

Comida e bebida nas literaturas portuguesa e brasileira: o projeto *ReadingFood*

Diana Santos

Linguatca & Univ. of Oslo
Postboks 1003 Blindern
N-0315 Oslo, Noruega
d.s.m.santos@ilos.uio.no

Eckhard Bick

Univ. of Southern Denmark
Campusvej 55
DK-5230 Odense M, Dinamarca
eckhard.bick@gmail.com

Cristina Mota

Linguatca & INESC
R. Alves Redol 9
1000-029 Lisboa, Portugal
cristina.mota@inesc-id.pt

Resumo

Neste artigo descrevemos brevemente o projeto *ReadingFood* sobre o campo semântico da comida e bebida na literatura, que pretende comparar as obras de quatro países no período 1840-1920, mas cingindo-nos a Portugal e ao Brasil. Após apresentar as infraestruturas já desenvolvidas, tornando pública a pesquisa neste domínio, apresentamos o trabalho já feito e alguns estudos preliminares: a criação de uma taxonomia do domínio na literatura, a desambiguação em contexto, e o estudo de refeições (ou eventos relacionados com comida e bebida).

1 Motivação

Desde os anos 90 do século passado que surgiu a leitura distante, uma forma de complementar os estudos literários com métodos computacionais de maior abrangências, que não necessitassem da leitura próxima de milhares ou mesmo milhões de obras, veja-se [Moretti \(2013\)](#). Dessa forma surgiram trabalhos dos mais variados géneros em linguística com corpos, em que os corpos eram constituídos por obras literárias.

Para o português existem infelizmente menos obras digitalizadas e acessíveis do que para o inglês, mas pelo menos as bibliotecas nacionais de Portugal e do Brasil, e projetos como a Biblioteca Digital de Literatura de Países Lusófonos¹, o projeto Gutenberg², o projeto LusoLivros³, o projeto Adamastor⁴ e o Internet Archive⁵, digitalizaram (e, em alguns casos, modernizaram) centenas ou mesmo milhares de obras em português. Veja-se [Schöch et al. \(2021\)](#) para mais informação de como coligir uma coleção de obras de Portugal, e algumas das diferenças entre estes projetos.

Depois existem projetos que tentam enriquecer essas obras em termos linguísticos, anotando-as e tornando a procura no conjunto de obras acessível através da internet. Um exemplo é a Lite-rateca ([Santos, 2019](#)), que associa às obras tanto informação linguística (morfossintática e semântica) sobre cada palavra ([Santos, 2014](#)), como informação que podemos considerar como informação literária, por exemplo as personagens ([Santos e Freitas, 2019](#); [Santos et al., 2026a](#)).

No projeto em apreço, *ReadingFood* ([Santos et al.](#)), associamos aos textos literários informação sobre comida e bebida, que reputamos um domínio fundamental para estudar a cultura e a literatura ([Meigs, 1997](#); [Boyce e Fitzpatrick, 2017](#); [Coghlan, 2020](#); [Barkan, 2021](#)). Neste projeto, em curso, além de identificar a presença do domínio e os vários usos em texto literário em diferentes autores e escolas em quatro literaturas (portuguesa, brasileira, norueguesa e italiana), pretendemos:

- investigar a presença de metáforas e expressões idiomáticas;
- identificar expressões convencionais associadas a bebida e comida (rituais de convite, de brinde, de início e fim de refeição, por exemplo), como elogiar ou descrever comida ([Traverso e Dimachki, 2017](#)) e/ou pessoas ([Korthals, 2008](#));
- localizar cenas associadas a refeições ou à sua preparação que façam parte do enredo ([Schank e Abelson, 1977](#)), estudando também a sua função ([Brown, 1984](#));
- comparar as quatro culturas envolvidas segundo todos estas vertentes.

No presente artigo, referimo-nos simplesmente ao português, e apresentamos as primeiras duas tarefas: a classificação das palavras na secção 2 e a sua desambiguação em contexto na secção 3, seguida

¹<https://literaturabrasileira.ufsc.br>

²<https://www.gutenberg.org/>

³aparentemente descontinuado

⁴<https://projectoadamastor.org/>

⁵<https://archive.org/>

de alguns estudos preliminares sobre a anotação e classificação de refeições, na secção 4.

