

Comida e bebida nas literaturas portuguesa e brasileira: O projeto ReadingFood

Diana Santos, Eckhard Bick, Cristina Mota

d.s.m.santos@ilos.uio.no, eckhard.bick@gmail.com, cristina.mota@inesc-id.pt

DHandNLP2026, Salvador, 13 de abril de 2026



Estudos literários computacionais

Estudar a literatura em português usando métodos e ferramentas de processamento de linguagem natural (PLN) e de linguística com corpos (LC)

- Os dados são literários
- As perguntas de pesquisa são literárias
- Os métodos são de PLN ou de LC
- O ambiente computacional é “gêmeo” do para estudar linguística ou língua: a Literateca é uma parte da Gramateca (Santos, 2014)

O projeto *ReadingFood*

Estudar o campo semântico da comida e bebida na literatura.

- Três línguas, quatro literaturas
- Norueguês, italiano, português (Portugal e Brasil)
- Literatura no domínio público: romances e novelas
- Neste artigo apenas nos referimos à literatura portuguesa e brasileira
- Não é um projeto de sociologia nem de culinária nem de estudos de comida, é um projeto de estudos literários (e linguísticos) computacionais

Veja-se Santos et al. (2026) para uma panorâmica geral, e para motivação.



Fungi e pirão.

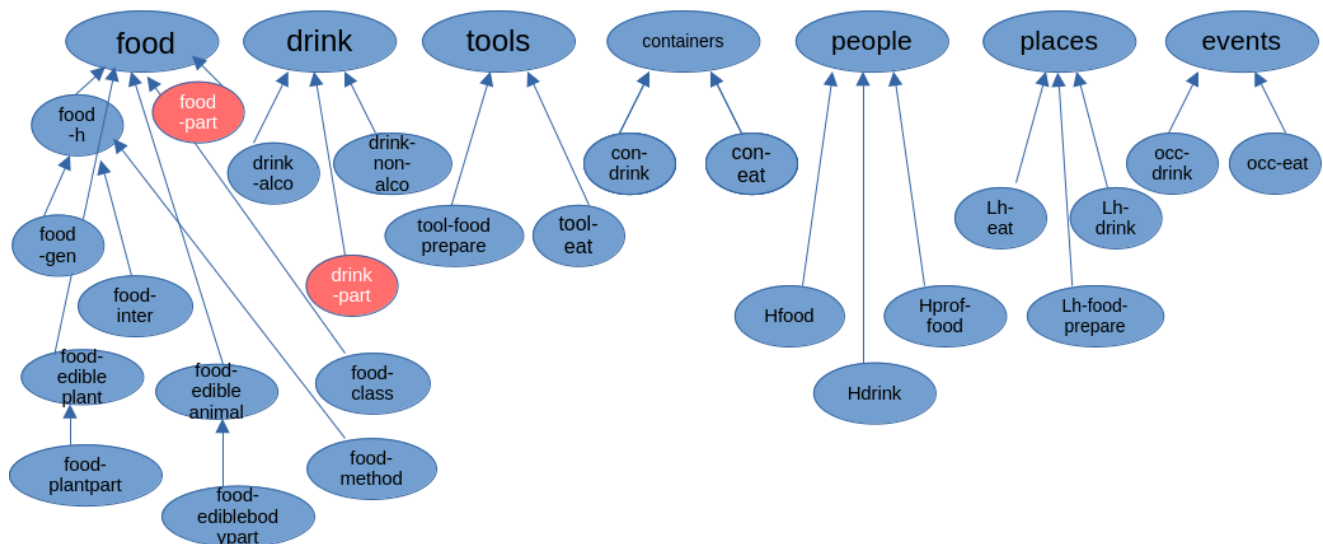
Organização do projeto

Para cada literatura

- 1 Identificação do vocabulário presente nas obras, e sua classificação (taxonomia)
- 2 Desambiguação em contexto
- 3 Classificação dos usos das palavras ou expressões de comida, assim como identificação dos seus usos metafóricos e de expressões convencionais associadas a esse domínio
- 4 Identificação de cenas de comida e bebida (refeições, preparação, visitas a estabelecimentos onde se come e bebe)
- 5 Objetivo dessas cenas no enredo

Comparação entre as quatro literaturas.

