

Reading Food:

Presenting its infrastructure and user interface(s)

Diana Santos, Elisaveta Khachaturyan, Michael Preminger, Åse
Kristine Tveit, Eckhard Bick

`d.s.m.santos@ilos.uio.no, elizaveta.khachaturyan@ilos.uio.no,
michaelp@oslomet.no, aasekt@oslomet.no, eckhard@bick.dk`

DHKO, 15-16 October 2025

Brazilian, Italian, Norwegian and Portuguese, mainly in the 1840-1920 period, due to copyright constraints, and stemming from COST action Distant Reading for European literary history (ELTeC corpora).

- Project submitted to Forskningsrådet – waiting for answer in week 46
- Continuation of an infrastructure grant by Univ. of Oslo

Created in 2004, completely restructured in 2023-2024, by GrammarSoft ApS, which also owns the constraint grammar parsers used to annotate the corpora.

Conclusion

1. **Large corpora** can be used not only for linguistic purposes, but also for a large variety of humanistic research goals relating to literary, cultural or social topics
2. **Linguistic annotation** (and a search interface exploiting it) make it possible to identify the nature and usage of
 - the word clusters, notions and concepts involved
 - possible stereotypes related to the area
 - the association strength between concepts and perceived attributes
3. **Qualitative inspection** of large text collections is only possible in a meaningful way by first running a **quantative analysis**: statistics, co-occurrences, visualisations ...

https://corp.visl.dk/Corpuseye_IKS.pdf

Why CorpusEye

- Long experience of collaboration with Eckhard Bick, CorpusEye developer, for Portuguese
- Good experience with Norwegian (Bokmål) for the COST action
- A multilingual and multifaceted corpus system, with shown successes
- Run by an external/private firm, with clear contracts and deadlines



The project in a nutshell

The project encompasses several different aims:

- 1 a lexical taxonomy of F&D in the three languages
- 2 a classification of metaphors and idiomatic expressions related to this domain
- 3 a novel annotation of uses of F&D in literary works (after all, a significant number of mentions of food and drink in the texts analysed are not literal)
- 4 an investigation of pragmatic issues associated with FD
- 5 the study of food-related episodes and their purposes in literary text

In a nutshell, both linguistic and literary issues associated with food and drink.

Mainly novels published in the period 1840 to 1920

- 97 works in Norwegian (Bokmål)
- 100 works in Italian
- 215 works in Portuguese (87 from Brazil), plus 10 + 10 P&B novels after that.

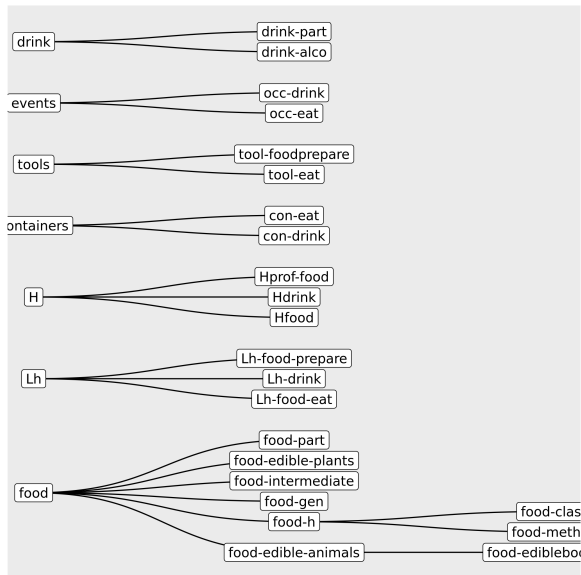
Several kinds of information

- Metadata, in XML
- Morpho-syntactic annotation with VISLCG3 parsers
- Semantic and frame annotation with VISLCG3 parsers
 - In this case, information about food and drink

Current annotation plans

First layer

- set of taxonomy categories
- disambiguation in context, using VISLCG3 disambiguation rules



Need for disambiguation, examples in Portuguese



copo

con-drink, drink-part, NO



café

drink, food-edible-plant, product, Lh-food, occ-eat (meal or part of a meal)



