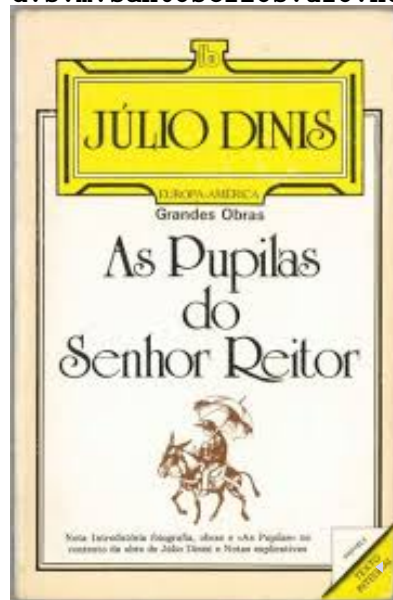


# Distant reading health

A pilot study on health and disease in lusophone literature

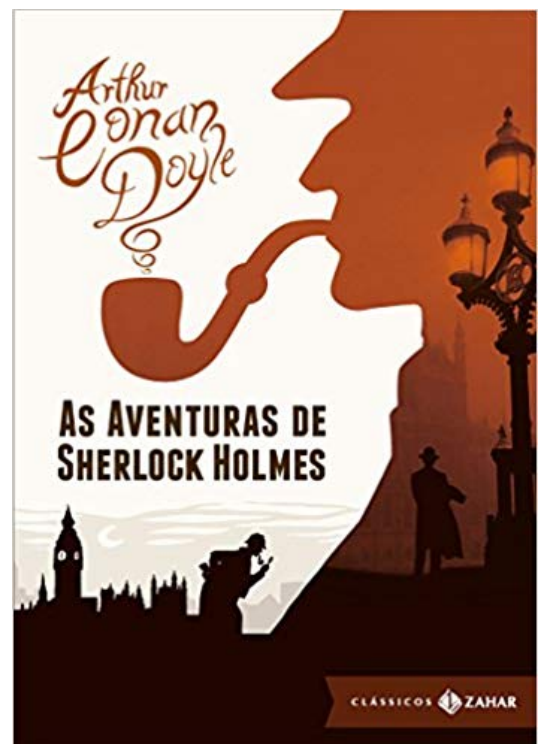
Diana Santos  
and the lusophone DR team

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



## Distant reading

- Proposed by Franco Moretti as the only alternative to stop studying only canonical works in few languages
- “world literature” + “the great unread”
- Use formal features in as many as possible texts and visualize them
- Bringing big data methods into literature studies



Moretti (2013), The slaughterhouse of literature

# First try: Detective novels, or the Rivals of Conan Doyle

Hypothesis: the invention/use of clues is the clue to success.

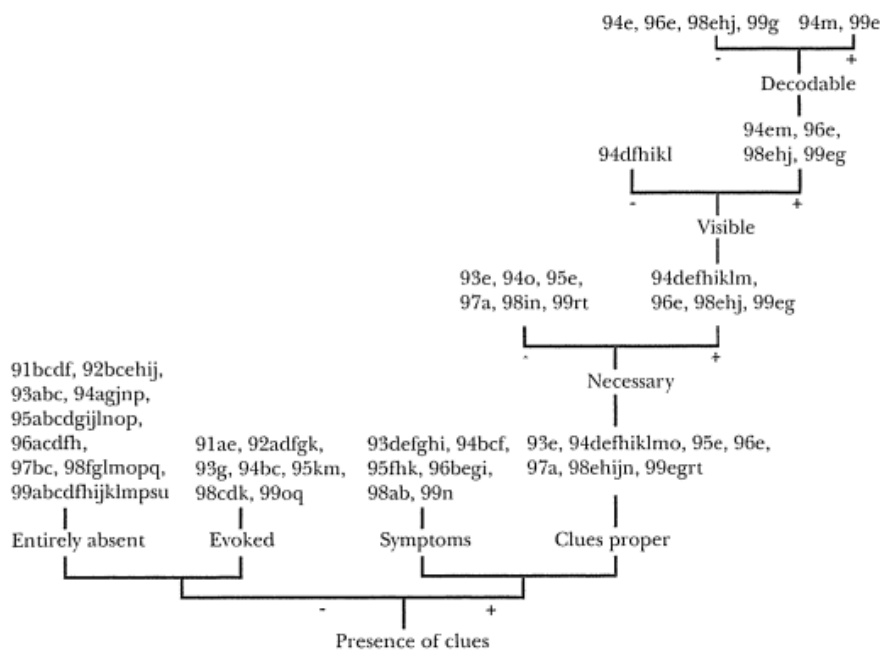


Figure 2: Clues in the Strand magazine, 1891–99

Flaw: still using the canon to find what to investigate.