Vocabulário do campo semântico da comida e bebida



alimentos, bebidas, locais, profissões, instrumentos, eventos, objetos, características humanas

Ainda falta: ações (verbos)

A vermelho: meronímia e não hiperonímia

<https://www.linguateca.pt/Gramateca/> → Literateca → Comida e bebida

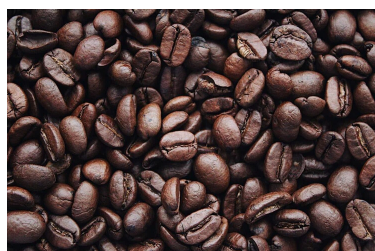


Desambiguação em contexto: a palavra *café*

Nas obras classificadas há 1838 lemas de *café*

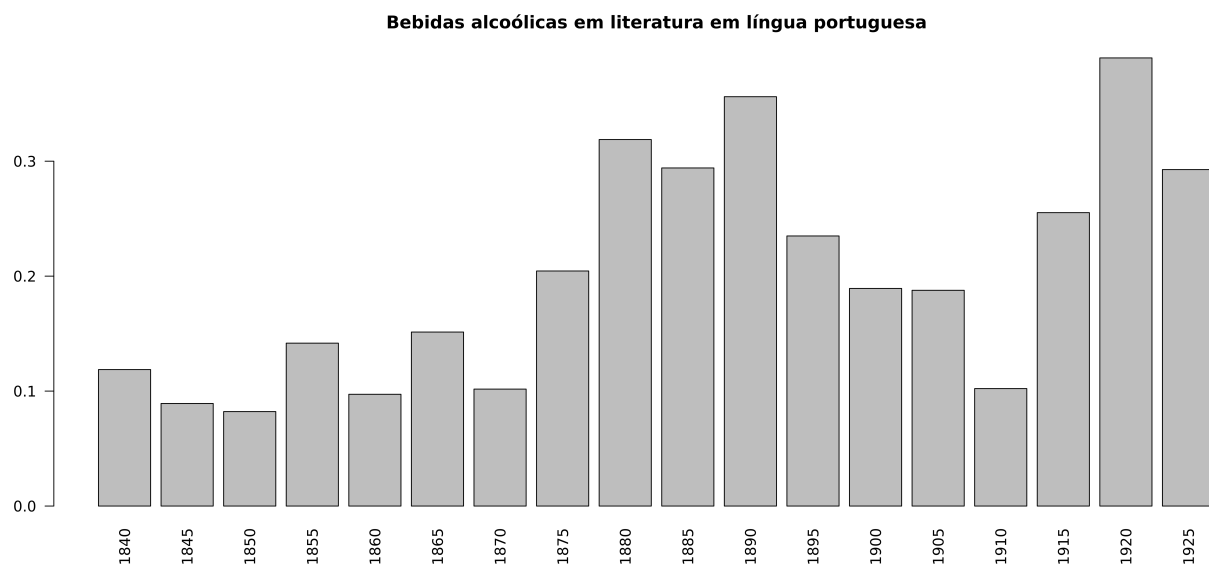
| | | |
|--------------------------------|-----|------|
| bebida | 851 | 46% |
| lugar de comida | 423 | 23% |
| refeição ou parte dela | 163 | 8,8% |
| outros (planta, produto, etc.) | 405 | 22% |

- o lema *café* aparece, relativamente, quatro vezes mais na literatura brasileira
- a referência aos cafés como loja é muito mais frequente em Portugal

