

Humanities Going Digital (HUGOD)

[2020-1-CZ01-KA226-HE-094363]

Organização de Dados
Sketch Engine

Sessão 11

07/11/2022

Chiara Barbero
Sílvia Barbosa



AULAS

Organização dos dados



07

08

Análise de Dados



09

10

Trabalho final escrito



11

12

Extração de dados



Dúvidas















Apresentação de trabalhos



DASHBOARD CORLEX - PT

CORLEX - PT

-  **Word Sketch**
Collocations and word combinations
-  **Thesaurus**
Synonyms and similar words
-  **Parallel Concordance**
Translation search
-  **N-grams**
Multiword expressions (MWEs)
-  **Trends**
Diachronic analysis, neologisms
-  **OneClick Dictionary**
Automatic dictionary drafting
-  **Word Sketch Difference**
Compare collocations of two words
-  **Concordance**
Examples of use in context
-  **Wordlist**
Frequency list
-  **Keywords**
Terminology extraction
-  **Text type analysis**
Statistics of the whole corpus
-  **Bilingual terms**
Bilingual terminology extraction



DASHBOARD portug × i 🔗

CORLEX - PT

- Word Sketch**
Collocations and word combin
- Thesaurus**
Synonyms and similar words
- Parallel Concordance**
Translation search
- N-grams**
Multiword expressions (MWEs)
- Trends**
Diachronic analysis, neologism
- OneClick Dictionary**
Automatic dictionary drafting

<input checked="" type="checkbox"/>	CORLEX - PT	Portuguese	6,609,678
<input type="checkbox"/>	G_2022	Portuguese	382,497
<input type="checkbox"/>	novo_vino	Portuguese	776,401
<input type="checkbox"/>	O_2022	Portuguese	549,738
<input type="checkbox"/>	teste	Portuguese	126,009
<input type="checkbox"/>	V_2022	Portuguese	46,449
<input type="checkbox"/>	VOG	Portuguese	704,772
<input type="checkbox"/>	Portuguese Web 2011 (ptTenTen11) ⌚	Portuguese	3,896,392,719
<input type="checkbox"/>	Araneum Portug allicum Maius [2015] 💰	Portuguese	862,134,902
<input type="checkbox"/>	Brazilian Portug uese corpus (Corpus Brasileiro) 💰	Portuguese	871,117,178

34 corpora

Show description

[ADVANCED SEARCH](#) [CREATE CORPUS](#)

5

Tarefa 01

Procurar *corpora* em português

- Quantos existem?
- A cada grupo será atribuído um *corpus*
- Observem as características do mesmo.
- Quantos *tokens*, *words*, documentos, frases?
- Quantos lemas?
- Quantos nomes, adjetivos, verbos, advérbios?
- Quais os 10 N, ADJ, V, ADV?

