üuch’,’d1007-u778-w3’,’milch’,’NN’
’9842020’,’milch’,’d1048-u92-w11’,’milch’,’NN’
’9842040’,’milch’,’d1048-u94-w14’,’milch’,’NN’

...
```

2.4.3 CSV

```
1_id 1_span 1_anno_default_ns:id 1_anno_default_ns:normalised 1_anno_default_ns:tag
9825126 m`üuch d1007-u778-w3 milch NN
9842020 milch d1048-u92-w11 milch NN
9842040 milch d1048-u94-w14 milch NN

...
```

2.4.4 Token

```
0. echli/d1007-u778-w1/ein klein/PIAT blòüi/d1007-u778-w2/blaue/ADJA [m`üuch]/d1007-u778-w3/milch/NN
1. bi/d1048-u92-w9/bei/APPR der/d1048-u92-w10/der/ART [milch]/d1048-u92-w11/milch/NN
2. pro/d1048-u94-w12/pro/APPR litter/d1048-u94-w13/liter/NN [milch]/d1048-u94-w14/milch/NN

...
```

2.4.5 Grid

```
0. tok  echli blòüi m`üuch
id  d1007-u778-w1[1-1] d1007-u778-w2[2-2] d1007-u778-w3[3-3]
normalised  ein klein[1-1] blaue[2-2] milch[3-3]
start  media_pointers#d1007-T1328[1-6]
tag  PIAT[1-1] ADJA[2-2] NN[3-3]
who  person_db#EJos1007[1-6]
```

```
1. tok  bi der milch
id  d1048-u92[1-3] d1048-u92-w9[1-1] d1048-u92-w10[2-2] d1048-u92-w11[3-3]
normalised  bei[1-1] der[2-2] milch[3-3]
start  media_pointers#d1048-T79[1-3]
tag  APPR[1-1] ART[2-2] NN[3-3]
who  person_db#WRos1048[1-3]

...
```

3 Sketch Engine

The second search tool that can be used for searching the Archimob corpus is *Sketch Engine*[2]. This tutorial captures only a fraction of queries possible with Sketch Engine, for more detailed information please refer to the software user guide at `https://www.sketchengine.co.uk/user-guide/`, where you can also find video tutorials `https://www.sketchengine.co.uk/user-guide/videos/sketch-engine-in-videos/`.

3.1 Access to Sketch Engine

To access the Archimob corpus on Sketch Engine, login on the website `https://www.sketchengine.co.uk/`. Please contact us[6] in order to get the login credentials. After logging in, you can select the corpus you want to search; choose *Archimob_Release1*. You will see a query window as in figure 10 where simple queries can be performed. Clicking "Query types", "Context" and "Text types" will extend the window to provide the respective query options (see figure 11).

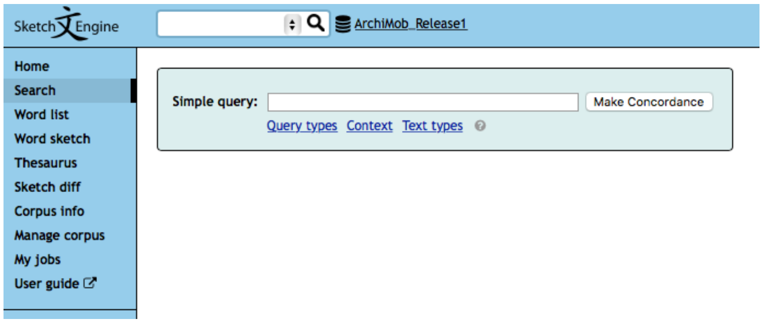


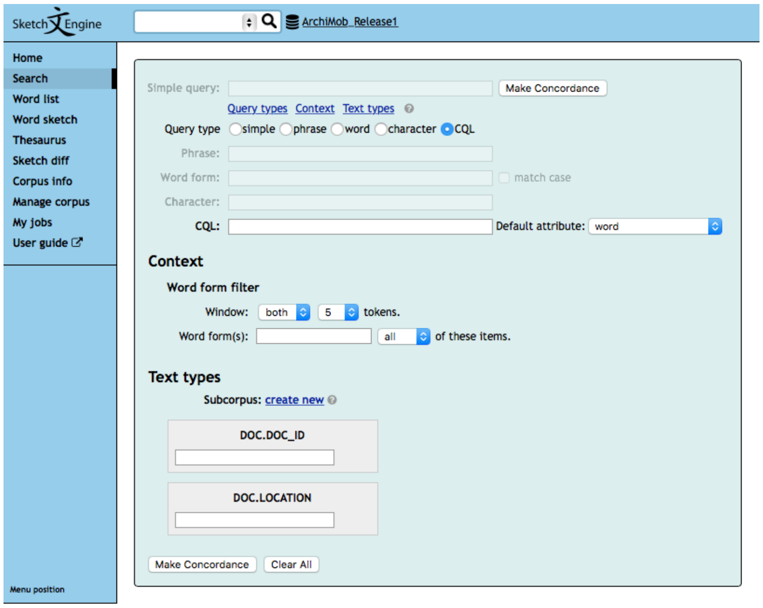
Figure 10: Query window for the simple search

3.2 Queries

Figure 11 shows the extended query window for a corpus search. The first third of the window ("Query types") displays the different types of queries which can be executed (see section 3.2.1). The other two sections provide options to limit the context or the text types (see section 3.2.3) on which the search is executed.