## Health and disease: privative opposition in Portuguese?

- Clues: words/terms that relate to the *saúde* domain (health).
- Now, this is not a genre, so the research question is: can we find something about culture and society in the way health and disease are thematized in (mostly canonical) literature?
- This is a pilot. If we find something interesting, we'll have to do a much larger and more serious study.

Privative opposition: *saúde* (health) can be good or bad, while *doença* (disease) is always bad. So *saúde* is unmarked, *doença* is marked.

# Investigating the (folk)lexicon

The first step is to categorize the clues: the words that point (belong to the semantic domain) of *saúde* can be, in addition to generic (*saúde*):

**disease names or symptoms** peste, lepra, febre, doente (sick), acamado (bedridden), tumor, inchaço, ...

**medical infrastructure** médico (doctor), enfermeiro (nurse), hospital, termas (spa), operar, ...

**progression** adoecer (fall ill), convalescença, restabelecer-se (recover), melhoras (improvement), epidemia, contágio, cura (healing), ...

**cause** bactéria, vírus, micróbio...

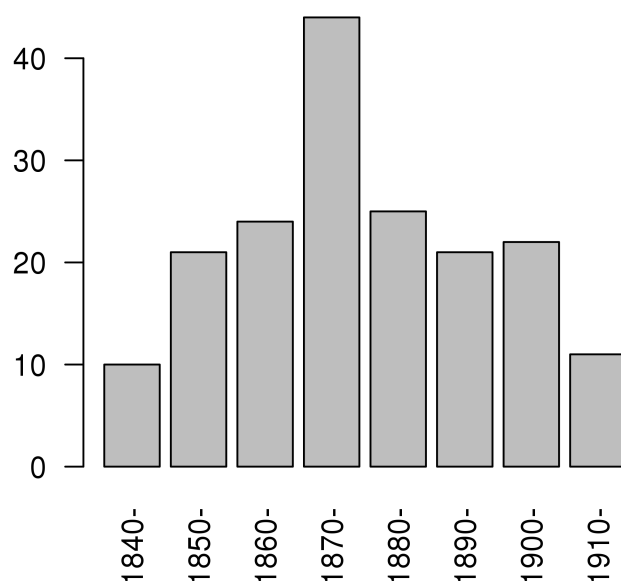
**instruments** termómetro, bisturi, estetoscópio, ...

and annotate all occurrences of these words within the literary texts.

## The material

- Total of 178 novels
- 58 Brazilian, 120 Portuguese
- Different origins
- Only 18 works by women
- Balanced canonicity

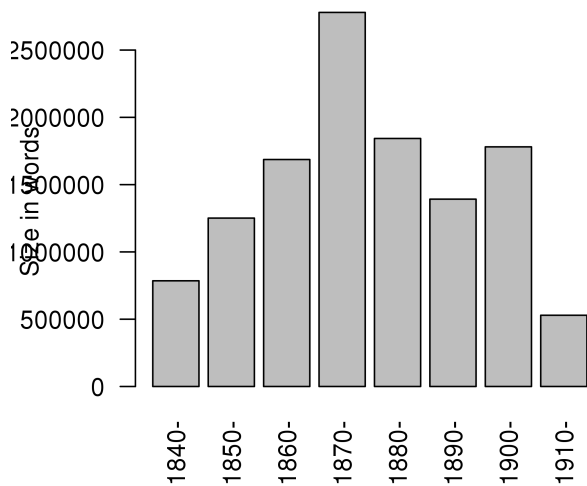
**Number of novels per decade**



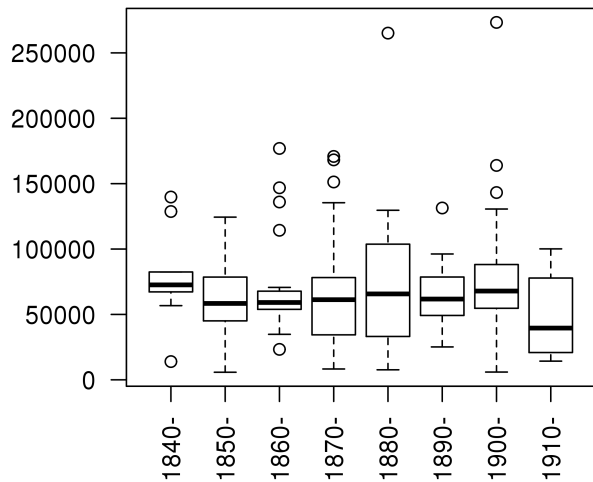
# The material (cont'd)