portug



Any language

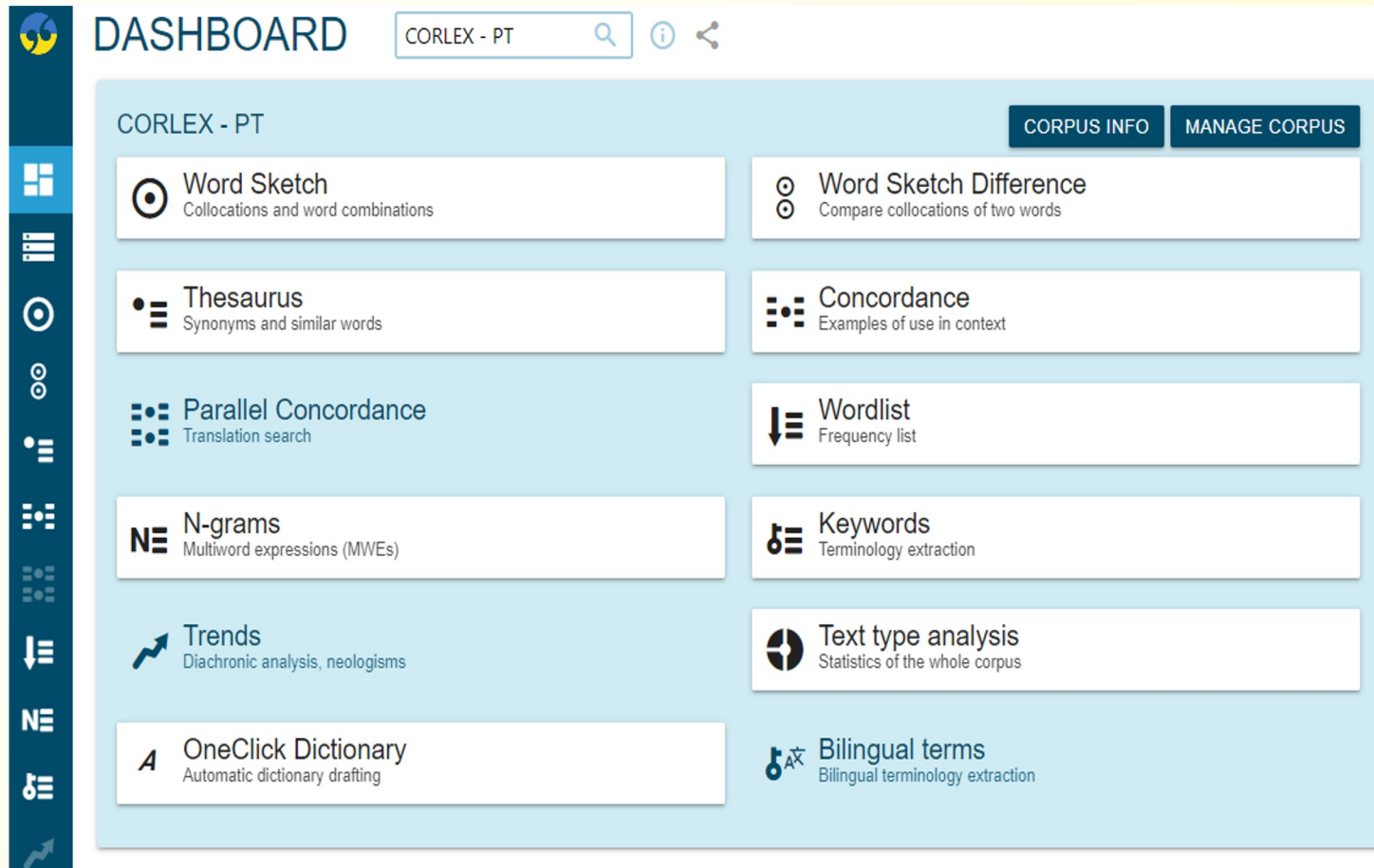


Only with word sketches



32 corpora 4 languages

Language	Name ↑	Words	
Portuguese	Araneum Portug allicum Maius [2015]	862,134,902	...
Portuguese	CHILDES Portug uese Corpus	216,407	...
Portuguese	DGT, Portug uese	53,950,705	...
Portuguese	ELEXIS Portug uese Web 2020	1,021,937,614	...
Portuguese	ELEXIS Portug uese Web 2020 (ptTenTen20) WSD sample	1,997,515	...
Portuguese	EUR-Lex judgments Portug uese 12/2016	35,412,936	...
Portuguese	EUR-Lex Portug uese 2/2016	595,066,701	...
Portuguese	EUROPARL7, Portug uese	53,778,766	...
Portuguese	Gutenberg Portug uese 2020	14,309,476	...
Portuguese	Newspapers in Portug uese (CetemPúblico, CetenFolha)	56,768,822	...
Portuguese	OpenSubtitles 2018 - Brazilian Portug uese	545,598,189	...
Portuguese	OpenSubtitles 2018 - Portug uese	466,021,603	...
Portuguese	OPUS2 Brazilian Portug uese	272,300,927	...
Portuguese	OPUS2 Portug uese	297,700,205	...
Portuguese	Portug uese Web 2011 (ptTenTen11, Palavras parsed)	2,757,635,105	...
Portuguese	Timestamped JSI web corpus 2014-2016 Portug uese	1,109,771,393	...
Portuguese	Timestamped JSI web corpus 2014-2021 Portug uese	4,685,199,909	...
Portuguese	Timestamped JSI web corpus 2021-03 Portug uese	108,540,406	...
Portuguese	Timestamped JSI web corpus 2021-04 Portug uese	85,486,841	...
Portuguese	CoPEP - The Corpus of Portug uese from Academic Journals (v. 1.4)	40,423,011	...



DASHBOARD CORLEX - PT

CORLEX - PT **CORPUS INFO** **MANAGE CORPUS**

- Word Sketch**
Collocations and word combinations
- Word Sketch Difference**
Compare collocations of two words
- Thesaurus**
Synonyms and similar words
- Concordance**
Examples of use in context
- Parallel Concordance**
Translation search
- Wordlist**
Frequency list
- N-grams**
Multiword expressions (MWEs)
- Keywords**
Terminology extraction
- Trends**
Diachronic analysis, neologisms
- Text type analysis**
Statistics of the whole corpus
- OneClick Dictionary**
Automatic dictionary drafting
- Bilingual terms**
Bilingual terminology extraction

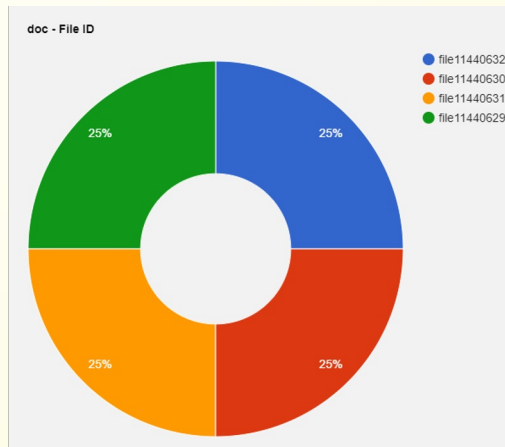
CORLEX

7

- Quantos *tokens*, *words*, documentos, frases?
- Quantos lemas?
- Quantos nomes, adjetivos, verbos, advérbios?
- Quais os 10 N, ADJ, V, ADV?

COUNTS i

Tokens	7,912,295
Words	6,609,678
Sentences	382,338
Documents	4



WORDLIST

CORLEX - PT



lemma (105,735 items | 7,912,295 total frequency)

WORDLIST

CORLEX - PT



noun (71,313 items | 1,851,728 total frequency)

WORDLIST

CORLEX - PT



adjective (16,496 items | 464,470 total frequency)

WORDLIST

CORLEX - PT



verb (12,268 items | 1,127,551 total frequency)

WORDLIST

CORLEX - PT



adverb (3,620 items | 415,224 total frequency)

8

Procurar o lema “humanidade”. Ocorre no *corpus*?

Procurar as *word* que começam/ que contém/terminam com “digital”

DASHBOARD CORLEX - PT

CORLEX - PT CORPUS INFO MANAGE CORPUS

- Word Sketch: Collocations and word combinations
- Word Sketch Difference: Compare collocations of two words
- Thesaurus: Synonyms and similar words
- Concordance: Examples of use in context
- Parallel Concordance: Translation search
- Wordlist: Frequency list
- N-grams: Multword expressions (MWEs)
- Keywords: Terminology extraction
- Trends: Diachron analysis, neologisms
- Text type analysis: Statistics of the whole corpus
- OneClick Dictionary: Automatic dictionary drafting
- Bilingual terms: Bilingual terminology extraction

BASIC ADVANCED ABOUT

find ?

- words
- lemmas
- adjective
- adverb
- conjunction
- noun
- numeral
- preposition

- all
- starting with
- ending with
- containing
- matching regex
- from this list:

	Word	Frequency ? ↓
1	humanidades	7 ...
2	deshumanidades	1 ...

	Lemma	Frequency ? ↓
1	digital	74 ...
2	digitalização	6 ...
3	digitalis	4 ...
4	digitalizar	3 ...

	Word	Frequency ? ↓
1	digital	30 ...
2	digitalização	6 ...
3	digitalis	4 ...
4	digitalizadas	1 ...
5	digitalizada	1 ...
6	digitalizados	1 ...
7	analógico-digital	1 ...

	Word	Frequency ? ↓
1	digital	30 ...
2	analógico-digital	1 ...