2 Taxonomia

Começámos por desenvolver uma taxonomia dos conceitos que nos pareceram estar associados às palavras de comida e bebida que encontrámos no nosso corpo de obras literárias, formado essencialmente por obras do século XIX e princípio do século XX. Os números referem-se ao número de palavras distintas já classificadas nessa categoria. Não sendo definitiva, apresentamo-la na Figura 1.

food-gen (31) palavras gerais que se referem a comida

food-h (200) palavras que se referem a uma comida medeada por um processo qualquer culinário, e que têm um nome

food-intermediate (54) palavras que se referem a ingredientes ou especiarias, mas que não são comidas separadamente

food-edibleplant (145) palavras que descrevem entidades do mundo vegetal que são comidas (cozinhadas ou cruas)

food-plantpart (3) palavras que descrevem partes de uma entidade do mundo vegetal comestível

food-edibleanimal (83) palavras que se referem a animais que são comidos

food-ediblebodypart (46) palavras que se referem a partes de animais que se comem

food-part (33) palavras usadas para definir uma parte de algo comestível

drink-part (10) palavras usadas para referir uma parte de bebida

food-meth (17) palavras usadas para descrever um método de tratar os alimentos

food-class (39) palavras usadas para descrever uma classe de alimentos

con-food (42) palavras que descrevem um contentor de comida

con-drink (39) palavras que descrevem contentores de bebida

tool-foodprepare (2) utensílios de cozinha

tool-eat (4) utensílios para comer

Hprof-food (18) profissões associadas a comida

Lh-foodprepare (10) locais de preparação ou armazenamento de comida

Lh-eat (5) locais onde se come

Lh-drink (7) locais onde se bebe

drink (9) bebidas em geral, alcoólicas ou não

drink-non-alco (31) bebidas não alcoólicas

drink-alco (126) bebidas alcoólicas

occ-eat (23) palavras que se referem a ocasiões em que se come

occ-drink (5) palavras que se referem a ocasiões em que se bebe

jfood (55) palavras que descrevem qualidades da comida ou bebida

Hfood (4) pessoas descritas em relação à sua atitude à comida

Hdrink (7) pessoas descritas em relação à sua atitude à bebida

Como será evidente, em muitos casos uma mesma palavra pertence a mais do que uma categoria, e é por isso que procedemos à desambiguação em contexto.

Assim, muitas palavras descrevem tanto uma quantidade de bebida como um contentor para bebidas: *copo*, *xícara*, *chávena*, *cálice*, ... assim como muitas palavras descrevem (fora do contexto) uma refeição, ou aquilo que se comeu ou bebeu à refeição: *jantar*, *café*... Também alguns nomes de "tratamentos" dados à comida ou à classe de comida são usados para referir aquilo que se comeu ou mesmo nomes de receitas: *assado*, *cozido*, *sopa*...

Além disso, é evidente que a maior parte das palavras do campo da comida e bebida não pertencem exclusivamente a esse campo, quer por homonímia, como *copo* parte de espada, ou *rancho* edifício, como pelas várias relações de extensão de sentido ou metafóricas: a) de comida para outros domínios, como *coco* para chapéu de coco, ou *apimentar as conversações*, ou b) doutros domínios para comida, como *ladrilhos* de marmelada, ou c) ainda casos mais gerais que se aplicam a vários domínios, um

