

# Presenting ReadingFood and its infrastructure

Diana Santos<sup>1,\*</sup>, Michael Preminger<sup>2</sup>, Elizaveta Khachatryan<sup>1</sup>, Åse Kristine Tveit<sup>2</sup> and Eckhard Bick<sup>3</sup>

<sup>1</sup>University of Oslo

<sup>2</sup>OsloMet

<sup>3</sup>University of Southern Denmark / GrammarSoft

## Abstract

In this paper, we provide general information on the *ReadingFood* project, discussing its architecture and infrastructure choice(s). After presenting the CorpusEye environment, we give some “appetizers” of the diverse studies that can be done with this system. First, we present a taxonomy developed to cover food and drink in the three languages, and we explain the need to disambiguate for a fine-grained classification in this domain. We then illustrate different kinds of studies that ReadingFood allows, namely a) lexical semantics of container words in Italian, b) changes in drinking habits in Norway, and c) differences between Portuguese and Brazilian societies in what concerns drinking. We finish by a short discussion on how other infrastructural challenges for other parts of the project can be solved.

## Keywords

computational literature, distant reading, food and drink, Italian, Norwegian, Portuguese

## 1. What is *ReadingFood* all about

Food and drink are an integral part of the life of living beings. Studies related to food describe eating and drinking not only as natural habits, but also through the lens of social rules and cultural traditions (Meigs 1997) and literary analysis (Boyce and Fitzpatrick 2017; Coghlan 2020; Barkan 2021).

Making use of comparable corpora of novels for most European literatures from 1840 to 1920 created by the COST action Distant Reading for European Literary History<sup>1</sup>, the *ReadingFood* project investigates how literary works in Norway, Italy, and Portugal – together with Brazil – deal with food and drink. This period is culturally interesting: it corresponds to nation-building periods for Italian, Norwegian, and Brazilian states, where literature played an important role; the time period encompasses both romanticism and realism/naturalism as well as budding modernism, and it reflects the impact of the Industrial Revolution on social structures and eating culture. Fictional texts provide access to subjective and imagined scenes of eating and drinking, and at the same time it is an ideal object to trace shared assumptions and social structure, as the sociology of literature teaches (Furuland 1997; Svedjedal 1997).

The choice of these four literatures and cultures has several reasons. It is reasonable to assume that food cultures vary depending on climate. First, climate dictates which plants and animals can be used agriculturally in a given region. Second, a warm climate facilitates social eating in open spaces, while a cold climate favours more private, indoor meal contexts. Also, for a country with a far-flung

---

*Digital Humanities in Norway: Infrastructure and Interface*

\*Corresponding author.

✉ d.s.m.santos@ilos.uio.no (D. Santos); michaelp@oslomet.no (M. Preminger); elizaveta.khachatryan@ilos.uio.no (E. Khachatryan); aasekt@oslomet.no (K. Tveit); eckhard.bick@gmail.com (E. Bick)

🌐 <https://www.hf.uio.no/ilos/english/people/aca/dssantos/index.html> (D. Santos);

<https://www.oslomet.no/om/ansatt/michaelp/> (M. Preminger);

<https://www.hf.uio.no/ilos/english/people/aca/elizavek/index.html> (E. Khachatryan);

<https://www.oslomet.no/om/ansatt/aasekt/> (K. Tveit); <https://portal.findresearcher.sdu.dk/en/persons/bick/> (E. Bick)

🆔 0000-0002-3108-7706 (D. Santos); 0000-0003-3524-4374 (M. Preminger); 0000-0002-3433-9833 (E. Khachatryan);

0000-0001-6634-6109 (K. Tveit); 0000-0002-5505-4861 (E. Bick)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



Digital Humanities in the Nordic and Baltic Countries Publications – ISSN: 2704-1441

<sup>1</sup><https://www.distant-reading.net>

empire like Portugal, food culture may vary due to the influx of goods from other parts of the globe, and finally, a culture's general wealth may influence the sophistication of its food and drink traditions. Our research hypothesis is that the Norway of the literature period in question is a good example of an unsophisticated food culture with climatically limited resources and an indoor meal setting. Mediterranean countries, on the other hand, stand for a food and drink tradition with a larger variety of agricultural resources and more social meal settings. However, the old Mediterranean center, Italy, was not a global empire in modern times and might not be representative for the whole region. Hence, the addition of Portugal that entered modernity with a sea-borne empire spanning the entire globe. The triple of Norwegian, Italian and Portuguese thus represents a minimal set of cultures to cover both the climate divide and the local/global divide, while Brazil represents a non-European location that shared with Portugal language and culture but not environment or history. It is worth noting that these four cultures are not very often (almost never) discussed in literary and social studies dedicated to food (if compared to English and French gastronomic traditions; see e.g. Boyce and Fitzpatrick (2017) or Stene-Johansen (2019)) and have never been compared.

We are mainly interested in the following research questions. How do social conditions correlate with mentions of food, and how do episodes of eating or preparing food in literature work in narratives? Which metaphors use eating and drinking as a source? To answer these questions and compare their answers across different cultures, we specify a formal description of the domain of food based on the corpora (i.e. a formal description anchored in the three languages), while the texts will be annotated in depth in context, providing a research environment with specific search facilities tailored to the domain of food and drink in literary texts.

This is a digital humanities project that connects research questions in language and literature, making it possible to jointly answer questions from both disciplines. How are food and drink employed in language and literature in these cultures during the period examined? Which language ingredients (lexical items, metaphors, and collocations) are used in the narrative text when dealing with food? The contrastive emphasis is also important: we are keen on ascertaining whether the findings are universal or national, whether they depend on the period or literary school, or are mainly a question of the author's individual style.

### 1.1. Task description

In a nutshell, *ReadingFood* aims are:

1. to contribute to establishing food and drink (F&D) as an important domain in literature research and hence contribute to a rewriting of European literary history ;
2. to develop a public digital humanities environment for further studies,
3. and, last but not least, to significantly advance linguistic and cultural knowledge about this domain in the three languages studied.

More concretely, some of the tasks which will have to be executed are the following. First, we will create and revise corpus annotation on three levels:

- identify which tokens correspond to the F&D domain, through disambiguation in context
- identify (multiword) expressions that correspond to objects in the F&D domain, also dealing with collocations and idioms, and pragmatic set phrases in this domain
- classify the purpose of employment of the F&D objects in discursive/literary text: as objects of eating, as products, in order to characterize literary characters or places, etc.

in tandem with, second, the creation of a lexical taxonomy for the domain of food and drink, which can be used as an index into the works.

Then we will identify F&D episodes, that is, meals or food preparation scenes, in the spirit of Schank and Abelson (1977), see Section 5.

## 1.2. Status

This project was submitted to the Research Council of Norway in 2025, after having received a small infrastructure grant in 2024 from the University of Oslo, which allowed us to do some preliminary work that will be reported here, namely an initial corpus-based drink ontology for the three languages, as well as some explorations on subjects related to drink.

During writing of the present paper we learned that funding from Research Council was not granted. We are therefore currently investigating other ways to proceed work on the subject.

## 1.3. Data used

First we will detail briefly the data material that we already have at our disposal, and on which the studies and resources reported are based. It is, however, possible that it will be enlarged during the project's timeline.

For Norwegian, 97 novels in Bokmål, from the period 1840 to 1920, more specifically from *Mordet på Maskinbygger Roolfsen* (Maurits Hansen, 1840) to *Kransen* (Sigrid Unset, 1920), corresponding to 40 authors and containing 5.893 million tokens (measured by CorpusEye's tools). These novels were reused from COST ELTeC-nor and *Bokselskapet*, and prepared by the second and fourth authors.

For Italian, the 100 novels collected by the Italian COST ELTeC-ita corpus, corresponding to 39 different authors, were reused. The oldest is *Promessi Sposi* by Alessandro Manzoni from 1827, and the newest is *Uno, nessuno e centomila* by Luigi Pirandello from 1926. CorpusEye's word count is 6.876 million tokens.

Finally, for Portuguese, in addition to the 100 novels from COST ELTeC-por (Schöch et al. 2021), we made use of other publicly available novels from Literateca (Santos 2019), namely further 10 from Portugal, and 95 from Brazil. Of these the earliest published were *O piolho viajante* (Policarpo da Silva, 1802) in Portugal and *A Moreninha* (Joaquim Manuel de Macedo, 1844) in Brazil. The total number of tokens is 13.427 millions for Portuguese literature and 8.554 millions for Brazilian literature. For the period 1840-1920, the latest published were *Húmus* (Raul Brandão) and *Amor crioulo* (Abel Botelho), in Portugal in 1919, and *A capital federal* (Coelho Neto) and *Numa e a Ninfa* (Lima Barreto), in Brazil in 1915. In the Portuguese corpus, there are in addition 20 novels (10 from Brazil and 10 from Portugal) that were published after 1920. We should probably at once explain that it is possible to "remove" those extra novels from search and processing, if/when one wants to compare just the period 1840-1920 across countries and languages. But it sounded relevant to also analyze other works which were available, especially because we envisage a newer version of the corpus with many more works soon.<sup>2</sup>

Unfortunately, there are no published quality assessments of ELTeC collections, and the fact that the different texts come from different digitization initiatives and practices results in texts with different quality as far as OCR is concerned. Still, one can at least state that all ELTeC texts were human revised after OCR. One possibly even more important issue is the lack of orthographical standardization across different times, and sometimes also across different publishers. So this has to be taken in consideration by the parsers and will be reflected in the taxonomies. At the current stage of the project, we have considered as distinct words like *aguardente* and *água-ardente*, or *champagne* and *champanhe*.

For more information on the ELTeC collections, see the ELTeC site<sup>3</sup>. Additionally, we provide an exhaustive list of the works in the four literatures in the appendix.

The *ReadingFood* corpora are hosted at <https://corp.visl.dk/>, and the figures and data were obtained on January 2026.

<sup>2</sup>But we agree with an anonymous reviewer that this motivation is very weak, and that it would have been better to stick to the 1840-1920 period.

<sup>3</sup>Specifically <https://distantreading.github.io/ELTeC/>.

## 2. CorpusEye as the main environment

One of the aims of *ReadingFood* is to provide an infrastructure to study food and drink in the aforementioned four literatures. For this, CorpusEye was the best environment we found, given its modern search capabilities, its connection with the parsers we used, and its multilingual capabilities.

This can be somehow substantiated by the discussion on distant reading in Bick (2026, chapters 10.3 and 10.5). As to doing distant reading with corpora, this is an obvious practice, and we redirect the readers to e.g. Gavin (2022 & 2023) for a justification and critique.

CorpusEye's interface supports the hermeneutic circle of reiterating a shift between close and distant reading, or qualitative and quantitative corpus inspection. That way, the scholarly process can be honed to better capture what is to be deduced from the data. Since it is not likely or even possible that a distant reading search is perfectly formulated at the first try, it is necessary to inspect result examples in a qualitative, sample-based fashion in order to perfect the quantitative/statistical query used for the distant reading approach as such.

We will describe here the current CorpusEye capabilities, but it should be mentioned that our intention is to create some new functionalities for the *ReadingFood* project, especially for the part of the project concerned with meals.

### 2.1. History

The original version of CorpusEye was developed by the VISL-project at the University of Southern Denmark to support linguistically informed corpus searches, concordances and statistics in a graphical fashion, with an original focus on grammar teaching and grammatical studies. Category-driven, linguistic searches, rather than text-only searches, were made possible by using Constraint Grammars (Bick 2003) to annotate the corpora for the various linguistic levels - lemma, morphology, syntactic function, dependency structure etc. The first version (Bick 2005) used the IMS<sup>4</sup> original CQP engine<sup>5</sup> and worked with 6 languages. The current version (Bick 2025) was reprogrammed for CQP's successor engine, Manatee (Rychlý 2007), in order to allow for utf-8 data and larger corpus sizes, and has data for most Germanic and Romance languages.

### 2.2. Functionalities

CorpusEye supports a combined quantitative-qualitative research strategy, where it is possible to iterate searches with categorically constrained conditions and subcorpus-restrictions, and to shift seamlessly between quantitative methods (e.g. frequency statistics and visualization) and qualitative methods (inspection of linguistically constrained concordances).

For instance, candidates for lexical studies can be identified using relative frequencies in specific linguistic contexts, and subsequently concordance-inspected or exported together with the containing sentences. If present, bibliographical meta-information can be used to create sub-corpora on the fly (e.g. by year or author) or to compute bar charts for the temporal or author distribution of certain words, categories or patterns. As the *ReadingFood* corpora have been annotated semantically, all of the above is also possible with semantic search categories or semantically constrained subcorpora. In addition, word vector similarity can be used to grade word similarity in 2D scatter plots with user-defined x- and y-axes (e.g. grading food or drink items against an x-axis defined as female-male and a y-axis defined as poor-rich), as in Figure 1.

