

Distant reading of food and drink in three languages

Diana Santos

Linguateca & University of Oslo
d.s.m.santos@ilos.uio.no



Image from <https://bokogbibliotek.no/debatt/heime-pa-biblioteket-4685/>

MADRI-PIN



What is distant reading

- a new turn in literary studies
- term coined by Franco Moretti
- an attempt to consider more than what can be close read
- using computerized tools
- advantages: challenges canons, allows one to concentrate in a much larger number of works (in different languages), allows one to study literary history in different ways

Our pilot project in 2023

MADRI-PIN: Mat og drikke i portugisisk, italiensk og norsk

- OsloMET organized two workshops
- and one open meeting at ILOS, afternoon 19 December 2023

Main participants: Michael Shalom Preminger, Åse Kristine Tveit, Elizaveta Khachatryan, Janicke Kaasa, Diana Santos, Eckhard Bick

- identified some initial research questions
- identified how to go about, specifically in terms of which annotation to perform
- human analysed some meal scenes in four different literatures (also Russian)

Our intentions to proceed this work

- corpus annotation of Italian, Portuguese and Norwegian literary texts with the VISL3 tools
- creation of an ontology of food and drink, based on the empirical data
- use of CorpusEye to do (comparable) corpus searches
- annotation of meals in the literary works (so far manual, but possibly automatable with machine learning)

Presentation of possible ontology

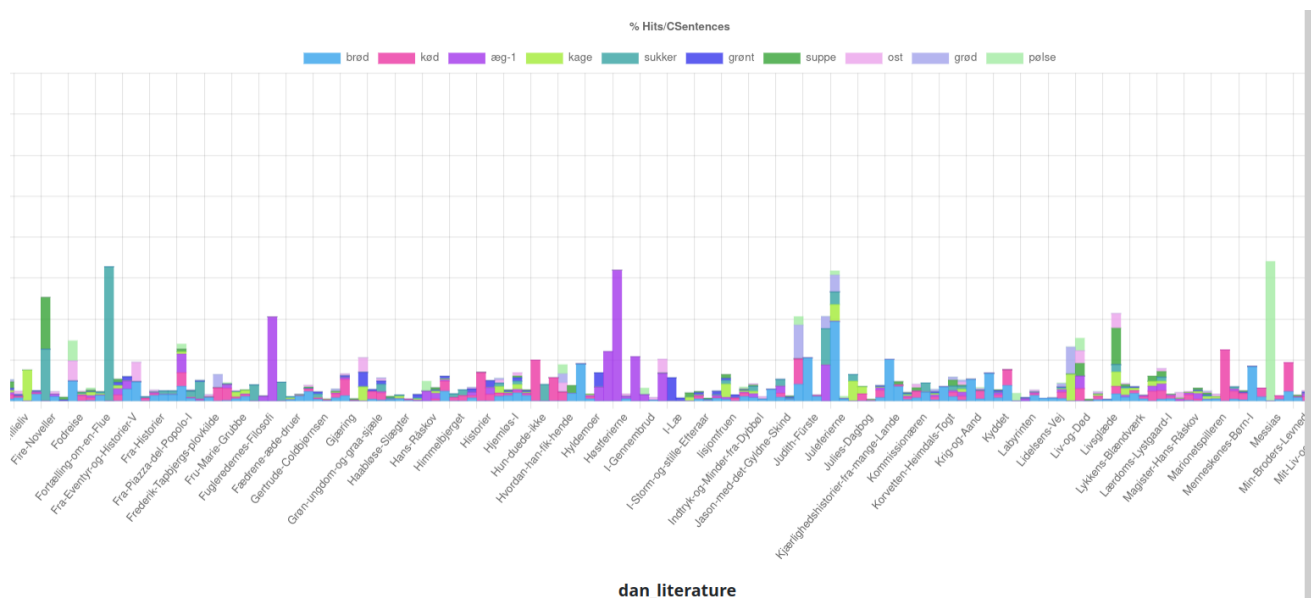
From Preminger et al. 2023:

- ◆ **<food>** **<food-h>** (human-processed), **<fruit>**, **<spice>**
 - subcategories: **<food-meat>**, **<food-fish>**, **<food-veg>**
 - **<food-liq>** (e.g. *soup*)
 - **<food-warm>** (cooked, fried, boiled), **<food-cold>** (e.g. *ice cream*)
 - **<food-sweet>**
- ◆ **<cm-chem>** (chemical substances), **<cm-rem>** (medical remedies)
- ◆ **<drink>**, subc.: **<drink-alco>**, **<drink-hot>** (e.g. *tea*), **<drink-cold>**
- ◆ events: **<occ-eat>** (banquet), **<occ-drink>** (pub crawl), **<act-food>**
- ◆ places: **<Leat>**, subc.: **<Lh-eat>** (kitchen), **<inst-eat>** (restaurant)
- ◆ containers: **<con-food>** (plate), **<con-drink>** (cup), **<con-store-food>**
- ◆ tools: **<tool>**, subc.: **<tool-eat>** (fork), **<tool-gastro>** (frying pan)
- ◆ humans: **<Halco>** (drunkard), **<Hprof-food>** (baker)
- ◆ others: **<sick-food>**, **<ism-food>**, **<percep-food>** (hunger, thirst)

categories can overlap, for instance 'cup' as **<tool-drink>** and **<con-drink>**



Presentation of CorpusEye



Intended outcomes

- the language(s) of food and drink in the three languages, and how they are used in literature (in the three literatures)
- clusters of novels according to their relationship to food and drink
- the meaning of meals for a plot
- public resources for further studies
- (initial) comparison of three different languages/cultures

Methodology

- corpus annotation with rule based parsers (with some statistical enhancements),
- human revision,
- exploratory statistics
- development of an automatic classification system for meal scenes in literary texts

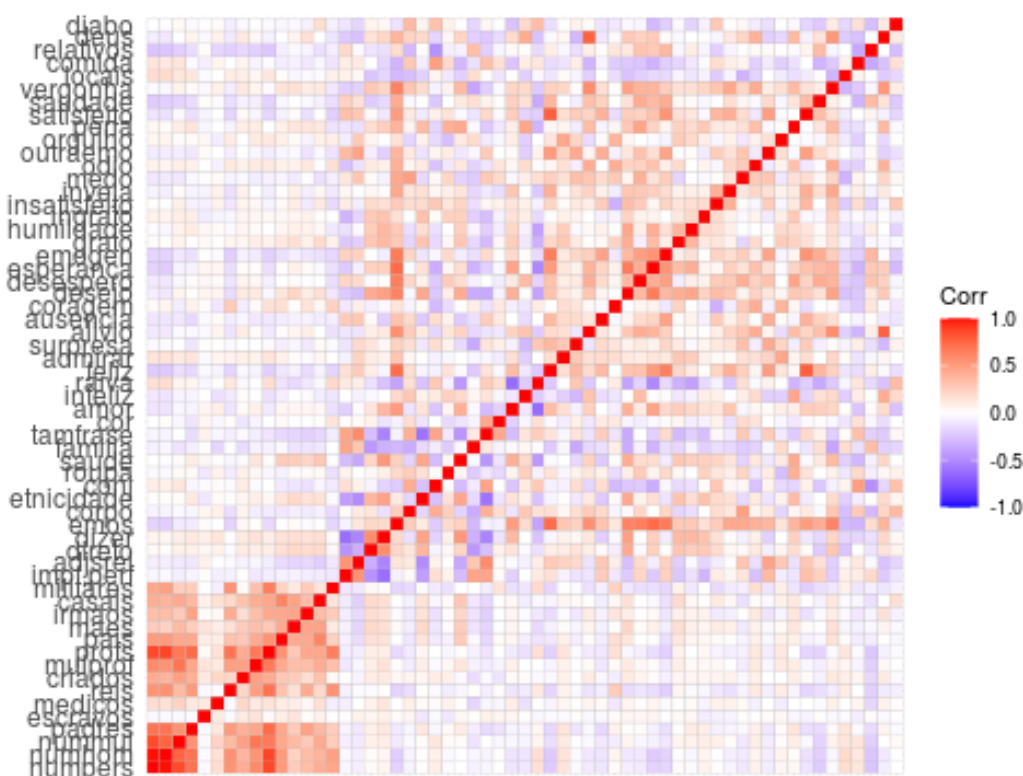
First results

One interesting question: is there a correlation between features about entire novels, and features describing their content? E.g.,

- will novels with medical doctors as characters feature a lot of health words?
- will novels with family relationships among characters feature a lot of family relationship words?
- will novels with a lot of eating/drinking words feature a lot of meal scenes in the plot?

