

## Assignments Lecture 9

21 Nov 2012

*Assignments are to be made by each student individually. These assignments count for the final grade. Hand in your assignment November 5, at the latest.*

Part A. Make a study of the words in the corpus labeled with this POS-tag: **PNM**

1. Look up the POS label in the CINTIL annotation manual (see course website, lecture 2). What grammatical category corresponds to this POS-tag?

Answer:

2a. How many words (tokens) in the corpus are labeled with this POS-tag? Answer:

2b. Which unix command(s) did you use to do this count?

\$

3a. How many word forms (types) in the corpus are labeled with this POS-tag? Answer:

3b. Which unix command(s) did you use to do this count?

\$

4a. Make a sorted frequency list of the words labeled with this POS-tag with the most frequent words first. Which command(s) did you use to make this list?

\$

4b. Write down the top 10 words and their frequency.

Part B. Study the following verb lemma: **ficar**

6a. How many verb forms (types) do you count in the corpus that have this verb lemma? Make sure that you convert the verb forms in your list to lower case. Answer:

6b. Which unix command(s) did you use to do this count?

\$

6c. Can you make concordances for this verb with one concordance per line and with a context of 4 words to left and 2 words to the right? Which command did you use?

\$

6d. Can you make concordances for this verb with one concordance per line and with a context of 4 words and POS tags to left and 2 words and POS tags to the right? Which command(s) did you use?

\$