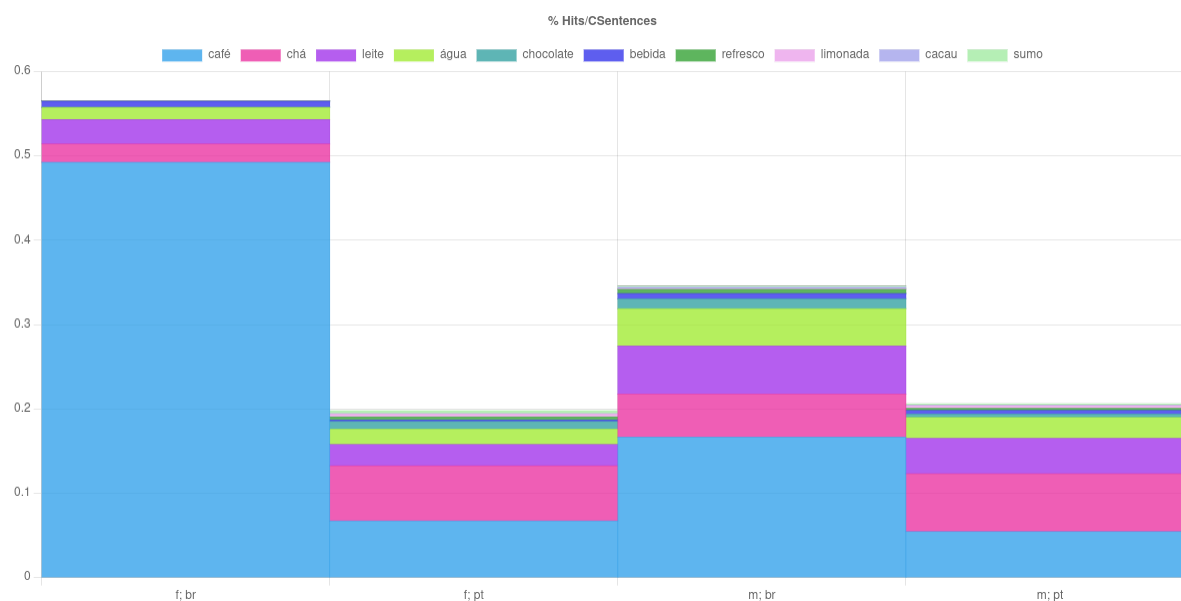


Visualização

Dois sistemas para visualização dos resultados da anotação: a Gramateca (AC/DC associado à linguagem R), e o CorpusEye (Bick, 2005), <https://corp.visl.dk/>.