Need for disambiguation, example of Norwegian *Skjænk*

- furniture** I den lille Sal mod Nord var der en **Skjænk**, og i den maatte der altid staa en Flaske Burgunder (Alexander L. Kielland. *Garman & Worse*, 1880)
- present** Vi give eder dem næsten som **Skjænk** og Gave for at I skulle lære vort Sprog (Jens-Andreas Friis. *Lajla*, 1890)
- drink** Der sad Madam Torsen med sin Rok og reiste sig saavist aldrig for at finde lidt at byde – ikke en Kop varm Ølost , ikke den elendigste **Skjænk** med Fusel i. (Ragnhild Jølsen. *Rikka-Gan*, 1904)
- verb** *Han lo. «Skjænk i fuld kop du.* (S. Undset. *Vaaren*, 1914)

furniture	8
alcoholic drink	5
present	3
verb of serving	3

8 out of 19 associated with drink...



Need for an (additional) different unit than word

- Food and drink entities can appear as strings of many words
 - *garrafa de aguardente de cana da Madeira* (bottle of akvavit from sugarcane from Madeira)
 - *travessas de frango ensopado* (plates of cooked chicken)
 - *bananas fritas em azeite* (bananas fried in olive oil)
 - *sopa de puré de arroz à la princesse* (soup made of rice puree a la princesse)
- Different sets of words or groups of words can point to the same entity
 - *uísqe* and *whisky*
 - *vinho da Madeira* and *Madeira*
- Issue of compound words in Germanic languages (in this case Norwegian) makes comparison across languages especially tricky
 - Is *venneskaal* different from *skaal*? Like *afskedspjolter*
 - Is *franskbrændevin* different from *brændevin*?
 - *elskovsdrik* or *yndlingsdrikk* kinds of drinks?

Preliminary drink taxonomy (just words), Portuguese

- (38) Seltz aluá café café-com-leite cajuada capilé chazada chibé chocolate chá cidra drinque garapa gasosa genebra gengibirra gingerbeer groselha infusa infusão laranjada leite limonada lúcia-lima mate néctar orchata peppermint refresco soda soda-water suco sumo umbuzada uvaia xarope zurrapa água
- (112) Alvaralhão Amontillado Apolinaris Bairrada Beirão Bordeaux Bordéus Bourgogne Bucelas Campolide Caparica Carcavelos Carls=Berg Cartaxo Chablis Chambertin Chartreuse Chateau-Lagrange Chateau=Iquem Chianti Chypre Château-Laffite Colares Corvelo Douro Guinness Madeira Marrasquino Porto Pá Reno Rioja Romanée-Conti Sauterne Setúbal St.=Emilion Stolichnaya Tokay Vidainhos Villar=d'Allen absintho aguardente ambrosia anis anisete aperitivo azougue bagaço bock borgonha branca brandy cachaça caiçuma cana carrascão cauim cerveja chabli champanhe chicha chope cidra cliquot cocktail cognac conhaque coquetel cordial curaçau drinque espumante garapa gim gin ginja grappa grogue hidromel kermann kummel laranjinha licor malvasia manzanilla mata-bicho mocosoró moscatel moscato parati patrícia pombe ponche punch reino rum sangria santerre tinto tiquira uísque verdasco vermute vinho vinho=palhete vodca vodka whisky xerez zito zurrapa água-pé

Preliminary drink taxonomy (just words), Norwegian

- k (33) Kostforagter Melk chokolade drik drikk drikke drikkevand drikkevare forfriskning formiddagskaffe ingefærøl kaffe kaffetaar kaffitaar kakao kamillete limonade mjølk morgenkaffe morsmelk most mælk nationaldrik saft selters smage-Toddy sodavand sommerdrik te tevandsknægt toddy vand yndlingsdrik
- o (92) Acachaca Arak Champagne Julepunch absint afskeds-pjolter afskedspjolter akevitt alkohol amontillado apotekerlikør apotekerpunch avec avskedsbæger bajer bayerøl bock bordeaux bordvin brennevin bryg brygg brændevin burgunder champagne cognac dram drik drikk drikke drikkevare flaskeøl franskbrændevin franskvin frugtvin genever gin gløg gløgg grog gudedrik hedvin hjemmebryg hjertestyrkning hvidtøl hvitvin hæggebærvin ispunch julebryg juleøl julpunch knert koldpunch konjak kornbrændevin laddevin landvin likør madeira marsala morgenøl nationaldrik nypelikør porter portvin punch rabarbervin rhinskvin rhum ribsvin risling rom rødvin sauterne sherry skænk snaps spirituosa spirituosu spiritus sprit stikkelsbærvin stout taar velkomstskænk vermouht vermut vin whisky whiskytoddy yndlingsdrik øl

