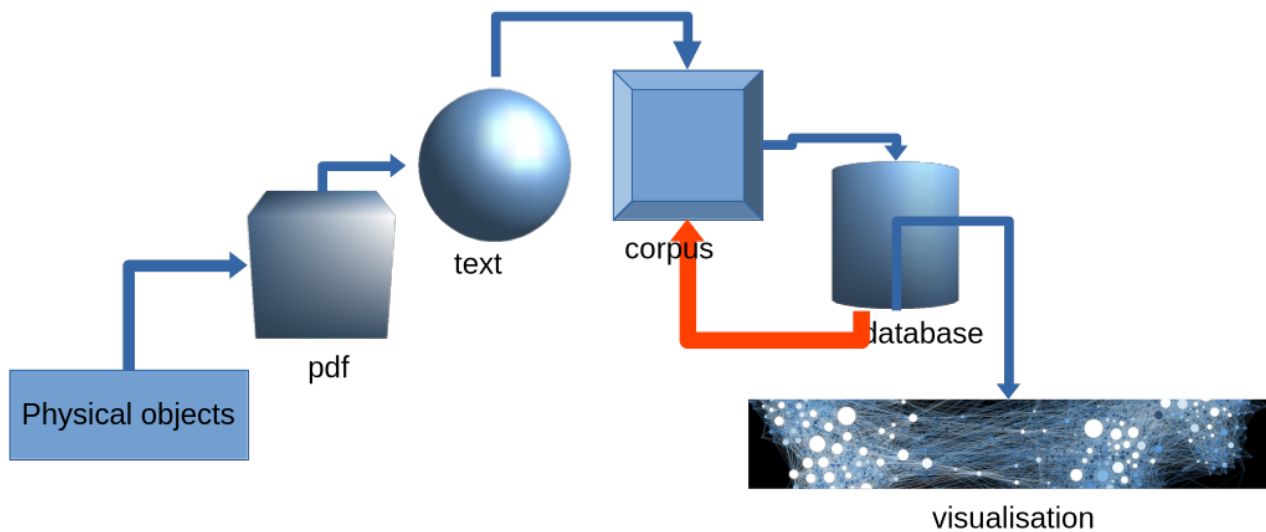


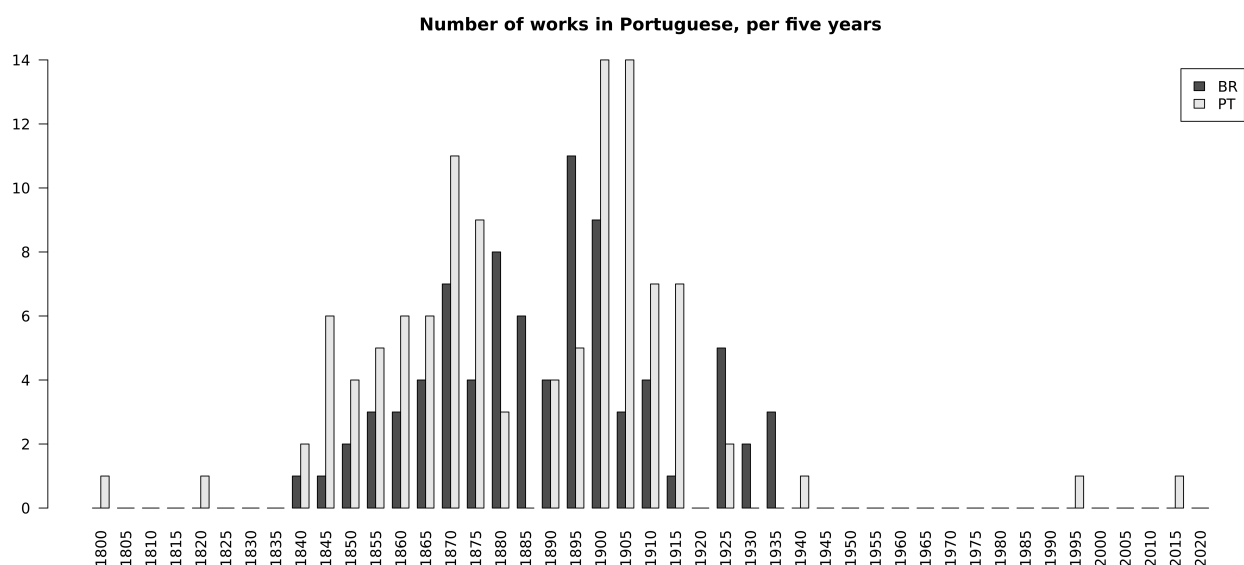
OR: Annotating food in (Portuguese) literature

Different phases and objects in DH



What Portuguese literature?

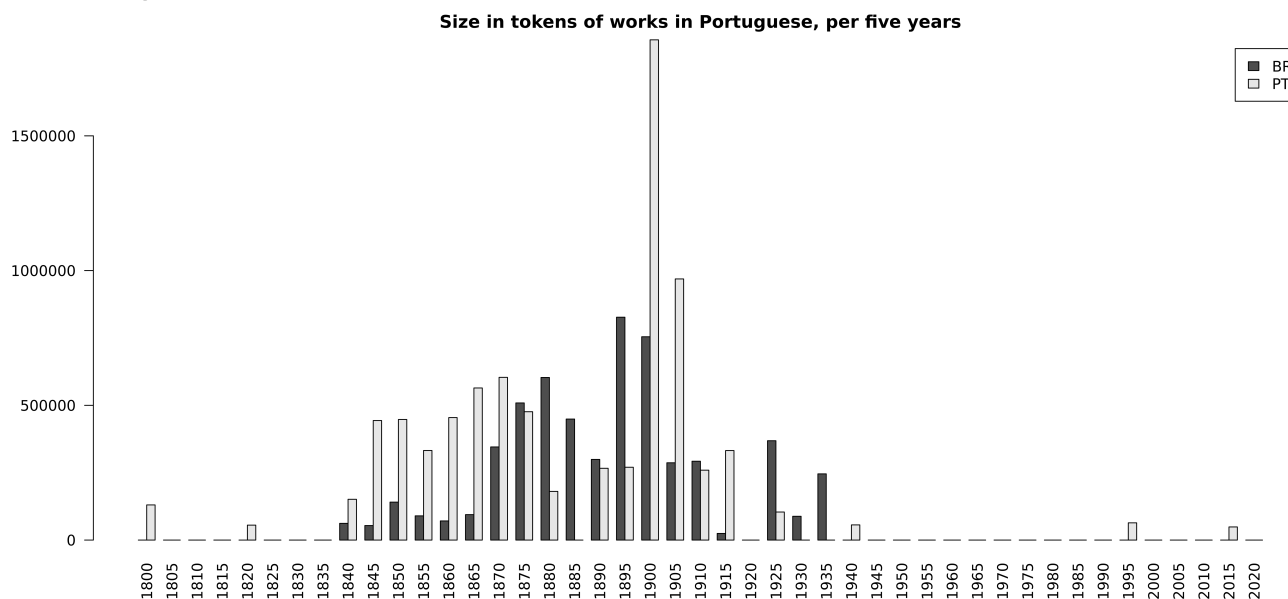
The corpus: Literateca from 1800 to now



199 novels or novellas from Portugal, 133 from Brazil, plus 6 excerpts from Portugal and 11 from Brazil, and 3 others (AN, CV, MZ)