WORDLIST

teste



Get more space



lemma (20,031 items | 377,770 total frequency)



Lemma	Frequency ? ↓	...	Lemma	Frequency ? ↓	...	Lemma	Frequency ? ↓	...	Lemma	Frequency ? ↓	...
1 o	29,927	...	15 p		...	35 homem	1,057	...	40 porque	896	...
2 de	25,228	...	16 n		...	36 quando	1,018	...	41 sem	883	...
3 que	15,701	...	17 d		...	37 dar	1,015	...	42 haver	880	...
4 e	11,056	...	18 e		...	38 vir	1,011	...	43 padre	872	...
5 em	8,789	...	19 ir		...	39 ele	996	...	44 filho	867	...
6 ser	7,841	...	20 c		...				45 mulher	853	...
7 não	7,331	...	21 n		...				46 quem	834	...
8 um	6,870	...	22 me	1,873	...				47 esse	824	...
9 a	6,668	...	23 mas	1,840	...				48 outro	799	...
10 seu	3,993	...	24 lhe	1,813	...				49 nem	773	...
11 com	3,820	...	25 fazer	1,751	...				50 á	732	...
12 se	3,796	...	26 estar	1,654	...						



Concordance



Concordance with macro



N-grams



Word Sketch



Thesaurus

First 20,000 rows are displayed. You can download whole list

Rows per page: 50

1-50 of 20,000



1 / 400



Criação de conta pessoal
Criação do Corpus

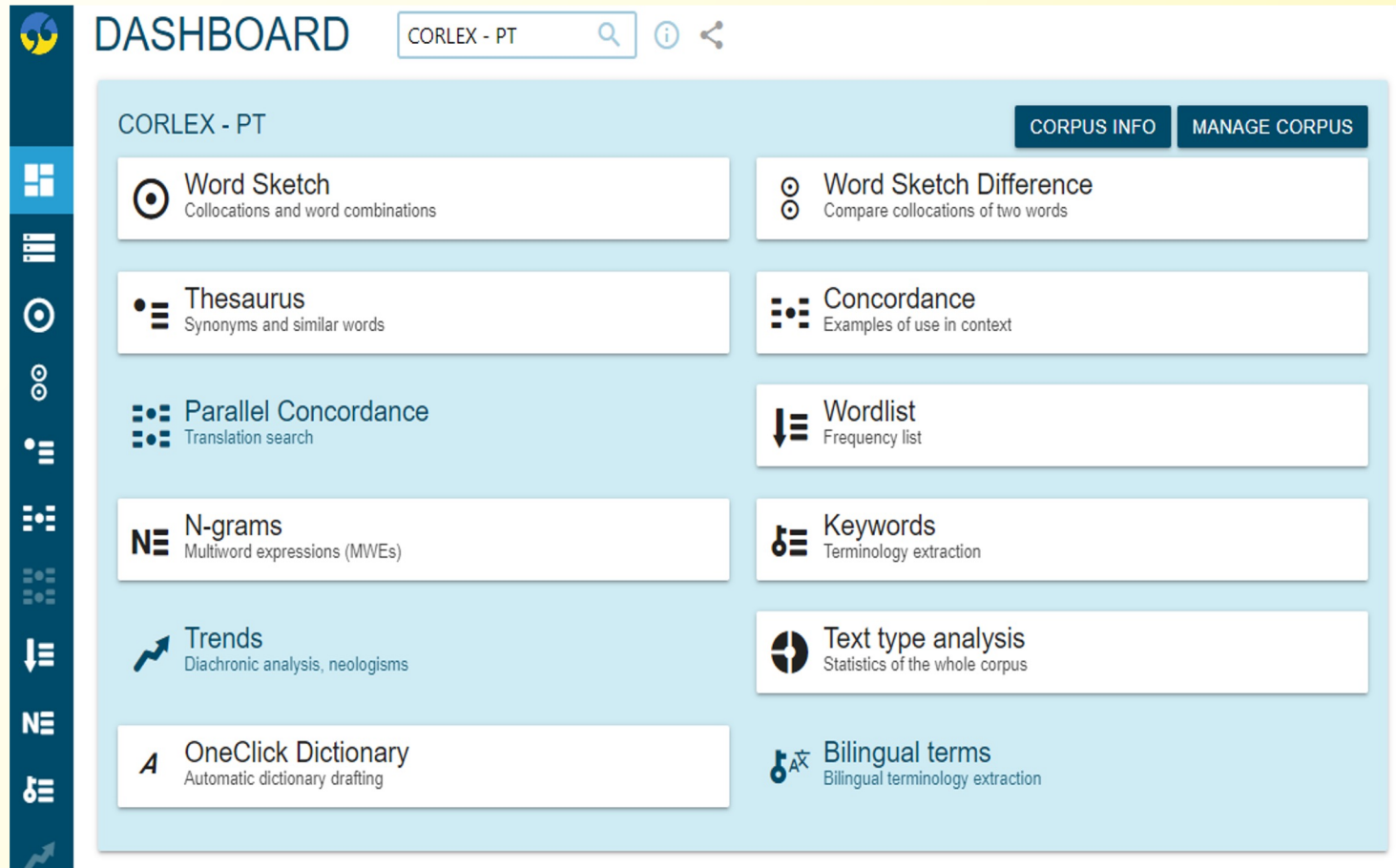
Corpus Info

Text Type Analysis

WordList

Concordance

Word Sketch



DASHBOARD CORLEX - PT

CORLEX - PT

CORPUS INFO **MANAGE CORPUS**

- Word Sketch**
Collocations and word combinations
- Word Sketch Difference**
Compare collocations of two words
- Thesaurus**
Synonyms and similar words
- Concordance**
Examples of use in context
- Parallel Concordance**
Translation search
- Wordlist**
Frequency list
- N-grams**
Multiword expressions (MWEs)
- Keywords**
Terminology extraction
- Trends**
Diachronic analysis, neologisms
- Text type analysis**
Statistics of the whole corpus
- OneClick Dictionary**
Automatic dictionary drafting
- Bilingual terms**
Bilingual terminology extraction

CONCORDANCE

teste



BASIC

ADVANCED

ABOUT

Query type ⓘ

simple

lemma

phrase

word

character

CQL

Part of speech

any

adjective

adverb

conjunction

noun

numeral

preposition

pronoun

Lemma

abc

✓ A = a ?