In terms of word size, total and per novel

Lusophone novels 1840-1919



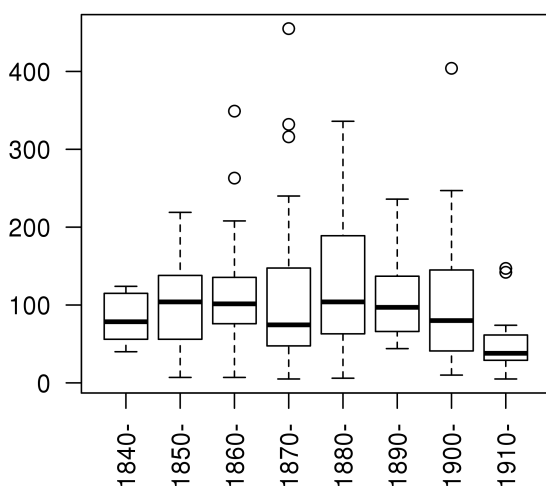
Size in words of lusophone novels



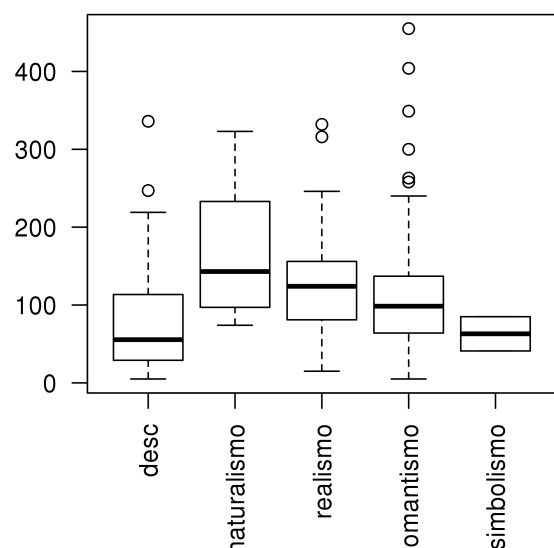
# Investigating the use of health terms in literature

The first pass is to identify their use in general, by date, and correlation with literary school:

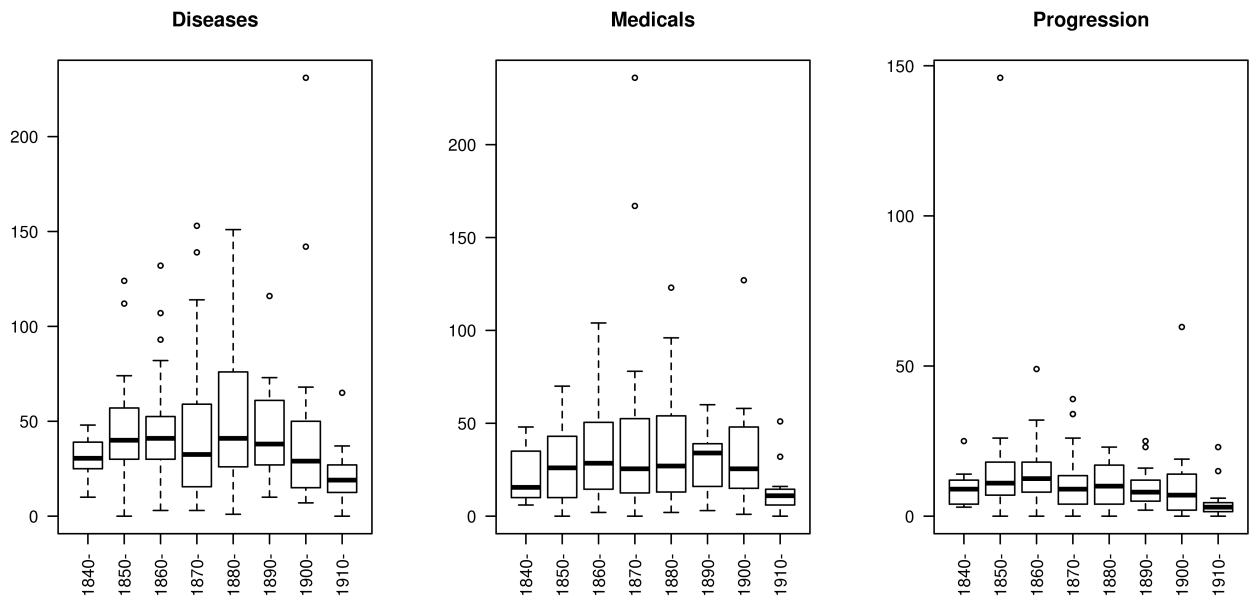
Presence of the health domain



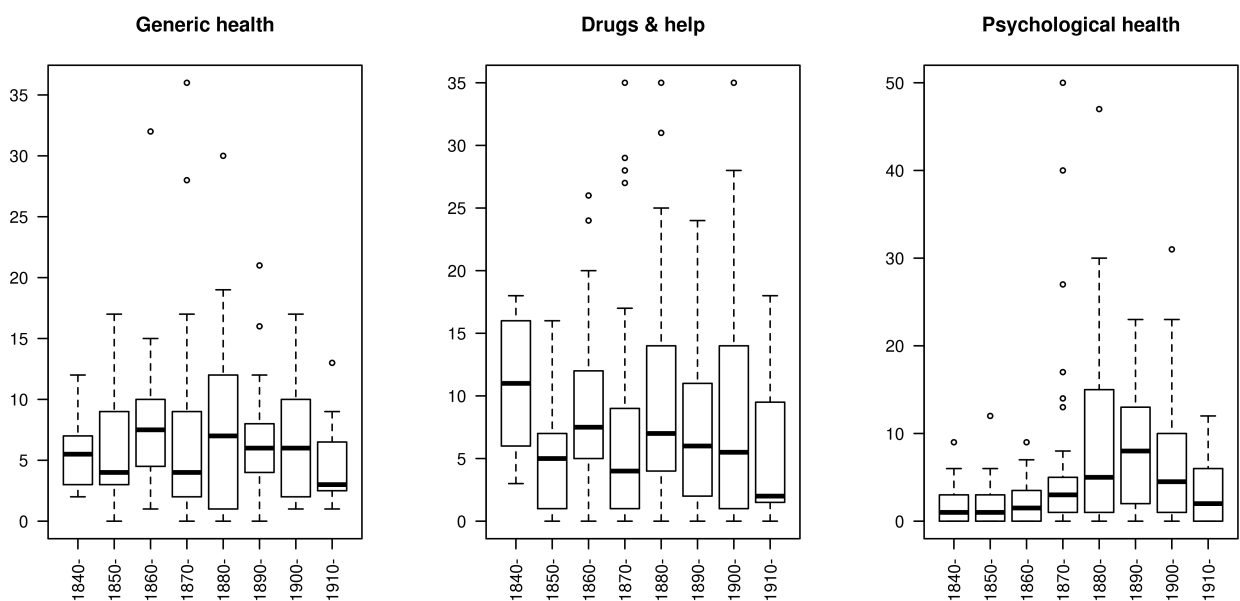
Correlation with literary school



# Distribution by these kinds of clues



# Distribution by these kinds of clues (cont'd)



Psychological health is only *nervoso* and derived terms...



## And what about *dor* (pain?)?

A heated discussion in the group

- Initially, *dor* was not considered a word from the health domain
- But this caused a lot of objections, especially because the word for illness in Portuguese (*doença*) comes etymologically from *dor*

So new inquiries were in place: the use of the word *dor*, separately, and together with the other words. Also:

- A random sample of 50 cases of *dor* was analysed to see how often it meant disease – 11 physical pain, – 28 sadness/moral, – 10 vague between the two.
- The topics in a topic models analysis of the same works were checked to see if health and *dor* co-occurred – NO