After each query, further options will be available in the left menu in order to sort (see section 3.3), analyse (see section 3.4) or save (see section 3.5) the result. In figures 12 and 13 you can see the left menu options.

[6] tanja.samardzic@uzh.ch



Query type ○simple ○phrase ○word ○character ●CQL

Default attribute: word

Window: both 5 tokens.

Word form(s): all of these items.

Figure 11: Query window showing all the options

3.2.1 Query Types

- **simple**: By default, the query is performed as flexible search. This means, if we search for *gäld* (eng. 'money'), which is normalised to *geld*, the result contains all kinds of variants such as *gäud* or *gält*, etc. which have the same normalisation. However, potentially different words not relevant to the search might be included in the result. In order eliminate unwanted examples, choose a different query type such as `word` or `CQL`.

- **word** finds the exact word form as it is typed (without orthographic flexibility).

- **character** finds a sequence of characters inside a token. *men* for instance corresponds to the RegEx `.*men.*`, meaning that any characters can surround *men*. The result will include words like *men*, *a****men***, *bi****men****e*, ***umen****and* etc.

- **phrase** finds examples of a sequence of `word`s exactly as typed.

- **CQL** allows more complex searches. The Corpus Query Language is explained in more detail in section 3.2.2. In this setting, the layer to be searched can be specified: `word` (as described above), `tag` (part of speech tags) or `normalised`. This option is to be used in combination with simple query in cases where we are not sure how a word should be written. In this case, we first perform a simple query typing the word in any of Swiss German variants. If the word is recognised, the system will return instances of the desired word, but probably some other words too.The resulting concordances will show (in grey) the normalised form of each returned instance, which we can then use to search for the desired normalisation using the "CQL" query type.

3.2.2 Corpus Query Language

```
[attribute="value"]
```

CQL allows to set more complex criteria in a search. The basic format is an attribute value pair as shown above. In order to search for a phrase, each token has to be surrounded by square brackets. The `attribute` can be `word`, `normalised` or `tag` as described above.

RegEx[5] can be used with values, e.g. words ending with *-zion*: `[word=".*zion"]`:

(3) *oder irgend än informa**zion** chönd si sich*

... or different variations of *haben wir*: `[word="h[ä|e]mm?er"]`

(4) *härdöpfu **hemer** sowisoo sälber ghaa*
*e schtund **hämmer** müse lauffe*

RegEx can also be useful when searching the part of speech tag layer. For example, to search for any auxiliary verb: `[tag = "VA.*"]`, any conjunction `[tag = "KO.*"]` or any particle `[tag = "PTK.*"]`.[7] Especially interesting in the case of the Archimob corpus might be to search for merged words, which are labelled with a part of speech tag ending with `+`.[8] To do so, the plus sign has to be escaped with a backslash (as otherwise it has a special meaning in regular expressions): `[tag = ".*\+"]`. The result are words like *wos*, *hets*, *wemme*, *dasch*, etc.

It is also possible to specify a part of speech tag for a specific word:
`[word = "äss[eä]" & tag = "NN"]`

(5) *mi händ das **ässe** ghaa wommer*

By placing a question mark after the square bracket, the token is made optional, i.e. it can appear once or not at all:

[7] The part of speech tags used in the Archimob coprus are an extended version of the Stuttgart Tübingen Tagset, please refer to the STTS guidelines `http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf` and the documentation of the Archimob corpus.