Let us give some explanation on how to make sense of these plots: A word's coordinates ("position") in the semantic scatter plot is calculated by comparing its own vector to the vectors of the four axis end points, in terms of "contextual" similarity (word embeddings similarity). A similarity value of 1 means exact synonyms, while 0 means no correlation. For words with mutually exclusive contexts, this value can also become negative. Since the x- and y-axis are user-defined concept vectors, the distance

<sup>4</sup>Institut für maschinelle Sprachverarbeitung, Universität Stuttgart

<sup>5</sup>Corpus Query Protocol, <https://www.ims.uni-stuttgart.de/forschung/projekte/corpus-workbench/> and Christ (1994)

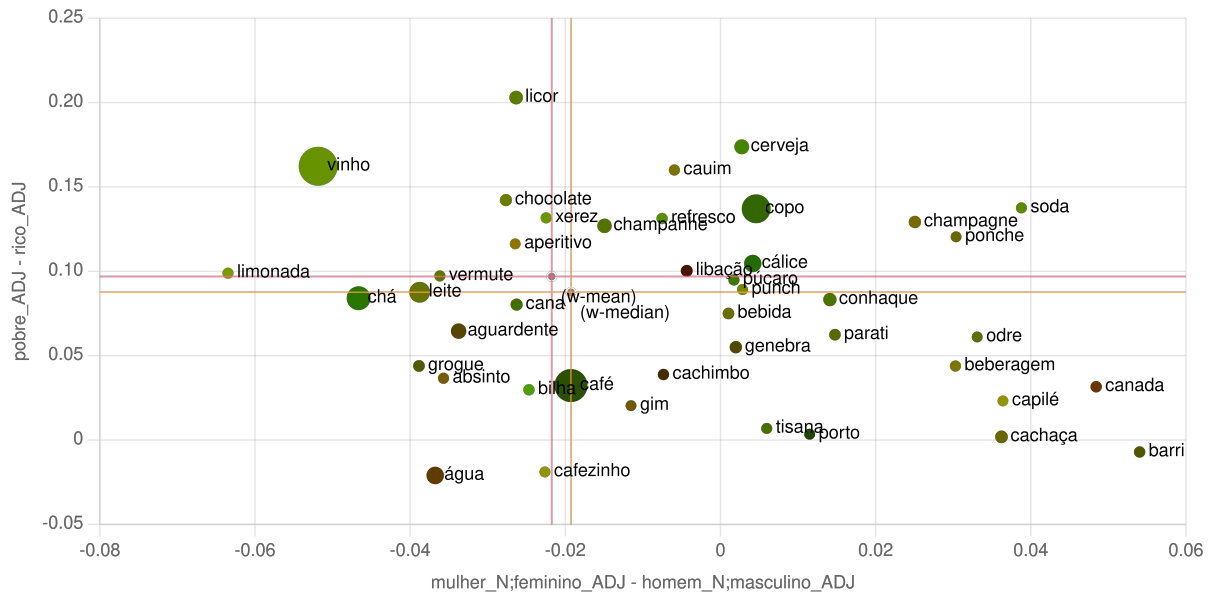


**Figure 1:** Portuguese drinks according to the Y axis pobre (poor) - rico (rich) and the X axis mulher /feminino (woman / feminine) - homem / masculino (man / masculine)

between two topic words in the scatter plot does not simply measure similarity between the words themselves, but rather how different they are with regard to the x- and y-concepts. This allows a whole group of semantically related words (e.g. fruits or beverages) to be graded according to various angles, such as gender, age or luxury. The size (surface area) of an individual word dot is a measure of the word's frequency in the corpus, and thus, indirectly, for the uncertainty of its position in the vector space. Big dots are statistically safer than small ones, as sample size correlates inversely with statistical certainty.

### 3. Preliminary work

As described above, the construction of a taxonomy for food and drink based on literary works assumes that we investigate a F&D lexicon in the context of the literary works contained in the corpus, and then (or at the same time) disambiguate the words or lexemes involved.

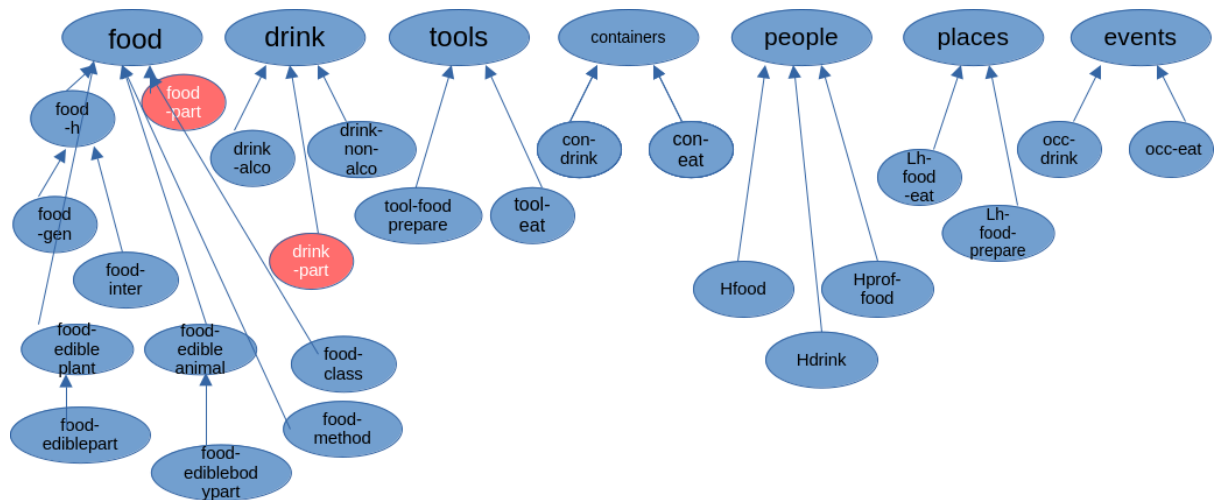


**Figure 2:** Current taxonomy, for the nominal domain (nouns and proper names)

Our initial work, complemented by corpus exploration and taking into consideration the general

semantic classification already present in the VISLcg3 grammars (Bick 2006)<sup>6</sup>, produced the taxonomy (with two instances of meronymy, in pink) in Figure 2, for the nominal domain. Verbs and adjectives, also relevant, are not in the figure. So far adjectives (marked with *j* food have a sentiment tag: positive or negative, while verbs are classified in a FrameNet style<sup>7</sup>, but so far have not been dealt with in our project.

Most of the labels are self-explanatory, except for some of food, which is by far the most complex domain: *food-h* means a human preparation with some name, *food-gen* labels generic ways of describing food (and not the food itself), *food-intermediate* (abbreviated in the figure to "food-inter") classifies products/ingredients that are used in food or drink, but not eaten or drunk separately, like sugar or vinegar, and finally *food-method* describes ways of cooking or preparing food, which are more general than just one dish, but more specific than just food, like stew or soup.

Also, some class names are deceptively parallel: while *Lh-drink* indicate places where you drink, like taverns or bars, in all *Lh-eat* you can eat and drink.

### 3.1. Taxonomical work

We started by the drink subdomain because we expected it to be simpler, and populated three first (sub)taxonomies – one for each language – of drink words. These are still preliminary for two reasons: 1) often the concepts that should be in the taxonomy are multiwords, and we have not yet dealt with this, and 2) because we have not yet decided how to handle systematically variant descriptions of the same drink.

The current drink taxonomies for beverages (alcoholic or not) for the three languages follow. Note that in this version *drink* encompasses also *drink-non-alco*.

#### 3.1.1. Italian

**drink (22)** acqua aranciata bevanda beverage bibita brodo caffè caffè camomilla cedrate cioccolata elisir infuso latte limonata nanfe seltz soda spremuta succo thè tisana

**drink-alco (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 vermute vinaccio vino wermuth

#### 3.1.2. Norwegian bokmål

**drink (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 tevand-sknægt toddy vand yndlingsdrik

**drink-alco (92)** Acachaca Arak Champagne Julepunch absint afskeds-pjolter afskeds-pjolter 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 spiritusosa spiritusoso spiritus sprit stikkelsbærvin stout taar velkomstskænk vermuth vermut vin whisky whiskytoddy yndlingsdrik øl

<sup>6</sup>Most specifically, with food and drink categories, see [https://edu.visl.dk/semantic\\_prototypes\\_overview.pdf](https://edu.visl.dk/semantic_prototypes_overview.pdf).

<sup>7</sup>See [https://framenet.dk/verbal\\_prototypes.pdf](https://framenet.dk/verbal_prototypes.pdf).

### 3.1.3. Portuguese

**drink (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

**drink-alco (112)** Alvaralhão Amontillado Apolinaris Bairrada Beirão Bordeaux Bordéus Bourgogne Bucelas Campolide Caparica Carcavelos Carls=Berg<sup>8</sup> Cartaxo Chablis Chambertin Chartreuse Chateau-Lagrange Chateau=Iquem Chianti Chypre Château-Laffite Colares Corvelo Douro Guiness 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 curaçuau drinque espumante garapa gim gin ginja grappa grogue hidromel kermann kummel laranjinha licor malvasia manzanilla mata-bicho mocroró moscatel moscato parati patrícia pombe ponche punch reino rum sangria santerre tinto tiquira uísque verdasco vermute vinho vinho=palhete vodka vodka whisky xerez zito zurrapa água-pé

### 3.1.4. Summing up

For all three languages there are almost three times as many denominations for alcoholic beverages than for non-alcoholic ones. There are a sizeable number of proper names describing alcoholic beverages, brand names often originating from place names.

In the Portuguese collection there may be more drink names because it encompasses two different countries, but there are few cases only occurring in one of the literatures.

In Norwegian, due to the compound words in Germanic languages, there are several drink words hard to classify as alcoholic or non-alcoholic, such as *yndlingsdrik* (favourite drink), *nationaldrik* (national drink), or drinks corresponding to special occasions like *afskedsdrik* (farewell drink), *velkomstskænk* (welcome drink) or *juleøl* (Christmas beer), that have to be investigated in context. Probably some of them should not appear in the final taxonomy.

Regardless, all the words mentioned in these three cases were annotated with *drink* or *drink-alco*, before disambiguation proceeds, which so far has only been done for Portuguese (drinks).

## 3.2. Disambiguation

The fact that drink disambiguation was only done for Portuguese means that the search results that one can get while searching for drink semantics in Norwegian or Italian have little precision (although perfect recall), and have to be considered with caution.

Let us give some examples to make this more concrete:

If we take the word *Skjænk* in Norwegian, we can see that it has at least four meanings in the corpus:

**furniture** *I den lille Sal mod Nord var der en Skjænk, og i den maatte der altid staa en Flaske Burgunder* (In the little room on the north side, there was a cupboard in which a bottle of Burgundy was always kept standing). (Alexander L. Kielland. *Garman & Worse*, 1880)

**present** *Vi give eder dem næsten som Skjænk og Gave for at I skulle lære vort Sprog* (We offer them to you as a sort of gift, so that you may learn our language). (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* (Mrs Torsen was sitting by her spinning wheel, never got to her feet to find something to offer – no cup of posset, not even a taste of the poorest hooch). (Ragnhild Jølsen. *Rikka-Gan*, 1904)

<sup>8</sup>Some items are multiword, and their lemmas are conveyed with = to mark a space. This is a feature of the parser and no mistake.

**verb** *Han lo. «Skjænk i fuld kop du.»* (He laughed. "Pour me a full cup, will you"). (Sigrid Undset. *Vaaren*, 1914)

If we look at all occurrences of this lemma, we find just 8 out of 19 cases associated with drink, as can be checked in Table 1.

|                 |   |
|-----------------|---|
| furniture       | 8 |
| alcoholic drink | 5 |
| present         | 3 |
| verb of serving | 3 |

**Table 1**

Classification of the lemma *Skjænk* in the Norwegian corpus

Incidentally, it is interesting to note that there is no such "serving" verb in Portuguese or Italian.

Two other examples of the need for disambiguation in Portuguese: the word *copo* can mean glass – which in itself can be used as a container for drink and as an amount of drink, 1206 occurrences – but it can also mean hilt (a part of a sword), absolutely unrelated to drink, in 48 occurrences.

And then the word *café* ('coffee'), which is probably more interesting due to the many meanings that appear in Portuguese and Brazilian literature. An example of each follows.

**plant** *Para levar a efeito este pensamento – o da destruição da planta abençoada, servem-se do de cultivar com largueza o café no interior das províncias onde até o presente se cultivou largamente a cana.* (to make effective this thought – destruction of the blessed plant, they use the cultivation by large of coffee in the interior of the provinces where until now cane has been cultivated) (Franklin Távora, *O Matuto*, 1888)

**drink** *Pois agora, colegas, disse o abade sorvendo o último gole de café, o que está a calhar é um passeio à fazenda.* (Well now, mates, said the abbott drinking the last slurp of coffee, what is agreeable is a walk to the farm) (Eça de Queirós, *O crime do Padre Amaro*, 1875)

**part of meal** *Rosa aparecia ao café, exalando do seu sorriso, ...* (Rosa would arrive at the coffee hour, with her smile...) (Eça de Queirós, *Os Maias*, 1888)

*Entre o queijo e o café, demonstrou-me Quincas Borba que o seu sistema era a destruição da dor.* (Between cheese and coffee, Q. B. demonstrated to me that his system was the destruction of pain) (Machado de Assis, *Quincas Borba*, 1891)

**coffee shop** *Dos cafés em geral, e de como são característicos da civilização de um país* (About coffee shops in general, and how they characterize the civilization of a country) (Almeida Garrett, *Viagens na minha terra*, 1846)

**product** *Quem, ao vêr esta furia commercial que agita as populações, este ardor com que trocam lãs por café e assucar, algodões por cacáu e colza, cereaes por carvão de pedra, não se compadece dos destinos da humanidade* (Who, seeing this trade fury that agitates the population, this haste with which they exchange wool for coffee and sugar, cotton for cocoa and colza, cereal by coal, does not feel sorry for human fate) (António Pedro Lopes de Mendonça, *Memórias dum doido*, 1849)

For the *ReadingFood* project, neither plant nor product are relevant, while the uses of *café* as non-alcoholic drink, meal or part of meal, and coffee shop are connected with food (part of meal: occ-eat, coffee shop: Lh-eat) and drink (drink-non-alco). See their distribution in Table 2, which shows how important disambiguation is: We could not measure the import of coffee as drink, and the mentioning of eating places, if the instances had not been annotated in context.