How many tokens

The corpus: Literateca from 1800 to now



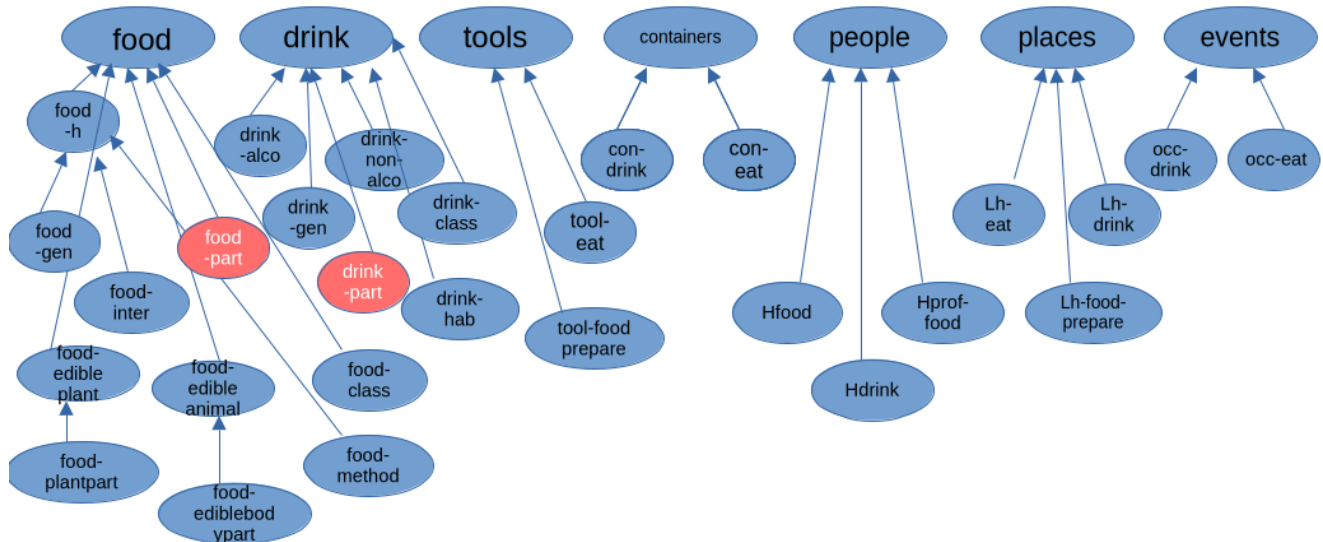
13,829,196 tokens from Portugal, 8,227,842 from Brazil, **22,070,645** in total.

The food and drink domain

- A semantic field present in (all) languages, for example according to Wierzbicka, but at least in Portuguese.
- We have annotated several (more dignified) semantic domains in AC/DC, such as colour, body, emotion, family, health...
- Why? Basically to compare cultures
- But it gives us lexical information, terminological information, and metaphors, which noone can say are outside linguistics

An empirical study

Assuming the following (initial) taxonomy, we classified words and expressions belonging to the F&D domain. Also annotated properties of F&D items.



Problems

Cruse (2004: 179) on real-life taxonomies

All this makes it virtually impossible to construct a well-formed hierarchy from clothing terms. The appearance of chaos can be mitigated if we bear in mind the following points:

- (i) Neat hierarchies appear only if the perspective is kept constant; if this is not the case, cross-classification can occur.*
- (ii) Each perspective potentially yields a separate hierarchy.*
- (iii) Different hierarchies can intersect in various ways*
- (iv) [...] the elements in taxonomic hierarchies are not full lexical senses, but contextually circumscribed microsenses*

In addition, there is the thorny aspect of contrastive aspects

The taxonomies of different languages can differ not only in the names of the categories but also in which categories are recognized.

Cruse (2004: 180)

As Jurafsky puts it, different cultures have different grammars of food.

Which we tried **not** to include, but obviously borders are fuzzy sometimes

- Animals and plants
- Agriculture
- Business
- Health
- Religion
- Cooking (food preparation) and storing

Examples

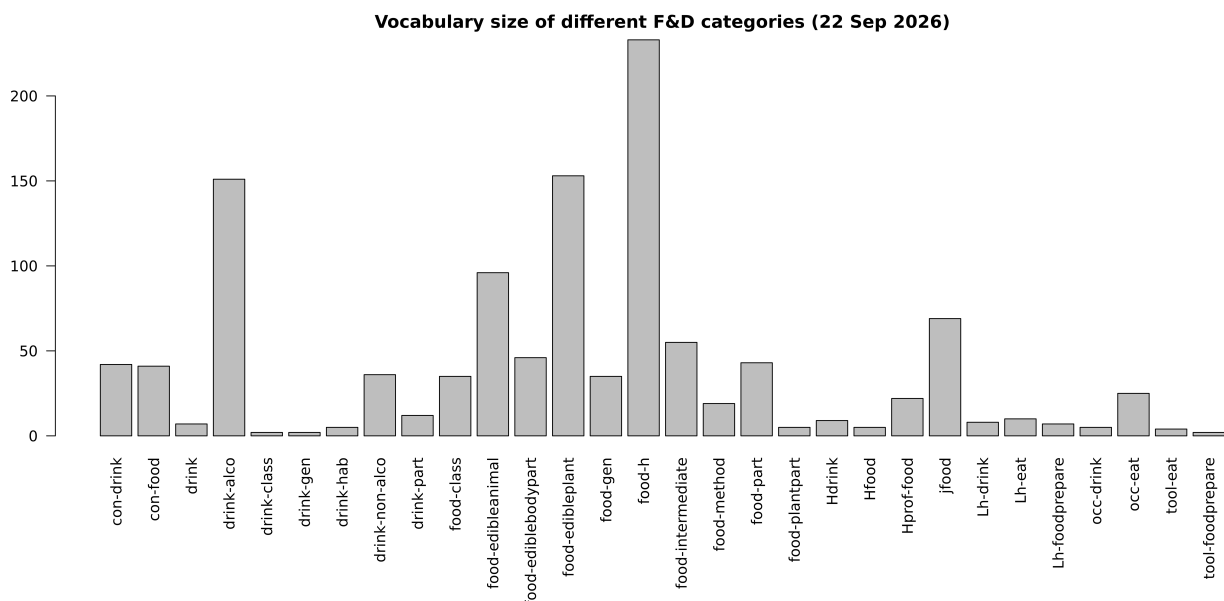
One needs to decide. Illustrating the arbitrariness of the classification:

- Is human milk food? No, except when used with tools (spoons).
*Até quasi fins de janeiro a criança sofreu ataques sucessivos, dez dias **alimentada com meia dúzia de colheres de leite**, adormecida sob a acção violenta dos calmantes (...)* (CarMDia-Filho-das-Ervas)
Until end of January the child suffered successive attacks, was for ten days **fed with half a dozen spoons of milk**, put violently into sleep by sedatives (...)
- Is soup a food or a drink? Drink only when the verb is *beber*, to drink.
*Bebia **uma colher de sopa**, erguia os olhos ao teto e suspirava profundamente.* (LuiGuiJ-Histórias-para-gente-alegre)
She drank **a spoon of soup**, looked up to the ceiling and sighed deeply.
- Whenever smell is discussed, it is not considered/annotated as food.
(For further work...) Documentation of the choices is essential!

