

Looking at Named Entities in ELTeC level 2

Training materials

Diana Santos

d.s.m.santos@ilos.uio.no

Distant [📄] Reading



COST action CA16204 is supported by the EU Framework Programme Horizon 2020

23 March 2022



Contents

- Some studies done with Portuguese places
- General comparison of different collections in ELTeC-2, in R (See R training materials)
- Napoleon and Christmas: Looking at one person and one event across collections

The three “modules” are quite separate, and can be watched separately.

Purpose of this training

- give an idea of what can be done for literary history using NE, in a preview for a paper for NorLit 2022
- get hands on with the collections
 - those with NE
 - those with level 2
- feel empowered for simple visualization in R
- understand several differences in the collections
- play with some cases

Students/attendees can engage at different levels, from simple observers to motivated R users.

Disclaimer(s)

- There is no common understanding of what and how it should be encoded in level 2, in general and specifically on named entities
- I am not trying to suggest one, nor to criticize other collections
- It is possible that the materials used here are already obsolete or will be until the end of the action
- The annotation quality is variable, and none is perfect
- It would be nice that students had heard my lecture “History of named entity recognition” for background

All the opinions voiced here are solely mine.

- Joint presentation with Eckhard Bick
- Created for December 2021
- Will possibly be updated for June 2022
- One purpose is to problematize the notion of place for literary studies
- Another purpose is to give examples of studies, when one has place annotation in place

Comparing NE annotation in different collections

- Learn about the NE-annotated collections: SLV, SRP and POR
- Learn about the other level-2 annotated collections
- Enter R
 - 1 Manipulating dataframes
 - 2 Boxplots
 - 3 Comparing number of NES
 - 4 Comparing number of proper names
 - 5 Correlation between proper names and NEs
 - 6 Specific kinds of NE

Specific instructions for the R part, so that students can follow and come back to it.

Looking at one person and one event

As an “aperitif” for culture studies. How important/shared are things in Europe, through literature?



- Napoleon: probably the most important character in Europe for the 1840-1920 period
- In a majority Christian Europe, how important is Christmas? (knowing that we have three Christian flavors: Catholic, Ortodox and Protestant)

Limitations of this “play” research

Looking for a word (lemma) is usually not precise enough. Maybe *Bonaparte* as well?

- *Napoleão/Napoleon* is a firstname in several languages, so not all are Bonaparte. Especially in France there are other rulers called Napoleon...
- *Noël* is a first name in French
- *Božič* is a surname in Serbian
- *Karácsony* is a surname in Hungarian
- *Natal* is a place name in Portuguese
- Languages with compounds have to be searched differently (anything that starts with?) It works for German (*Weihnacht...*) but not for Norwegian (*jul...* can be the beginning of many other words)
- Should one also look for *napoleonic*-like words in different languages?

Of course the classification of EVENT for Christmas would solve some of the issues, but not all: one could be interested in “Christmas trees”, obviously not EVENTS

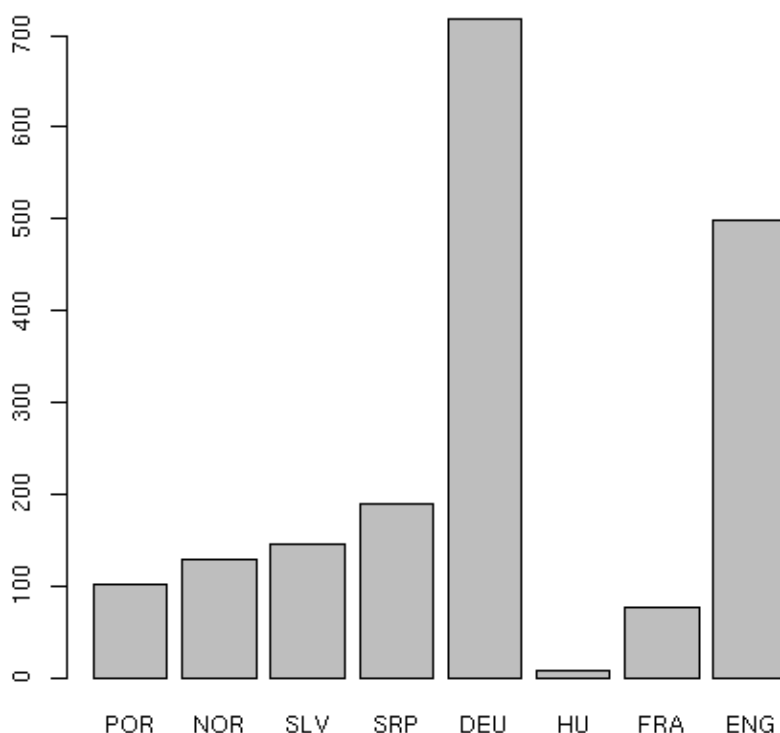
Very important to distinguish between

- sheer frequency in the total collection
- number of different works which mention it (the same relevant issue)

Remember that the sizes of the collections differ in number of words, and also that Norwegian has only 62 works so far.