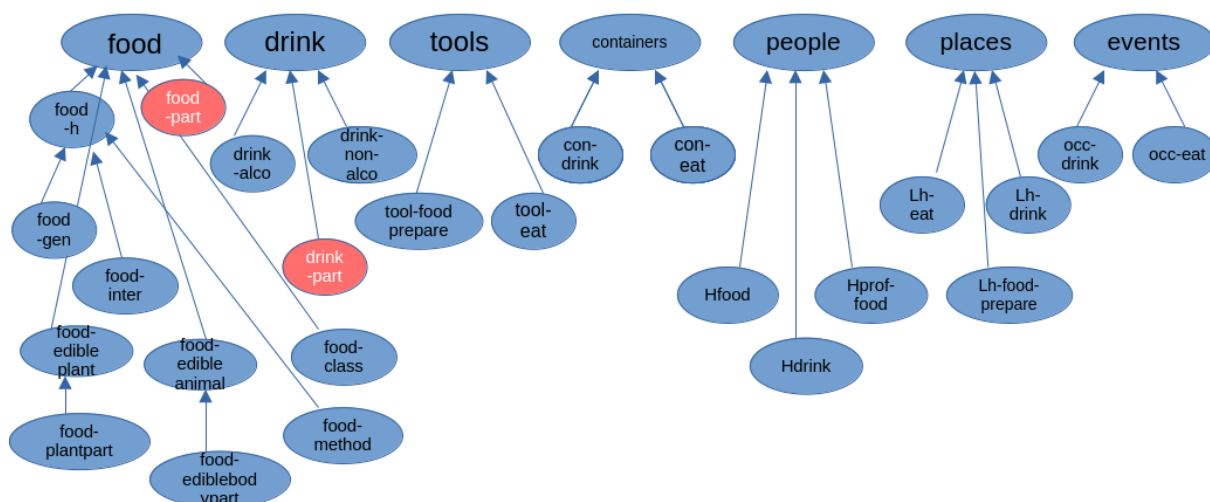


Figura 1: Categorias de comida e bebida

deles a comida e bebida, como *delícia* ou *acompanhamento*. Além disso, não podemos esquecer que os nomes dos animais e das plantas comestíveis se podem referir aos ditos animais e plantas sem ser num contexto de alimentação, e o mesmo se refere à palavra *água*, que forma rios e mares, e é usada por exemplo para lavar, e portanto não só para beber.⁶

Seja como for, os números das diferentes palavras que estão classificadas em cada categoria exigem uma clarificação: ainda não juntámos casos flagrantes de um mesmo conceito com várias grafias, quer devido a aportuguesamento de nomes estrangeiros (*uísque* e *whisky*, *chabli* e *Chablis*, *gim* e *gin*, *pudim* e *pudding*), como a variações ortográficas (*fêvera* e *febra*, *taverna* e *taberna*, *toicinho* e *toucinho*, *sande* e *sanduíche*, *água-ardente* e *aguardente*), algumas delas diferentes em Portugal e no Brasil. Também até agora considerámos casos de diminutivos ou aumentativos como potenciais palavras diferentes (*feijão-frade* e *feijão-fradinho*, *panela* e *panelão*), e aceitámos como unidades lexicais alguns casos de unidades multi-palavras, sobretudo quando tinham uma grafia alternativa com hífen, como *café da manhã* e *café-da-manhã*, *pé de moleque* e *pé-de-moleque*. Tudo isto terá de ser objeto de regras claras, numa próxima fase.

3 Desambiguação em contexto

Todas as palavras marcadas como potencialmente representando comida ou bebida têm de ser de-

sambiguadas para se obter uma panorâmica deste campo semântico na literatura. Esse trabalho, em progresso, é feito usando regras automáticas de disambiguação escritas pelos autores, usando a filosofia de colaboração humano-máquina descrita em Santos e Mota (2010).

Aqui vamos apresentar a desambiguação da palavra *café*, palavra interessante não só por ser muito frequente, mas por ter muitas acepções que interessa distinguir, três das quais no campo semântico em que estamos interessados.

Vejam-se exemplos de usos distintos da palavra *café* em obras literárias:

planta *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.* (Franklin Távora, *O Matuto*, 1888)

bebida *Pois agora, colegas, disse o abade sorvendo o último gole de **café**, o que está a calhar é um passeio à fazenda.* (Eça de Queirós, *O crime do Padre Amaro*, 1875)

parte de refeição *Rosa aparecia ao **café**, exalando do seu sorriso, ...* (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.* (Machado de Assis, *Quincas Borba*, 1891)

loja *Dos **cafés** em geral, e de como são característicos da civilização de um país* (Almeida Garrett, *Viagens na minha terra*, 1846)

⁶Um estudo preliminar dessa palavra na literatura portuguesa e brasileira (Santos, 2024) apontou para apenas 7,7% casos de *água* como bebida.