Further examples of choices

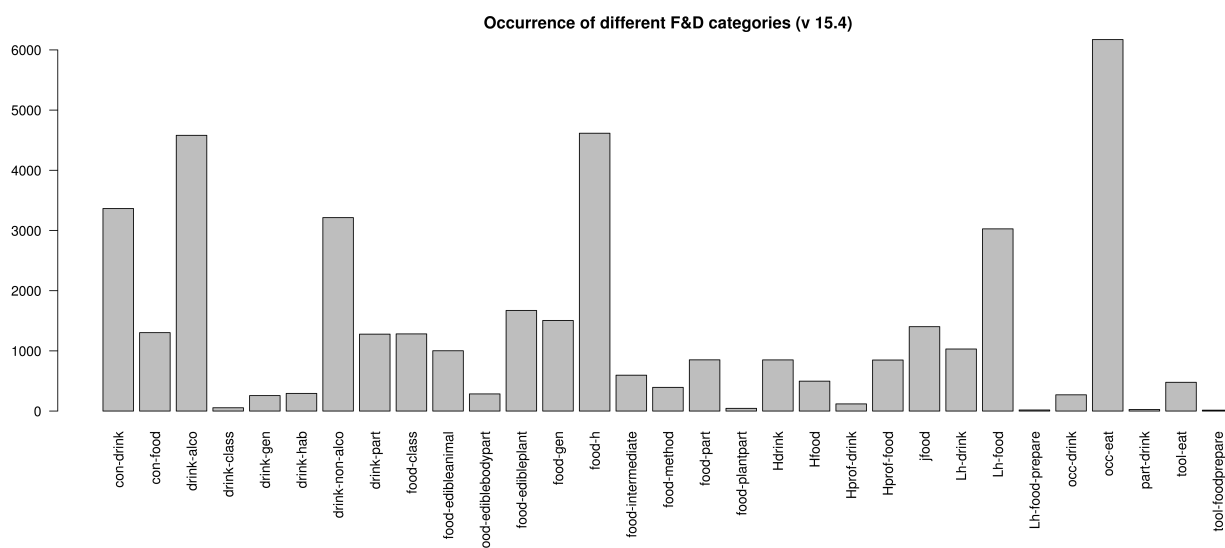
- *Arroz* (rice) is marked as food-edibleplant, although in most cases it could or should be interpreted as food-h
- *vinho do Porto* (Port wine) is marked as an alcoholic drink MWE, with *vinho* also marked as drink-alco, while *Porto* is marked as drink-alco
- *colher de chá* (tea spoon), *colher* is marked as tool-eat and *chá* is NOT marked as drink-non-alco
- *garrafa de vinho* (wine bottle), *garrafa* is marked as con-drink and *vinho* is NOT marked as drink-alco
- *garrafa de vinho* (bottle of wine), *garrafa* is marked as con-drink and *vinho* is marked as drink-alco
- No marking whatsoever in
 [...] e Simão Peixoto apostou perder cinquenta **garrafas de vinho** de 1770 (CCB-A-Doida-do-Candal)
 (And S P made a bet that he would lose fifty **bottles of wine** from 1770)

The F&D data in a nutshell



See the data (1154 terms) in
<https://www.linguateca.pt/Gramateca/ComidaBebida.html>

The F&D data in a nutshell II: tokens



Currently 40,030, corresponding to 1082 different lemmas.

Work involved

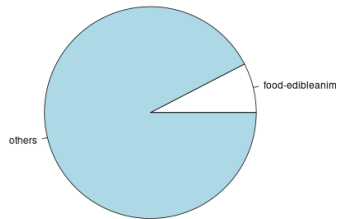
To give an idea of the human / manual work involved, all occurrences had to be inspected and classified. For example

- 1304 occurrences of lemma *copo* (glass, hilt, ...)
- 1893 occurrences of lemma *café* (coffee, coffee house, ...)
- 460 occurrences of lemma *cálice* (a special kind of cup, also used metaphorically, ...)
- 2109 occurrences of lemma *carne* (meat, flesh, ...)
- 1627 occurrences of lemma *pão* (bread, also used metaphorically, ...)
- 2275 occurrences of lemma *Porto* (port, Porto, ...)

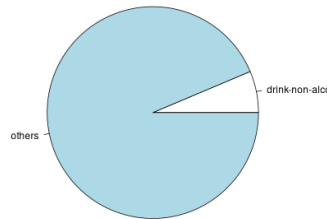
More than 128,000 occurrences manually investigated, giving rise to ca. 40,500 cases of food and drink annotated occurrences, and to ca. 9,800 VISLCG rules

Some examples of distribution of words

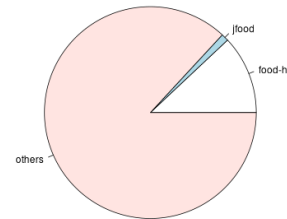
vaca (369) (cow)



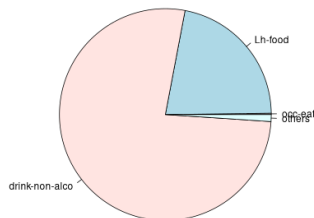
água (6418) (water)



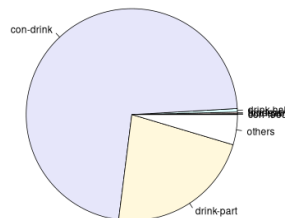
doce (3232) (sweet)



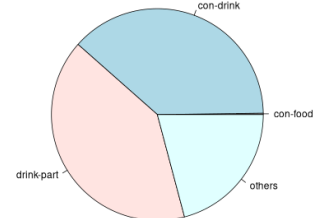
café (1894) (coffee)



copo (1304) (glass)



cálice (460) (chalice)



Regular polysemy in the food and drink domain

Identify cases where classes of lexical items can mean different things depending on context:

- drink amount (part) vs. drink container: *copo*, *escudela*, *garrafa*, *malga*, *taça*, *chávena*, *cálice*, *xícara*
- food portion vs. food container: *prato*, *escudela*
- occasions of eating, and what is eaten at these occasions: *sobremesa*, *jantar*, *café*