Preliminary drink taxonomy (just words), Italian

- k (22) acqua aranciata bevanda beverage bibita brodo caffè café camomilla cedrate cioccolata elisir infuso latte limonata nanfe seltz soda spremuta succo thè tisana
- o (31) Chianti acquavita alcool amaretto assenzio assenzo beverage birra birrone chambertin champagne grappa idromele liquore malvasia maraschino marsala mosto poncio punch rosolio rosso rum sciampagna sciampagnone sherry spumante vermutte vinaccio vino wermuth

Drink containers in Italian

In modern language, *tazza* is the most commonly used word when speaking about liquids and the action of drinking; while *ciotola* is more specific. The word *tazzina* is used in modern Italian to refer to a small cup for drinking coffee.



ciatola

tazza



tazzina

Drink containers in Italian - dictionary entries

Vocabolario degli Accademici della Crusca (1st edition: 1612, 4th edition: 1729-1738), <http://accademiadellacrusca.it>

Ciotola (1st edition)

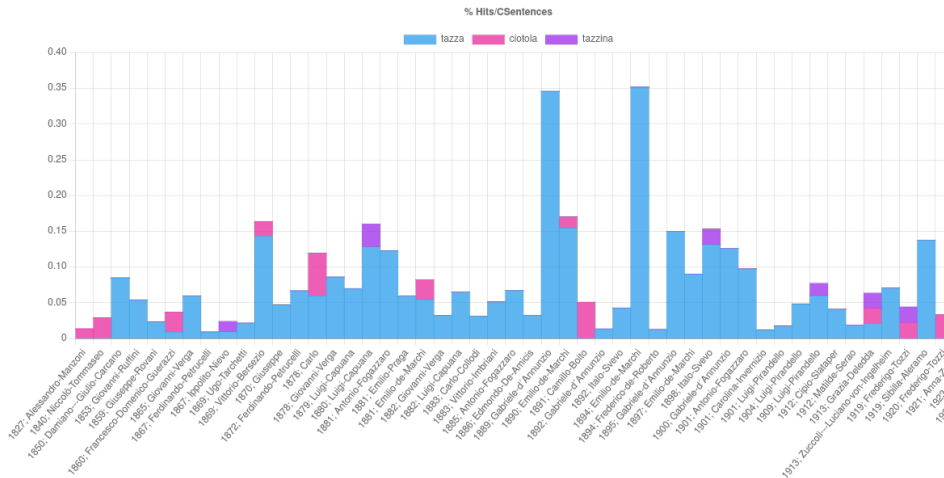
Definiz:	vasetto da bere, senza piede, di tenuta di poco più d'un comun bicchiere, da alcuni in lat. detta trulla. Flos. 13.
Esempio:	<i>Cr. 4. 32. In una metadella di dolce mosto, mescola una ciótola d'aceto.</i>
Definiz:	E ad un altro vaso, alquanto maggiore, col piede, ma di forma più piatta, e di varie materie, per diversi usi, diciamo. TAZZA.