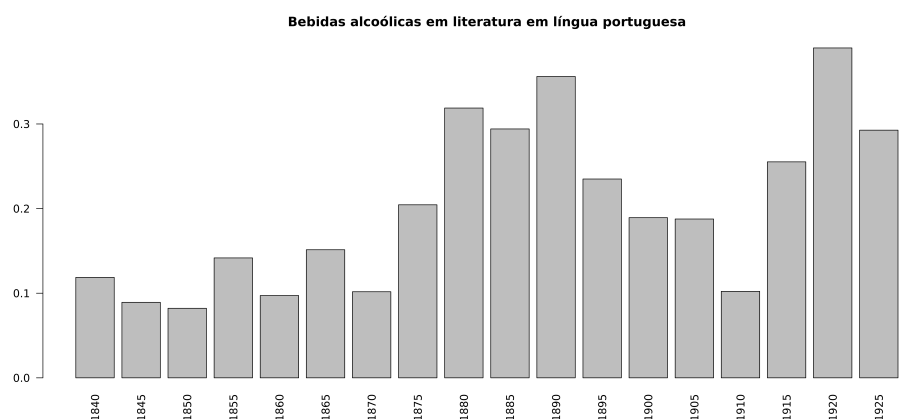


Figura 2: A frequência relativa das palavras referentes a bebidas alcoólicas, vezes 1000

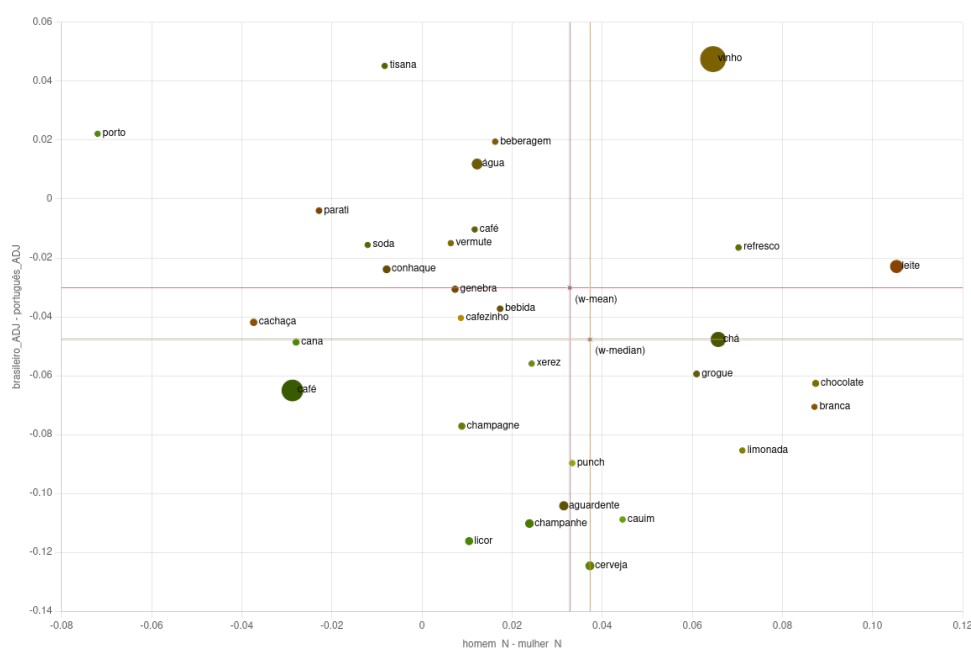


Figura 3: Bebidas, no eixo dos XX homens vs. mulheres, no eixo dos YY Brasil e Portugal

produto *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* (António Pedro Lopes de Mendonça, *Memórias dum doido*, 1849)

Os casos de *café* presentes nas obras estudadas podem ser resumidos na Tabela 1.