Subcorpus ⓘ

none (the whole corpus)



Macro ?

none



Filter context ⓘ ▾

Text types ? ▾

GO

CONCORDANCE

teste

lemma **senhor** • 1,284
2,779.85 per million tokens • 0.28%

Get more space

Details
 Left context
 KWIC
 Right context

1	<input type="checkbox"/>	doc#0 i cauda, e começou a ladrar festivamente. </s><s>-- Até que enfim! exclamou a senhora Engracia com satisfação. </s><s>Já não é sem tempo! </s><s>E pousou a meia
2	<input type="checkbox"/>	doc#0 a! exclamou alegremente a tia Engracia, vendo entrar a neta. </s><s>-- E' que a senhora , voltou Margarida, disse-me que mettesse a uso o fato da terra porque, quando
3	<input type="checkbox"/>	doc#0 ando voltasse, a vestir taes fatos cá na terra. </s><s>-- E ella? </s><s>-- E' boa senhora . </s><s>Não vai longe d'isso. </s><s>-- Põe a ceia na mesa. </s><s>-- A Joanini
4	<input type="checkbox"/>	doc#0 icia, esmigalhando no caldo um pedaço de brôa. </s><s>-- Ainda não. </s><s>A senhora não está em estado. </s><s>Tem uma tosse que é mesmo uma dôr d'alma ouvi-
5	<input type="checkbox"/>	doc#0 abana a cabeça com um modo desanimado que é uma pena vêr. </s><s>-- E o senhor ? </s><s>-- Chora pelos cantos, e ao pé d'ella mostra-se muito satisfeito. </s><s>
6	<input type="checkbox"/>	doc#0 rra. </s><s>Uma gente com tanto de seu! </s><s>-- Isso mesmo dizia hontem o senhor padre capellão. </s><s>E conversaram longo tempo, comendo sobre o caldo ca
7	<input type="checkbox"/>	doc#0 rei; mas quem se habitua a viver n'um palacio, em que até os criados parecem senhores , como ha do voltar de gosto a uma casa pequena e pobre? </s><s>-- Basta a c
8	<input type="checkbox"/>	doc#0 a sentar-se. </s><s>Uma ideia, que ainda não tivera, veio sobressalta-la: -- Se a senhora morre, não verei Lisboa. </s><s>Não me levarão para lá </s><s>E o seu cuidad
9	<input type="checkbox"/>	doc#0 i: -- O Victor que monte a cavallo e vá a toda a brida chamar o doutor. </s><s>A senhora está peor. </s><s>E, voltando cosias, desapareceu deixando os criados consti
10	<input type="checkbox"/>	doc#0 os braços um cadaver. </s><s>Vendo entrar Estevam, bradou angustiada: -- Ai! senhor marquez, parece-me que o medico já não vem cá fazer nada. </s><s>-- Um esp
11	<input type="checkbox"/>	doc#0 m o seu nome. </s><s>Foi preferivel assim. </s><s>No estado em que estava a senhora marquez, a vida era-lhe um verdadeiro martyrio. </s><s>Não seja egoista, me
12	<input type="checkbox"/>	doc#0 asatou a chorar, dizendo por entre soluços: -- Não tem nada que me agradecer, senhor marquez. </s><s>Cumpri o meu dever: V. Ex.a não deseja mais nada? </s><s>--
13	<input type="checkbox"/>	doc#0 nante, ouvindo a perfida insinuação da criada franceza. </s><s>Já no tempo da senhora , que Deus haja, eu notei varias cousas que não me agradaram... Enfim, calla-l
14	<input type="checkbox"/>	doc#0 i que morre o peixe. </s><s>-- E' bem verdade. </s><s>Mas então já em vida da senhora ?... -- É o que lhe digo. </s><s>Elle não queria ninguem no quarto senão ella. </s><s>
15	<input type="checkbox"/>	doc#0 </s><s>Elle não queria ninguem no quarto senão ella. </s><s>-- Mas isso era a senhora marquez... -- O' filha, sempre és d'uma innocencia! </s><s>Começou, pois, Ma
16	<input type="checkbox"/>	doc#0 >Foi ter com a governante e disse-lhe sem preambulos: -- Quero-me ir embora, senhora Thomazia. </s><s>Faça o favor de me fazer as contas. </s><s>-- Queres-le ir er
17	<input type="checkbox"/>	doc#0 a, a sr.a Thomazia perguntou-lhe em tom de mofa: -- Então não te despedes do senhor marquez? </s><s>-- Acha que devo? </s><s>-- Decerto. </s><s>Não tens estadc
18	<input type="checkbox"/>	doc#0 ma cadeira de vorgia, lia os jornaes de Lisboa. </s><s>-- Venho dizer-lhe adeus, senhor marquez. </s><s>-- Dizer-me adeus? </s><s>Então onde é que vai? </s><s>-- Vc
19	<input type="checkbox"/>	doc#0 resolvi retirar-me. </s><s>-- Então não estava bom aqui? </s><s>-- Estava sim, senhor ; mas minha avó anda doente... a canceira da casa é muita e... -- Isso não é raz
20	<input type="checkbox"/>	doc#0 elha e com a voz balbuciante, respondeu: -- Eu não posso acceitar o seu favôr, senhor marquez, porque a minha intenção é não voltar. </s><s>Estevam ficou um insta

Rows per page: 20 1-20 of 1,284 < > 1 / 65 > > |

CONCORDANCE teste

lemma **senhor** • 1,284
2,779.85 per million tokens • 0.28%

Left context KWIC Right context

1 ⓘ doc#0 i cauda, e começou a ladrar festivamente. </s></s>-- Até que enfim! exclamou a **senhora** Engracia com salvação. </s></s>Já não é sem tempo!</s></s>E pousou a meia

2 ⓘ doc#0 a! exclamou alegremente a tia Engracia, vendo entrar a neta. </s></s>-- E' que a **senhora** , voltou Mariana da, disse-me que mettesse a uso o fato da terra porque, quando

3 ⓘ doc#0 ando voltasse, a vestir taes fatos cá na terra. </s></s>-- E ella? </s></s>-- E' boa **senhora** . </s></s> não vai longe d'isso. </s></s>-- Põe a ceia na mesa. </s></s>-- A Joanin!

FILTER

BASIC ADVANCED

Keep lines [?] containing

Quick filters
HIDE SUB-HITS
ONLY 1ST HIT IN DOCUMENT

Query type ⓘ
simple
lemma
phrase
word
character
CQL

Part of speech
any
adjective
adverb
conjunction
noun
numeral
preposition
pronoun

Lemma
Antônio

✓ A = a ?