It should be highlighted that it is not always possible to unambiguously interpret a word into one sole category. See e.g. the following instance, where it is unclear whether we are in presence of a product or a drink, and so both were marked:

*Eu bem a vi desfazer-se em atenções comigo, dando-me o melhor quarto, a melhor manteiga, e o café, depois do jantar, fora do ajuste; mas os olhos do meu coração andavam desvairados em contemplações de mais poéticas provas de amor, e não podiam baixar ao devido apreço da boa manteiga e do café de Cabo Verde, como amorosos mimos e demonstração de ternura. (I was well aware that she was treating me well, giving me the best room, the best butter, and coffee, after dinner, outside the program; but my heart's eyes were mad from contemplating the most poetic love demonstrations, and could not look down to duly praise the good butter and Cape Verde's coffee as loving and indulging signs of tenderness.) (Camilo Castelo Branco, *Coração, Cabeça e Estômago*, 1862)*

This means that there may be more tags than word occurrences in our material.

## 4. Examples of use of CorpusEye

In this section we will look at different subject matters, in the three languages, illustrating the visualization capabilities of CorpusEye.

We start by investigating the lexical semantics of drink containers in Italian over the period represented by the Italian corpus in subsection 4.1, go on addressing Norwegian drinking habits in subsection 14 and end by comparing Brazilian and Portuguese literatures as far as gender and drinking profiles are concerned, in subsection 4.3.

### 4.1. Italian lexical semantics

The material collected within the corpus can be analyzed in a lexical-semantic perspective. This section will illustrate that approach. For this purpose, the word *tazza* ('cup') will be described.

The main scope of lexicology is to describe words as a system, whereas semantics investigates the meaning in general. Lexical semantics deals with words seen as part of the lexical system and provides lexicology the information needed to create this system. Research conducted at the cross-roads of these two fields focuses on answering two crucial questions: What does a word mean? What is its place in a lexical system? This section shows how a literary corpus, limited in time, can contribute to this investigation and what further insights it can provide.

#### 4.1.1. What does a word mean?

In a dictionary of modern Italian, the word *tazza* is described based on its several distinguishing properties that help to set it apart from other words of the same semantic field, i.e. other liquid containers (for example, *bicchieri* 'glass' and *bottiglia* 'bottle'). This approach is known as componential analysis (for a more detailed analysis, see Khachatryan 2025). In dictionary definitions, among important properties, material, shape, and function are mentioned.

*piccolo recipiente di ceramica, porcellana o altro materiale, a bocca rotonda, di solito provvisto di manico ad ansa, usato per bere o servire alimenti liquidi (De Mauro 2025)* (a small ceramic, porcelain, or other container with a round mouth, usually with a looped handle, used for drinking or serving liquid foods)

|                               |     |      |
|-------------------------------|-----|------|
| drink                         | 849 | 46%  |
| coffee shop                   | 423 | 23%  |
| meal or part of meal          | 162 | 8.8% |
| others (plant, product, etc.) | 406 | 22%  |

**Table 2**

Classification of the 1838 occurrences of lemma *café* in the Portuguese and Brazilian corpus



*legno*) and is used to hold money (five occurrences) or with liquids (such as *acqua*, *latte*, or *minestra*). In the corpus, the word is used to refer to this type of object, independently on the period or on the author's regional belonging (as we can see from Figure 3: Manzoni is from Lombardy (Milan) whereas Pirandello is from Sicily (Agrigento)). In case of *tazza* more solid materials are mentioned (e.g., *bronzo*, *vetro*, or *cristallo*). It is worth noting that *tazza* can also refer to a tub: an object of similar shape but a different function. Twice *tazza* is used as quantifier with immaterial content in an expression: *una tazza di felicità* ('a cup of happiness').

An analysis of various types of liquids which *tazza* contains can lead to a further investigation of food traditions in that period. For example, the combination *una tazza di brodo* occurs in different authors. In the example below, *brodo* (broth) is offered/served to guests on a silver tray:

*La cameriera mi portava una tazza di brodo su un vassoio d'argento e mi parlava con grande rispetto.* (The waitress brought me a cup of broth on a silver tray and spoke to me with great respect.) (Grazia Deledda, *Canne al vento*, 1913)

Alcoholic drinks that are normally served in glasses can also be served in a *tazza*, although the word *bicchiere* appears in the corpus (303 occurrences). For example, a combination *una tazza di birra* (of beer) seems to be possible as well (five occurrences, in different authors), whereas a combination with the word *vino* (wine) appears once in a literal meaning and once in a comparison:

*Cercai di spegnere la sua infelicità con un certo numero di tazze di birra, ch'egli ingoiò molto volentieri* (I tried to quench his unhappiness with a certain number of cups of beer, which he swallowed very willingly) (Emilio de Marchi, *Redivivo*, 1894)

*...il sorriso della diva si sente inebriare dall'orgoglio come da una tazza di vin troppo forte* (...the diva's smile feels intoxicated by pride, like a cup of too-strong wine) (Gabriele D'Annunzio, *Il piacere*, 1889)

The place of a word in a lexical system can be investigated in two directions: the paradigmatic (vertical) relationship and the syntagmatic (horizontal) relationship. Above, we compared the word *tazza* with the word *ciotola*, considered as synonyms by the *Vocabolario della Crusca*. Among other elements within the paradigmatic relationship, other liquid containers can be mentioned, e.g. *bicchiere*. Both appear in the same context, but the content is different (like in (b)):

(a) ... *un lieve rumore di tazza o di bicchiere posato sul marmo* (...) (...a light sound of a cup or glass placed on the marble) (Luigi Capuana, *Giacinta*, 1879)

(b) *Sì, e verrò qualche sera a prendrmi mi in casa tua una tazza di caffè o a bere un bicchier di vino alla la vostra salute.* (Yes, and I'll come over some evening to your house to have a cup of coffee or a glass of wine to your health.) (Luigi Pirandello, *Il fu Mattia Pascal*, 1904)

In some texts, a *tazza* can be placed on or close to a *sottocoppa* (lit. 'undercup'), an object that is rarely mentioned in modern language. Its definition appears in De Mauro's (2025) dictionary, but not in Sabatini & Coletti's (Sabatini, Coletti, and Manfredini 2022).

Among the meronyms of *tazza*, the word *orlo* ('edge, border') appears, but a *manico* ('handle') – though part of dictionary definitions – does not. *Orlo* is important for describing the drinking process.

*Ad ogni momento, Maria accostava le labbra all'orlo della tazza per provare* (At every moment, Maria brought her lips to the border of the cup to test) (Gabriele D'Annunzio, *Il piacere*, 1889)

*Nicoletta, che aveva recato alla la bocca la tazza, guardò sua madre di sopra l'orlo di quella, assaporando la cioccolata che rimaneva.* (Nicoletta, who had raised the cup to her mouth, looked at her mother over the border of it, savoring the chocolate that remained) (Luciano Zuccoli, *La freccia nel fianco*, 1913)

This kind of description appears in several texts, in some cases the word *orlo* is mentioned, in some other cases not, but other details may be added. For example, other objects are introduced (e.g., a *sottocoppa* ‘undercup’) or various actions are described (e.g., *trangugiare* or *ingollare*, both mean ‘to swallow fast and greedily’).

It seems that a description of the action of drinking is used to slow down narrative time in a conversation or to describe character’s emotional states. This observation needs further investigation.

Regarding the syntagmatic relationship, it is also interesting to observe which verbs are combined with the word *tazza*, e.g., *ricolmare*, *ingollare*, *porgere*, *posare*, *accostare*, *offrire*, *rifiutare*, *pagare*, see Figure 4. Some of these verbs become rare in modern language and may be discussed from the point of view of verbal semantics and linguistic development. For example, in modern language, the verbs *posare* and *accostare* occur in a limited type of contexts and are often replaced by the verb *mettere* (or *mettere accanto*).

#### 4.1.3. Some final remarks

The analysis presented here can be developed in two directions. First, examining a word in a literary context provides new insights for historical lexicology and historical semantics. Both disciplines study the evolution of word meanings and usages, but this evolution is often described based on the analysis of shorter contexts (see e.g. Markus (2008)).

A literary context helps situate a word not only in linguistic, but also in cultural and social contexts. When doing that, it establishes a new parameter: the speaker who utters the words and the hearer who

| Token       | G: freq <sup>2</sup> /norm | C: freq <sup>2</sup> /norm | C: freq/corp · 10 <sup>6</sup> | freq/conc | num | Compare?                            |
|-------------|----------------------------|----------------------------|--------------------------------|-----------|-----|-------------------------------------|
| prendere    | 24.77                      | 13.42                      | 357.54                         | 11.4%     | 21  | <input checked="" type="checkbox"/> |
| essere      | 0.14                       | 0.15                       | 187.28                         | 5.9%      | 11  | <input checked="" type="checkbox"/> |
| porgere     | 180.56                     | 27.80                      | 187.28                         | 5.9%      | 11  | <input checked="" type="checkbox"/> |
| bere        | 183.23                     | 20.58                      | 170.26                         | 5.4%      | 10  | <input checked="" type="checkbox"/> |
| portare     | 3.57                       | 2.76                       | 136.20                         | 4.3%      | 8   | <input checked="" type="checkbox"/> |
| posare      | 253.68                     | 12.15                      | 119.18                         | 3.8%      | 7   | <input checked="" type="checkbox"/> |
| versare     | 30.70                      | 14.85                      | 102.15                         | 3.2%      | 6   | <input checked="" type="checkbox"/> |
| recare      | 5.12                       | 3.26                       | 85.13                          | 2.7%      | 5   | <input checked="" type="checkbox"/> |
| dare        | 0.56                       | 0.26                       | 68.10                          | 2.2%      | 4   | <input checked="" type="checkbox"/> |
| fare        | 0.19                       | 0.08                       | 68.10                          | 2.2%      | 4   | <input checked="" type="checkbox"/> |
| preparare   | 4.73                       | 2.95                       | 68.10                          | 2.2%      | 4   | <input type="checkbox"/>            |
| tenere      | 0.90                       | 0.59                       | 68.10                          | 2.2%      | 4   | <input type="checkbox"/>            |
| offrire     | 1.37                       | 1.70                       | 51.08                          | 1.6%      | 3   | <input type="checkbox"/>            |
| trangugiare | 4914.81                    | 46.80                      | 51.08                          | 1.6%      | 3   | <input type="checkbox"/>            |
| vedere      | 0.47                       | 0.09                       | 51.08                          | 1.6%      | 3   | <input type="checkbox"/>            |
| accettare   | 0.51                       | 0.74                       | 34.05                          | 1.1%      | 2   | <input type="checkbox"/>            |
| accostare   | 14.83                      | 0.99                       | 34.05                          | 1.1%      | 2   | <input type="checkbox"/>            |

Figure 4: Most frequent verbs preceding *tazza* in the Italian corpus

interprets (addresser-addressee, using Jakobson’s terminology (Jakobson 1957/1987)). This parameter is important for creating a complete picture of language: “If we remember that language is a shared system of paradigmatic and syntagmatic meaning-generating signs, then we understand that there can be no complete theory of language and brain that does not include the cultural/collective context. Meaning is, thus, a product of a dynamic and constant dialectic that is enacted uniquely in any speech act” (Andrews 2014, 54). This cultural/collective context can be retrieved from the corpus.

The analysis of words used in literary texts from a concrete historical period provides modern speakers not only with ethnographic and cultural knowledge about the objects to which those words refer, but also with a more detailed picture and better understanding of past speakers. This image of the speaker can be investigated from a narratological perspective, i.e., by considering how the narration is organized, which details are important for the writer and which role they play in narration. For example, one can observe that the word *tazza* is often used in descriptions providing additional information on the characters and their behaviour in conversation. This place of everyday objects in narration deserves further investigation.

## 4.2. Drink in Norwegian literature across time

The drinking habits to be traced in the corpus indicate, as could be expected, differences in class as well as demographically and geographically, but even the changed drinking habits over time are possible to track down.

### 4.2.1. Background

The Norwegian history of drinking habits was complicated in the times when the novels were written. Alcoholic drinks (mainly beer and liquor) were produced and consumed in large quantities in both the countryside and urban areas, while political authorities and growing ideological movements strove to limit alcohol use, either by import restrictions, serving restrictions or prohibited home production of liquor. Throughout the period, alcohol was a hot topic in public discussions. In the decades before 1900, drinking alcohol and notably liquor was to an increasing degree regarded as sinful among the growing lay Christian movement. During WW1, import of grain became difficult, and to secure the food situation a law was passed in 1917, prohibiting the total production and use of beer, liquor, and even wine. Concerning liquor, the law lasted until 1926. Paradoxically, Norwegian trade agreements from the same period, with Portugal, among other countries, instructed Norway to import rather large quantities of alcohol, in trading with fish (Hauge 1996, 19-20). In the following, we will look at if and how this situation is made visible in the corpus.