E Péricles, desde a sopa de legumes até a goiabada, falou de fotografia – dos grandes progressos da arte, de uma objetiva que encomendará, de certas placas de sensibilidade prodigiosa, do futuro fotográfico do mundo – todo o progresso contido entre as quatro paredes negras de uma câmara escura (CN-Esfinge)

And P, **from the soup to the dessert**, talked about photography

- cooked animals and plants and living animals and plants

Irregular polysemy in the F&D domain, or simple unspecification

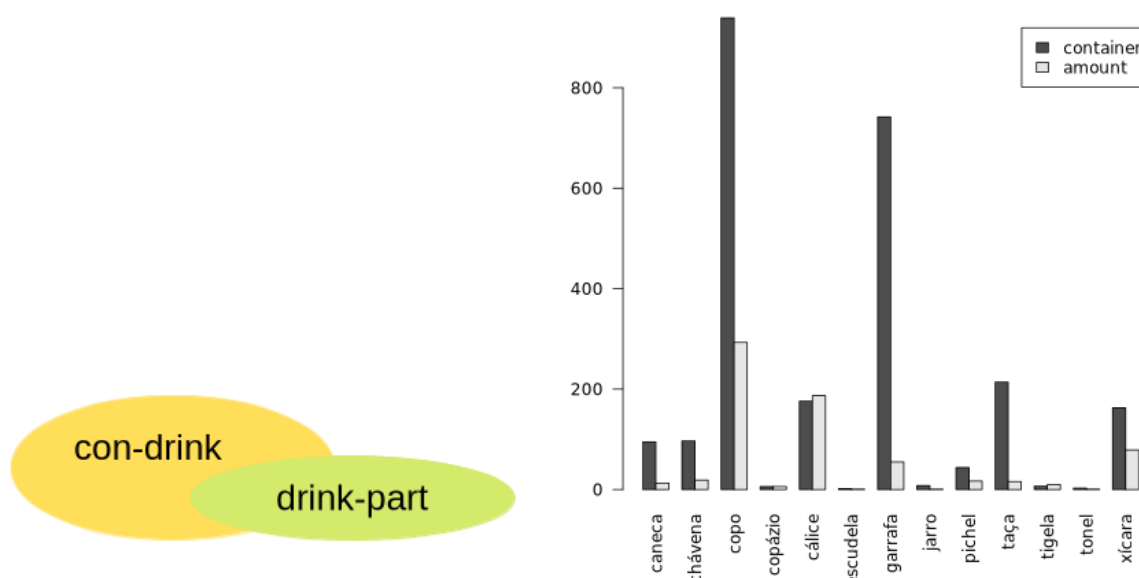
- some forms of presenting/cooking food can be salty or sweet, like *pastel* (pastry), *bolo* (cake), *torta* (pie)
- Some drinks can be alcoholic or not, like *gengibirra* (gingerbeer) or *cidra* (sider), or just unspecified about their contents, like *zurrapa* (bad tasting drink) or *refresco* (refreshment)
- Several foods (food-h) derive from other members of the food domain
 - foods: *lampreia de ovos*, *toucinho do céu*, *ovos de chocolate*, *salame de chocolate*
 - tools/containers: *frigideiras de Braga*, *covilhetes de picado*

Of course, there are hundreds of foods that are just named because of similar form, such as *cavacas*, *guardanapos*, *fios (de ovos)*, as well as objects that are named after food – usually plants, such as *coco* (coconut), from *chapéu de coco* (bowler hat)

Navigation icons: back, forward, search, etc.

Regular polysemy: findings

One can assess the extent of it quantitatively, and even try to identify clues for choosing the right sense. We found 13 cases in 56 where the same word did double duty:



Navigation icons: back, forward, search, etc.

Using LDOCE as the basis for populating an LKB (lexical knowledge base), Kilgarriff notes i.a.

gin (a glass of) a colourless strong alcoholic drink...

martini (a glass of) an alcoholic drink...

*In the dictionary the glass-of sense of **bourbon** is not stated because it is predictable (and of low frequency).* (Kilgarriff, 1995:322)

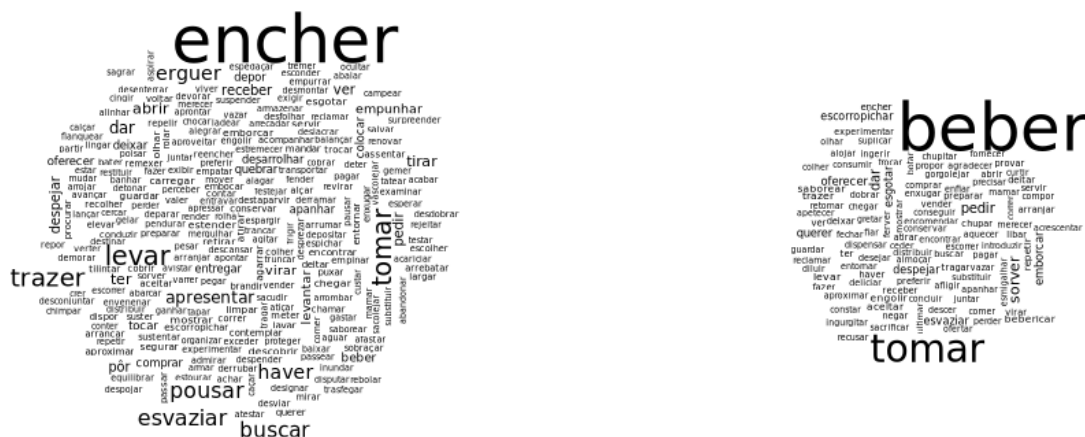
This raises a relevant question: How much one can/should predict (as in an LKB), or, alternatively, require empirical evidence (as lexicographers apparently do)?

Regular polysemy: verbs associated with the two senses

“Associated” here means having as direct object after the verb, automatically computed by PALAVRAS

con-drink (1249) (out of 3364)

drink-part (757) (out of 1276)



Regular polysemy: verbs associated with two food senses

con-food (370) (out of 1303)

food-part (397) (out of 850)



Diana Santos (UiO)

29/9/2026 21 / 46

21 / 46

Pustejowsky's qualia

When confronted with real corpus data, one can see at once how complex the clustering procedure is and how corpus investigation cannot be conducted successfully without an appropriate architecture of the lexicon as a base. (Pustejowsky & Jesek, 2008)

(5) $\left[\begin{array}{l} \text{sandwich}(x) \\ \text{CONST} = \{\text{bread}, \dots\} \\ \text{FORMAL} = \text{physform}(x) \\ \text{TELIC} = \text{eat}(P, w, x) \\ \text{AGENTIVE} = \text{make_activity}(z, x) \end{array} \right]$

A better way to formalize/explain the regular polysemy described above is to postulate that words like *prato* (dish) are complex types, and then specific instances are dot exploitation.

- 1 recipe
- 2 container
- 3 amount

This solves the equally natural use of the word in these meanings, as well as the sentences where more than one can be invoked (co-predication).

