

Gender depiction in Portuguese Distant reading Brazilian and Portuguese literature

Cláudia Freitas
maclaudia.freitas@gmail.com

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

2nd Annual Conference of Computational Literary Studies, 22-23 June 2023

Presentation



Introduction

- Two aims with this study
- Two goals with this presentation
- Disclaimer

The material

- Distribution in time
- Genre description
- Size

Discussion

Basic results

- List of characterizing expressions
- Relative characterisation



- ▶ Study characterizing descriptions of people in Portuguese
 - ▶ in fiction
 - ▶ and in other texts by the same authors



- ▶ Study characterizing descriptions of people in Portuguese
 - ▶ in fiction
 - ▶ and in other texts by the same authors
- ▶ See how they are distributed when depicting females and males



- ▶ to present the methodology and the resources,
- ▶ to bring one other language to computational literary studies: Portuguese



We have no strong opinions as to gender, and no agenda

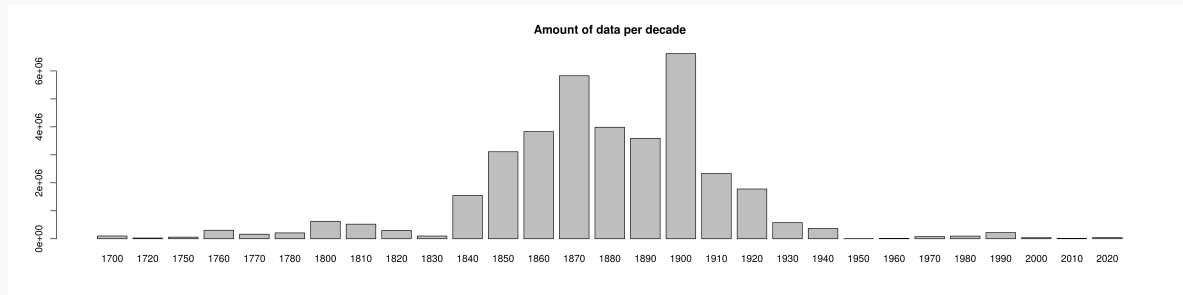
- ▶ We are aware that gender is nowadays considered different from sex. Gender is a social construct, so no need for gender to be binary
- ▶ We are aware that "cultural analytics" works have been blamed for naively reinforcing binarity in gender studies by e.g. Mandell (2019)
- ▶ But we are also aware that most of the literature we survey and study precludes these ideas and analyses, that is, in those works and those times, there were two sexes and two genders, something which is very easy to trace given (binary) morphological gender in Portuguese
- ▶ By making our data available we give other researchers the opportunity to create other (more advanced) classificaton schemes where e.g. different genders are taken into consideration

We strongly believe that it is important to also tap into other languages before categorical decisions are made, such as that this kind of studies are wrong