Range [?]
● Token
○ Sentence
○ Custom

-5 -4 -3 -2 -1 KWIC 1 2 3 4 5

Text types [?] v

GO

CONCORDANCE teste

lemma **senhor** • 1,284
2,779.85 per million tokens • 0.28%

filter **Antônio (-3..3,+KWIC)** • 1
2.16 per million tokens • 0.00022%

Sort word x

Left context KWIC Right context ↑

1 ⓘ doc#2 o se arranjam não sei: o caso é que eu fiquei sem o dinheiro e o meu amigo, o **senhor Antônio** Pisco, apareceu-me com uma hospedaria na Rua do Arsenal, onde um

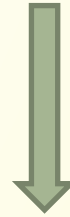
CONCORDANCE teste

lemma **senhor** • 1,284
2,779.85 per million tokens • 0.28%

Left context KWIC Right context

1 doc#0 | cauda, e começou a ladrar festivamente. </s></s>-- Até que enfim! exclamou a **senhora** Engracia com satisfação. </s></s>Já não é sem tempo! </s></s>E pousou a meia

2 doc#0 | ! exclamou alegremente a tia Engracia, vendo entrar a neta. </s></s>-- E' que a **senhora** , voltou Margarida, disse-me que mettesse a uso o fato da terra porque, quando



CONCORDANCE teste

lemma **senhor** • 1,284
2,779.85 per million tokens • 0.28%

Sort word

COLLOCATIONS

BASIC ADVANCED ABOUT

Attribute ? lemma

Range ? -5 -4 -3 -2 -1 KWIC 1 2 3 4 5

Custom range

Show functions ?

- T-score
- MI
- MI3
- log likelihood
- min. sensitivity
- logDice
- MI.log_f

Minimum frequency in corpus ? 5

Minimum frequency in given range ? 3

Single column ?

GO

CONCORDANCE

teste

lemma **senhor** • 1,284
2,779.85 per million tokens • 0.28%

Sort word

Get more space

Collocations

CHANGE CRITERIA BACK TO CONCORDANCE

	Lemma	Cooccurrences ?	Candidates ?	T-score	MI	LogDice ↓
1	<input type="checkbox"/> duquesa	71	105	8.39	7.93	10.71
2	<input type="checkbox"/> --	418	10,080	19.07	3.90	10.24
3	<input type="checkbox"/> !	161	3,588	11.90	4.01	10.08
4	<input type="checkbox"/> condessa	43	54	6.53	8.16	10.04

Only this

Whole concordance but this

CONCORDANCE

teste

lemma **senhor** • 1,284
2,779.85 per million tokens • 0.28%

filter **duquesa** -3, 3 • 71
133.71 per million tokens • 0.015%

Sort word

Get more space

Details

Left context KWIC Right context ↑

- doc#2 </s><s>-- Não, senhor.</s><s>Esse vulto era um homem que jantou hoje com a **senhora ... Duquesa ...** eu não sei se é duquesa... seja lá quem for... -- Seria o barão de
- doc#2 aridade do crepúsculo, passou por Alberto como quem receia ser conhecido.-- A **senhora Duquesa ?** -- perguntou a um criado. -- A senhora que a vossa Excelência hoje
- doc#2 r quem deixa este dinheiro, eu vou deixar aqui o meu nome.</s><s>-- É ocioso, **senhora Duquesa** de Clifton.</s><s>-- Ah! conhece-me... É muita honra, senhora Dona E
- doc#2 ras escritas com sangue, e não com fogo... -- Isso é muito bonito em romances, **senhora Duquesa** ; mas aqui... repare... uma pouca de saliva e uma luva faz desaparece
- doc#2 esa.</s><s>-- Quero uma vingança!... -- já sei que é esse o vosso pensamento, **senhora Duquesa** </s><s>... Mas quero a pronta hoje a lá </s><s>... Aumentaram os e
- doc#2 lo do fel, que o infame me verteu no coração... há de bebê-lo... -- De que modo, **sen**
- doc#2 rros varonilmente, e trocou por outros os fulminantes.</s><s>-- Sois prevenida, **sen**
- doc#2 as suas variantes... Cada vez mais ininteligível... -- Sou um hieroglífico humano, **sen**

CONCORDANCE

teste

lemma **senhor** • 1,284
2,779.85 per million tokens • 0.28%

filter **meu** -3, 3 • 126
272.79 per million tokens • 0.027%

Sort word

Get more space

Details

Left context KWIC Right context ↑

- doc#0 de longe: -- Ehl tia Engracial</s><s>Pode chegar aqui?</s><s>-- Lá vou, **meu senhor** ! é um ai enquanto desço.</s><s>E, com um vagar que contrastava com as su
- doc#1 for para os grandes, risonho no seu ar de pobre diabo.</s><s>-- "Vivam, **meus senhores** !</s><s>-- "Adeus, Neves, então o que ha lá por fóra?</s><s>-- "Não sei nada .
- doc#1 :.</s><s>E' um jogo delicioso, não acha mr. Vilhegas?</s><s>-- "Divino, **minha senhora** !... VII Primeiro do que ninguem tinha o dr. José Ramalho chegado á estação pi
- doc#1 scapara -- quer acompanhar-me até junto de meu primo?</s><s>-- "Oh, **minha senhora** ! -- gaguejou o ricaço, oferecendo presuroso o braço onde ella mal tocou com i
- doc#1 do: -- "O quê?!</s><s>O sr. Telles não bebe vinho?!</s><s>-- "Oh! não, **minha senhora** !</s><s>-- Não posso tomar nada que me excite os nervos doentamente emotivo
- doc#1 a abraçando-a, com as lagrimas a assomarem-lhe aos olhos.</s><s>-- "**Minha senhora** ! -- curvou-se elle, tambem vencido pela emoção.</s><s>-- "Deixa, Isabella, o s

CORLEX

16

- Quais os contextos onde ocorre “humanidade”?
- Quais os contextos onde ocorre “digital”?