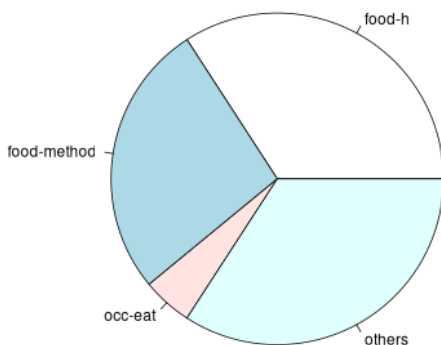
Diana Santos (UiO)

29/9/2026 22 / 46

22 / 46

Derivational morphology

- container+*ada* = large part measured by container: *colherada* (23), *garfada* (27), *pratada* (6), *gamelada* (1), *panelada* (0)
- past participle of a cooking method: *assado* (186), *frito* (35), *guisado* (34), *cozido* (22), *churrasco* (21), *ensopado* (21), *requeentado* (9), *picado* (5), *moqueado* (4), *corado* (3), *cristalizado* (3), *estufado* (2), *estrugido* (1), *refogado* (0)
 - can be a method, usually describing the way the dish is brought about
 - or the dish itself (regular polysemy)
 - or the occasion in the meal that it is eaten: *cozido* (82)



Connection of cooking methods with cooking targets

- Which methods are most mentioned? What about *cozinhado* (cooked, prepared)? 8 cases N food-gen, 8 cases V
- Which ingredient/dish has the highest number of different methods mentioned? *galinha* (assada, ensopada, corada, cozida), *carne* (assada, cozida, ensopada), *peixe* (frito, cozido, moqueado, assado), *vitela* (assada, estufada), *frango* (assado, ensopado)...
- Or ask the opposite question: What is most fried?

peixe 8
linguado 5
sardinha 5
fígado 4
salpicão 3
batata 2
carapau 2
cigarra 2
bacalhau 1
faneca 1
presunto 1

Rosch's basic level categories

The basic level of classification, the primary level at which cuts are made in the environment, appears to result from the combination of these two principles; the basic categorization is the most general and inclusive level at which categories can delineate real-world correlational structures. (Rosch et al., 1976, p. 384)

basic category cuts in the world should be those which yield the most information for the least cognitive load. Category cuts should provide the most inclusive categories which can follow the correlational structures perceived in the world. Basic objects are the categories at the level of abstraction for which the cue validity of categories is maximized. (Rosch et al., 1976, p. 428)

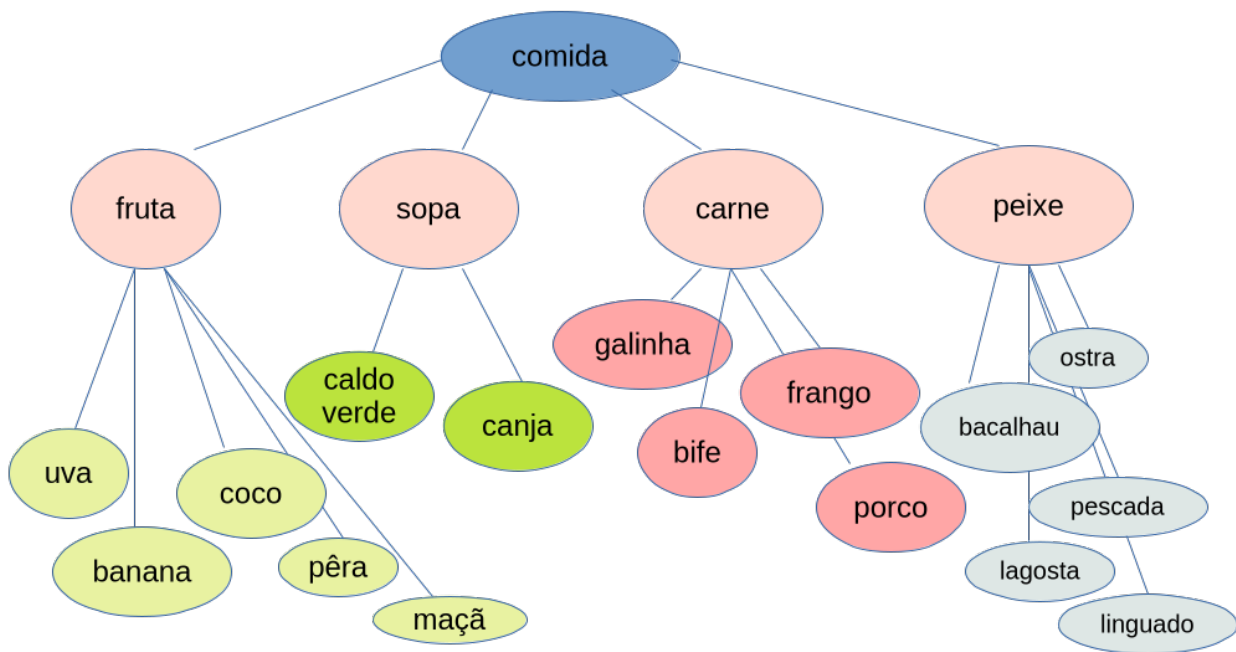
Given 9 taxonomies, subjects were required to list attributes of objects... Categories of concrete nouns in common use in English: frequency of ten or more in the Brown corpus... but removing the category “if the superordinate cross-cut a large number of other taxonomic structures (e.g., *food*)”. (Rosch et al., 1976, p. 387-388)

So, are there basic categories in the food domain?

Just restraining to food-h, food-edible*, food-class and food-gen,
and using token frequency.

<i>fruta</i>	fruit	505	<i>pão</i>	bread	121
<i>comida</i>	food	489	<i>prato</i>	dish	114
<i>alimento</i>	food	423	<i>bife</i>	beefsteak	103
<i>caldo</i>	soup, potage	409	<i>peixe</i>	fish	97
<i>doce</i>	sweet	388	<i>bacalhau</i>	codfish	95
<i>sopa</i>	soup	346	<i>queijo</i>	cheese	94
<i>mel</i>	honey	299	<i>torrada</i>	toast (bread)	79
<i>ovo</i>	egg	241	<i>frango</i>	young chicken	75
<i>bolo</i>	cake	207	<i>presunto</i>	ham	69
<i>arroz</i>	rice	202	<i>porco</i>	pork	64
<i>iguaria</i>	good food	147	<i>manjar</i>	food	61
<i>carne</i>	meat	142	<i>batata</i>	potato	59
<i>feijão</i>	beans	140	<i>chouriço</i>	sausage	52
<i>galinha</i>	chicken	121	<i>laranja</i>	orange	52

Possible partial taxonomies?



Differences between Portugal and Brazil

Kilgarrieff (1996) criticized Hofland & Johansson (1982) comparison of American vs. British English, Brown/LOB comparison, because “high chi-square values are not necessarily associated with a pre-theoretical notion of distinctiveness”, and suggested using the Mann-Whitney rank test (or the Wilcoxon test). [Basically, highly frequent words have higher chi-square values.]