[8] For more information about the use of the plus sign in the Archimob corpus, please refer to the documentation of the corpus.

```
[word = "für"][]?[(word = "z" |  word = "zum")][tag = "V.*"]
```

(here: `[]?` meaning any word).

Parentheses can be used for grouping. For example `[(word = "z" |  word = "zum")]` means there must be either the word *z* or the word *zum*.

(6) *tischli inschtaliirt* ***für z ässe*** *druf*
aagfrogt woorde ***für cho z singe*** *sìì het*
öpper haa ***für der z hälffe*** *gwüss geng*

Curly brackets `{m,n}` allow `m` to `n` repetitions of the preceding token:

```
[word = "l[aoò]+"][]{1,3}[word = "l[aoò]+"]
```

(7) *dä jung pürschtel da* ***la furt z laa*** *i däre ziit*

3.2.3 Text Types

The query can be limited to a specific document ID[9] or to a specific location which is available in the corpus (e.g. *NW (Stans)*). The information about the document ID or location can also be accessed by clicking the blue number on the left in any search result (see figure 12).

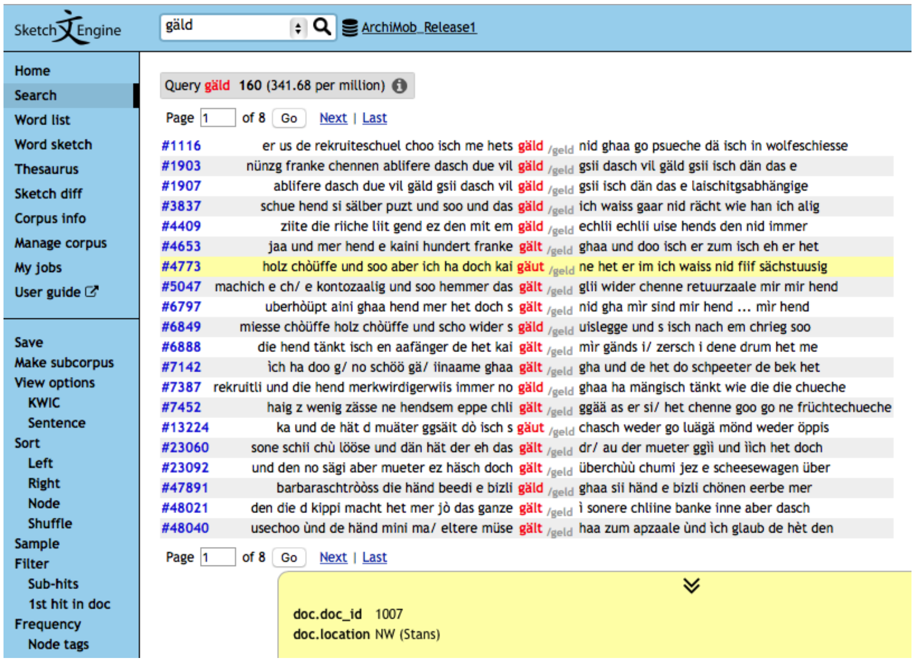
3.3 Sort

The results of the performed query can be sorted. Check the sort options in the menu on the left (see figure 13). By clicking “Left” under the point “Sort”, the resulting sentences will be sorted alphabetically according to the word appearing on the left side of the queried word. “Node” will sort the according to the searched word itself. By clicking “Sort” itself, more complex sorting options can be specified.

3.4 Analysis

In order to compute frequencies of the performed search, click “Frequency” in the menu on the left (see figure 13). A window like in figure 14 will pop up to specify the criteria for the frequency analysis. Figure 15 shows an example

[9] Please refer to the Archimob corpus documentation.