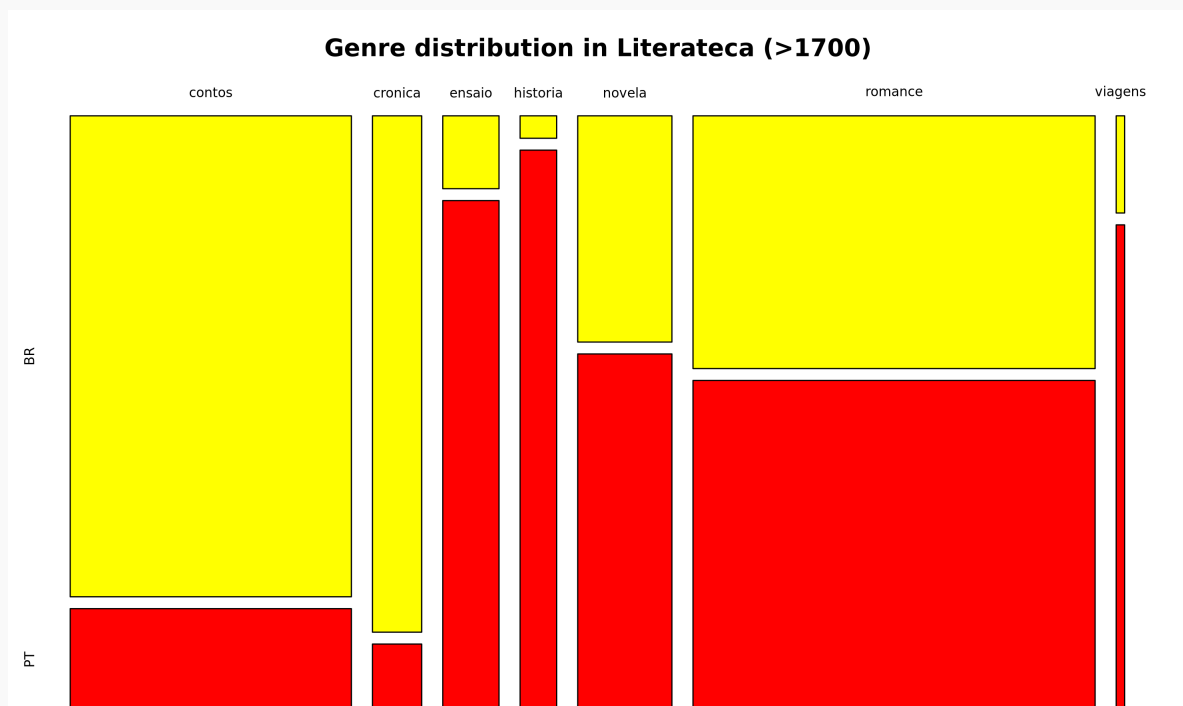


We use in this study **Literateca**

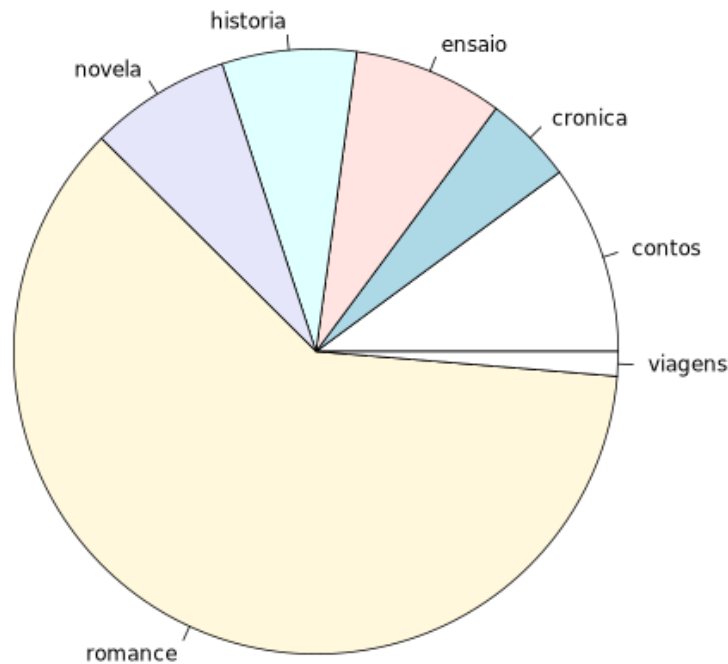
- ▶ literary works and/or works written by literary authors
- ▶ including Portuguese and Brazilian authors, but also parts of works by other writers in Portuguese (Angola, East-Timor, Cape Verde, etc.)
- ▶ imbalanced according to country, to time, to author gender



Genre distribution per works



Genre distribution per words



Size of the (sub)corpora



Literature	no. of tokens	no. works	no. authors
Total	31,991,537	640	198
Portuguese	19,919,769	296	126
Brazilian	12,071,768	346	73
Prose material from 1700 to the present			
Literature	no. of tokens	no. works	no. authors
Total	24,720,246	529	178
Portuguese	19,919,769	296	126
Brazilian	19,523,921	307	69

Fiction from 1840



1. we have not included drama nor poetry, for two reasons: fear of different syntactic structures and therefore poorer parser performance, and gross imbalance across varieties (most came from Portugal)
2. we have included other text written by the same authors when it was available (literature scholars had done the digitizing work) (but we also did studies only on fiction)
3. we removed translated material and material from varieties other than Brazilian and Portuguese
4. we used 1700 as a cutoff point, except when the studies compared the two varieties, in which case we used 1840 (Brazil became independent in 1822)

All these decisions are arguable, but readers can redo the work with other subsets, since the material (and the commands) are available.