Além disso, podemos observar diferenças entre a literatura portuguesa e brasileira, simplesmente dividindo as acepções entre as duas literaturas. Em primeiro lugar, o lema *café* aparece relativamente quatro vezes mais na literatura brasileira. Por outro lado, torna-se imediatamente claro que a referência

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

Tabela 1: Classificação dos 1838 lemas *café*

aos cafés como loja é mais frequente em Portugal (apenas 13% na literatura brasileira) e que a referência ao café outros é mais frequente no Brasil (29%).

3.1 Visualização dos dados

Usando o AC/DC (Santos e Bick, 2000) para fazer procuras complexas, e tratando o resultado com a ajuda da linguagem/ambiente R (R Development

Obra	Data	Autor	Palavras	Capítulos	Refeições	Comida	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	

Tabela 2: Informação sobre as dez obras analisadas, representando um total de 77 refeições em 273 capítulos. AA - Aluísio Azevedo, EQ - Eça de Queirós, JD - Júlio Dinis. Comida relativa foi multiplicada por 1000, ou seja, representa o número de palavras de comida por mil palavras.

Core Team, 2008), podemos ver a distribuição das bebidas alcoólicas por períodos de cinco anos na Figura 2.

Usando o CorpusEye (Bick, 2005), um ambiente especialmente desenvolvido para visualizar grandes conjuntos de dados anotados, podemos apreciar a relação entre bebidas, género do autor e contexto brasileiro ou português, usando palavras pulverizadas ("word embeddings"), na Figura 3. Embora este processo de visualização não distinga os diferentes sentidos das palavras, podemos ver que o vinho é preferencialmente português e o café e a cerveja brasileiros. Mais inesperadamente, o vinho é mais feminino e o café aparece em contextos mais masculinos.

4 Identificação de refeições

Uma refeição num romance é uma cena a partir da qual um leitor pode inferir período, classe social, às vezes até religião, com base no que e como as pessoas comem, e em que lugares. E pode servir para caracterizar protagonistas e relações entre eles, além de ilustrar tipos de convívio, e normas de etiqueta de uma época ou classe.

De acordo com Brown (1984, nossa tradução), "Refeições na ficção são acima de tudo signos literários: por isso, são sujeitos ao mesmo tipo de análise que outros fenómenos literários". McGee (2001) insiste na importância na ficção escrita por mulheres em inglês no princípio do século XX dos jantares ("dinner"), que, segundo ela, são a refeição mais social e aquela à qual estão associados mais rituais e expectativas. De acordo com esta autora, a reputação de uma mulher, na época,

dependia geralmente das suas qualidades de cozinheira, ou, em estratos sociais mais elevados, da capacidade de organizar reuniões sociais à volta da comida.

Começamos por anotar manualmente as refeições (ou cenas à volta da comida) em dez romances em português, descritos na Tabela 2, para observar a relação entre o número de palavras do campo semântico comida e bebida e a existência de uma cena (refeição, preparação) envolvendo comida.

Na Figura 4, mostramos a densidade de palavras de comida e bebida por refeição por autor.

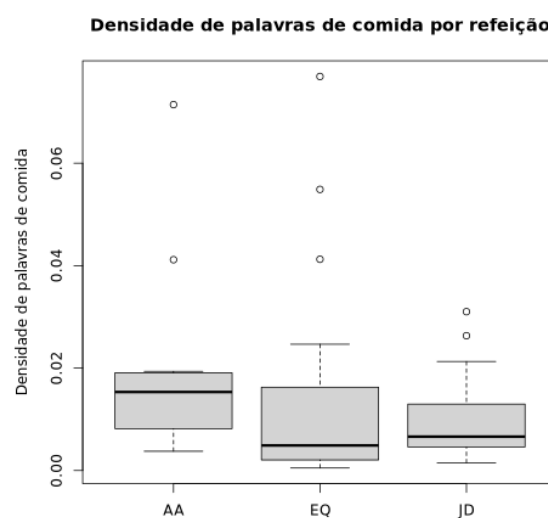


Figura 4: Quantas palavras do campo semântico comida e bebida por refeição, para cada autor