Tazza (4th edition)

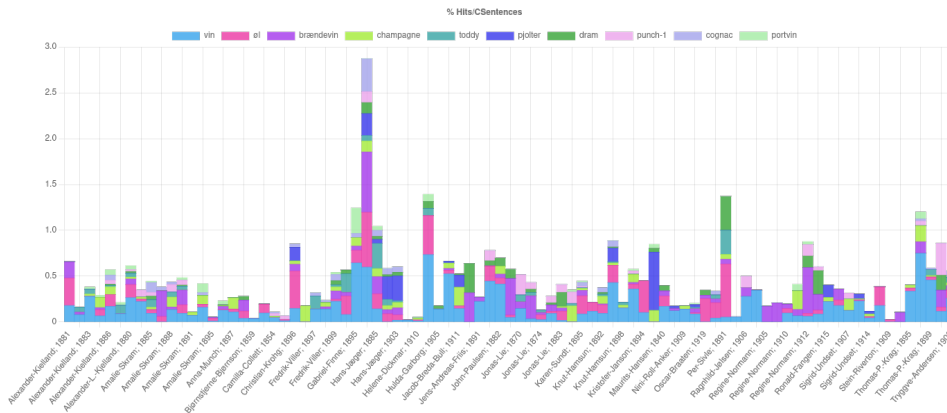
Definiz:	Sorta di vaso di forma piatta col piede di diverse maniere. Lat. patera, cotyla. Gr. κρατήρ, κοτύλη.
Esempio:	<i>Franc. Sacch. nov. 221. tit. A messer Ilario Doria ec. è tolta una tazza d'argento di valuta di trenta fiorini. </i>

When looking in a diachronic perspective, the definition of the word *tazza* appears only in the fourth edition, while in the first edition the reader is redirected to the word *ciotola*. This shows that the word *tazza* and the object to which it is referred entered in use later. According to Sabatini & Coletti's dictionary, the use of *tazzina* is attested since 1879.

Drink containers in Italian - in the corpus



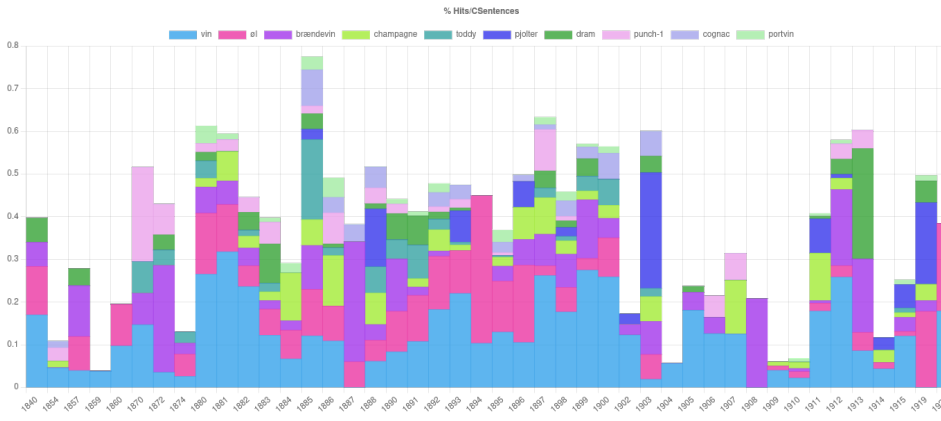
Main alcoholic drinks in Norwegian literature



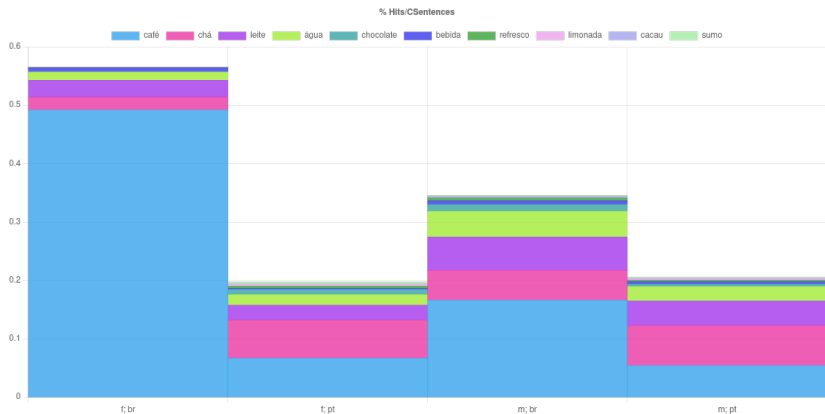
Main alcoholic drinks in Norwegian literature - preliminary remarks