gäld

Query gäld 160 (341.68 per million)

Page 1 of 8

#1116	er us de rekruiteschuel choo isch me hets	gäld	/geld	nid ghaa go psueche dä isch in wolfeschiesse
#1903	nünzg franke chennen ablifere dasch due vil	gäld	/geld	gsii dasch vil gäld gsii isch dän das e
#1907	ablifere dasch due vil gäld gsii dasch vil	gäld	/geld	gsii isch dän das e laischitgsabhängige
#3837	schue hend si sälber puzt und soo und das	gäld	/geld	ich waiss gaar nid rächt wie han ich alig
#4409	ziite die riiche liit gend ez den mit em	gäld	/geld	echlii echlii uise hends den nid immer
#4653	jaa und mer hend e kaini hundert franke	gält	/geld	ghaa und doo isch er zum isch eh er het
#4773	holz chòüffe und soo aber ich ha doch kai	gäut	/geld	ne het er im ich waiss nid fiif sächstuusig
#5047	machich e ch/ e kontozaalig und soo hemmer das	gält	/geld	glii wider chenne retuurzaale mir mir hend
#6797	uberhòüpt aini ghaa hend mer het doch s	gält	/geld	nid gha mìr sind mir hend ... mìr hend
#6849	miesse chòüffe holz chòüffe und scho wider s	gäld	/geld	uislegge und s isch nach em chrieg soo
#6888	die hend tänkt isch en aafänger de het kai	gält	/geld	mìr gänds i/ zersch i dene drum het me
#7142	ìch ha doo g/ no schöö gä/ iinaame ghaa	gält	/geld	gha und de het do schpeeter de bek het
#7387	rekruitli und die hend merkwirdigerwiis immer no	gäld	/geld	ghaa ha mängisch tänkt wie die die chueche
#7452	haig z wenig zässe ne hendsem eppe chli	gält	/geld	ggää as er si/ het chenne goo go ne früchtechueche
#13224	ka und de hät d muäter ggsäit dò isch s	gäut	/geld	chasch weder go luägä mönd weder öppis
#23060	sone schii chù lööse und dän hät der eh das	gält	/geld	dr/ au der mueter ggìi und ìich het doch
#23092	und den no sägi aber mueter ez häsch doch	gält	/geld	überchùù chumi jez e scheesewagen über
#47891	barbaraschtròòss die händ beedi e bizli	gäld	/geld	ghaa sii händ e bizli chönen eerbe mer
#48021	den die d kippi macht het mer jò das ganze	gält	/geld	ì sonere chliine banke inne aber dasch
#48040	usechoo ùnd de händ mini ma/ eltere müse	gält	/geld	haa zum apzaale ùnd ìch glaub de hèt den

Page 1 of 8

doc.doc_id 1007
doc.location NW (Stans)

Figure 12: Query result with further information about the document of one of the resulting sentences

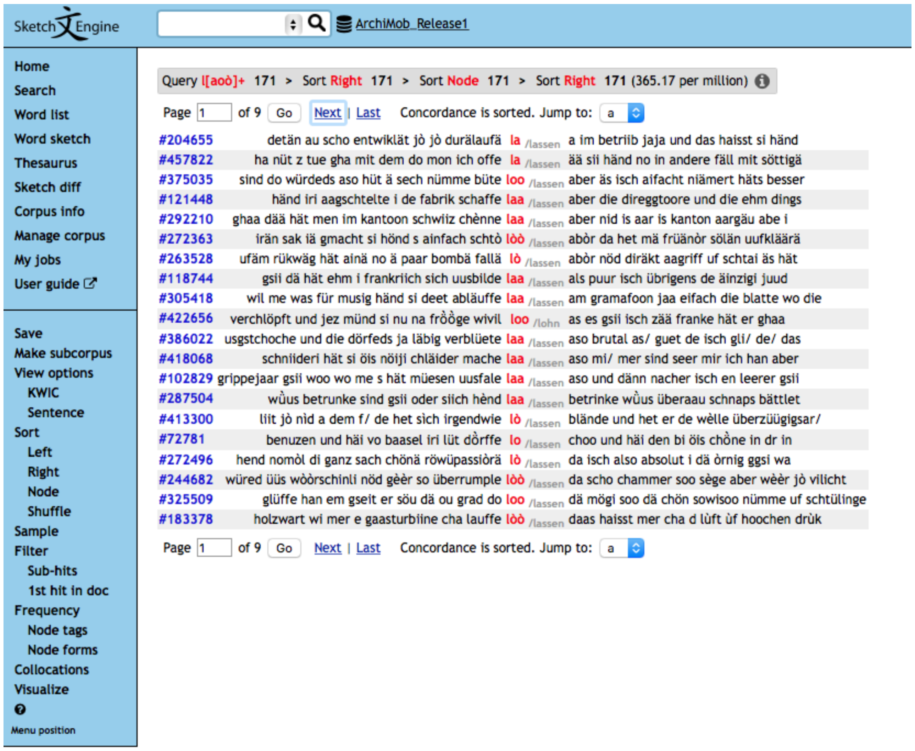
frequency analysis for the simple query *füf* (eng. 'five'). In the left menu you can choose further options for processing the frequency analysis or saving it.