See Santos (2024), “Literary similarity of novels in Portuguese” in PROPOR 2024, for more information.

Correlations



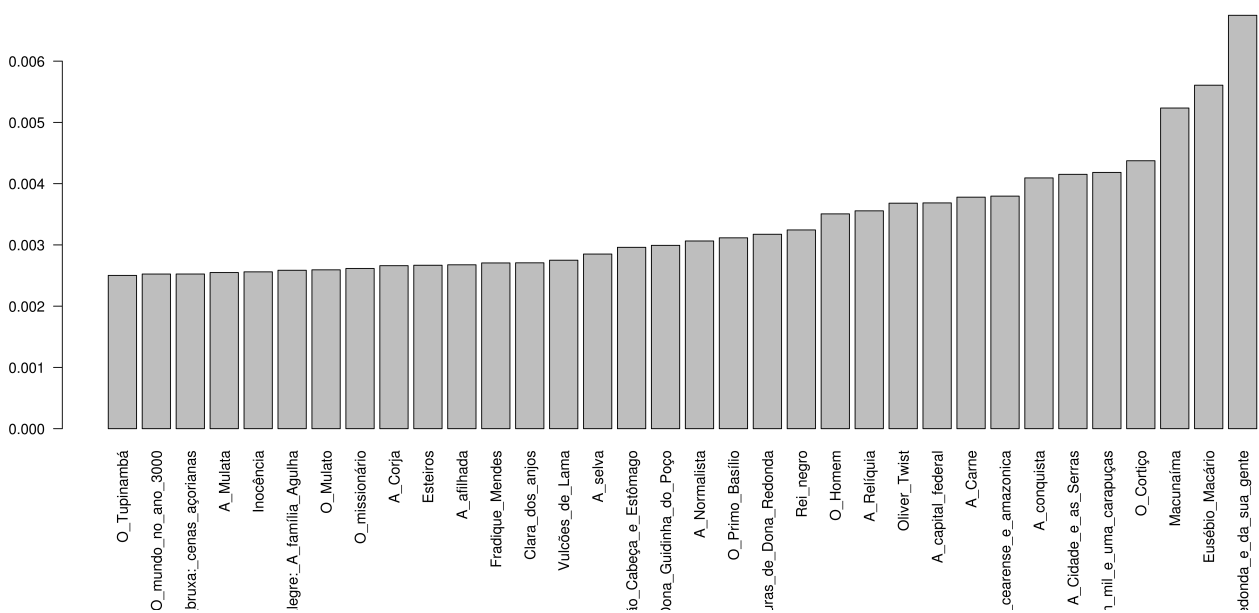
The language of food in Portuguese

In 366 novels, 23 million words, annotation not yet revised

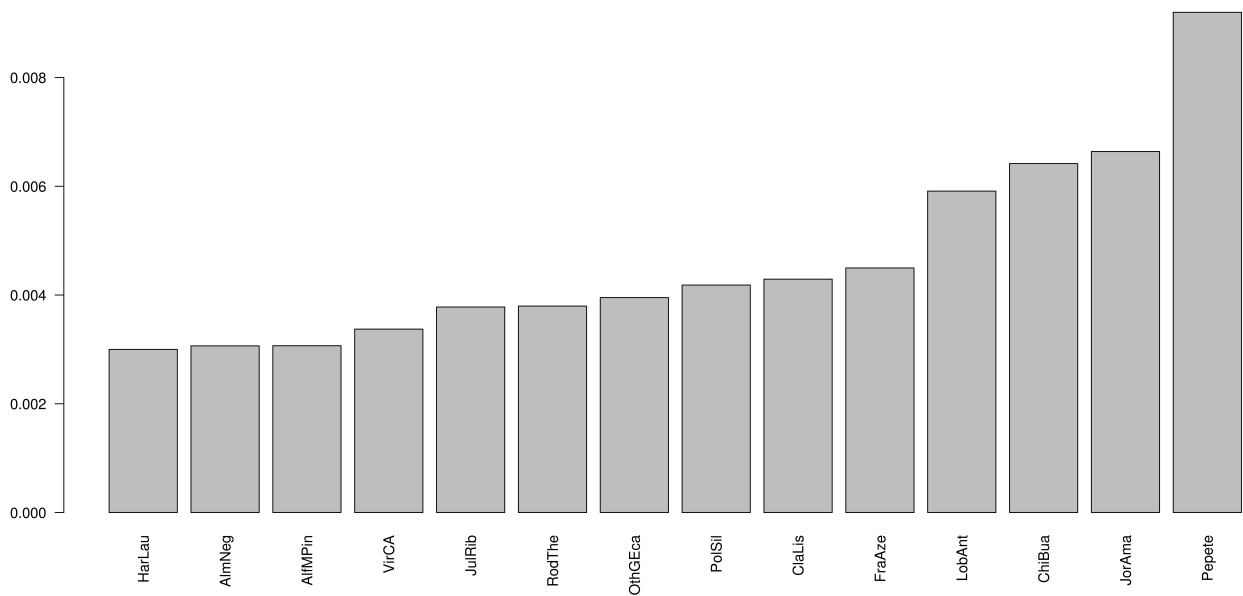
comer (V)	eat	3588	café	coffee	449
beber (V)	drink	3137	cear (V)	dinner	431
carne	meat	2198	ovo	egg	427
vinho	wine	1953	milho	mais	421
pão	bread	1863	sorver (V)	sip	387
chá	tea	967	sopa	soup	377
prato	dish/plate	814	mel	honey	367
leite	milk	741	bolo	cake	355
doce	sweet	696	farinha	flour	341
fruta	fruit	545	aguardente	brandy	339
comida	food	519	cerveja	beer	320
tomar (V)	take	513	arroz	rice	279
água	water	511	champanhe	champagne	253
caldo	broth	479	queijo	cheese	252
alimento	food	455	conhaque	cognac	196