The screenshot displays the CORLEX concordance tool interface. At the top, the search query is 'CORLEX - PT'. Below the search bar, a summary shows 'CQL [lc=="humanidades"] • 7' and '0.88 per million tokens • 0.000088%'. The interface includes a 'Details' tab and a table of search results. The table has columns for 'Left context', 'KWIC', and 'Right context'. The KWIC column shows the word 'Humanidades' in red. A large text box on the right provides a detailed view of the KWIC entry, showing the full sentence: 'a ser um pouco menos verdadeira... AMÉRICO DA COSTA RAMALHO, MEIO SÉCULO DE ESTUDOS CLÁSSICOS EM COIMBRA: BREVES NOTAS, in HISTÓRIA E DESENVOLVIMENTO DA CIÊNCIA EM PORTUGAL NO SÉC. XX, III VOLUME, A.C.L., LISBOA, 1992 <> A falta de tempo no momento presente não me permite ocupar-me senão de Coimbra, a Universidade portuguesa que melhor conheço. Isto não quer dizer que outras universidades não disponham de florescentes Departamentos de Estudos Clássicos, como a Universidade de Lisboa, e que em certas universidades, onde não existe tal departamento, se não encontrem notáveis latinistas, e até helenistas, como no Porto, em Braga, em Viseu, em Aveiro ou em Évora. Um estudo circunstanciado sobre a sua actividade ficará para outra ocasião. Em Coimbra, há cinquenta anos, as Humanidades Clássicas concentravam-se na actividade pedagógica, primeiro do Doutor Carlos Simões Ventura, catedrático de Grego; e a partir de 1939-1940, e por alguns anos, simultaneamente com o Professor Ventura, na actividade do Doutor Francisco da Luz Rebelo Goncalves, catedrático de Latim. O Professor

Criação de conta pessoal
Criação do Corpus

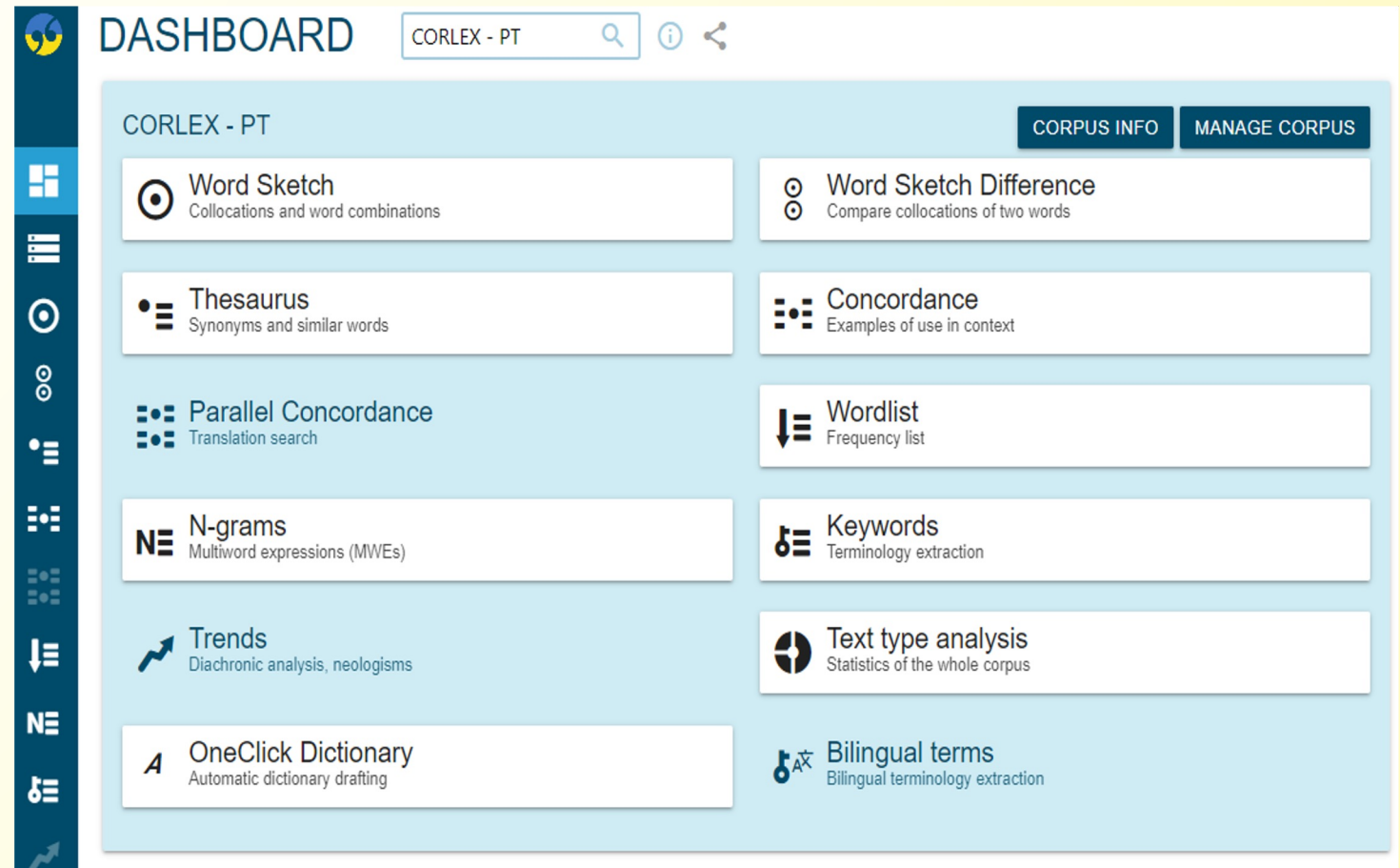
Corpus Info

Text Type Analysis

WordList

Concordance

Word Sketch



DASHBOARD CORLEX - PT

CORLEX - PT

CORPUS INFO **MANAGE CORPUS**

- Word Sketch**
Collocations and word combinations
- Word Sketch Difference**
Compare collocations of two words
- Thesaurus**
Synonyms and similar words
- Concordance**
Examples of use in context
- Parallel Concordance**
Translation search
- Wordlist**
Frequency list
- N-grams**
Multiword expressions (MWEs)
- Keywords**
Terminology extraction
- Trends**
Diachronic analysis, neologisms
- Text type analysis**
Statistics of the whole corpus
- OneClick Dictionary**
Automatic dictionary drafting
- Bilingual terms**
Bilingual terminology extraction

