Frequency Analysis

When you query for different tokens or different constructions, it might be interesting to see their frequencies instead of just the list of results. In that case, ANNIS allows you to perform an automatic analysis of frequencies.

Let us look at an example and assume that we want to know how often the regularly built French past participle is spelt with the correct ending *-é* (ignoring feminine and plural agreement) and how often it is spelt wrongly with the infinitive ending *-er*. In order to get these results, we formulate the query and then select “Frequency Analysis” from the option “More” as seen in Figure 1.

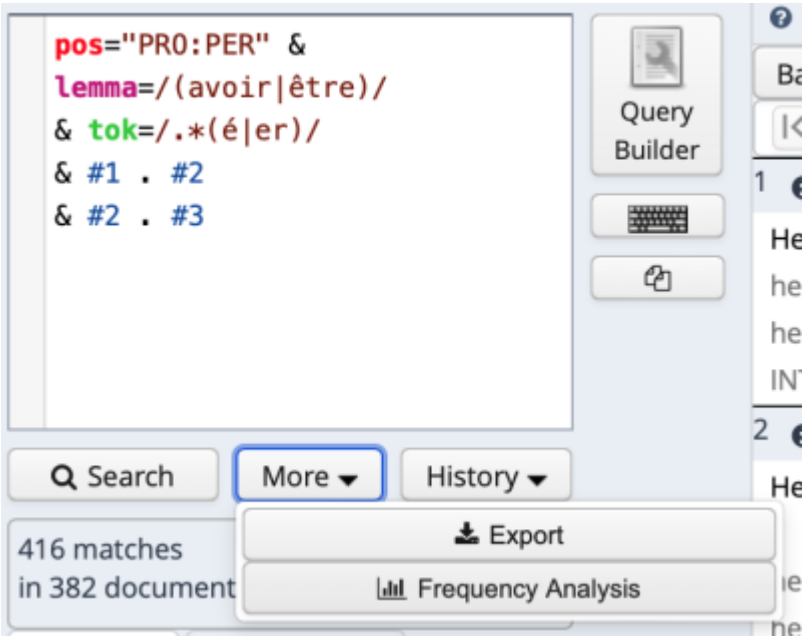
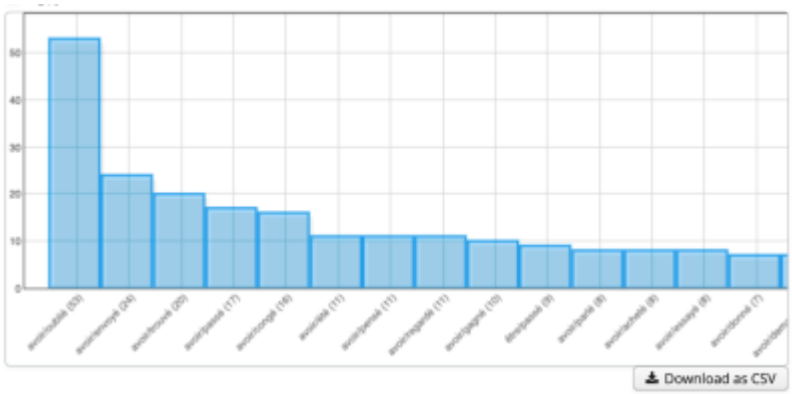


Figure 1: Formulating the query and selecting `Frequency Analysis` from `More`

In the following screen you could adjust your query and add query factors/categories, but we strongly discourage you from doing so, since the query might get really slow. If you want to add factors, you better start anew, i.e. formulate a new query and then start the Frequency Analysis again.

If you click on "Perform frequency analysis" at the bottom, you see your results as in Figure 2.



Download as CSV

146 items with a total sum of 416 (query on fra-tagged)

rank	#2 | lemma	#3 | spanned text	count
1	avoir	oublié	53
2	avoir	envoyé	24
3	avoir	trouvé	20
4	avoir	passé	17
5	avoir	congé	16

Figure 2: Results of frequency analysis

As you can see in Figure 2, we receive a full list of auxiliaries and verb forms with their respective frequencies. It is obvious that there are false positives, but if you perform the query, you will also findinteresting examples like *[avoir] demander*. If you now check the query `lemma="avoir" & tok="demander" & #1 . #2`, you will see, that this is in fact a misspelling.

This list can now also be downloaded as a text file for further processing by clicking "Download as CSV".

From:
https://sms.linguistik.uzh.ch/ -

Permanent link:
https://sms.linguistik.uzh.ch/02_browsing/05_support/02_frequency

Last update: **2022/06/27 07:21**