3.5 Export

The result of the query can be exported as a text file or in XML format. In order to do so, click "Save" in the menu on the left and specify the export options shown in the window as in figure 16.

3.6 Word Sketch: Collocations

Along with the above described "Search" options, Sketch Engine also offers the option "Word sketch". A word sketch shows the word's collocates categorised by grammatical relations including statistical analyses. Figure 17 shows an example word sketch for *ässe* (eng. 'eat'). The blue number shows the frequency of the collocation. The source concordances can be accessed by clicking on the number. Grey phrases represent the most typical use of the collocation.



Query l[aoò]+ 171 > Sort Right 171 > Sort Node 171 > Sort Right 171 (365.17 per million)

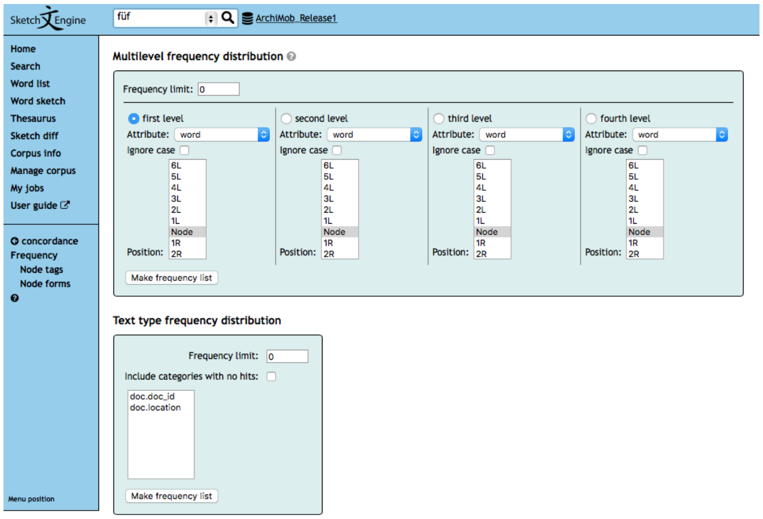
Page 1 of 9 Go Next | Last Concordance is sorted. Jump to: a

#204655	detän au scho entwiklät jò jò durälaufä	la /lassen	a im betriib jaja und das haisst si händ
#457822	ha nüt z tue gha mit dem do mon ich offe	la /lassen	ää sii händ no in andere fäll mit söttigä
#375035	sind do würdeds aso hüt ä sech nümme büte	loo /lassen	aber äs isch aifacht niämert häts besser
#121448	händ iri aagschtelte i de fabrik schaffe	laa /lassen	aber die direggtoore und die ehm dings
#292210	ghaa dää hät men im kantoon schwiiz chènne	laa /lassen	aber nid is aar is kanton aargäu abe i
#272363	irän sak iä gmacht si hönd s ainfach schtò	lòò /lassen	abòr da het mä früänòr sölän uufkläärä
#263528	ufäm rükwäg hät ainä no ä paar bombä fallä	lò /lassen	abòr nöd diräkt aagriff uf schtai äs hät
#118744	gsii dä hät ehm i frankriich sich uusbilde	laa /lassen	als puur isch übrigens de äinzigi juud
#305418	wil me was für musig händ si deet abläuffe	laa /lassen	am gramafoon jaa eifach die blatte wo die
#422656	verchlöpft und jez münd si nu na fröö̀ge wivil	loo /lohn	as es gsii isch zää franke hät er ghaa
#386022	usgstchoche und die dörfeds ja läbig verblüete	laa /lassen	aso brutal as/ guet de isch gli/ de/ das
#418068	schniideri hät si öis nöiji chläider mache	laa /lassen	aso mi/ mer sind seer mir ich han aber
#102829	grippejaar gsii woo wo me s hät müesen uusfale	laa /lassen	aso und dänn nacher isch en leerer gsii
#287504	wu̍us betrunke sind gsii oder siich hènd	laa /lassen	betrinke wu̍us überaau schnaps bättlet
#413300	liit jò nìd a dem f/ de het sìch irgendwie	lò /lassen	blände und het er de wèlle überzüügigsar/
#72781	benuzen und häi vo baasel iri lüt dö̀rffe	lo /lassen	choo und häi den bi öis chö̀ne in dr in
#272496	hend nomòl di ganz sach chönä röwüpassiòrä	lò /lassen	da isch also absolut i dä òrnig ggsi wa
#244682	würed üüs wòòrschinli nöd gèèr so überrumple	lòò /lassen	da scho chammer soo sège aber wèèr jò vilicht
#325509	glüffe han em gseit er söu dä ou grad do	loo /lassen	dä mögi soo dä chön sowisoo nümme uf schtülinge
#183378	holzwart wi mer e gaasturbiine cha lauffe	lòò /lassen	daas haisst mer cha d lùft ùf hoochen drùk

Page 1 of 9 Go Next | Last Concordance is sorted. Jump to: a

Figure 13: Right-sorted result result of the query `[word = "l[aoò]+"]`

In the left menu you can choose options for further processing the word sketch or saving it.