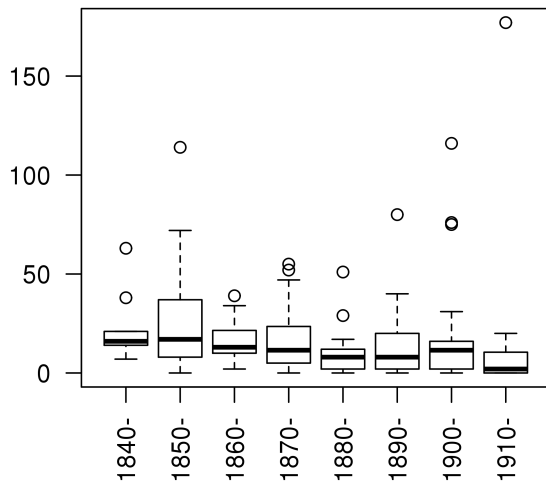
## Why different languages matter

But a little intermezzo here

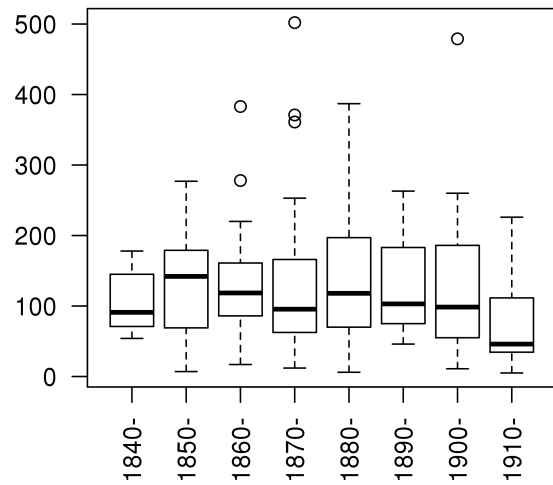
- I am not discussing the English word *pain*, nor the French words *douleur* and *mal*...
- and no better advocate of the importance of these differences is the chapter *Pain: is it a human universal? The perspective from cross-linguistic semantics* by Wierzbicka and Goddard (1994)

*The English word pain does not 'cut nature at its joints'. Its meaning is a conceptual construct, perhaps not uniquely English, but not truly universal either.*

### Distribution of *dor*

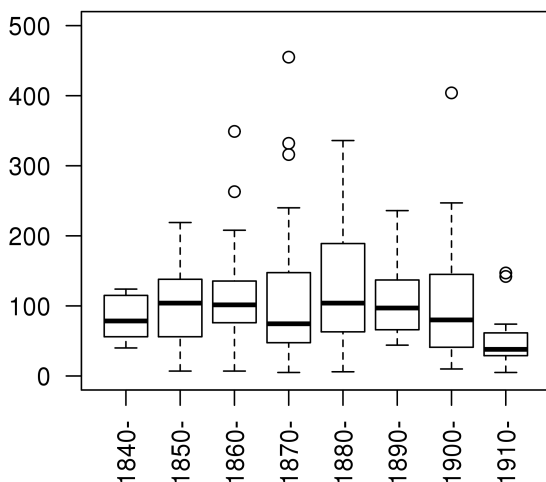


### Distribution of health with *dor*

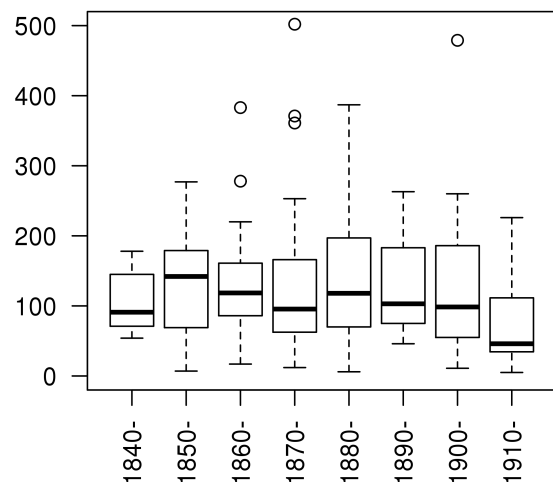


## Data on *dor*, cont'd

### Health in lusophone novels



### Health and *dor* in lusophone novels



Note that words corresponding to injuries (like *ferida* or *ferimento*) have not (yet?) been addressed/included.

## But do all health terms used concern health?

The second pass is to look at examples in context (we looked at 300 random cases), and noted that:

- Although illness may be relevant to the plot, most of the times it isn't
- Some cases can refer to plot characters (we already knew that, but got this immediately confirmed), main characters or types
- Others can be used to characterize a person or a place
- They are often used metaphorically
- A very common case is hypothetical mention
- There are totally irrelevant references
- There are several general truisms or considerations

So, a lot of additional work is in place...