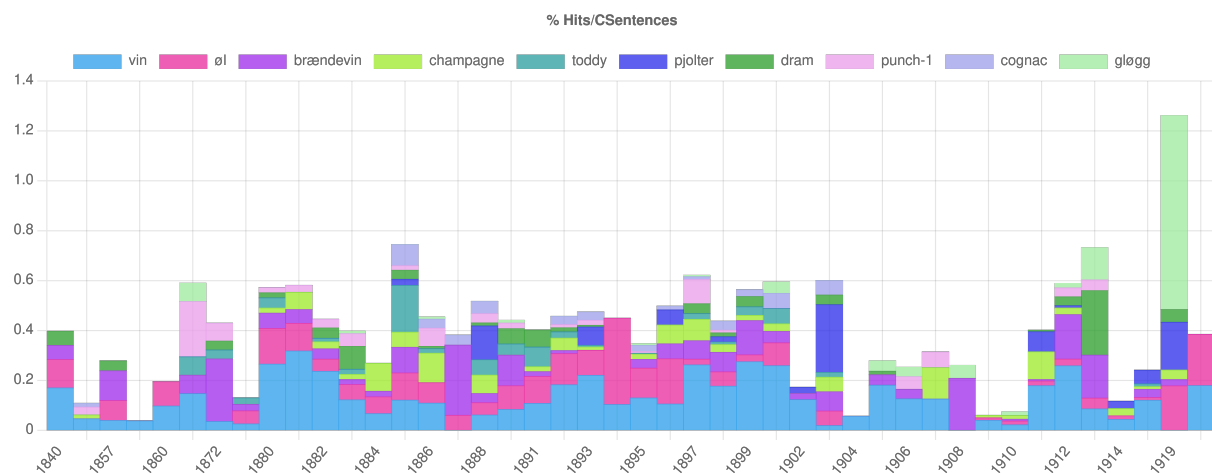


**Figure 5:** Frequency of alcoholic drinks mentioned in social scenes and description of meals in the Norwegian novels

#### 4.2.2. Changing drinking habits

*Laddevin*, bayer beer, and porter beer are drinks introduced in Norway by mid-1800. They were increasingly popular drinks in the latter half of the century, as was the mix of beer and liquor: *pjølter* as can be observed in Figure 5. Several of the novels in the Norwegian corpus are written within the prohibition era, hence we could presume that alcohol would be less prominent in social scenes, such as descriptions of meals or café visits in these novels. However, so far the material does not suggest so.

#### 4.2.3. The coffee case

Ethnologist Astrid Riddervold writes that coffee, as an initially urban drink, spread to the countryside and, to some extent, replaced alcohol as a social drink (Riddervold 1997). The historian Jorunn Sem Fure (Fure 2024) underlines the exchange of drinking habits from milk or alcohol to coffee as an important part of daily life from mid-1800. Coffee beans were expensive and were used mainly for special occasions, while substitutes were homemade and commonly used. It is surprising to see that Bjørnson, in one of his early novels, *Arne* (1859), writes this scene from a rural setting, at a farm of a relatively poor family:

*Moderen truede til gutten : du ser da vel , at fader ikke er i godlaget sitt , mente hun . « Vil du ikke have litt sterk kaffe med sirup i ,sagde hun ; hun vilde gøre ham god igen . Dette var en drikk , som bestemoderen havde likt , og de andre med henne . Nils likte den slett ikke , men havde dog drukket den , fordi de andre gjorde det . « Vil du ikke have litt sterk kaffe med sirup i ? » gentok Margit ; (“You can see very well father’s out of spirits to-day,” the mother said to Arne, by way of warning. “Shouldn’t you like some strong coffee with treacle?” she then said, turning to Nils, trying to drive away his ill-temper. Coffee with treacle had been a favorite drink with the grandmother and Margit, and Arne liked it too. But Nils never liked it, though he used to take it with the others. “Shouldn’t you like some strong coffee with treacle?” Margit asked again, for he did not answer the first time.)<sup>11</sup> (Bjørnstjerne Bjørnson, *Arne*, 1859)*

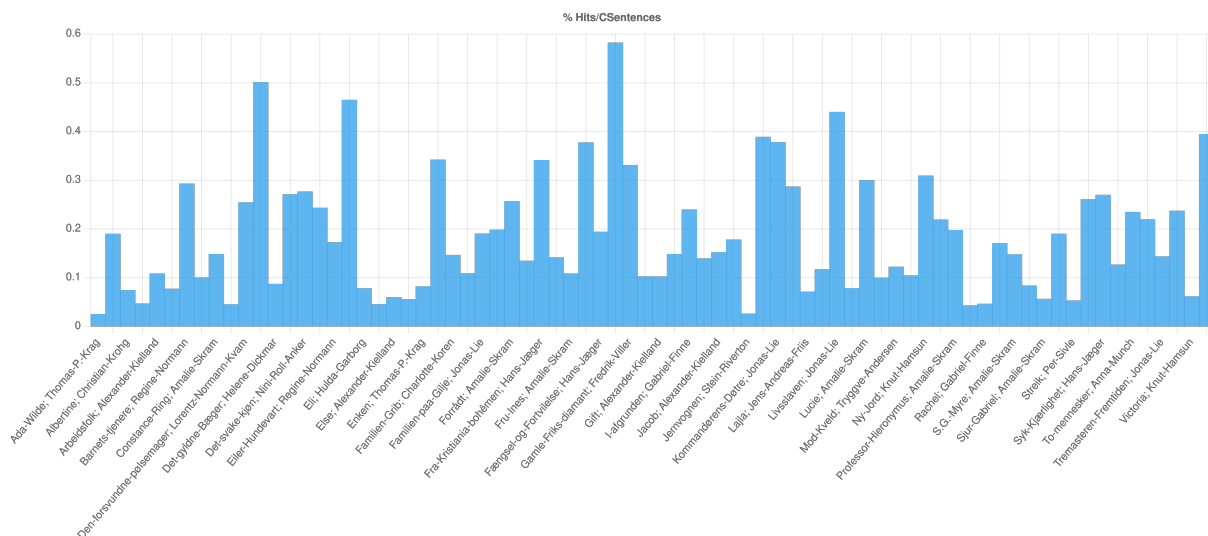


Figure 6: Coffee in the Norwegian novels

An overview of the occurrence of coffee in the Norwegian corpus (see Figure 4.2.3) shows that this drink is often a part of daily routine and social life from the 1890s onward. Contrary to what Riddervold (1997) writes, coffee seems to be more widespread at an earlier point in Norwegian history, at least

<sup>11</sup>Translation from Bjørnson (1869, page 28).

in the novels in our corpus. The scene in Bjørnson's novel is an early example of coffee drinking as a treat among poor people. Coffee is mentioned even in Camilla Collett's *Amtmandens Døttre* (1855), set in an upper class family, where it is underlined by Mrs. Ramm in a comment to Mr. Brandt that if he wants to have breakfast at their house, there will only be served coffee, meaning there will be no alcohol. These two examples, and others in CorpusEye, indicate that coffee was in daily use well before 1900. In Hans Jæger's novel of an urban artist's milieu; *Fra Kristianiabohemen* (1885), coffee is repeatedly mentioned as a drink consumed in relaxing situations, often combined with a cigar, and even accompanied by liquor. The overview in Figure 6 makes visible the use of coffee in the different social settings in the Norwegian novels. According to Figure 5, however, it cannot be concluded that alcohol was replaced by coffee in novels published during the prohibition period.

Further work with these findings could uncover patterns of drinking connected to gender, or to ideological movements in this period of time.

### 4.3. Comparing Brazilian and Portuguese literature

In Figure 7 we see the most frequent non-alcoholic drinks mentioned in Portuguese, namely coffee, tea, milk, water, chocolate, drink, refreshment, lemonade, cocoa and juice, and their distribution by female and male writers, in Brazilian and Portuguese literature. Each of the four columns represents the frequency in respectively female Brazilian writers, female Portuguese writers, male Brazilian writers and male Portuguese writers.

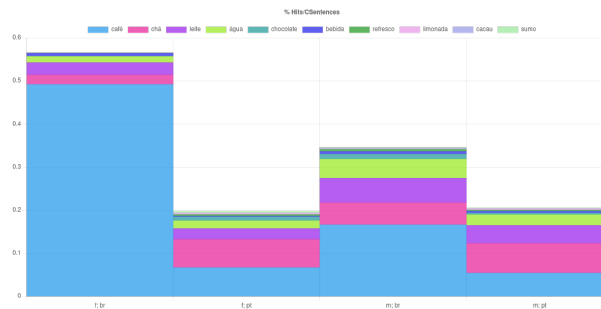


**Figure 7:** Non alcoholic drinks in Portuguese by country and gender. Note: n=4 works for Brazilian female authors

We observe that coffee is much more important as a drink in Brazil (or at least in Brazilian literature) while in Portuguese literature tea is more widespread. It is somewhat more surprising to see that women writers (in both literatures) have a higher proportion of coffee than their male counterparts. Note, however, that there is a very low proportion of women writers in the material, so that one should be wary to generalize: four works in Brazilian literature, and 19 works in Portuguese literature, respectively by 3 and 12 distinct authors.

In addition, there is the problem that most of the data has not been disambiguated, so it is possible that other instances of coffee have also been taken into consideration in Brazilian literature.

As to when and where the different drinks are used, we used the CorpusEye functionality based on word embeddings, asking for female vs. male in one axis, and Brazilian vs. Portuguese context, as displayed in Figure 8. Although the axes are not fool-proof – after all, in most works the country or nationality is not explicitly mentioned, nor need feminine contexts contain words like women or girl – the results seem to indicate some sensible, or predictable distinctions:

- contexts with women are associated mainly with non-alcoholic beverages, especially in Brasil
- wine is far more common in Portuguese contexts (although very much “feminine,” too)
- coffee is Brazilian and male (which may indicate that even in the works written by women it is mainly men who drink coffee)
- also beer is absolutely Brazilian (at least in this period), but does not distinguish between sexes



the one most often regarded as a ritual with attached social expectations:

The dinner or dinner party is crucial not only in defining the characters, their world, and their relationship to that world, but also in structuring the novel or the story and, for the reader, in understanding the author's relationship to her own world and her own historical times

## 5.2. Initial explorations

Being a social event confined in place and time, a meal or another food-related scene should in principle be detectable in a novel. In the natural language processing (NLP) literature, we find two related alleys, namely script analysis (Wanzare 2019; Wanzare, Roth, and Pinkal 2019) and scene annotation (Alrashid and Gaizauskas 2021). Scripts are stereotypical descriptions of a human action such as a restaurant visit (Schank and Abelson 1977), while scenes are sequences where characters, time and location are roughly constant.

Human or automatic detection of such scenes entails identifying their start and end points. Here the challenge is not only “closing in” on a text region in the novel where the density of food words is particularly high, because empirically this is not a straightforward given. Meals can stretch over text sequences where food words are fairly scarce (as in a conversation taking place and referred to at length during a meal). In addition, in most novels we have analysed there is no clear end of meal; some meals “start” (from the point of view of the narrator) in the middle; and meals rarely if ever correspond to more standard literary divisions such as chapters.

Santos, Bick, and Mota (2026) have done an initial exploration of meals in ten novels in Portuguese that illustrates and substantiates the above claims of delimitation difficulty based on concentration of food words.

From a computational point of view, meals correspond to smaller or larger chunks of text including several sentences, requiring an interface to be visualized, analysed and searched which is different from those based on words or sentences. In addition, the features that should classify each meal are not consensual: in Santos et al. (2026) we discuss what parameters would make sense to store and recover. It is as yet unclear whether or not CorpusEye should be extended or a completely different module/program developed for meal research. In any case, one of the subprojects planned by *ReadingFood* is a user requirements workshop with literary scholars, in order to identify relevant questions and corresponding visualization solutions.

## 6. Concluding remarks

Although the future of this research is currently unknown due to lack of funding, the existence of these corpora, marked with food and drink features (although not necessarily disambiguated), allowed us to look at this semantic domain – specifically drink – in a fresh and refreshing way, as we hope the previous sections demonstrated. Furthermore, the use of CorpusEye as a powerful infrastructure for this kind of research was vindicated. We would like to underline that all material is openly available for anyone interested, and can be used for other types of research.

## References

- Alrashid, Tarfah, and Robert Gaizauskas. 2021. “A Pilot Study on Annotating Scenes in Narrative Text using SceneML.” In *Proceedings of the Text2Story’21 Workshop, Online, 1 April 2021*, edited by Ricardo Campos, Alípio Jorge, Adam Jatowt, Sumit Bhatia, and Mark Finlayson, 7–14.
- Andrews, Edna. 2014. *Neuroscience and Multilingualism*. Cambridge University Press.
- Barkan, Leonard. 2021. *The Hungry Eye: Eating, Drinking and European Culture from Rome to the Renaissance*. Princeton University Press.