Multilevel frequency distribution

Frequency limit: 0

● first level	○ second level	○ third level	○ fourth level
Attribute: word	Attribute: word	Attribute: word	Attribute: word
Ignore case □	Ignore case □	Ignore case □	Ignore case □
Position: 6L 5L 4L 3L 2L 1L Node 1R 2R	Position: 6L 5L 4L 3L 2L 1L Node 1R 2R	Position: 6L 5L 4L 3L 2L 1L Node 1R 2R	Position: 6L 5L 4L 3L 2L 1L Node 1R 2R

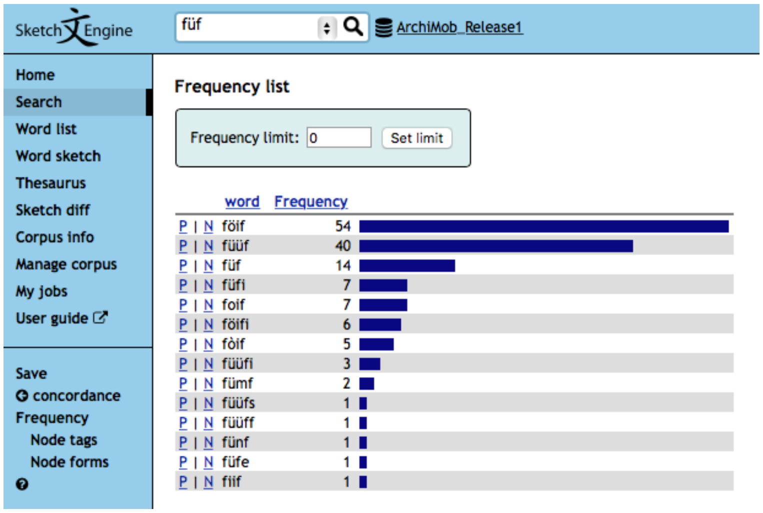
Text type frequency distribution

Frequency limit: 0

Include categories with no hits: □

doc.doc_id
doc.location

Figure 14: Options to be set for the frequency analysis

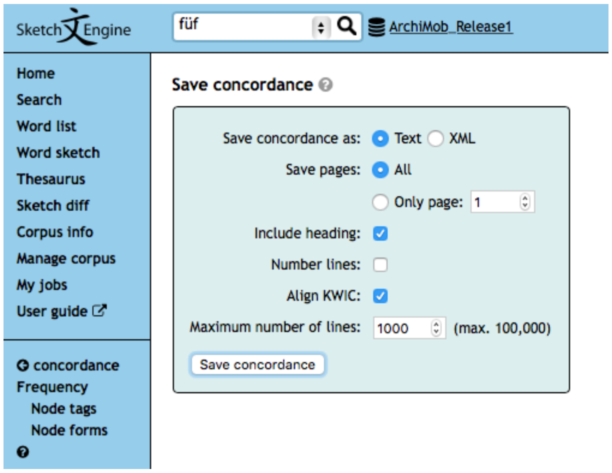


Frequency list

Frequency limit: 0

	word	Frequency
P | N	föif	54
P | N	füüf	40
P | N	füf	14
P | N	füfi	7
P | N	foif	7
P | N	föifi	6
P | N	fòif	5
P | N	füüfi	3
P | N	fümf	2
P | N	füüfs	1
P | N	füüff	1
P | N	fünf	1
P | N	füfe	1
P | N	fiif	1

Figure 15: Frequency analysis for the query *füf*



Save concordance

Save concordance as: ● Text ○ XML

Save pages: ● All

○ Only page: 1

Include heading: ☑

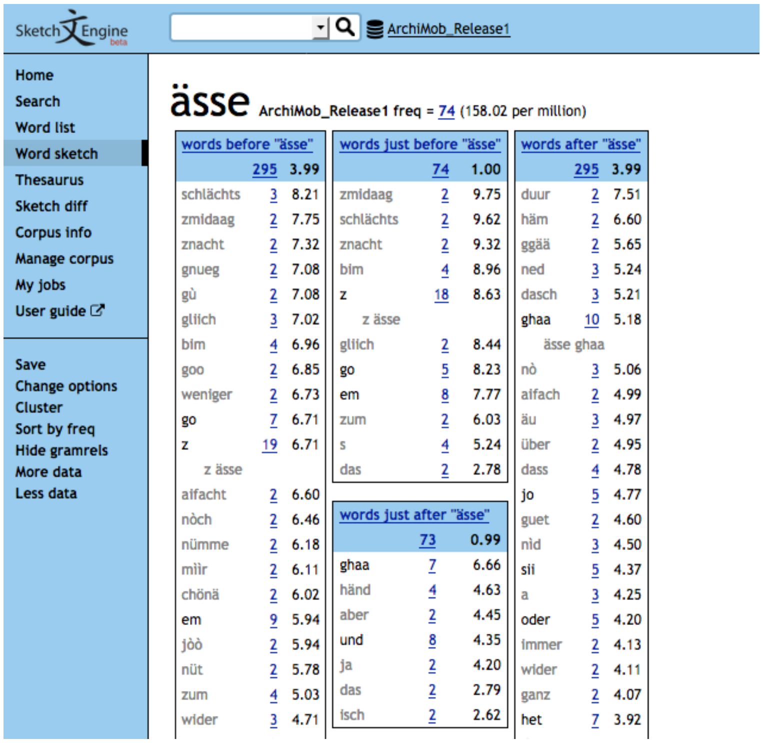