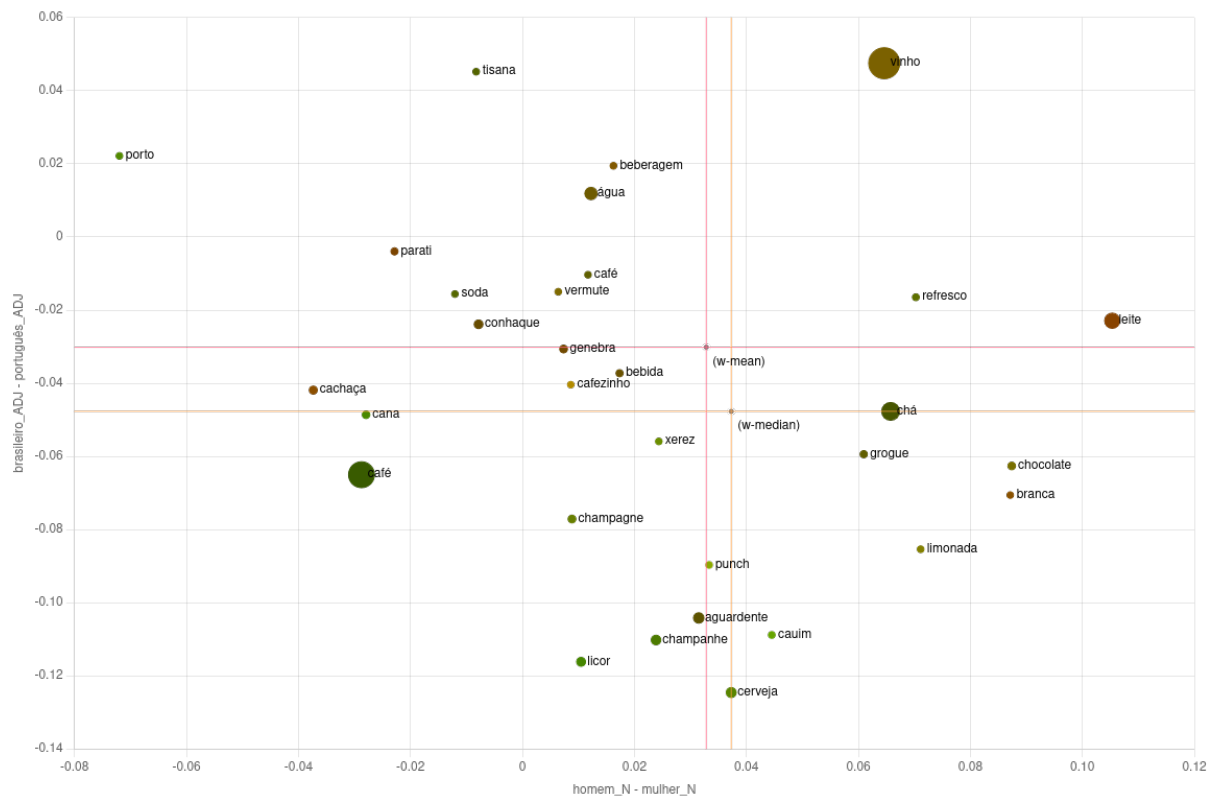


Visualização, CorpusEye



As escritoras brasileiras são as que mencionam mais café; os escritores brasileiros são os que mencionam mais leite.





Identificação de refeições (e outras cenas de comida e bebida)

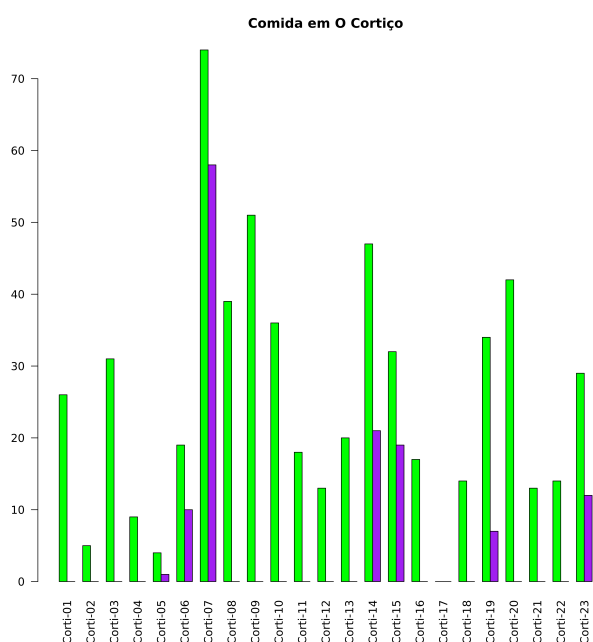
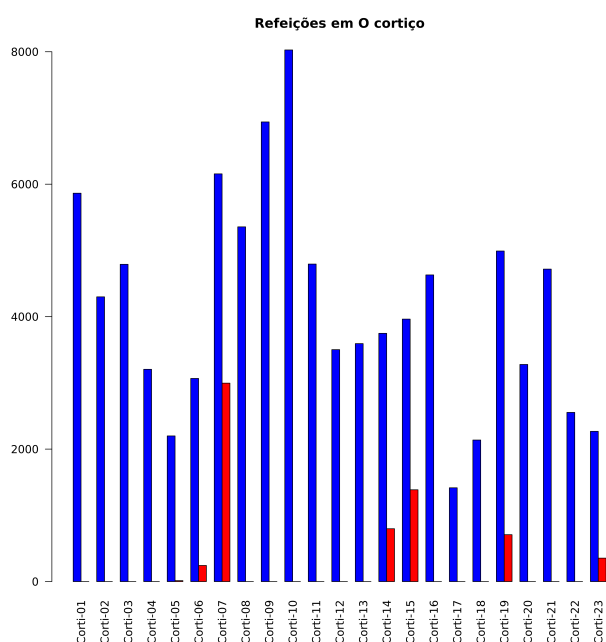
- Aqui já não estamos a pensar numa anotação lexical ou de oração, mas sim identificar episódios (subtextos) numa obra, que podem ter uma duração (em termos de palavras) apreciável.
- Inspirados por Schank e Abelson, vamos tentar criar (ou descobrir) um guião (script) das refeições nos períodos a que se referem as obras. S&A explicam como se compreendem narrativas baseadas em conhecimento específico, por exemplo, o que significa uma ida a um restaurante.
- A primeira tarefa foi identificar manualmente todas as refeições em dez obras, também para ver se a densidade de palavras associadas a comida e bebida era superior nesses casos.