- Bick, Eckhard. 2003. "VISL CG-3: Constraint Grammar on the Move: An application-driven paradigm." In *Rule-Based Language Technology*, edited by Arvi Hurskainen, Kimmo Koskenniemi, and Tommi Pirinen, 112–140. University of Tartu.
- Bick, Eckhard. 2005. "CorpusEye: Et brugervenligt web-interface for grammatisk opmærkede korpore" In *10. Møde om Udforskningen af Dansk Sprog 7.-8.okt.2004, Proceedings*, edited by Peter Widell and Mette Kunøe, 46–57. Århus University.
- Bick, Eckhard. 2006. "Noun Sense Tagging: Semantic Prototype Annotation of a Portuguese Treebank." In *Proceedings of the Fifth Workshop on Treebanks and Linguistic Theories (December 1-2, 2006, Prague, Czech Republic)*, edited by Jan Hajic and Joakim Nivre, 127–138.
- Bick, Eckhard. 2025. "CorpusEye: Et brugervenligt korpusværktøj til forskning og undervisning." In *20. Møde om Udforskningen af Dansk Sprog (10-11. Oct. 2024), Proceedings*, edited by Winni Collin, Inger Hanse Schoonderbeek, Tina Thode Hougaard, and Karen Schriver, 57–70. Aarhus University.
- Bick, Eckhard. 2026. *CorpusEye: Manual and Use Case Scenarios*. Syddansk Universitetsforlag.
- Bjørnson, Bjørnstjerne. 1869. *Arne: A sketch of Norwegian country life*. Translated by Augusta Plesner and S. Rugeley-Powers. Gutenberg project. <https://www.gutenberg.org/ebooks/35246>.
- Boyce, Charlotte, and Joan Fitzpatrick. 2017. *A History of Food in Literature From the Fourteenth Century to the Present*. Routledge.
- Brown, James W. 1984. *Fictional Meals and their Function in the French Novel 1789– 1848*. University of Toronto Press.
- Christ, Oliver. 1994. "A modular and flexible architecture for an integrated corpus query system." In *Proceedings of the 3d Conference on Computational Lexicography and Text Research (COMPLEX'94, Budapest 1994)*, 23–32.
- Coghlan, J Michelle. 2020. *The Literature of Food*. Cambridge University Press.
- Crusca, Academia della. 1st edition: 1612, 4th edition: 1729-1738. *Vocabolario degli Accademici della Crusca*. Academia della Crusca. <http://accademiadellacrusca.it>.
- De Mauro, Tullio. 2025. *Il Nuovo De Mauro*. <https://dizionario.internazionale.it/>.
- Fure, Jorunn Sem. 2024. *Kaffedrikkingens moralske metamorfoser*. Telemark Museum. <https://www.telemarkmuseum.no/fagartikler/kaffedrikkingens-moralske-metamorfoser/>.
- Furuland, Lars. 1997. "Litteratur och samhälle." In *Litteratursociologi: texter om litteratur och samhälle*, edited by Lars Furuland and Johan Svedjedal. Studentlitteratur.
- Gavin, Michael. 2022 & 2023. "Why distant reading works." *New Literary History* 53 & 54 (4): 613–633.
- Hauge, Ragnar. 1996. "Alkohol i norsk historie." *Norsk Epidemiologi* 6 (1): 13–21.
- Jakobson, Roman. 1957/1987. "Linguistics and poetics." In *Language and literature*, edited by Krystyna Pomorska and Stephen Rudy. Belknap Press of Harvard University Press.
- Khachaturyan, Elizaveta. 2025. *Italian Words: a Practical Introduction to the Italian Lexical System*. Routledge.
- Markus, Manfred. 2008. "The Topos of Foul/Filth in Middle English Homilies: a Study in Historical Lexicology." *Neuophilologische Mitteilungen* 109 (2): 223–240.
- McGee, Diane. 2001. *Introduction: A Time to Eat*. In *Writing the Meal: Dinner in the Fiction of Twentieth-Century Women Writers*. University of Toronto Press.
- Meigs, Anna. 1997. "Food as cultural construction." In *Food and culture: a Reader*, edited by Carole Counihan and Penny van Esterik, 95–106. Routledge.

- Riddervold, Astrid. 1997. *Drikkeskikker*. Teknologisk forlag.
- Rychlý, Pavel. 2007. “Manatee/Bonito - A Modular Corpus Manager.” In *Proceedings of RASLAN 2007 - First Workshop on Recent Advances in Slavonic Natural Language Processing, Karlova Studánka, December 14-16, 2007*, edited by Petr Sojka and Aleš Horák, 65–70.
- Sabatini, Francesco, Vittorio Coletti, and Manuela Manfredini. 2022. *Dizionario italiano Sabatini Coletti*. <https://www.elexico.com/users/disc/>.
- Santos, Diana. 2019. “Literature studies in Literateca: between digital humanities and corpus linguistics.” In *Humanists and the digital toolbox: In honour of Christian-Emil Smith Ore*, edited by Martin Doerr, Øyvind Eide, Oddrun Grønvik, and Bjørghild Kjelsvik, 89–109. Novus forlag.
- Santos, Diana, Eckhard Bick, and Cristina Mota. 2026. “Comida e bebida nas literaturas portuguesa e brasileira: O projeto ReadingFood.” In *Proceedings of the 17th International Conference on Computational Processing of Portuguese (PROPOR 2026)*, edited by Marlo Souza, Iria de-Dios-Flores, Diana Santos, Larissa Freitas, Jackson Wilke da Cruz Souza, and Eugénio Ribeiro, 266–274. ACL. <https://aclanthology.org/2026.propor-2.35.pdf>.
- Santos, Diana, Michael Preminger, Åse Kristine Tveit, and Elizaveta Khachaturyan. 2026. *Identifying food episodes in literary texts*. In preparation.
- Schank, Roger C., and Robert P. Abelson. 1977. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Routledge.
- Schöch, Christof, Tomaz Erjavec, Roxana Patras, and Diana Santos. 2021. “Creating the European Literary Text Collection (ELTeC): Challenges and Perspectives.” *Modern Languages Open* 1:1–19. <https://doi.org/10.3828/mlo.v0i0.364>.
- Stene-Johansen, Knut. 2019. *Smakens politikk*. Scandinavian Academic Press.
- Svedjedal, Johan. 1997. “Det litteratursociologiska perspektivet.” In *Litteratursociologi: texter om litteratur och samhälle*, edited by Lars Furuland and Johan Svedjedal. Studentlitteratur.
- Wanzare, Lilian D. A., Michael Roth, and Manfred Pinkal. 2019. “Detecting Everyday Scenarios in Narrative texts.” In *Proceedings of the Second Workshop of Storytelling (StoryNLP@ACL 2019), Florence, Italy*, 90–106. <https://aclanthology.org/W19-3410.pdf>.
- Wanzare, Lilian Diana Awuor. 2019. “Script Acquisition: A Crowdsourcing and Text mining approach.” PhD diss., Univ. Saarbrücken.

## A. Works in the four literatures

The works are presented chronologically, using the titles that correspond to the orthography of the text. Size is in tokens, computed by CorpusEye.

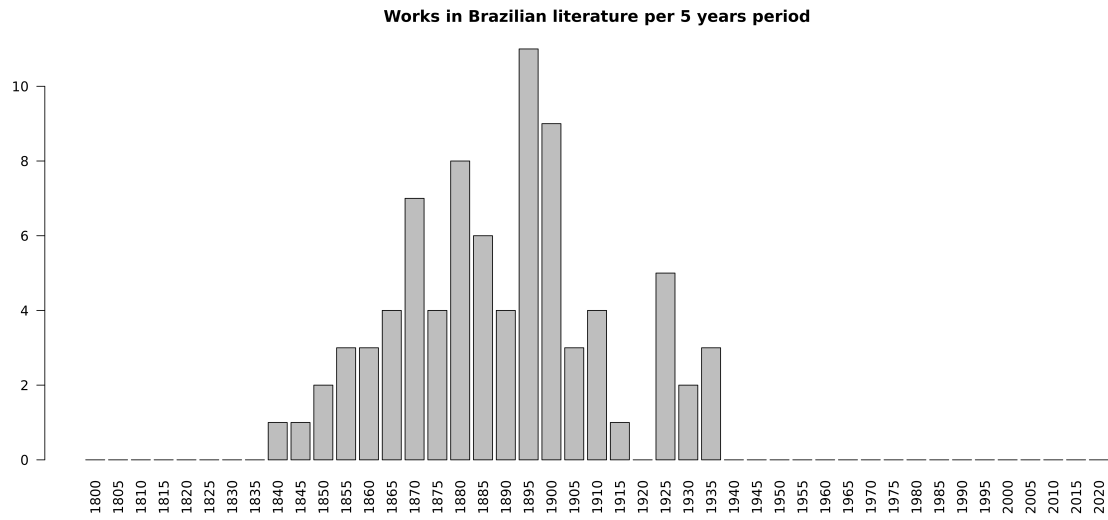


**Figure 9:** Brazilian literature per 5 years period

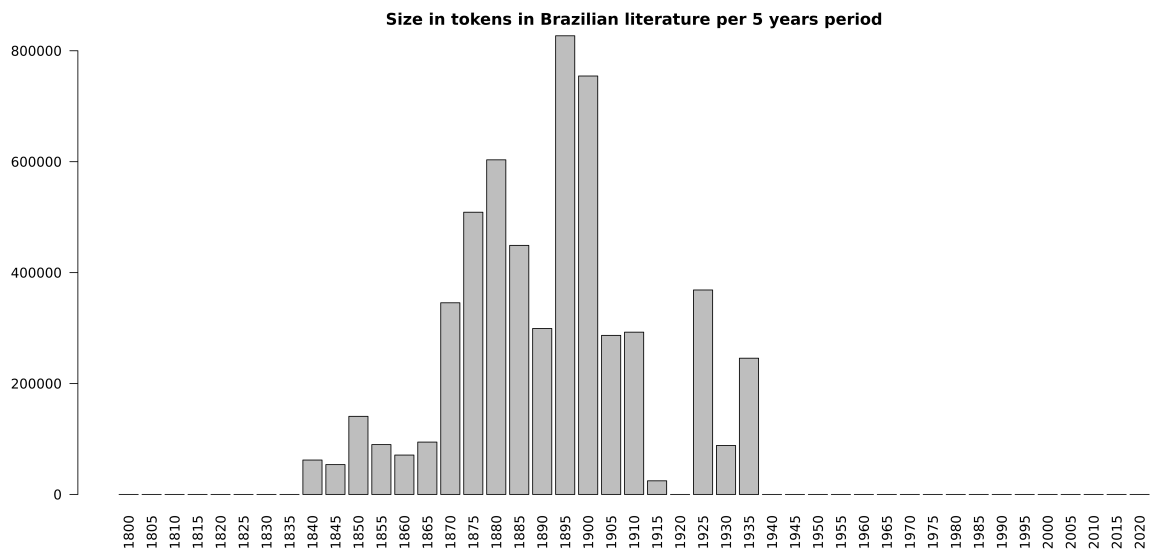


**Figure 10:** Size of Brazilian literature per 5 years period

**Table 3**

## Brazilian literature

| Title                                          | Author                             | Year | Size   |
|------------------------------------------------|------------------------------------|------|--------|
| A Moreninha                                    | Joaquim Manoel de Macedo           | 1844 | 62078  |
| Arca açoriana                                  | Almiro Caldeira da Andrade         | 1849 | 53867  |
| Maria ou a menina roubada                      | Antônio Gonçalves Teixeira e Souza | 1852 | 65632  |
| Memórias de um sargento de milícias            | Manuel Antonio de Almeida          | 1852 | 75206  |
| Carolina                                       | Casimiro de Abreu                  | 1856 | 7292   |
| A viuvinha                                     | José de Alencar                    | 1857 | 22757  |
| Úrsula                                         | Maria Firmina dos Reis             | 1857 | 59858  |
| As Obras do Diabinho da Mão Furada             | Antônio José da Silva              | 1861 | 33461  |
| Paulo                                          | Bruno Seabra                       | 1861 | 28720  |
| A cruz de fogo                                 | José da Rocha Leão                 | 1862 | 8857   |
| Ida                                            | Viriato Bandeira Duarte            | 1865 | 15195  |
| Jovens interessantes                           | Paulina Filadélfia                 | 1865 | 5816   |
| Os Canibais                                    | Álvaro do Carvalho                 | 1868 | 17496  |
| O Ermitão do Muquém                            | Bernardo Guimarães                 | 1869 | 55927  |
| Adelaide de Sargans                            | Caetano Alves de Sousa Filgueiras  | 1870 | 19415  |
| Histórias para gente alegre: A família Agulha  | Luís Guimarães Júnior              | 1870 | 96156  |
| Os franceses no Rio de Janeiro                 | Moreira de Azevedo                 | 1870 | 28092  |
| Inocência                                      | Visconde de Taunay                 | 1872 | 67362  |
| O Garimpeiro                                   | Bernardo Guimarães                 | 1872 | 55191  |
| O Seminarista                                  | Bernardo Guimarães                 | 1872 | 48499  |
| O vaqueano                                     | Apolinário Porto Alegre            | 1872 | 30892  |
| A escrava Isaura                               | Bernardo Guimarães                 | 1875 | 68752  |
| O selo da roda                                 | Pedro Ivo                          | 1876 | 86635  |
| Maurício                                       | Bernardo Guimarães                 | 1877 | 145091 |
| Os retirantes                                  | José do Patrocínio                 | 1879 | 208408 |
| Uma Lágrima de Mulher                          | Aluísio de Azevedo                 | 1880 | 43118  |
| O Mulato                                       | Aluísio de Azevedo                 | 1881 | 130760 |
| A Condessa Vésper                              | Aluísio de Azevedo                 | 1882 | 140760 |
| As Jóias da Coroa                              | Raul Pompéia                       | 1882 | 36643  |
| O alienista                                    | Machado de Assis                   | 1882 | 21764  |
| Uma Tragédia no Amazonas                       | Raul Pompéia                       | 1882 | 31678  |
| Casa de pensão                                 | Aluísio Azevedo                    | 1884 | 127345 |
| Filomena Borges                                | Aluísio de Azevedo                 | 1884 | 71290  |
| Hóspede                                        | Pardal Mallet                      | 1887 | 31541  |
| O Homem                                        | Aluísio de Azevedo                 | 1887 | 68192  |
| A carne                                        | Júlio Ribeiro                      | 1888 | 76722  |
| O Ateneu                                       | Raul Pompéia                       | 1888 | 82435  |
| No declínio                                    | Visconde de Taunay                 | 1889 | 46331  |
| O coruja                                       | Aluísio Azevedo                    | 1889 | 143834 |
| O Cortiço                                      | Aluísio de Azevedo                 | 1890 | 105465 |
| Dona Guidinha do Paço                          | Manuel de Oliveira Paiva           | 1891 | 75910  |
| Quincas Borba                                  | Machado de Assis                   | 1891 | 104192 |
| Rose Castle                                    | Virgílio Várzea                    | 1893 | 13705  |
| A Mortalha de Alzira                           | Aluísio de Azevedo                 | 1895 | 63571  |
| A viúva Simões                                 | Júlia Lopes de Almeida             | 1895 | 46933  |
| Livro de uma Sogra                             | Aluísio de Azevedo                 | 1895 | 67105  |
| O Bom Crioulo                                  | Adolfo Caminho                     | 1895 | 49393  |
| Tentação                                       | Adolfo Caminha                     | 1896 | 54462  |
| Flor de sangue                                 | Valentim Magalhães                 | 1897 | 81459  |
| O paroara: scenas da vida cearense e amazonica | Rodolpho Teophilo                  | 1899 | 104352 |
| Dom Casmurro                                   | Machado de Assis                   | 1899 | 87413  |
| Mocidade morta                                 | Luiz Gonzaga Duque Estrada         | 1899 | 112472 |
| O cacaolista: scenas da vida do Amazonas       | Inglês de Sousa                    | 1899 | 51340  |
| Simá                                           | Lourenço Amazonas                  | 1899 | 108510 |
| A normalista                                   | Adolfo Caminha                     | 1900 | 78412  |
| Girândula de Amores                            | Aluísio de Azevedo                 | 1900 | 123429 |