- *Punsj* (punch) is less used after 1890, maybe correlated with the change from the wide realist novels with a lot of society to the sadder naturalist novels)
- *Øl* (beer) appears less after 1900, when *pjølter* (a cocktail made from liquor and soda) does a strong entrance, a cheap drink with wide adherence in cities. Maybe because there are more urban novels in this period
- *Vin* (wine) is evenly distributed, but *dram* (shot) appears less after 1885
- The graph with overview on alcohol drinks over time shows a variety of drinks in the novels from 1880-1900. Many types are mentioned, while there are clearly fewer occurrences in the years after 1900. This can again be linked to the literary development, where the tendency is a turn from the broad depictions of the external surroundings of human life to an increased interest in the human inner life.

Main alcoholic drinks in Norwegian literature by date only

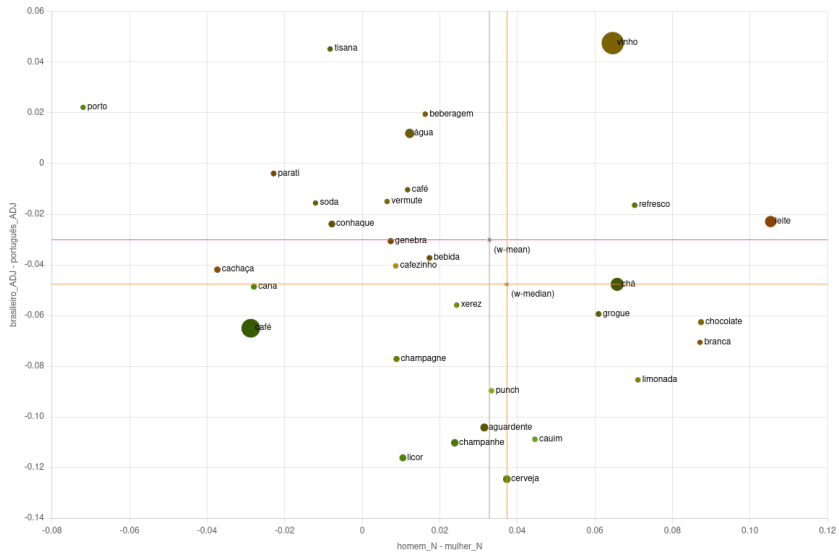


Non-alcoholic drinks in Portuguese according to gender and language variety



Female Brazilian writers mention most coffee; male Brazilian writers mention most frequently milk.

Food and drink in gender and language variety axes



A preview of what can be done in CorpusEye

In sum, we gave a short preview of what CorpusEye can do in distant reading food and drink in literary corpora:

- diachronic lexical semantic studies
- contexts of drinking
- distribution by authors, year and other metadata
- complex vizualizations based on word embeddings

The detection of meals in literary works

Scene detection, inspired by Schank's frames, but with the goal of understanding the several purposes of including such scenes in literary works.

- Objective features of “meals or preparation scenes”: length in words, number of participants, number of food and drink entities or words, location, existence of direct speech
- Subjective features: purpose in the plot, (emotional) ambiance

How would a literary scholar go about finding / searching for eating/drinking/cooking episodes in a large corpus? How would s/he like to see the results? We intend to do one user requirements workshop with as many literary scholars as possible.

Pragmatic issues associated with eating and drinking

Another angle of the ReadingFood project is the investigation of cultural traditions (and set phrases) associated with eating and drinking, and how often are they present, or implicit, in literary texts. For example, learn

- how often phrases like “takk for maten” are present
- how often it is described what one does at the beginning of a meal
- how often one does “skjænker” or serves him/herself
- which subparts of a meal are more or less conventional and thus mentioned as such (dessert, coffee, apertif etc.)

There may be wide variety in time, region and social class. And this may be of interest as well.

- What is the weight of different F&D words/concepts in common language?
- What are the most common metaphors associated with food and drink?
- Are they common to the three languages, or very culture-specific?
- What other domains of experience are related to F&D? For example, we have seen at least colour, consistency and (human) character.