Creating 10 arbitrary disjoint subcorpora of 500,000 words for PT and BR, I applied the `wilcoxon.test()` in R to identify distinctiveness of some words when marked as belonging to the drink domain: (BR/PT: 8M/13M)

café	W 67, p = 0.01876	W 75, p = 0.01327	W 73, p = 0.02202	958/537
chá	W 21, p = 0.02881	W 25, p = 0.06301	W 24, p = 0.05243	263/624
vinho	W 26, p = 0.07526	W 30, p = 0.14310	W 36, p = 0.31500	602/1235
xícara	W 95, p = 0.00021	W 94, p = 0.00032	W 89, p = 0.00210	210/30
água	W 71, p = 0.12300	W 56, p = 0.68420	W 78, p = 0.03546	256/267
cerveja	W 77, p = 0.04326	W 83, p = 0.01150	W 79, p = 0.02881	184/96
copo	W 40, p = 0.48130	W 50, p = 1.00000	W 44, p = 0.68420	507/718
chávena	W 23, p = 0.04326	W 34, p = 0.24750	W 38, p = 0.39300	13/100

Similes and metaphors

- Cases where the concept is shared between domains: smell, colour, consistency, temperature, quantity
- Cases where the origin domain is food or drink: taste, digestibility, capacity to become drunk/elated, (generalized) hunger and thirst, literary creation
- Cases where the target domain is food (not found!)

Conventional vs. creative

- The former are important to teach as a foreign language
- The latter are relevant to understand authors and their mindset
- Of course there is a middle ground that is diachronic: creative turns of phrase that become ingrained

Examples of conventional idiomatic expressions

<i>a pão e laranjas</i>	on bread and oranges	almost nothing to eat
<i>a pão e água</i>	on bread and water	in prison
<i>às sopas de</i>	on soups of	economically dependent on
<i>tirar nabos da púcara</i>	remove turnips from the mug	find out
<i>cheirar a esturro</i>	smell burnt	something is wrong
<i>não ter papas na língua</i>	no porridge on the tongue	say things directly
<i>pão, pão, queijo, queijo</i>	bread, bread, cheese, cheese	says things directly
<i>favas contadas</i>	fava beans counted	inevitable, known thing
<i>pagar as favas</i>	pay the fava beans	get the burden
<i>mandar à fava</i>	send to the fava bean	dismiss
<i>pôr tudo em pratos limpos</i>	put everything in clean dishes	open up, not hide anything
<i>esperto como um alho</i>	clever as garlic	very clever
<i>em/de carne e osso</i>	flesh and bone	real
<i>mostarda ao/no nariz</i>	mustard in the nose	angry
<i>a taça transbordar</i>	the cup brimmed over	too much
<i>alhos * bugalhos</i>	garlic * gallnut	easy to confuse

Examples of conventional expressions or themes

- *Dá Deus nozes a quem não tem dentes!* = lit. God gives nuts to toothless people! = of a person who cannot profit from a situation
- *Por aí é que vai o gato às filhoses!* = lit. In that way/path comes the cat to eat the cakes = of something that explains a complex situation

Common metonymies

- *pão* (bread) as maintenance, everything one needs to live
- *talher* (cutlery) as place at table
- *carne* (flesh and meat) as body

Common metaphors

- enthusiasm as *embriaguez* (drunkenness)
- passion or love as wine or nectar
 - vinho acre e delicioso da paixão* (passion's acid and delicious wine)
(CarMDia-A-Mulata)
- victory is like an alcoholic drink
- revenge is like an alcoholic drink
 - deliciosos tragos daquela vingança* (delicious sips of that revenge)
(AG-O-Arco-de-Santana)

Examples of creative similes and metaphors

About human relationships

- *Isto de casamento é como a sopa; não se deixa esfriar* (marriage is like soup, one should not let it cool) (JosdAle-Senhora)
- *O amor é como um cacho de uvas: toda vez que a carne assoberba devora-lhe uma, até que só resta a vinha da amizade, se não lhe dá a filoxera do esquecimento, do tédio, da repulsa, da saciedade, da aversão* (love is like a bunch of grapes) (MdOP-A-afilhada)
- *para quem a mulher outra coisa não é senão uma espécie de fruto amargo que a gente prova e deita fora* (for whom woman is just a kind of sour fruit which one tastes and throws away) (AdoCam-A-Normalista)
- *porque a mulher é como a laranja: se muito a apertam logo amarga* (because woman is like an orange, if you squeeze it a lot it sours) (AntFBar-Um-duelo-nas-sombras)

Examples of creative similes and metaphors (cont'd)

Comparing other things to the result of spices

- *ceias amostardadas por dançarinas* (dinners mustarded by female dancers) (CCB-Carrasco-de-Vitor-Hugo)
- *remetendo o artigo, recomendando-lhe «que o apimentasse»* (sending back the article, suggesting “he peppered it”) (EQ-Os-Maias)
- *gracejos viperinos com que apimentava as conversações* (snaky jokes with which he peppered the conversation) (AntATVas-A-ermida-de-Castromino)
- *um leve acento adocicado do Brasil* (a slight sweetened Brazilian accent) (EQ-A-Tragédia-da-Rua-das-Flores)
- *podem adoçar o amargo da velhice* (they can sweeten the sourness of old age) (CamCB-O-olho-de-vidro)

Examples of creative similes and metaphors (cont'd 2)

To do with language, writing and thought (3 first) or social issues

- *é a canela com que os taquígrafos, umas vezes por ironia e outras por lisonja, polvilham a insípida aletria de certos discursos* (this is the cinammon with which stenographers, sometimes ironically, other times flattering, spread/powder the insipid “aletria” of some speeches.) (JdA-O-Garatuja)
- *as minhas ideias entouriam o estômago intelectual como se fossem castanhas cozidas* (my ideas would fill the intellectual stomach as if they were boiled chestnuts) (CCB-Coração-Cabeça-e-Estômago)
- *Embora travem á hervilhaca, é o que temos, e o que nos dão os Watteis dos fricassés litterarios* (although they taste vetch, that's what we have, and what we are given by the Watteis of the literary fricassés) (CamCBra-Carlota-Angela)
- *A melhor mostarda é a fome* (the best mustard is hunger) (CCB-A-Filha-do-Doutor-Negro)
- *onde a nobre carne de Eva se vende [...] como a de vaca* (where Eve's noble flesh is sold [...] like cow meat) (EQ-A-Cidade-e-as-Serras)

Cooccurrence: expressions vs. words

What is the lexicon? Of course spaces are not the only factor. So, how to decide that something is a lexical item?