**Table 4**

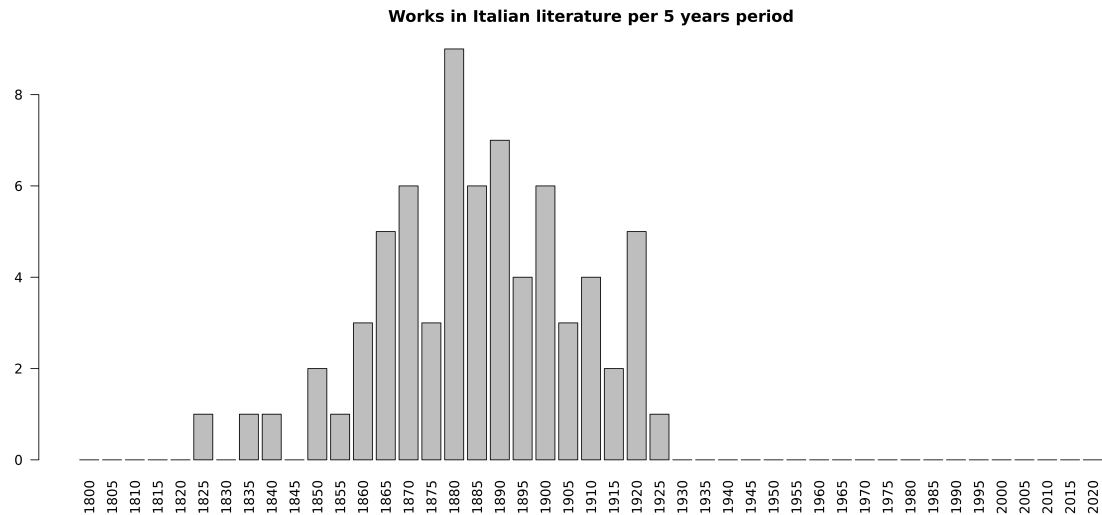
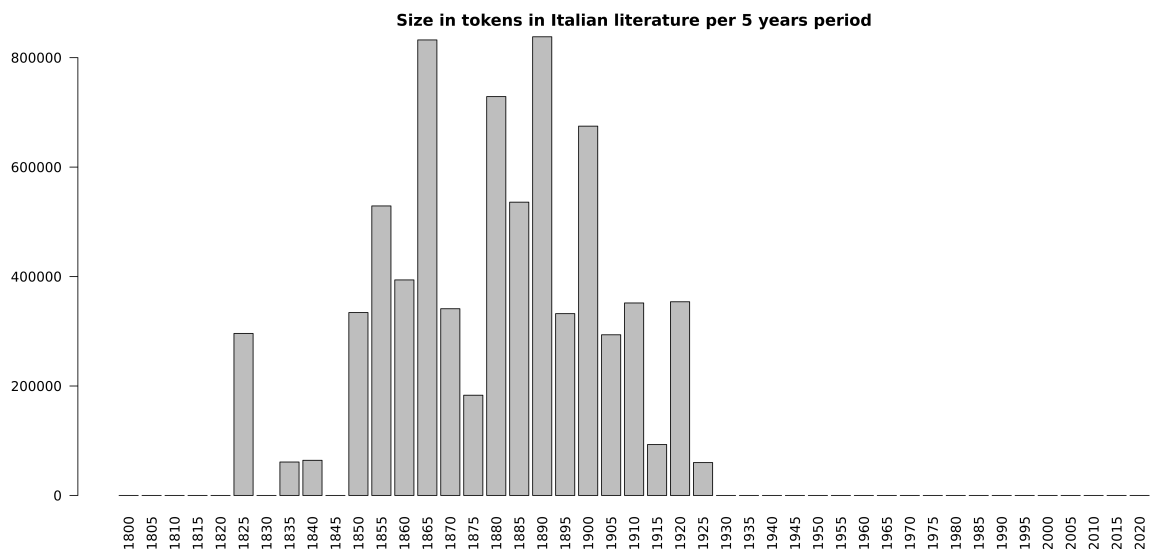
Brazilian literature (cont.)

| Title                                                                   | Author                       | Year | Size   |
|-------------------------------------------------------------------------|------------------------------|------|--------|
| Os Caramurús: romance histórico da descoberta e independência do Brasil | Artur Lobo de Ávila          | 1900 | 86508  |
| George Marcial                                                          | Virgílio Várzea              | 1901 | 144612 |
| Luzia Homem                                                             | Domingos Olímpio             | 1903 | 83344  |
| A Noiva do Paladino                                                     | Virgílio Várzea              | 1904 | 10172  |
| O bandido do Rio das Mortes                                             | Bernardo Guimarães           | 1904 | 78366  |
| O brigue flibusteiro                                                    | Virgílio Várzea              | 1904 | 49484  |
| Turbilhão                                                               | Coelho Neto                  | 1904 | 100164 |
| O trovador da infanta                                                   | Oliveira Mascarenhas         | 1906 | 87505  |
| A Casa dos Fantasma                                                     | Luís Augusto Rebelo da Silva | 1908 | 119587 |
| Miss Kate                                                               | Cosme Velho                  | 1909 | 79744  |
| A Profissão de Jacques Pedreira                                         | João do Rio                  | 1910 | 67576  |
| Maria Dusá                                                              | Lindolfo Rocha               | 1910 | 88115  |
| A luta                                                                  | Emília Bandeira de Melo      | 1911 | 48656  |
| Triste Fim de Policarpo Quaresma                                        | Lima Barreto                 | 1911 | 88278  |
| Cinza e bruma                                                           | Othon Gama d'Eça             | 1918 | 24608  |
| A Marquesa de Santos                                                    | Paulo Setúbal                | 1925 | 82637  |
| Amar, verbo intransitivo                                                | Mário de Andrade             | 1927 | 49596  |
| Brás, Bexiga e Barra Funda                                              | Antônio de Alcântara Machado | 1927 | 14510  |
| O Príncipe de Nassau                                                    | Paulo Setúbal                | 1928 | 70598  |
| A viagem maravilhosa                                                    | Graça Aranha                 | 1929 | 151341 |
| O Tupinambá                                                             | Souza Lima                   | 1931 | 49297  |
| Os irmãos Leme                                                          | Paulo Setúbal                | 1933 | 38986  |
| O sonho das esmeraldas                                                  | Paulo Setúbal                | 1935 | 32810  |
| Mana Maria                                                              | Antônio de Alcântara Machado | 1936 | 158147 |
| Memorial de um passageiro de bonde                                      | Amadeu Amaral                | 1938 | 54716  |

**Table 5**

## Italian literature

| Title                                                 | Author                      | Year | Size   |
|-------------------------------------------------------|-----------------------------|------|--------|
| Promessi Sposi                                        | Alessandro Manzoni          | 1827 | 296019 |
| Il Duca d'Atene                                       | Niccolò Tommaseo            | 1837 | 61113  |
| Fede e bellezza                                       | Niccolò Tommaseo            | 1840 | 64265  |
| Storia di una povera famiglia                         | Damiano - Giulio Carcano    | 1850 | 175709 |
| Lorenzo Benoni ovvero Scene della vita di un italiano | Giovanni Ruffini            | 1853 | 158498 |
| Cent anni                                             | Giuseppe Rovani             | 1859 | 529049 |
| Pasquale Paoli ossia, la rotta di Ponte Nuovo         | Francesco Domenico Guerazzi | 1860 | 258065 |
| I moribondi del Palazzo Carignano                     | Ferdinando Petrucci         | 1862 | 49256  |
| La Scapigliatura e il 6 febbraio                      | Carlo Righetti              | 1862 | 86540  |
| Una peccatrice                                        | Giovanni Verga              | 1865 | 50279  |
| Le confessioni d'un italiano                          | Ippolito Nievo              | 1867 | 423470 |
| Memorie di Giuda                                      | Ferdinando Petrucci         | 1867 | 188940 |
| Fosca                                                 | Ugo Tarchetti               | 1869 | 75880  |
| Povera Giovanna!                                      | Vittorio Bersezio           | 1869 | 93823  |
| Cantoni il Volontario                                 | Giuseppe                    | 1870 | 62463  |
| Clelia ovvero il governo dei preti                    | Giuseppe Garibaldi          | 1870 | 92252  |
| Gli Incendiari Della Comune                           | Ulisse Barbieri             | 1871 | 45952  |
| Lucifero                                              | Ulisse Barbieri             | 1871 | 23568  |
| Il sorbetto della regina                              | Ferdinando Petrucci         | 1872 | 94379  |
| La colonia felice, Utopia                             | Carlo Dossi                 | 1874 | 22593  |
| La desinenza in A                                     | Carlo                       | 1878 | 70646  |
| Tigre reale                                           | Giovanni Verga              | 1878 | 41204  |
| Giacinta                                              | Luigi Capuana               | 1879 | 71334  |
| Profumo                                               | Luigi Capuana               | 1880 | 82538  |
| Il Cappello Del Prete                                 | Emilio de Marchi            | 1881 | 73252  |
| Malombra                                              | Antonio Fogazzaro           | 1881 | 186432 |
| Memorie del Presbiterio                               | Emilio Praga                | 1881 | 91358  |
| Il marchese di Roccaverdina                           | Luigi Capuana               | 1882 | 103559 |
| Il marito di Elena                                    | Giovanni Verga              | 1882 | 61698  |
| Dio ne scampi dagli Orsenigo                          | Vittorio Imbriani           | 1883 | 60460  |
| Le avventure di Pinocchio                             | Carlo Collodi               | 1883 | 54165  |
| Senso                                                 | Camilo Boito                | 1883 | 15426  |
| Daniele Cortis                                        | Antonio Fogazzaro           | 1885 | 130260 |
| In basso                                              | Ulisse Barbieri             | 1885 | 67356  |
| Cuore                                                 | Edmondo De Amicis           | 1886 | 112290 |
| Lydia                                                 | Neera - Anna Radius Zuccari | 1888 | 59608  |
| Il piacere                                            | Gabriele d'Annunzio         | 1889 | 138876 |
| La trovatella di Milano                               | Carolina Invernizio         | 1889 | 27547  |
| Demetrio Pianelli                                     | Emilio de Marchi            | 1890 | 129677 |
| Processi verbali                                      | Frederico de Roberto        | 1890 | 48738  |
| Il maestro di setticlavio                             | Camillo Boito               | 1891 | 43147  |
| L'innocente                                           | Gabriele d'Annunzio         | 1892 | 108152 |
| Una vita                                              | Italo Svevo                 | 1892 | 143393 |
| I viceré                                              | Frederico de Roberto        | 1894 | 278157 |
| Redivivo                                              | Emilio de Marchi            | 1894 | 86884  |
| Le vergini delle rocce                                | Gabriele d'Annunzio         | 1895 | 73593  |
| La testa della vipera                                 | Vittorio Bersezio           | 1896 | 52654  |
| Giacomo l'idealista                                   | Emilio de Marchi            | 1897 | 120674 |
| Senilità                                              | Italo Svevo                 | 1898 | 85320  |
| Il bacio d'una morta                                  | Carolina Invernizio         | 1900 | 119337 |
| Il fuoco                                              | Gabriele d'Annunzio         | 1900 | 129598 |
| I misteri delle soffitte                              | Carolina Invernizio         | 1901 | 125981 |
| L'esclusa                                             | Luigi Pirandello            | 1901 | 78775  |
| Piccolo mondo moderno                                 | Antonio Fogazzaro           | 1901 | 120007 |
| Il fu Mattia Pascal                                   | Luigi Pirandello            | 1904 | 101163 |
| Una donna                                             | Sibilla Aleramo             | 1906 | 80255  |
| Un Vampiro                                            | Luigi Capuana               | 1907 | 13106  |

**Figure 11:** Italian literature per 5 years period**Figure 12:** Size of Italian literature per 5 years period**Table 6**

Italian literature (cont.)

| Title                    | Author                          | Year | Size   |
|--------------------------|---------------------------------|------|--------|
| I vecchi e i giovani     | Luigi Pirandello                | 1909 | 200251 |
| Il mio carso             | Cipio Slataper                  | 1912 | 36979  |
| La mano tagliata         | Matilde Serao                   | 1912 | 153816 |
| Canne al vento           | Grazia Deledda                  | 1913 | 80426  |
| La freccia nel fianco    | Zuccoli - Luciano von Ingelheim | 1913 | 80514  |
| Con gli occhi chiusi     | Frederigo Tozzi                 | 1919 | 63845  |
| Il passaggio             | Sibilla Aleramo                 | 1919 | 29205  |
| Tre croci                | Frederigo Tozzi                 | 1920 | 38880  |
| Una passione             | Anna Zuccari                    | 1921 | 73041  |
| Il dio dei viventi       | Grazia Deledda                  | 1922 | 54356  |
| Trasfigurazione          | Sibilla Aleramo                 | 1922 | 8304   |
| La coscienza di Zeno     | Italo Svevo                     | 1923 | 179350 |
| Uno, nessuno e centomila | Luigi Pirandello                | 1926 | 60098  |