WORD SKETCH

teste



Get more space



homem as noun 1,057x



homem + verbo	sintagma preposicional	verbo + homem	homem + adjetivo	adjetivo + homem	homem + adjetivo participial
ser homem é	...de homem	parecer parecia um homem	superior um homem superior	pobre pobre homem	honrar homem honrado
chamar homem , que o mundo chamava	homem de substantivo	falar fala com um homem	pobre um homem pobre	d'este d'este homem	matar um homem morto
estar homem não está	...por homem	conhecer Conheces ainda pouco os homens	rico homens ricos	único o único homem	
poder homem que pode	...com homem	amar amava muito esse homem	bom passos de um homem bom	d'um d'um homem	
ter homem tem	homem em substantivo	d'esse d'esse homem	extraordinário um homem extraordinário	primeiro o primeiro homem	
deixar homem que os deixa	...para homem	ver viu um homem	estranho um homem estranho	grande grande homem	
ir homem que foi	homem a substantivo	fossar fossava positivamente um mau homem	feliz homem mais feliz do globo	bom bom homem	
fazer homem , não faça	...por+os homem	espiritar espírito do homem			
procurar homens procuram		haver há homens			
dever homem devia		matar matar este homem			
levar homem de espírito , levou-a		encontrar encontrar um homem			
haver homem , há		viver vivia desde muito para um homem			
substantivo ser-estar homem	verbo com se + homem	preposição+infinitivo			
alberto Alberto era um homem	vestir vestiu de homem	... homem para Vinf			
dinis Dinis era um homem	sentir sentia-se o mais infeliz dos homens				

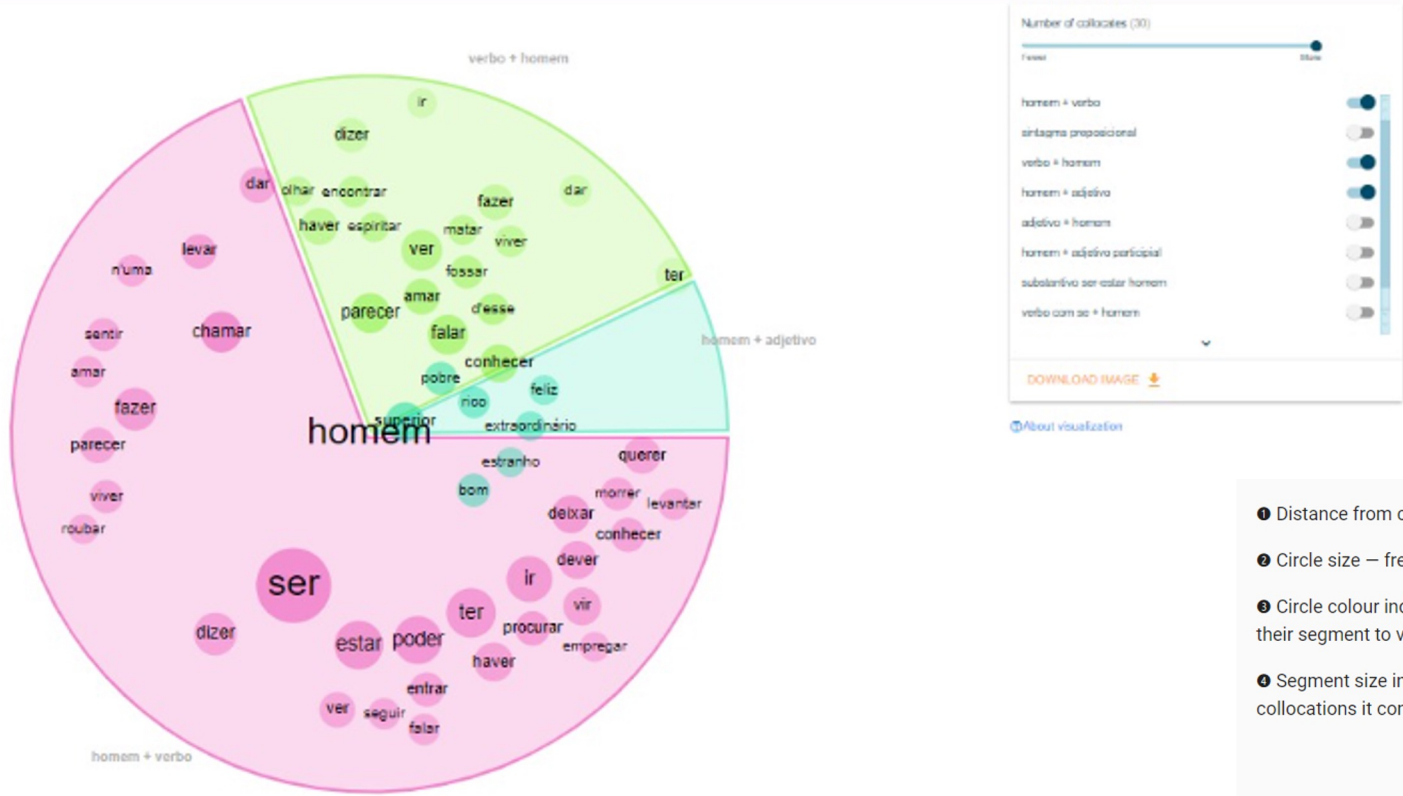
WORD SKETCH

teste

homem: as noun 1,057x

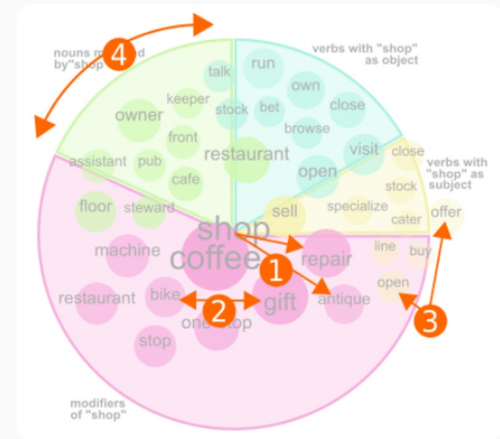
CHANGE VIEW OPTIONS

- Show frequencies?
- Show collocation examples?
- Show scores?
- Combine grammatical relations?
- Cluster similar items?
- Sort results by?
 - Score
 - Frequency



visualization by SKETCH ENGINE

- 1 Distance from centre – typicality (score). *repair shop* is more typical than *antique shop*.
- 2 Circle size – frequency. *gift shop* is more frequent than *bike shop*.
- 3 Circle colour indicates which segment (grammatical relation) they belong to. Sometimes circles may be placed outside their segment to visualize the other statistics correctly.
- 4 Segment size indicates the size of the grammatical relation in relation to the other visualized relations, i.e. how many collocations it contains in total. (Not how many appear in the visualisation.)