## Some examples of metaphorical use

- *Se na minha mão estivesse o remédio* (If I had the solution in my hands)
- *a peste da crioula está aí* (she is there like the plague)
- *uma expansão febril de concupiscência* (a feverish expansion of greed)
- *saúde da vossa alma* (health of your soul)
- *peço uma grande saúde* (I toast to his health)
- *dito que V. S<sup>a</sup> estava muito doente de desgostos* (said that you were sick from sadness)



## Some examples of hypothetical mention

- *cheguei a supor-vos enfermo!* (I even imagined you had fallen ill)
- *Mas que precisão tem elle de expôr a sua saude n'um clima tão perigoso?* (But why expose his health in such dangerous climate?)
- *o receio contínuo de doenças fantasiadas* (the steady fear of imagined diseases)
- *Mas Amélia recusara-se a que o seu filho usasse cueiros alheios, trazendo-lhe talvez um contágio de doença* (A. did not allow her child ... for fear of contagious illness)

## Some examples of generical mention

- *um simplório incapaz de destrinçar entre doença e pecado* (a simpleton unable to distinguish between illness or sin)
- *A saude não se póde comprar com ouro* (health cannot be bought with gold),
- *o que não tem remedio remediado está* (what has no remedy, we do not need to do anything about)
- *E mau médico o que sofre que o doente o interrogue sobre a moléstia e o tratamento* (bad is the doctor who suffers questions from the patient about the disease and the treatment)
- *Na Allemanha e na Suissa comem-n'as para curar as doenças gastro-intestinaes.* (In Germany and Switzerland they eat them to cure gastro-intestinal diseases.)

## Some examples of actual mention of diseases in the plot

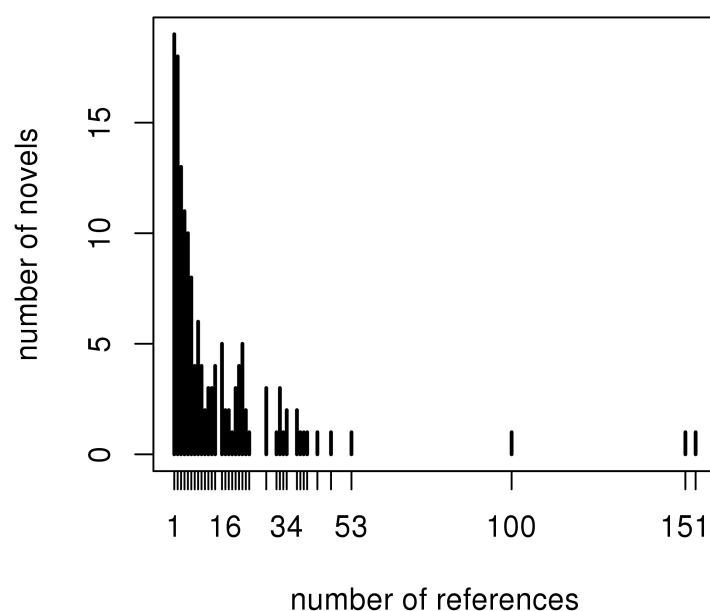
- *o medico constatara que a doente soffria de uma lesão cardiaca* (the doctor realized that the ill (woman) suffered from a heart lesion)
- *desaparecendo naquele ar de febre que enchia a alcova* (disappearing in that feverish air that filled the room)
- *mas a senhora tem febre* (but you have feber)
- *Velha assim mesmo e doente, como estou* (old and ill as I am)
- *É preciso levá-la daqui para o hospital* (you must take her to the hospital)

## When a doctor is mentioned

- generic doctor
- specific doctor
- character
- charlatan

There are 21 novels where there are more than 25 mentions to a doctor (*médico* or *facultativo*), out of the 151 where at least it is mentioned once, out of 181 total (83%).

### References to a medical doctor in novels



# Why doctors?

My interest in doctors was triggered by my belief that

- There is a huge proportion of Portuguese authors who are doctors or at least trained as doctors: Lobo Antunes, Miguel Torga, Fernando Namora, Júlio Dinis, Prista Monteiro, Fialho de Almeida, ...
- This is something that was called to my attention by my English teacher at the British institute, who claimed that the status difference between MD, GPs in England and in Portugal was huge
- The importance that literature has for the medical profession is again higher in Portugal – for example, our Nobel laureate in medicine, Egas Moniz, was also the editor (and main scholar) of the works by Júlio Dinis, who in turn immortalized the figure of the country doctor, João Semana...

So I thought this was something special about Portuguese literature... and thus the next question was: would that imply that doctors would occur frequently in the plots?