**Table 7**

## Norwegian literature

| Title                           | Author                | Year | Size   |
|---------------------------------|-----------------------|------|--------|
| Mordet paa Maskinbygger Roofsen | Maurits Hansen        | 1840 | 27771  |
| Den Eensomme                    | Andreas Munch         | 1846 | 33584  |
| Amtmannens døtre                | Camilla Collett       | 1854 | 113961 |
| Synnøve Solbakken               | Bjørnstjerne Bjørnson | 1857 | 40836  |
| Arne                            | Bjørnstjerne Bjørnson | 1859 | 40773  |
| En glat gut                     | Bjørnstjerne Bjørnson | 1860 | 34452  |
| Et Levnedsløb                   | Marie Wexelsen        | 1866 | 32239  |
| Min Bedstemoders Fortælling     | Magdalene Thoresen    | 1867 | 14531  |
| Den Fremsynte                   | Jonas Lie             | 1870 | 37525  |
| Tremasteren Fremtiden           | Jonas Lie             | 1872 | 62617  |
| Lodsen og hans hustru           | Jonas Lie             | 1874 | 77573  |
| Garman & Worse                  | Alexander L. Kjelland | 1880 | 81745  |
| Else                            | Alexander Kielland    | 1881 | 26830  |
| Farlige Folk                    | Kristian Elster       | 1881 | 83660  |
| En Professor                    | Johan Vibe            | 1882 | 30371  |
| Familien Pehrson                | John Paulsen          | 1882 | 75071  |
| Gaa paa!                        | Jonas Lie             | 1882 | 73916  |
| Skipper Worse                   | Alexander Kielland    | 1882 | 63817  |
| Familien paa Gilje              | Jonas Lie             | 1883 | 53586  |
| Gift                            | Alexander Kielland    | 1883 | 60149  |
| Livsslaven                      | Jonas Lie             | 1883 | 47676  |
| Fortuna                         | Alexander Kielland    | 1884 | 63641  |
| Constance Ring                  | Amalie Skram          | 1885 | 126407 |
| Fra Kristiania bohømen          | Hans Jæger            | 1885 | 152250 |
| Arbeidsfolk                     | Alexander Kielland    | 1886 | 70604  |
| Kommandørens Døtre              | Jonas Lie             | 1886 | 65213  |
| Sankt Hans Fest                 | Alexander Kielland    | 1886 | 37367  |
| Sjur Gabriel                    | Amalie Skram          | 1887 | 31950  |
| To venner                       | Amalie Skram          | 1887 | 57999  |
| Lucie                           | Amalie Skram          | 1888 | 59691  |
| Syk Kjærlighet                  | Hans Jæger            | 1888 | 82785  |
| Doktor Wangs børn               | Gabriel Finne         | 1890 | 44865  |
| Lajla                           | Jens Andreas Friis    | 1890 | 60358  |
| S.G. Myre                       | Amalie Skram          | 1890 | 119921 |
| Sult                            | Knut Hamsun           | 1890 | 76605  |
| Ella                            | Jens Andreas Friis    | 1891 | 38697  |
| Fru Ines                        | Amalie Skram          | 1891 | 42666  |
| Jacob                           | Alexander Kielland    | 1891 | 55625  |
| Streik                          | Per Sivle             | 1891 | 31295  |
| Eit frit Forhold                | Hulda Garborg         | 1892 | 38273  |
| Forrådt                         | Amalie Skram          | 1892 | 49422  |
| Mennesker                       | Alvilde Prydz         | 1892 | 77201  |
| Mysterier                       | Knut Hamsun           | 1892 | 143484 |
| Tidlig Høst                     | Otto Valseth          | 1892 | 84131  |
| Familien Grib                   | Charlotte Koren       | 1893 | 94560  |
| Ny Jord                         | Knut Hamsun           | 1893 | 115427 |
| De Fredløse                     | Kristofer Janson      | 1894 | 94609  |
| Kjærlighedens Forvildelser      | Karen Sundt           | 1895 | 203844 |
| Kobberslangen                   | Thomas P. Krag        | 1895 | 46352  |
| Paa Sct. Jørgen                 | Amalie Skram          | 1895 | 56544  |
| Professor Hieronymus            | Amalie Skram          | 1895 | 72826  |
| Rachel                          | Gabriel Finne         | 1895 | 35241  |
| Albertine                       | Christian Krohg       | 1896 | 51257  |
| Korset                          | Sigbjørn Obstfelder   | 1896 | 21447  |
| Ungt Sind                       | Dikken Zwilgmeyer     | 1896 | 35822  |
| Ada Wilde                       | Thomas P. Krag        | 1897 | 70544  |
| I cancelliraadens dage          | Tryggve Andersen      | 1897 | 73719  |
| Karl Monks Oplevelser           | Fredrik Viller        | 1897 | 50887  |

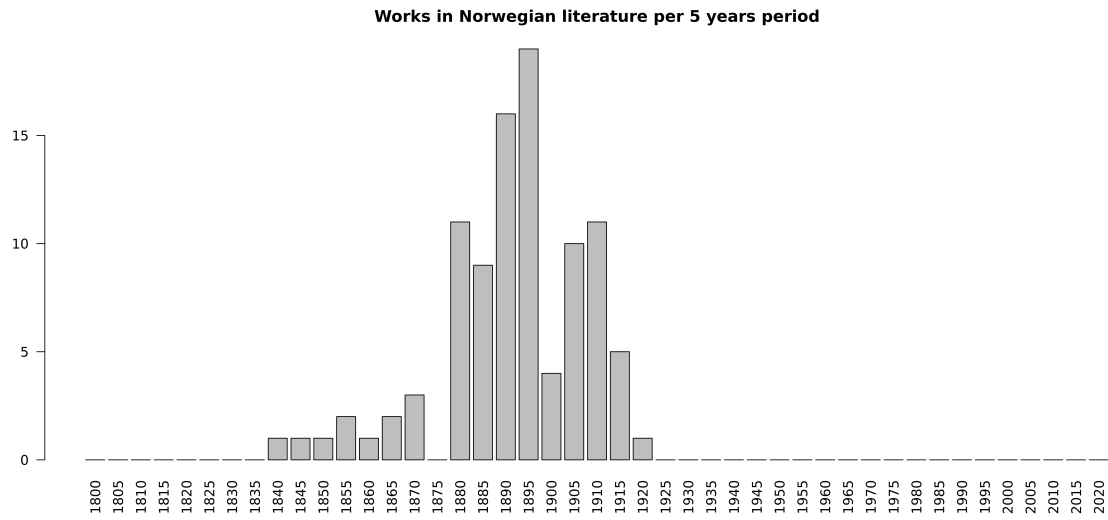


**Figure 13:** Norwegian literature per 5 years period

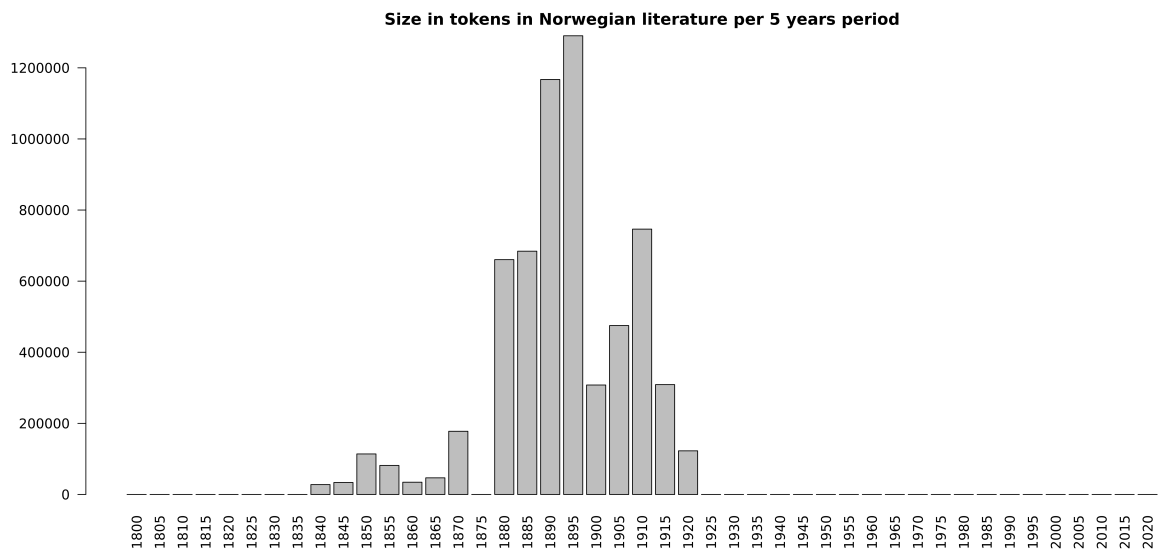


**Figure 14:** Size of Norwegian literature per 5 years period

**Table 8**

Norwegian literature (cont.)

| Title                               | Author               | Year | Size   |
|-------------------------------------|----------------------|------|--------|
| To mennesker                        | Anna Munch           | 1897 | 72431  |
| Afkom                               | Amalie Skram         | 1898 | 138915 |
| Gamle Friks diamant                 | Fredrik Viller       | 1898 | 72108  |
| I afgrunden                         | Gabriel Finne        | 1898 | 21097  |
| Victoria                            | Knut Hamsun          | 1898 | 39816  |
| Cargadør Sahl                       | Jonas Dahl           | 1899 | 58161  |
| Det hemmelighedsfulde skib          | Fredrik Viller       | 1899 | 75645  |
| Enken                               | Thomas P. Krag       | 1899 | 93506  |
| Mod Kveld                           | Tryggve Andersen     | 1900 | 109592 |
| Harriet Blich                       | Nils Collett Vogt    | 1902 | 78584  |
| Fængsel og Fortvilelse              | Hans Jæger           | 1903 | 91136  |
| Rikka Gan                           | Ragnhild Jølsen      | 1904 | 28547  |
| Feranda Mona                        | Ragnhild Jølsen      | 1905 | 47378  |
| Fru Evas Dagbog                     | Hulda Garborg        | 1905 | 47055  |
| Krabvaag                            | Regine Normann       | 1905 | 31407  |
| Hollases Kronike                    | Ragnhild Jølsen      | 1906 | 57592  |
| Psyche                              | Helene Dickmar       | 1906 | 74887  |
| Fru Marta Oulie                     | Sigrid Undset        | 1907 | 31431  |
| Stængt                              | Regine Normann       | 1908 | 36963  |
| Benedicte Stendal                   | Nini Roll Anker      | 1909 | 41853  |
| Fortællingen om Viga Ljot og Vigdis | Sigrid Undset        | 1909 | 52676  |
| Jernvognen                          | Stein Riverton       | 1909 | 53841  |
| Angsten                             | Sven Elvestad        | 1910 | 32522  |
| Barnets tjenere                     | Regine Normann       | 1910 | 36190  |
| Det gyldne Bæger                    | Helene Dickmar       | 1910 | 113659 |
| Faafængt                            | Regine Normann       | 1911 | 26850  |
| Jenny Winge                         | Sigrid Undset        | 1911 | 126846 |
| Livets Triumf                       | Jacob Breda Bull     | 1911 | 76522  |
| De fortæptes hus                    | Sven Elvestad        | 1912 | 41790  |
| Dengang                             | Regine Normann       | 1912 | 57218  |
| Eli                                 | Hulda Garborg        | 1912 | 70393  |
| Eiler Hundevart                     | Regine Normann       | 1913 | 43834  |
| Vaaren                              | Sigrid Undset        | 1914 | 120510 |
| De svake                            | Ronald Fangen        | 1915 | 39889  |
| Det svake kjøen                     | Nini Roll Anker      | 1915 | 125833 |
| Danserinden                         | Helene Dickmar       | 1918 | 53103  |
| Den forsvundne pøsemager            | Lorentz Normann Kvam | 1919 | 32678  |
| Ulvehiet                            | Oscar Braaten        | 1919 | 57441  |
| Kransen                             | Sigrid Undset        | 1920 | 122700 |