1. people?: one reviewer questioned the conflation of (fictional) people with real people, and argued that descriptions of these two kinds of humans should not be confused. However, they are descriptions by the same people (authors). Anyway, we also study separately descriptions in fiction.
2. gender vs. sex: while we say we are interested in gender, we use mainly sex (morphological gender) as identifiable categories. So maybe we should say that this is a study that starts from a binary (sex) distinction and intends to identify the cultural gender that is used while describing these two sexes.
3. our study may be a way to uncover bias, and also a way to find especially interesting descriptions that may have changed with time, but mainly it is a first step in providing the linguistic means for further, and more literary inspired, studies.



- ▶ We have not done this study because we were sure about a particular situation. Rather, we wanted to identify what was possible to find using rather simple descriptions, which may be interpreted in different ways and with different biases by different researchers.
- ▶ This is not the final verdict about gender in Portuguese (literature), rather, it is just a first step that can be improved and criticized by its readers, and hopefully foster more research on the subject with a distant reading approach.
- ▶ We are also aware that literature is not a true description of society, but that literary authors do have an important influence in the minds of the readers of their time, namely the elites in Portugal and Brazil.



Some other themes that we find interesting to pursue, in future work:

- ▶ Are there authors specially misogynous in lusophone literature?
- ▶ Are there intrinsically gender-related adjectives that should be looked into more specifically? Such as *varonil*, *feminino*, *masculino*, *efeminado*, *mulheril*, ...
- ▶ Can one identify gay people using adjectives/characterizing words?
- ▶ Have there been clear changes in depicting the two sexes through time?
- ▶ Are there obvious differences in gender depiction between translated and original literature?
- ▶ Are there other depicting categories, like age, or civil status, that should be studied separately?



A lexicon of characterizing expressions

Circa 4,500 ways of characterizing one person, divided in 4 categories, plus one other category. Classified out of context.

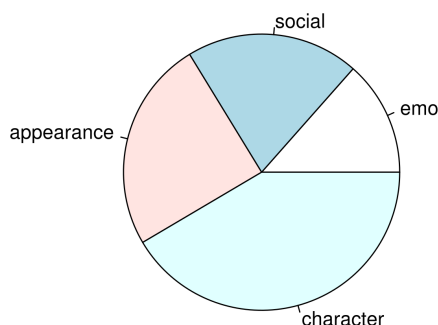
Number of depicting word (types)

type	size
social	1391
appearance	672
emotional	514
character	1578
other	326
total	4481