The distribution of food by lusophone novels

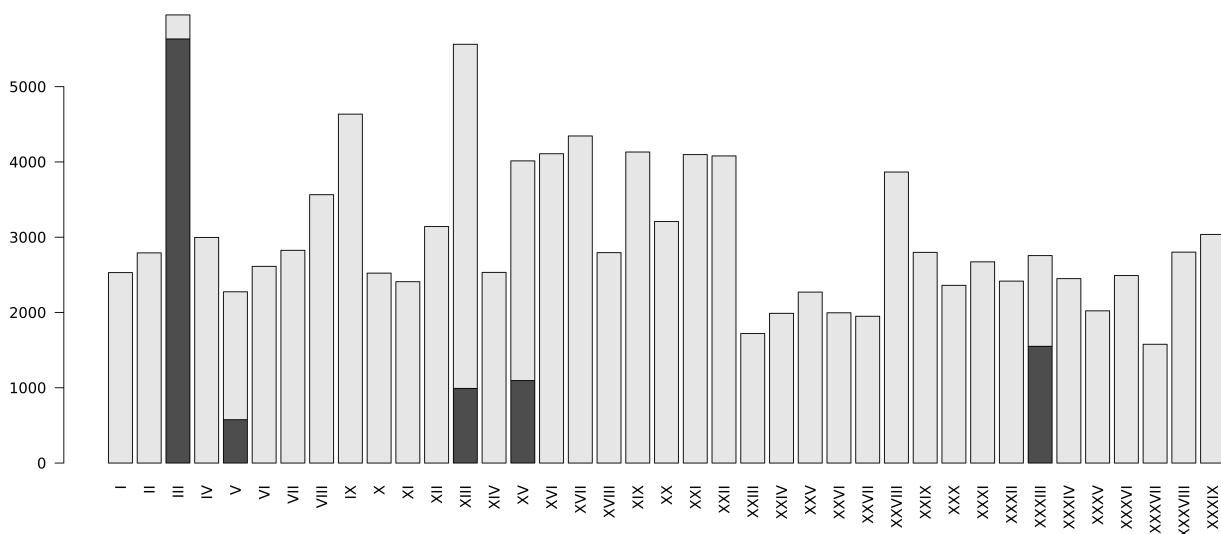


The distribution of food by lusophone authors



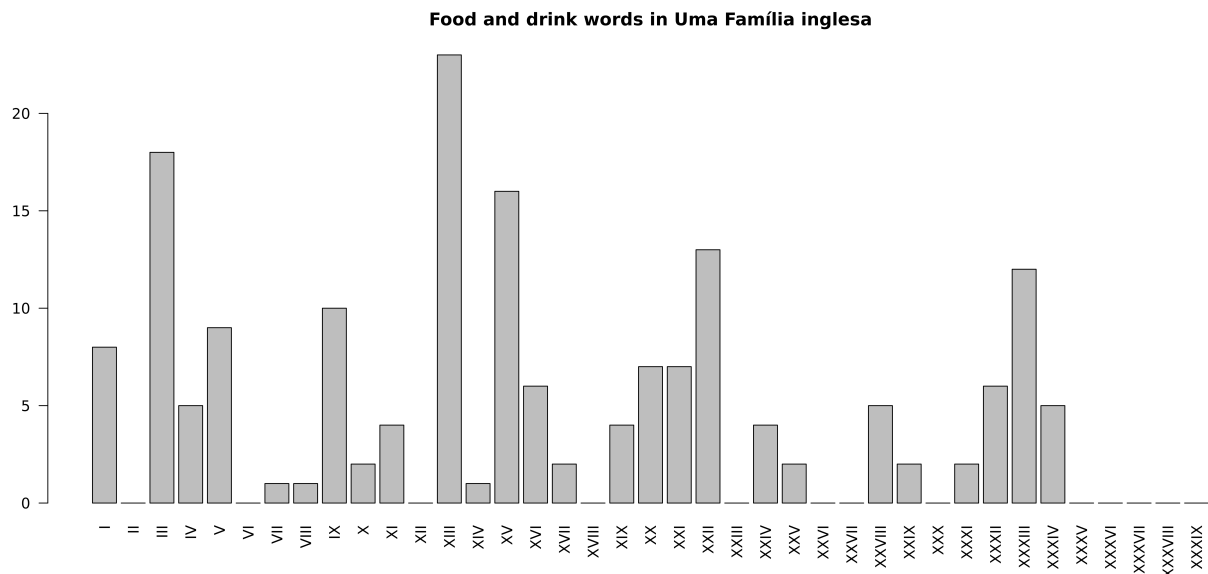
Meal episodes in one novel

Meals in Uma família inglesa



First, identify meals in the full text, then count the words.

Food and drink words in the same novel



Count food and drink words per chapter.

The question of water

How often is *água* something we drink? (As opposed to sea, rivers, rain, washing, cooking material, ...)

	Size in tokens	<i>água</i>	drink	%	
Colonia	5.361.152	1730	<u>189</u>	160	0.092
Tycho Brahe	3.550.064	823	<u>91</u>	61	0.074
NOBRE	9.621.352	2111	<u>92</u>	70	0.033
OBras	11.179.736	4762	<u>380</u>	318	0.067
Vercial	15.912.885	5624	<u>554</u>	442	0.079
PANTERA	701.199	<u>312</u>	19	33	0.106
Literateca (novels)	23.161.262	6664	-	511	0.077

(Underlined the human revised cases, for which correction rules were written.)

Concluding remarks

- A project in its beginning
- We have some idea of the research questions
- The methodology is also defined
- A lot of work (annotation and human revision) is required
- We believe we will produce an important resource that can be used by many others to study food and drink in literature(s), as well as an example to be followed