- If it only occurs together: *maçã reineta*
- If it has a definitely separate meaning: *pão, pão, queijo, queijo*
- if it describes a coordination, an ingredient, an origin or brand: *café com leite, vinho do Porto, queijo Chester, vinho verde, caldo verde*
- if it varies with or without the “classifier” (showing that it is becoming lexicalized): *Madeira*
 - dois caixotes com **vinho velho da Madeira**
 - Como, sem Porto ou **Madeira**, sem Lisboa, sem Cartaxo, ousa um súdito britânico erguer a voz, naquela harmoniosa desafinação insular que lhe é própria e que faz parte do seu respeitável carácter nacional – faz; não se riam: o inglês não canta senão quando bebe... aliás quando está BEBIDO.
 - A moça então por sua conta e risco pediu uma salada de ananás cozido em **madeira**
- If it occurs significantly more often than random (collocation)?

Let us look at cooccurrences (with *de*)

<i>copo de vinho</i>	89	glass of wine
<i>xícara de café</i>	81	cup of coffee
<i>copo de água</i>	74	glass of water
<i>garrafa de champanhe/champagne</i>	27+17	bottle of champagne
<i>garrafa de vinho</i>	38	bottle of wine
<i>cálice de vinho</i>	34	chalice of wine
<i>cálice de conhaque/cognac</i>	26+6	chalice of cognac
<i>copo de cerveja</i>	32	glass of beer
<i>caldo de galinha</i>	32	chicken soup
<i>gole de café</i>	24	sip of coffee
<i>garrafa de conhaque/cognac</i>	14+9	bottle of cognac
<i>garrafa de cerveja</i>	22	bottle of beer
<i>copo de champanhe</i>	17	glass of champagne
<i>lombo de porco</i>	17	pork sirloin
<i>gole de vinho</i>	16	sip of wine
<i>xícara de chá</i>	15	cup of tea
<i>gole de chá</i>	14	sip of tea
<i>copo de aguardente</i>	14	glass of aquavit
<i>garrafa de aguardente</i>	14	bottle of aquavit
<i>cálice de genebra</i>	13	chalice of genebra
<i>gota de vinho</i>	13	drop of wine
<i>trago de vinho</i>	12	sip of wine

Surprisingly few multiword expressions (MWE)

NB! MWE are defined here as non-compositional

<i>sala de jantar</i>	293	dining room
<i>vinho do Porto</i>	63	Port wine
<i>vinho verde</i>	20	green wine
<i>caldo verde</i>	11	"green soup"
<i>vinho branco</i>	10	white wine
<i>couve galega</i>	9	Galician cabbage
<i>vinho da Madeira</i>	6	Madeira wine
<i>arroz doce</i>	6	"sweet rice"
<i>carnes frias</i>	5	cold meat dish
<i>trouxa de ovos</i>	5	"egg package"
<i>ovo quente</i>	4	"hot egg"
<i>vinho do Douro</i>	4	Douro wine
<i>queijo flamengo</i>	4	Flemish cheese
<i>ninhos de andorinha</i>	3	"swallow nest"
<i>carne verde</i>	3	"green meat"
<i>couve tronchuda</i>	3	"tronchuda" cabbage
<i>vinho de Bordéus</i>	2	Bordeaux wine
<i>arroz de forno</i>	2	oven rice
<i>doce de tijolo</i>	2	"brick sweet"

Navigation icons: back, forward, search, etc.

Diana Santos (UiO)

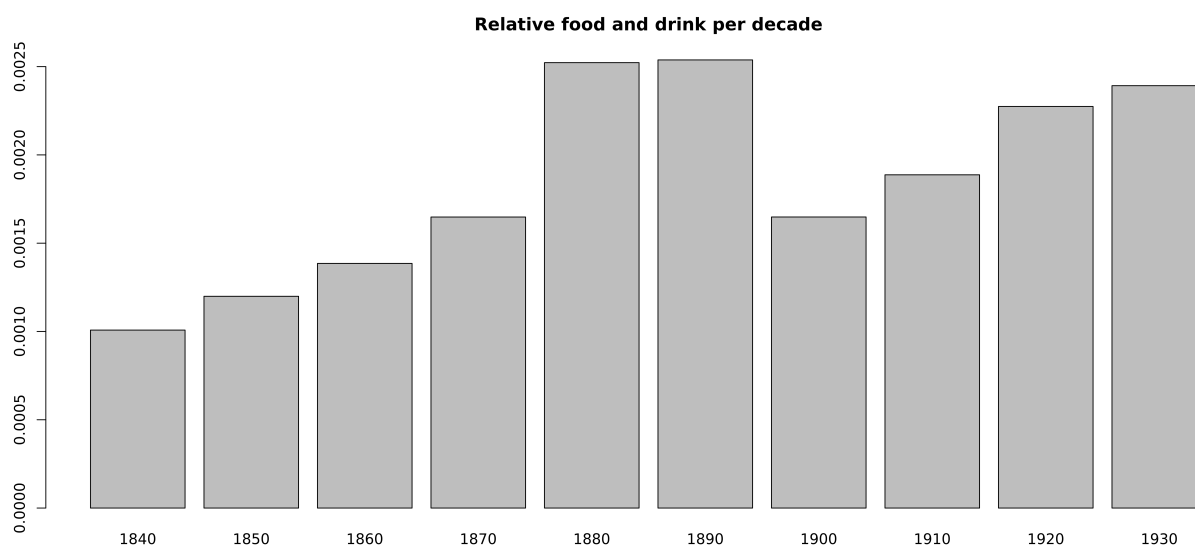
29/9/2026

37 / 46

<i>queijo londrino</i>	2	London cheese
<i>queijo suíço</i>	2	Swiss cheese

Time changes

Is there a difference of interest for food and drink throughout time?



It does not appear so.

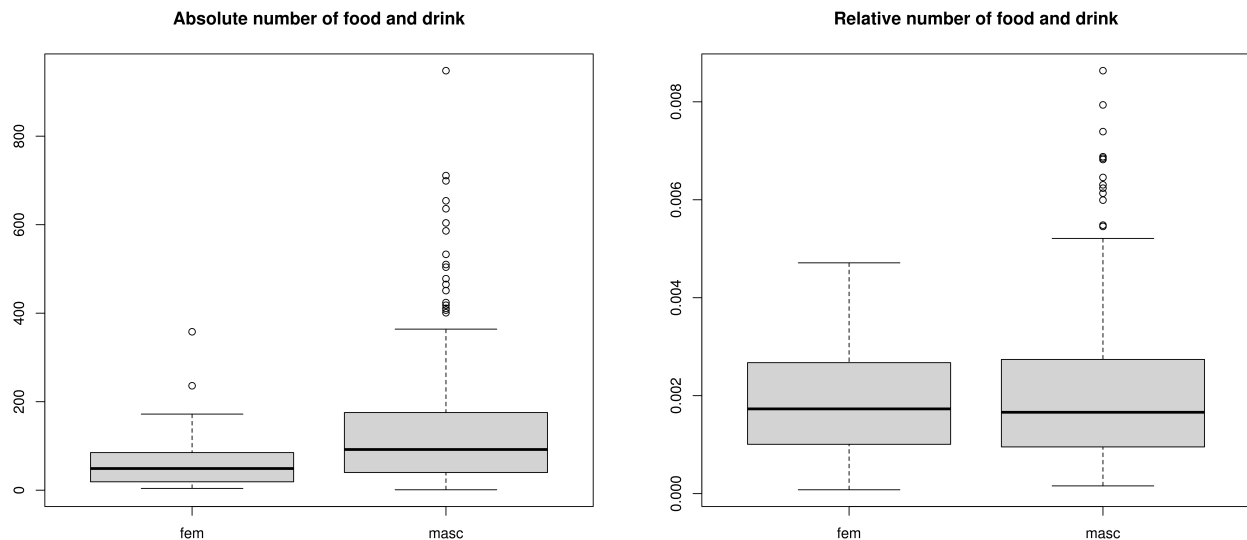
Navigation icons: back, forward, search, etc.