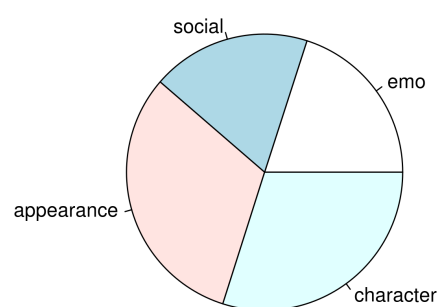
In the whole material



Masculine depiction

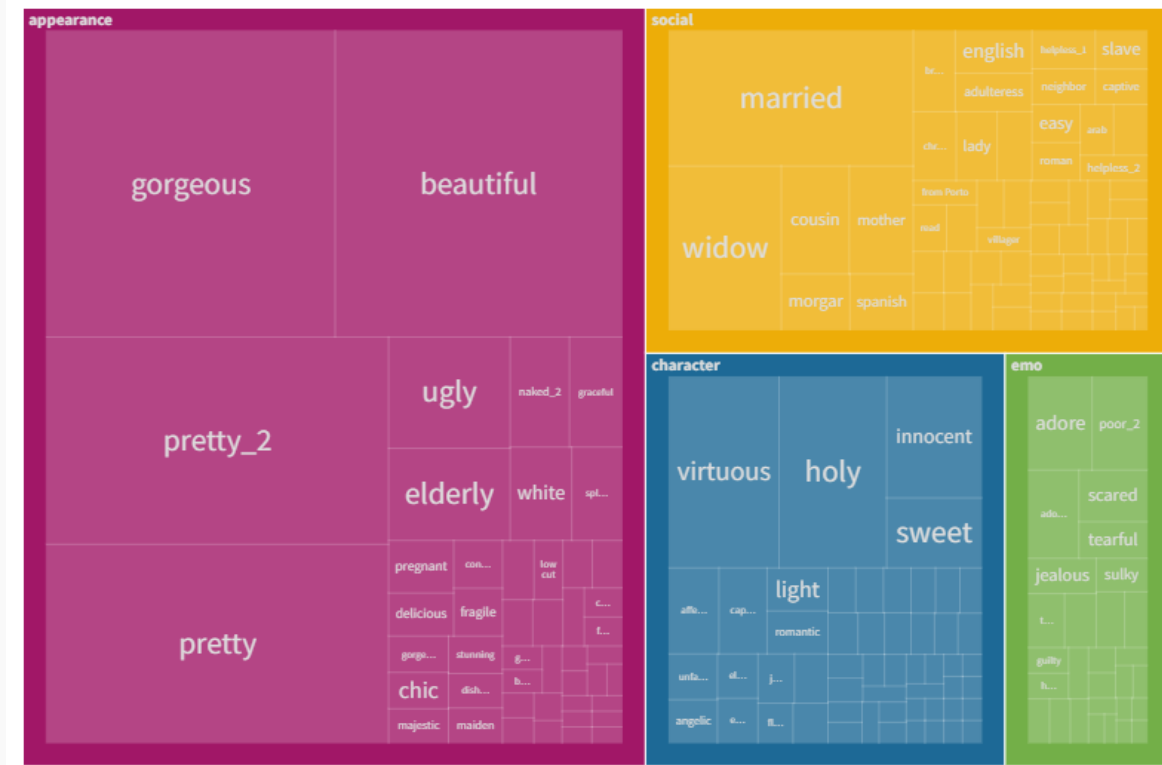


Feminine depiction

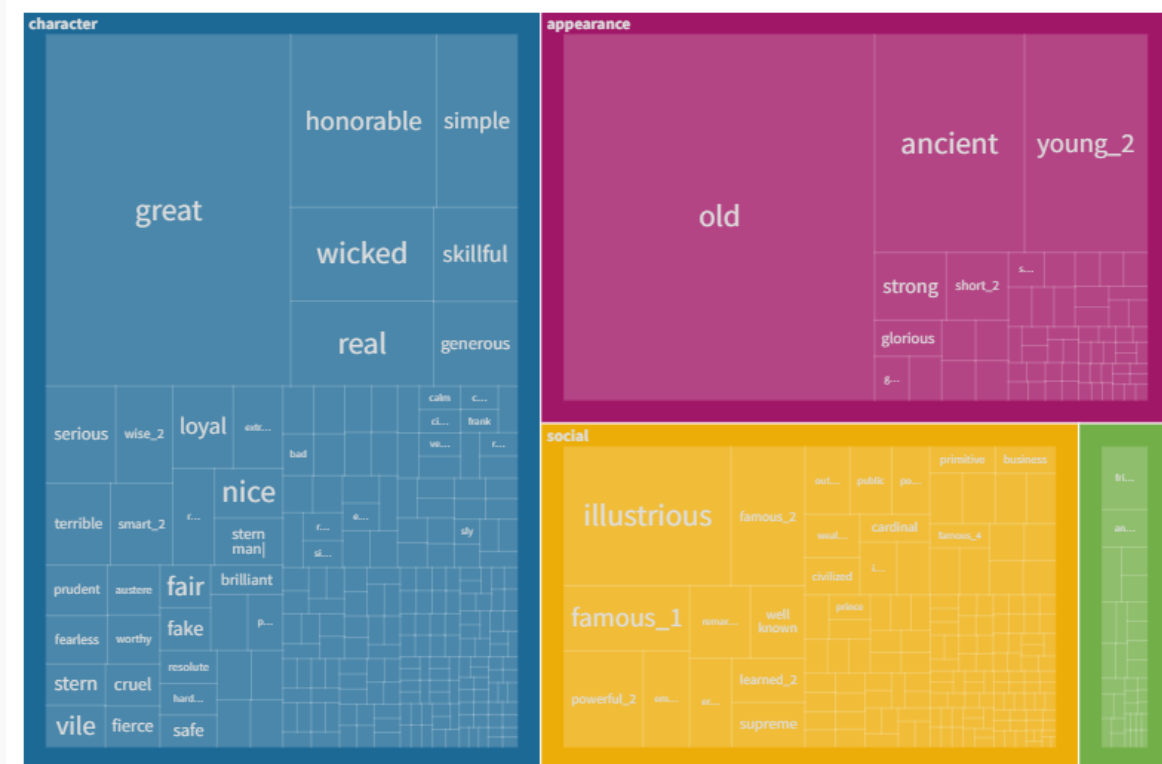


Women are more frequently described by their appearance; men by their character.

But is this all? Other differences



But is this all? Other differences



Thank you for your attention!

Only fiction



Genre and author gender in fiction (>1700)