Nas Figuras 5 e 6, os gráficos do lado esquerdo mostram o tamanho das refeições na obra, por capítulo, enquanto que os gráficos do lado direito



Figura 5: Tamanho das refeições em número de palavras, comparado com o tamanho dos capítulos, e palavras de comida por capítulo e por refeição, em cinco obras de Júlio Dinis



Figura 6: Tamanho das refeições em número de palavras, comparado com o tamanho dos capítulos, e palavras de comida por capítulo e por refeição, em duas obras de Aluísio Azevedo and duas obras de Eça de Queirós

ilustram as palavras de comida e bebida por capítulo (a verde) e dentro de uma refeição (a roxo).

Estes dados mostram que uma densidade alta de palavras de comida e/ou bebida não é (não pode ser) o único critério para identificar uma refeição. Pelo contrário, até vemos que quanto maior (mais longa) é uma refeição literária (e, portanto, provavelmente mais importante para o enredo), menor é a densidade de palavras de comida nesse episódio. Além disso, vemos que existem muitas referências a palavras associadas ao campo de comida fora de refeições.⁷ Pensamos explorar outras alternativas (ou em conjunção com a densidade) para identificar refeições literárias (Santos et al., 2026b):

- Procurar pistas sobre o início ou fim de uma refeição, por exemplo na forma de convite, ou indicação de que "o jantar está servido", ou de "levantar-se da mesa", como forma de sinalizar o começo ou fim de uma refeição. (Contudo, na maior parte das vezes estes sinais estão ausentes do texto, que muitas vezes "começa" já no meio de uma refeição.)
- Procurar pistas sobre sequências conhecidas. Uma refeição pode começar por um aperitivo ou terminar no café. Se isso for explicitado, podemos concluir que estamos dentro de uma refeição, e não uma simples menção de comida no texto.
- Proceder parágrafo a parágrafo, começando pelos parágrafos com alta densidade, e tentar juntá-los aos precedents ou seguintes se não houver mudanças radicais, por exemplo nos intervenientes ou no tempo.
- Usar os próprios nomes das refeições, se aparecerem na obra.

Outro elemento importante deste projeto é, além da identificação dos episódios associados a comida, a sua caracterização, para permitir posteriormente uma leitura distante.

Como tal, contamos associar, a cada episódio, várias características. Algumas objetivas, que reputamos serem possíveis de automatizar mais tarde, como

- número de participantes
- tempo do repasto/preparação

- nome da refeição (se mencionado)
- local (dentro ou fora de casa, num espaço privado ou público (café, taberna, restaurante...))
- tipo: preparação, refeição, brinde
- presença de criados (que servem)
- discurso direto durante a refeição

e outras, mais complexas mas provavelmente as mais relevantes de um ponto de vista literário, como

- objetivo (literário) da refeição no enredo (por exemplo, encontro de personagens, segredos revelados, alteração)
- objetivos do autor (caracterização das personagens, caracterização do local, do tempo, da classe social)

Em relação a este tipo de informação sobre as cenas, só depois de reunir um extenso grupo de casos anotados por peritos de literatura (e com suficiente consenso) poderemos avaliar se é possível obter uma semi-automatização desta tarefa,

5 Observações finais

Pensamos que este recurso – corpos anotados e desambiguados sobre o tema da comida e bebida – pode servir para estudar essa parte da cultura. Porque como diz Eagleton (1997, página 25) em *Edible écriture*, "se há algo garantido sobre comida, é que não é nunca apenas comida"⁸.

Este recurso será útil não só em termos lexicais, mas também para detetar metáforas associadas, para compreender o uso das palavras de comida e bebida em texto literário: caracterização económica, psicológica, histórico-social, assim como para a identificação de episódios no enredo (refeições ou cenas de preparação de comida), e qual o seu objetivo.