Diana Santos (UiO)

29/9/2026

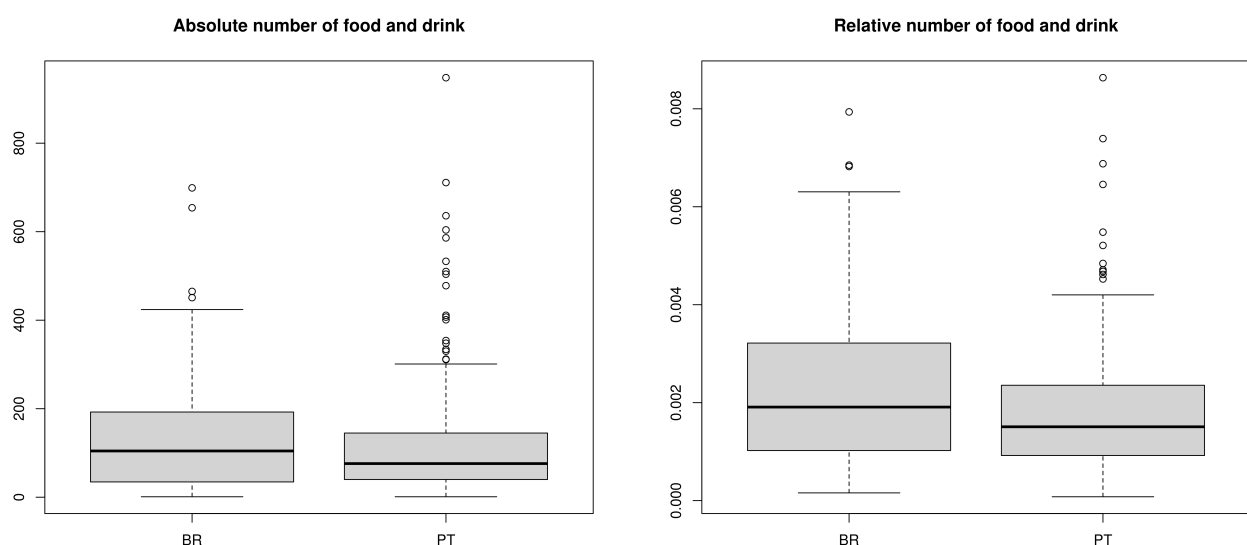
38 / 46

Food and gender



Absolutely no difference in the relative density of food and drink in women and men's writings (313 work by men and 24 by women; 162 authors, only 18 women): $t = -0.66981$, $df = 29.667$, $p\text{-value} = 0.5082$

Food and variety of Portuguese



Greater use of food and drink in Brazilian Portuguese: $t = 2.6504$, $df = 283.15$, $p\text{-value} = 0.008491$

Further work

- Add short stories (already in Literateca 221 works, 3,805,714 tokens)
- Add translated texts into Portuguese (6 works in Literateca, 562,653 tokens)
- Study the verb domain (frames are already automatically annotated, but human revision is required)
- Study several other cases of regular polysemy
- Systematically study differences between varieties, and through time
- Study idiomatic aspects of food, following Philip (2011), devising a systematic annotation of figurative uses (from similes to idioms)

In addition to the literary aspects: literary school, work, author, genre... and much more

Some accounting: how long does it take

- To do the first annotation (332 novels + 30 excerpts) it took approx. 3 months full-time.
- Added in September a new Portuguese historical novel: 3 days.
 - 207,062 new tokens
 - added 20 to 1311 food types (parrot, lizard, snake, monkey...)
 - added 359 food tokens

Request for feedback, criticism and suggestions

- Can this kind of work be useful?
- Something you would have done differently or additionally?
- Problems in the presentation or in the underlying assumptions?
- Forms of improving it significantly?

Thank you !

Frequency of edible animals

Total (337)		Brazilian (131)		Historical (46)	
galinha	121	frango	26	frango	9
bacalhau	95	peru	25	galinha	9
frango	75	galinha	23	papagaio	7
porco	63	pirarucu	21	carneiro	6
peru	45	porco	20	porco	6
leitão	40	bacalhau	18	coelho	5
vitela	39	leitão	18	leitão	5
ostra	37	ostra	13	boi	4
carneiro	36	<i>caça</i>	12	cabrito	4
boi	35	carneiro	11	peru	4
cabrito	30	lagosta	9	pescada	4
pescada	29	sardinha	8	sardinha	4
vaca	28	vaca	7	vaca	4
lagosta	25	boi	6	vitela	4
perdiz	25	pacuera	6	<i>caça</i>	3
sardinha	24	anta	5	caracol	3
pirarucu	22	tambaqui	4	lebre	3
<i>caça</i>	16	tucunaré	4	linguado	3
linguado	14	veado	4	pombo	3

Frequency of edible plants

arroz	175
feijão	140
batata	59
laranja	52
milho	39
limão	33
figo	32
azeitona	31
uva	29
banana	25
coco	24
couve	23
pêra	23
maçã	22
mandioca	21
cebola	20
castanha	19
ervilha	18
alface	17
fava	17
guaraná	17

References

- Cruse, Alan. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford University Press, 2nd ed., 2004.
- Jurafsky, Dan. *A linguist reads the menu*. W. W. Norton & Co, 2015.
- Kilgarriff, Adam. "Why chi-square doesn't work, and an improved LOB-Brown comparison". *ALLC-ACH Conference*, June 1996, Bergen.
- Kilgarriff, Adam. "Inheriting Polysemy". In Patrick St. Dizier and Evelyne Viegas (Eds.), *Computational Lexical Semantics*. Cambridge University Press, 1995, pp. 319-335.
- Philip, Gill. *Colouring Meaning: Collocation and connotation in figurative language*. John Benjamins Publishing Company, 2011.
- Pustejovsky, James & Elisabetta Jezek. "Semantic coercion in Language: Beyond Distributional Analysis". *Italian Journal of Linguistics*, 20, 1, 2008, pp. 181-214.
- Rosch, E., C.B. Mervis, W. Gray, D. Johnson & P. Boyes-Braem. "Basic Objects in Natural Categories", *Cognitive Psychology*, 8, 1976, pp. 382-439.