Valores para as dez obras

| Obra | Data | Autor | Palavras | Capítulos | Refeições | Comida | Relativa |
|-------|------|-------|----------|-----------|-----------|--------|----------|
| Pupil | 1867 | JD | 114426 | 42 | 3 | 211 | 1.84 |
| Famil | 1868 | JD | 147041 | 39 | 4 | 228 | 1.55 |
| Morga | 1868 | JD | 177122 | 33 | 8 | 204 | 1.16 |
| Justi | 1870 | JD | 34808 | 9 | 3 | 99 | 2.84 |
| Fidal | 1871 | JD | 168262 | 37 | 7 | 150 | 0.89 |
| Padre | 1875 | EQ | 171371 | 25 | 18 | 496 | 2.89 |
| Filom | 1884 | AA | 64569 | 23 | 3 | 112 | 1.73 |
| Maias | 1888 | EQ | 266017 | 18 | 21 | 857 | 3.22 |
| Corti | 1890 | AA | 95472 | 22 | 10 | 585 | 6.13 |
| Sogra | 1895 | AA | 61809 | 25 | 0 | 61 | 0.99 |
| Total | | | 1300897 | 273 | 77 | 30003 | |

JD - Júlio Dinis; EQ - Eça de Queirós; AA - Aluísio Azevedo

Exemplo em *O Cortiço* de Aluísio de Azevedo



- Um corpo relativamente grande de prosa literária portuguesa e brasileira anotado e desambiguado (neste momento a 75%) com o campo semântico da comida e da bebida, acessível publicamente em duas plataformas (AC/DC-Gramateca e CorpusEye)
- Dez obras marcadas com refeições e/ou outros episódios envolvendo comida ou bebida, como subconjunto da Literateca
- Uma primeira taxonomia de nomes e adjetivos usados na literatura
- Em curso alguns primeiros estudos sobre usos metafóricos de termos de comida e bebida
- No futuro pretendemos escrever um artigo sobre deteção automática de refeições, mas já sabemos, a partir deste estudo preliminar, que não basta observar a densidade de termos do domínio
- Um objetivo fundamental é caracterizar, como leitura distante, obras literárias e autores com base neste campo semântico. Como acepipe, podemos indicar que as autoras falam um pouco mais sobre comida que os autores, e que os brasileiros falam mais do que os portugueses

Referências

- Eckhard Bick. "CorpusEye: Et brugervenligt web-interface for grammatisk opmærkede korpora." Em *10. Møde om Udforskningen af Dansk Sprog 7. - 8.okt.2004, Proceedings*, 2005, pp. 46–57. Århus University.
- Diana Santos. "Gramateca: corpus-based grammar of Portuguese". In Jorge Baptista et al., *Proceedings of PROPOR 2014*, pp. 214-219.
- Diana Santos, Elizaveta Khachatryan, Michael Preminger, Åse Kristine Tveit & Eckhard Bick. "Presenting Reading Food and its infrastructure". *Digital Humanities in the Nordic and Baltic Countries Publications*, 2026. <https://www.linguateca.pt/Diana/download/DHK0ReadingFood.pdf>
- Roger C. Schank & Robert P. Abelson. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Routledge, 1977.