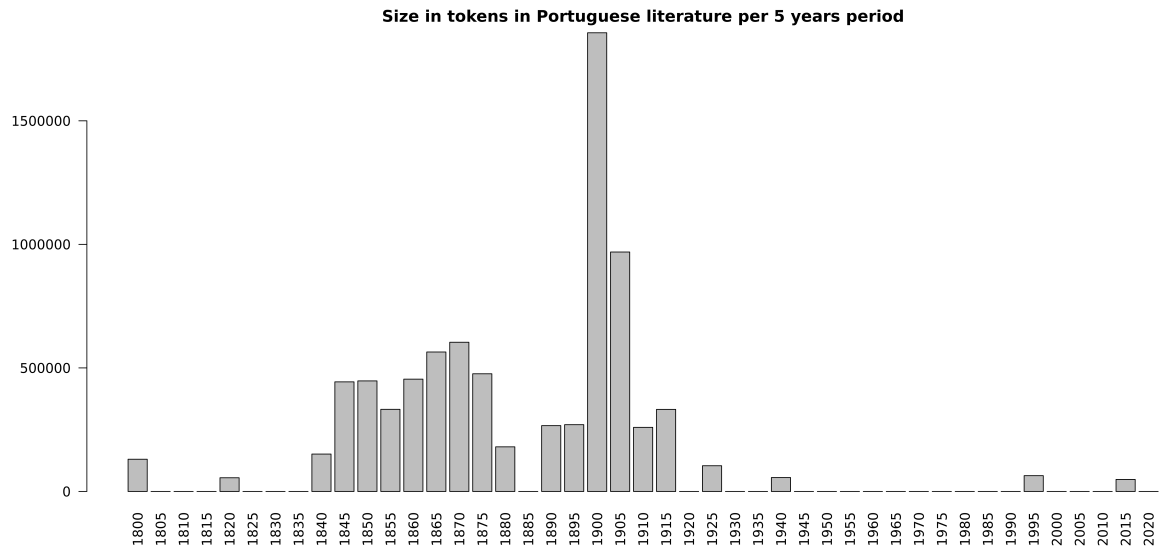


**Figure 15:** Size of Portuguese literature per 5 years period

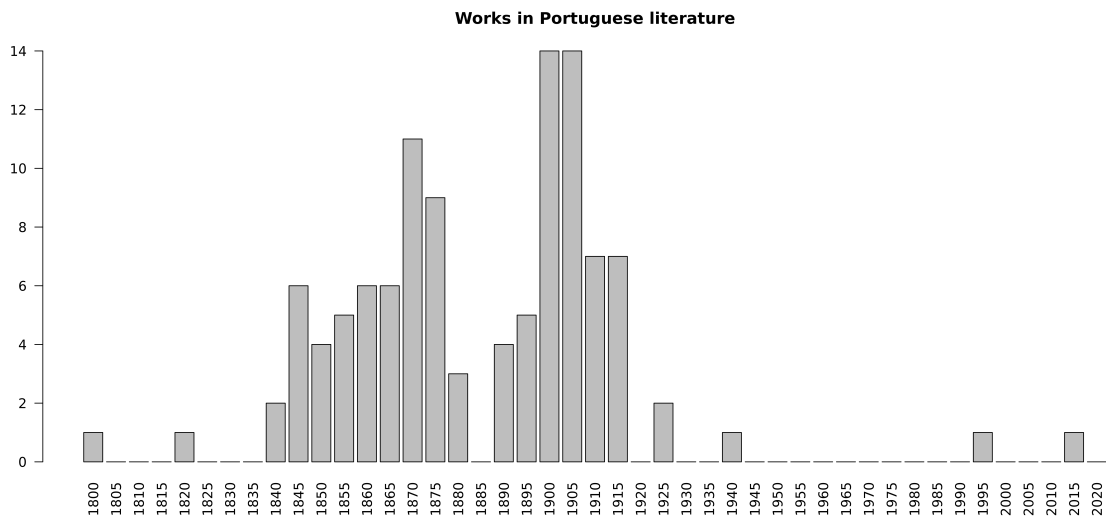


**Figure 16:** Portuguese literature per 5 years period

**Table 9**

## Portuguese literature

| Title                                             | Author                                                    | Year | Size   |
|---------------------------------------------------|-----------------------------------------------------------|------|--------|
| O piolho viajante: Viagens em mil e uma carapuças | Policarpo da Silva                                        | 1802 | 130427 |
| Arzilla                                           | Bernardino Pinheiro                                       | 1822 | 55448  |
| Eurico, o Presbítero                              | Alexandre Herculano                                       | 1843 | 75517  |
| O bobo                                            | Alexandre Herculano                                       | 1843 | 75837  |
| O Arco de Sant'Ana                                | Almeida Garrett                                           | 1845 | 81923  |
| O Galego                                          | Alexandre Herculano                                       | 1845 | 15318  |
| Viagens na minha terra                            | Almeida Garret                                            | 1846 | 91819  |
| O Pároco de Aldeia                                | Alexandre Herculano                                       | 1847 | 46178  |
| O Monge de Cister                                 | Alexandre Herculano                                       | 1848 | 165893 |
| Memorias d'um doido: romance contemporâneo        | António Pedro Lopes de Mendonça                           | 1849 | 42432  |
| Anátema                                           | Camilo Castelo Branco                                     | 1851 | 119553 |
| Uma eleição perdida                               | Conde de Ficalho                                          | 1852 | 35184  |
| Helena                                            | Almeida Garrett                                           | 1854 | 30920  |
| Mistérios de Lisboa                               | Camilo Castelo Branco                                     | 1854 | 261759 |
| Livro Negro de Padre Dinis                        | Camilo Castelo Branco                                     | 1855 | 72602  |
| O Cura de São Lourenço                            | M.M.S.A. e Vasconcelos                                    | 1855 | 68522  |
| O soldado de Aljubarrota                          | Matilde Isabel de Santana e Vasconcelos Moniz Bettencourt | 1857 | 43941  |
| Os tripeiros                                      | António José Coelho Lousada                               | 1857 | 48836  |
| A filha do Arcedíago                              | Camilo Castelo Branco                                     | 1858 | 98516  |
| Adelina                                           | Ana Plácido                                               | 1863 | 57720  |
| Luz coada por ferros                              | Ana Plácido                                               | 1863 | 57078  |
| Providência                                       | Augusto Sarmento                                          | 1863 | 93224  |
| A pálida estrela                                  | Bulhão Pato                                               | 1864 | 14928  |
| O médico do deserto                               | José Joaquim Rodrigues de Bastos                          | 1864 | 59562  |
| Os miseráveis da aristocracia                     | A. Varela                                                 | 1864 | 172030 |
| Os Brâmanes                                       | Francisco Luís Gomes                                      | 1866 | 63506  |
| As Pupilas do Senhor Reitor                       | Júlio Dinis                                               | 1867 | 124730 |
| Febo Moniz                                        | Oliveira Martins                                          | 1867 | 90201  |
| Henriqueta                                        | Maria Peregrina de Sousa                                  | 1867 | 26131  |
| Tempestades do coração                            | João Baptista de Mattos Moreira                           | 1867 | 95895  |
| Uma família inglesa                               | Júlio Dinis                                               | 1867 | 163960 |
| A Rosa do Adro                                    | Manuel Maria Rodrigues                                    | 1870 | 76995  |
| A ermida de Castromino                            | Antonio Aufusto Teixeira de Vasconcellos                  | 1870 | 103168 |
| O Mistério da Estrada de Sintra                   | Ramalho Ortigão e Eça de Queirós                          | 1870 | 90059  |
| A Filha do Cabinda                                | Alfredo Campos                                            | 1873 | 37441  |
| Astúcias de namorada                              | M. Pinheiro Chagas                                        | 1873 | 12624  |
| Christo não volta                                 | Alberto Pimentel                                          | 1873 | 8660   |
| O Manuelinho de Évora                             | António Francisco Barata                                  | 1873 | 59588  |
| Um conto português                                | Miguel J. T. Mascarenhas                                  | 1873 | 73886  |
| Um melodrama em Santo Tirso                       | Manuel Pinheiro Chagas                                    | 1873 | 8649   |
| A Lenda da Meia Noite                             | M. Pinheiro Chagas                                        | 1874 | 63354  |
| Mathilde                                          | Ana Maria Ribeiro de Sá                                   | 1874 | 69431  |
| Os selvagens                                      | Francisco Gomes de Amorim                                 | 1875 | 71578  |
| Um duelo nas sombras                              | António Francisco Barata                                  | 1875 | 61833  |
| A última ceia do Doutor Fausto                    | Alberto Pimentel                                          | 1876 | 13587  |
| As noites do Asceta                               | Alberto Pimentel                                          | 1876 | 12181  |
| El-rei dinheiro                                   | Arnaldo Gama                                              | 1876 | 115492 |
| O Cristão novo                                    | Diogo de Macedo                                           | 1876 | 31858  |
| A infâmia de Frei Quintino                        | Urbano Loureiro                                           | 1878 | 77804  |
| Da Parte d'El-Rei                                 | A. M. da Cunha e Sá                                       | 1878 | 75293  |
| O Romance da Rainha Mercedes                      | Alberto Pimentel                                          | 1879 | 16628  |
| A divorciada                                      | José Augusto Vieira                                       | 1881 | 69080  |
| O frade arrábido                                  | Joaquim Augusto de Oliveira Mascarenhas                   | 1881 | 35513  |
| A Brasileira de Prazins                           | Camilo Castelo Branco                                     | 1882 | 75997  |
| Severina                                          | Guiomar Torrezão                                          | 1890 | 5596   |

**Table 10**

Portuguese literature (cont.)

| Title                                            | Author                                        | Year | Size   |
|--------------------------------------------------|-----------------------------------------------|------|--------|
| O Barão de Lavos                                 | Abel Botelho                                  | 1891 | 146070 |
| O último cartuxo da Scala Caeli de Évora         | António Francisco Barata                      | 1891 | 86854  |
| O Dr. Luís Sandoval                              | Alice Moderno                                 | 1892 | 27893  |
| A guerrilha de Frei Simão                        | Alberto Pimentel                              | 1895 | 88515  |
| O Mundo no Ano Três Mil                          | Pedro José Supico de Moraes                   | 1895 | 103714 |
| Diário duma Criança                              | Ana de Castro Osório                          | 1897 | 22735  |
| Pero da Covilhã: Episódio Romântico do Século XV | Zeferino Norberto Gonçalves Brandão           | 1897 | 45123  |
| Aventuras do sr. Criptogamo: romance humorístico | Alfredo de Moraes Pinto                       | 1899 | 10189  |
| A ilustre Casa de Ramires                        | Eça de Queirós                                | 1900 | 141257 |
| Filho das ervas                                  | Carlos Malheiro Dias                          | 1900 | 129724 |
| Gomes Freire                                     | Rocha Martins                                 | 1900 | 511767 |
| O exilado                                        | Maurícia C. de Figueiredo                     | 1900 | 54723  |
| A bruxa: cenas açorianas                         | Augusto Loreiro                               | 1901 | 56531  |
| Amanhã                                           | Abel Montelho                                 | 1901 | 180245 |
| A carteira de um neurastênico                    | Antonio Lobo                                  | 1903 | 67001  |
| Ambições                                         | Ana de Castro Osório                          | 1903 | 99989  |
| O Conde de Castel Melhor                         | João da Câmara                                | 1903 | 259862 |
| O Conde de S. Luiz                               | Tomaz de Melo                                 | 1903 | 68908  |
| Alice                                            | Maria Amália Vaz de Carvalho                  | 1904 | 18513  |
| Os filhos do padre Anselmo                       | António da Costa Couto Sá de Albergaria       | 1904 | 152771 |
| Tríplice Aliança                                 | Raul de Azevedo                               | 1904 | 41197  |
| Viriato                                          | Teófilo Braga                                 | 1904 | 73801  |
| A Ala dos Namorados                              | Antonio Campos Júnior                         | 1905 | 298631 |
| A rua do ouro                                    | Alfredo de Mesquita                           | 1905 | 71760  |
| Dramas da Corte                                  | Alberto de Castro                             | 1905 | 42572  |
| Lisboa no ano 2000                               | Melo de Matos                                 | 1906 | 10831  |
| Os Pobres                                        | Raul Brandão                                  | 1906 | 61703  |
| Os bravos do Mindelo: Romance histórico          | Faustino da Fonseca                           | 1906 | 65563  |
| A Feiticeira                                     | Ana de Castro Osório                          | 1908 | 10094  |
| A vida por um prejuízo                           | Alice Pestana                                 | 1908 | 17102  |
| A vinha                                          | Ana de Castro Osório                          | 1908 | 6421   |
| Noites de Cintra                                 | Alberto Pimentel                              | 1908 | 42831  |
| O Marquês da Bacalhoa                            | António de Albuquerque                        | 1908 | 72563  |
| Os quatro reis impostores                        | Marcelino Mesquita                            | 1908 | 170928 |
| Sacrificada                                      | Ana de Castro Osório                          | 1908 | 21572  |
| O Salústio Nogueira                              | Teixeira de Queirós                           | 1909 | 76454  |
| Loucura...                                       | Mário de Sá Carneiro                          | 1912 | 18356  |
| Marina                                           | Maria Benedicta Mousinho de Albuquerque Pinho | 1912 | 34276  |
| Nova Sapho Tragedia Extranha                     | Visconde de Vila Moura                        | 1912 | 61493  |
| Ao de Leve                                       | Brito Camacho                                 | 1913 | 25674  |
| A Confissão de Lúcio                             | Mário de Sá Carneiro                          | 1914 | 34346  |
| A Marquesa de Vale Negro                         | Maria O'Neill                                 | 1914 | 63450  |
| Por bom caminho                                  | Maria O'Neill                                 | 1914 | 21821  |
| Decameron                                        | Virginia de Castro e Almeida                  | 1916 | 44276  |
| Inocente                                         | Virginia de Castro e Almeida                  | 1916 | 16198  |
| O Solar dos Pavões                               | Virginia de Castro e Almeida                  | 1916 | 25097  |
| A Engomadeira: Novela Vulgar Lisboeta            | José de Almada Negreiros                      | 1917 | 15413  |
| Amor crioulo                                     | Abel Botelho                                  | 1917 | 107607 |
| Húmus                                            | Raul Brandão                                  | 1917 | 84075  |
| O Pobre de Pedir                                 | Raul Brandão                                  | 1917 | 39487  |
| A Farsa                                          | Raul Brandão                                  | 1926 | 60091  |
| A morte do Palhaço                               | Raúl Brandão                                  | 1926 | 43992  |
| A História da Dona Redonda e da sua Gente        | Virginia de Castro e Almeida                  | 1942 | 56295  |
| Euronovela                                       | Miguel Vale de Almeida                        | 1998 | 63929  |
| mISTério@Tagus                                   | Luísa Marques da Silva                        | 2015 | 48729  |