## Roque Gameiro, 1904, cited in *Acta Med Port* 2014



# With a little help from European literature

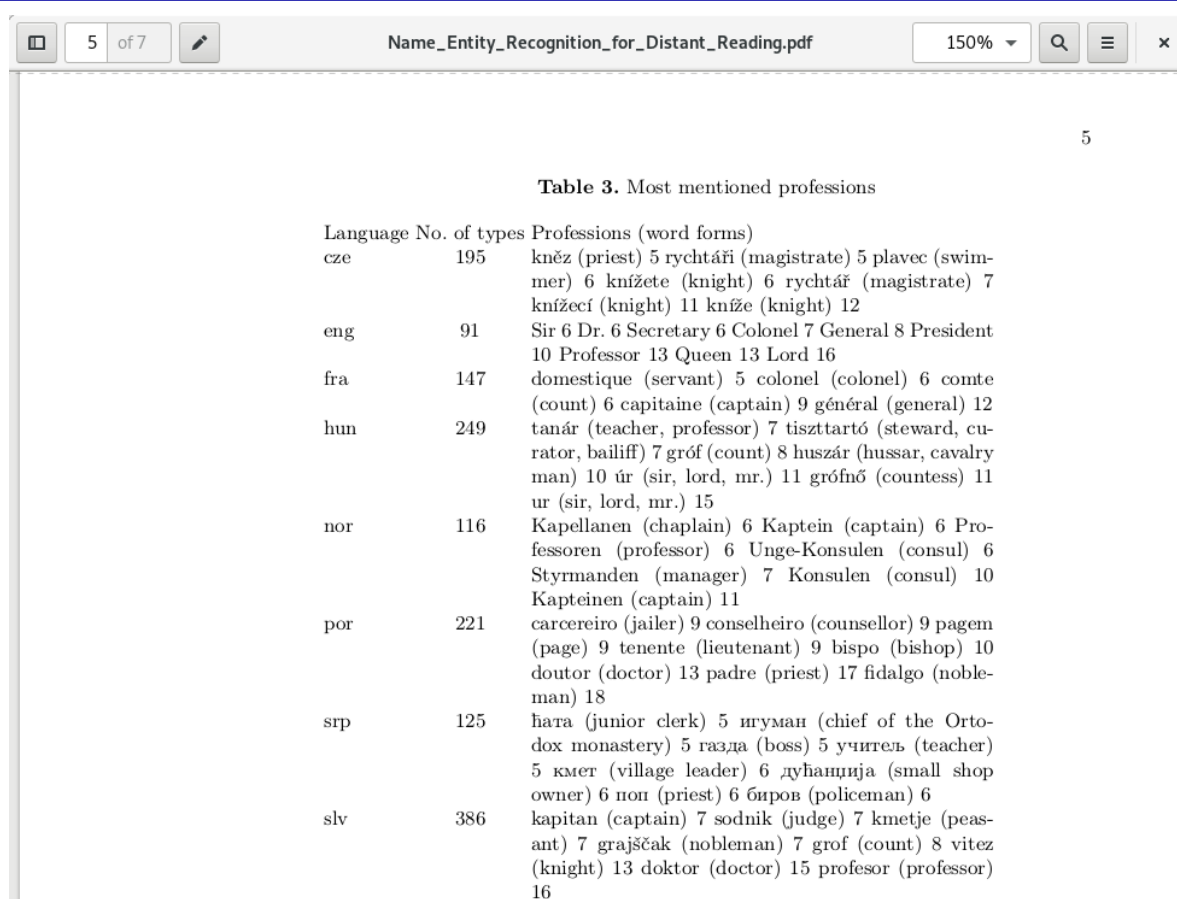
This interest was also boosted by the fact that I represent Norway in the COST action “Digital Reading for European literature”:

## Distant Reading

- period 1840-1919
- in the computational group we are looking at NER across languages based on excerpts from 20 randomly chosen novels in each literature
- we are including professions/roles: and doctors have appeared frequently

Now, one problem of course is to guarantee that the title “dr.” relates to medicine...