Number lines: □

Align KWIC: ☑

Maximum number of lines: 1000 (max. 100,000)

Save concordance

Figure 16: Save the result of a query



ässe ArchiMob_Release1 freq = 74 (158.02 per million)

words before "ässe"	295	3.99
schlächts	3	8.21
zmidaag	2	7.75
znacht	2	7.32
gnueg	2	7.08
gù	2	7.08
gliich	3	7.02
bim	4	6.96
goo	2	6.85
weniger	2	6.73
go	7	6.71
z	19	6.71
z ässe		
aifacht	2	6.60
nòch	2	6.46
nümme	2	6.18
miìr	2	6.11
chönä	2	6.02
em	9	5.94
jòò	2	5.94
nüt	2	5.78
zum	4	5.03
wider	3	4.71

words just before "ässe"	74	1.00
zmidaag	2	9.75
schlächts	2	9.62
znacht	2	9.32
bim	4	8.96
z	18	8.63
z ässe		
gliich	2	8.44
go	5	8.23
em	8	7.77
zum	2	6.03
s	4	5.24
das	2	2.78

words just after "ässe"	73	0.99
ghaa	7	6.66
händ	4	4.63
aber	2	4.45
und	8	4.35
ja	2	4.20
das	2	2.79
isch	2	2.62

words after "ässe"	295	3.99
duur	2	7.51
häm	2	6.60
ggää	2	5.65
ned	3	5.24
dasch	3	5.21
ghaa	10	5.18
ässe ghaa		
nò	3	5.06
aifach	2	4.99
äu	3	4.97
über	2	4.95
dass	4	4.78
jo	5	4.77
guet	2	4.60
nìd	3	4.50
sii	5	4.37
a	3	4.25
oder	5	4.20
immer	2	4.13
wider	2	4.11
ganz	2	4.07
het	7	3.92

Figure 17: Word sketch for *ässe*