Neste artigo, descrevemos a identificação manual de refeições em textos literários em português no espírito da leitura distante e propusemos a sua automatização através de várias estratégias diferentes, nomeadamente: concentração de palavras referentes ao domínio da alimentação; identificação de sequências naturais numa refeição; e referência à própria refeição, além de debatermos como caracterizar as próprias refeições em texto literário.

⁷Embora tenhamos de reconhecer que estes dados ainda não são baseados numa completa desambiguação, mas sim de apenas 70% dos casos.

⁸"If there is one sure thing about food, it is that it is never just food".

Referências

- Leonard Barkan. 2021. *The Hungry Eye: Eating, Drinking and European Culture from Rome to the Renaissance*. Princeton University Press.
- Eckhard Bick. 2005. CorpusEye: Et brugervenligt web-interface for grammatisk opmærkede korpora. Em *10. Møde om Udforskningen af Dansk Sprog 7.-8.okt.2004, Proceedings*, páginas 46–57. Århus University.
- Charlotte Boyce e Joan Fitzpatrick. 2017. *A History of Food in Literature From the Fourteenth Century to the Present*. Routledge.
- James W. Brown. 1984. *Fictional Meals and their Function in the French novel: 1789-1848*. Univ. of Toronto Press.
- J Michelle Coghlan. 2020. *The Literature of Food*. Cambridge University Press.
- Terry Eagleton. 1997. Edible écriture. *The Times*, Oct. 24, 1997.
- Michiel Korthals. 2008. Food as a Source and Target of Metaphors: Inclusion and Exclusion of Foodstuffs and Persons through Metaphors. *Configurations*, 16:77–92.
- Diane McGee. 2001. *Introduction: A Time to Eat*. In *Writing the Meal: Dinner in the Fiction of Twentieth-Century Women Writers*. University of Toronto Press.
- Anna Meigs. 1997. Food as cultural construction. Em *Food and culture: a Reader*, páginas 95–106. Routledge.
- Franco Moretti. 2013. *Distant Reading*. Verso.
- R Development Core Team. 2008. R: A language and environment for statistical computing.
- Diana Santos. 2014. *Corpora at Linguatca: Vision and Roads Taken*, páginas 219–236. Bloomsbury.
- Diana Santos. 2019. Literature studies in literateca: between digital humanities and corpus linguistics. Em *Humanists and the digital toolbox: In honour of Christian-Emil Smith Ore*, páginas 89–109. Novus forlag.
- Diana Santos. 2024. [Distant reading of food and drink in three languages](#).
- Diana Santos e Eckhard Bick. 2000. Providing Internet access to Portuguese corpora: the AC/DC project. Em *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC 2000)*, páginas 205–210.
- Diana Santos e Cláudia Freitas. 2019. [Estudando personagens na literatura lusófona](#). Em *STIL - Symposium in Information and Human Language Technology*.
- Diana Santos, Elizaveta Khachatryan, Michael Preminger, Åse Kristine Tveit, e Eckhard Bick. Presenting Reading Food and its infrastructure. *Digital Humanities in the Nordic and Baltic Countries Publications*.
- Diana Santos, Luisa Lima, e Emanuel Pires. 2026a. Marcação de correferência para a caracterização de personagens em obras literárias em português. Em *Proceedings of PROPOR 2026*.
- Diana Santos e Cristina Mota. 2010. Experiments in human-computer cooperation for the semantic annotation of Portuguese corpora. Em *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2010)*, páginas 1437–1444. European Language Resources Association.
- Diana Santos, Michael Preminger, Åse Kristine Tveit, e Elizaveta Khachatryan. 2026b. Identifying food episodes in literary texts. In preparation.
- Roger C. Schank e Robert P. Abelson. 1977. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Routledge.
- Christof Schöch, Tomaz Erjavec, Roxana Patras, e Diana Santos. 2021. [Creating the European Literary Text Collection \(ELTeC\): Challenges and Perspectives](#). *Modern Languages Open*, 1:1–19.
- Véronique Traverso e Loubna Dimachki. 2017. “ktîr Tajjibe ce plat! fî garlic?”: Compliments and assessments in French and Lebanese dinner talk. *Intercultural Pragmatics*, 4(2):137–163.