## In seven European literatures, very preliminary data



Name\_Entity\_Recognition\_for\_Distant\_Reading.pdf 150% Q ☰ x

5

**Table 3.** Most mentioned professions

Language	No. of types	Professions (word forms)
cze	195	kněz (priest) 5 rychtáři (magistrate) 5 plavec (swimmer) 6 knížete (knight) 6 rychtář (magistrate) 7 knížecí (knight) 11 kníže (knight) 12
eng	91	Sir 6 Dr. 6 Secretary 6 Colonel 7 General 8 President 10 Professor 13 Queen 13 Lord 16
fra	147	domestique (servant) 5 colonel (colonel) 6 comte (count) 6 capitaine (captain) 9 général (general) 12
hun	249	tanár (teacher, professor) 7 tiszttartó (steward, curator, bailiff) 7 gróf (count) 8 huszár (hussar, cavalry man) 10 úr (sir, lord, mr.) 11 grófnő (countess) 11 ur (sir, lord, mr.) 15
nor	116	Kapellanen (chaplain) 6 Kaptein (captain) 6 Professoren (professor) 6 Unge-Konsulen (consul) 6 Styrmenden (manager) 7 Konsulen (consul) 10 Kapteinen (captain) 11
por	221	carcereiro (jailer) 9 conselheiro (counsellor) 9 pagem (page) 9 tenente (lieutenant) 9 bispo (bishop) 10 doutor (doctor) 13 padre (priest) 17 fidalgo (nobleman) 18
srp	125	hara (junior clerk) 5 игуман (chief of the Orthodox monastery) 5 газда (boss) 5 учитељ (teacher) 5 кмет (village leader) 6 дућаница (small shop owner) 6 поп (priest) 6 биров (policeman) 6
slv	386	kapitan (captain) 7 sodnik (judge) 7 kmetje (peasant) 7 grajščak (nobleman) 7 grof (count) 8 vitez (knight) 13 doktor (doctor) 15 profesor (professor) 16

☰ Q ↻

## An initial screening for distant reading health

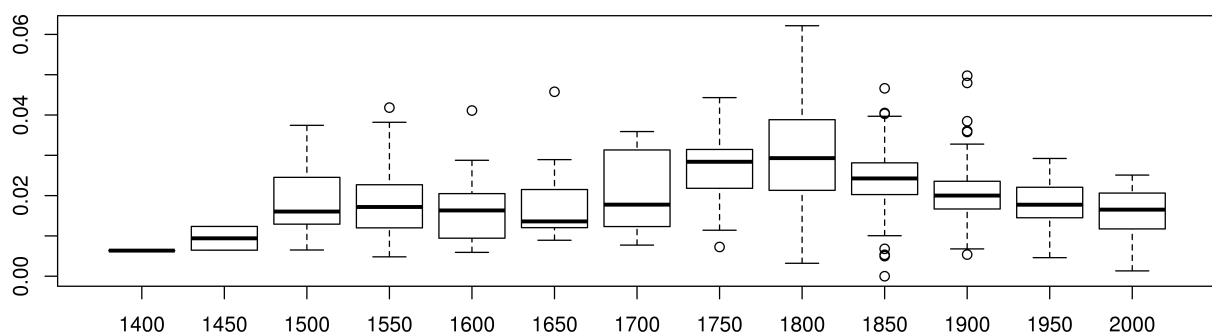
- the lexicon and its distribution
- the (varied) uses of same lexicon
- the issue of what is related/included in health
- health professions in the novel

No diagnostics yet. But a lot of interesting research alleys face us.

## One way from here

### Correlation of emotions and health

- Are there high correlations, or are these two different axes in a story?
- How are lexicons shared? *Dor* is a good example, but there are all sorts of metaphorical and shared vocabulary: *desmaio*, *nervos*, *deprimida*...
- What is the folk understanding regarding health and body which is embedded in language? *maus fígados*, *pálida*, etc.



- Goddard, Cliff & Anna Wierzbicka. Pain: is it a human universal?", in Cliff Goddard & Anna Wierzbicka (eds.). *Semantic and Lexical Universals*, John Benjamin Publishing, 1994, pp. 127-155.
- Moretti, Franco. "Conjectures on world literature", *New Left review* 1, Jan-Feb 2000, pp. 54-68.
- Moretti, Franco. "The slaughterhouse of literature". *Modern Language Quarterly*, 61:1. Reprinted in *Distant reading*, Verso, 2013, pp. 63-90.
- Stanković, Ranka, Diana Santos, Francesca Frontini, Tomaž Erjavec & Carmen Brando. "Named entity recognition for Distant Reading in several European literatures". Submitted.

### About the images:

- Picture from <http://dp.uc.pt/conteudos/entradas-do-dicionario/item/255-joao-semana>
- Cruz, Jorge. "Este cavalheiro era João Semana". *Acta Med Port* 2014 Jan-Feb;27(1):148-150.