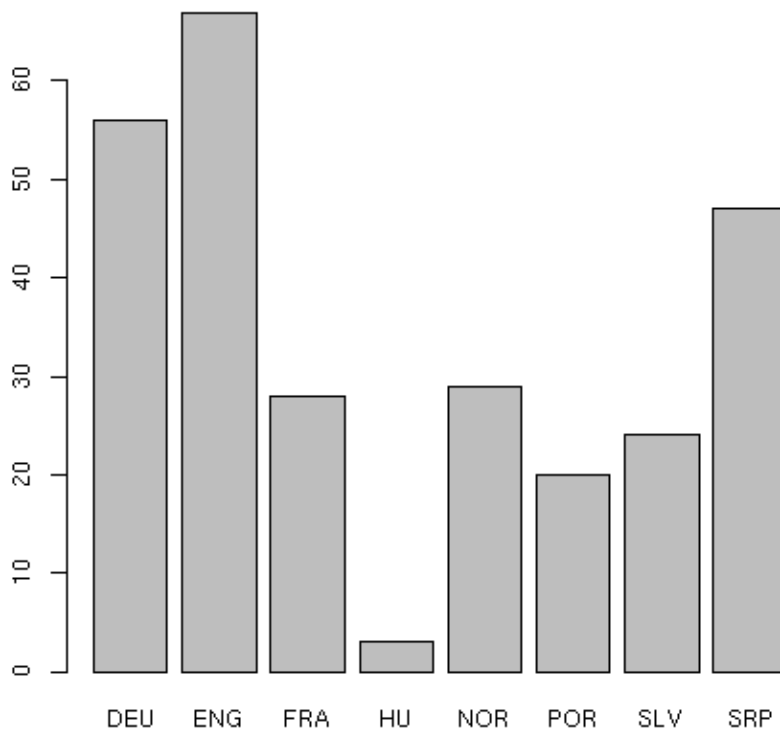
Christmas

Christmas in 8 European literatures: amount of references



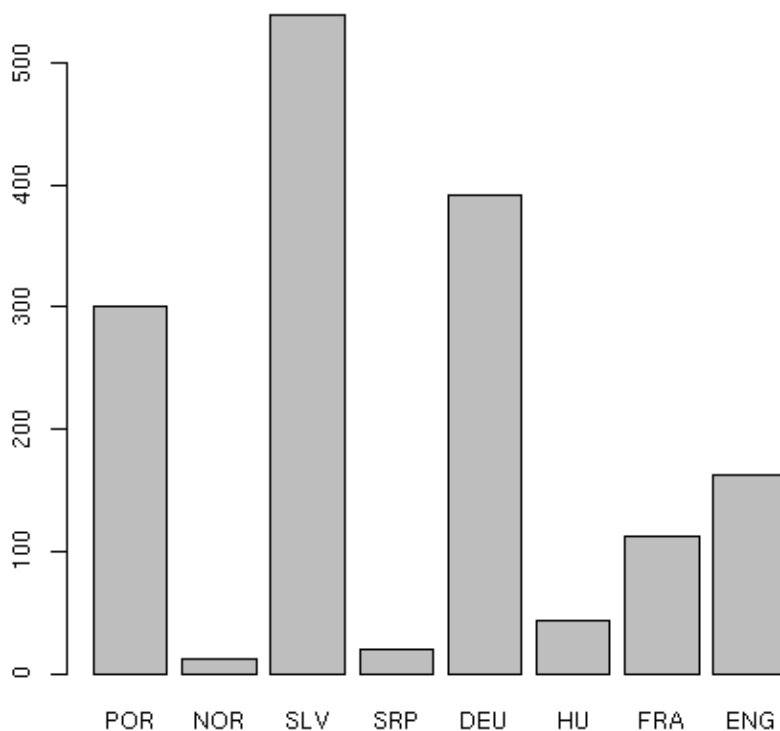
Christmas

Christmas in 8 European literatures: number of works which cite it



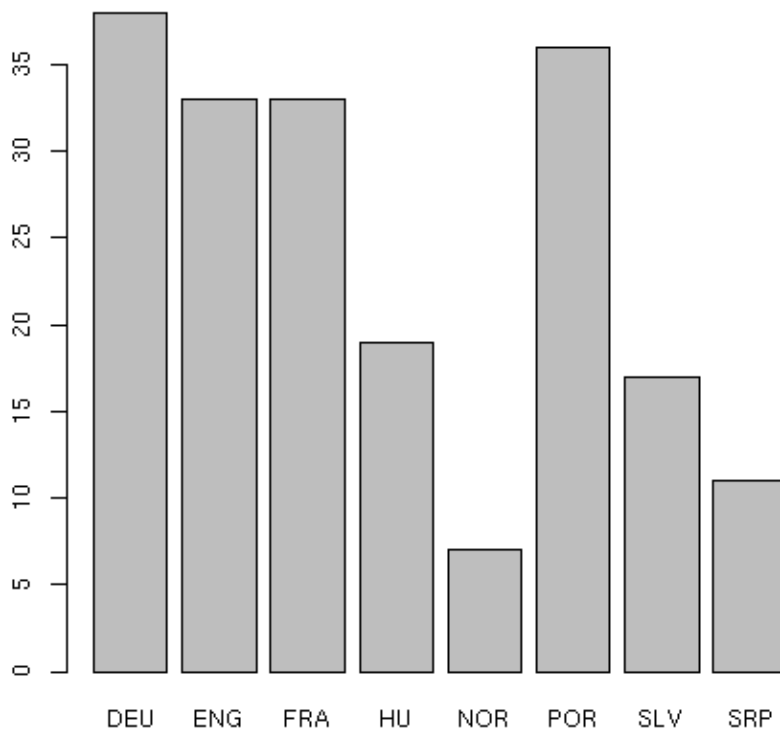
Napoleon

Napoleon in 8 European literatures: amount of references



Napoleon

Napoleon in 8 European literatures: number of works citing him



Final remarks

Goals of this morning

- a better understanding and familiarity with the different ELTeC collections
- some examples that may inspire other studies
- some knowledge on how to use basic R capabilities to visualize and test hypotheses
- further understanding of what may be involved when one talks about named entities for literature, and of the many different options that one can take