Word Sketch de humanidades / digital

WORD SKETCH CORLEX - PT

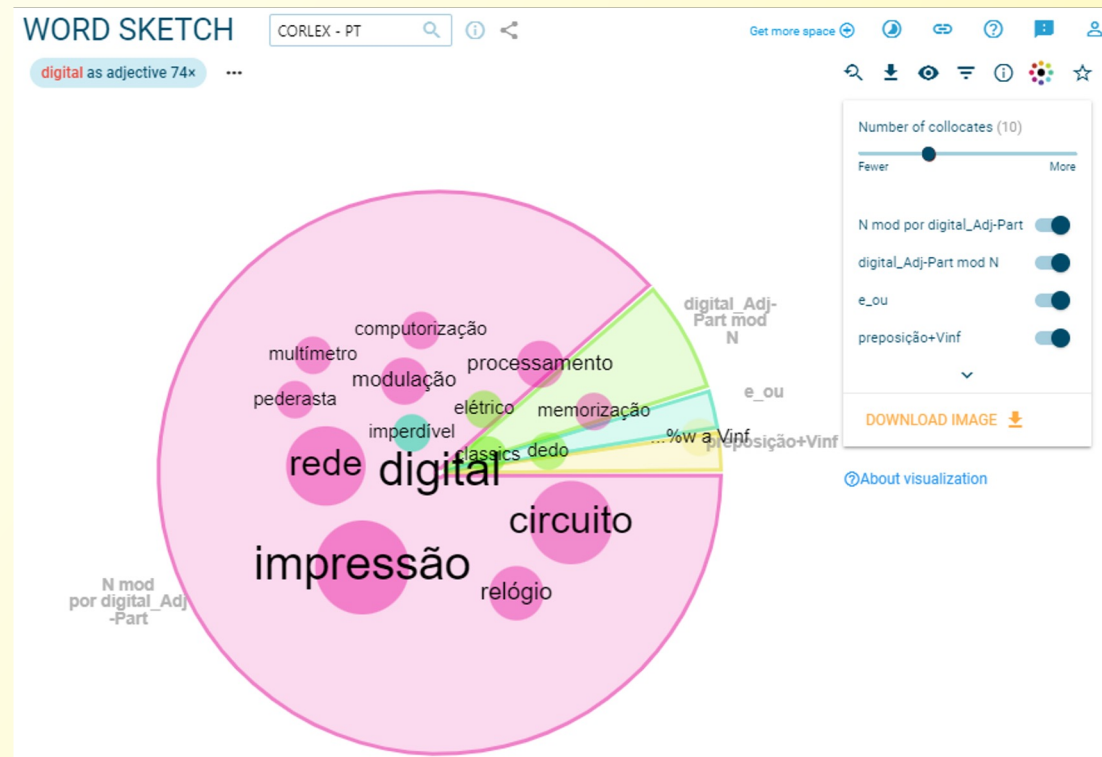
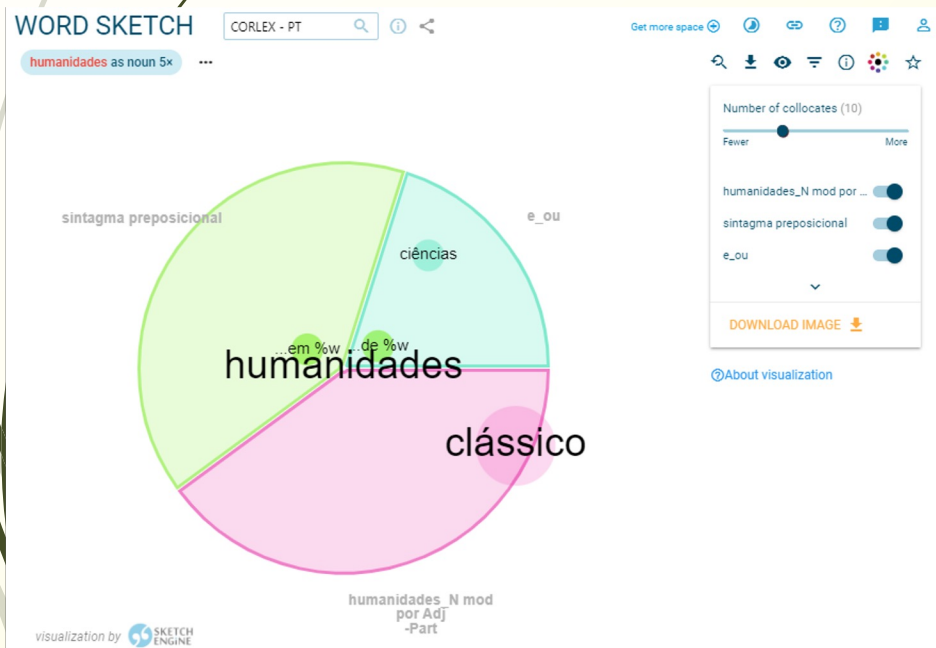
humanidades as noun 5x

humanidades_N mod por Adj-Part	sintagma preposicional	e_ou
clássico	...de humanidades	ciências
	...em humanidades	

WORD SKETCH CORLEX - PT


digital as adjective 74x


N mod por digital_Adj-Part	digital_Adj-Part mod N	e_ou	preposição+Vinf
impressão	classics	imperdível	...digital a Vinf
impressões digitais	elétrico		
rede	dedo		
as redes digitais			
modulação			
processamento			
relógio			
círculo			
nos círculos digitais			
multímetro			
pederasta			
memorização			
computorização			
computador			
vídeo			





- Consultar os Guias de Ajuda


HELP & SUPPORT VIDEO LESSONS KEYBOARDS SHORTCUTS


 **Quick start guide**
Basics in a few minutes

 **User guide**
Everything about Sketch Engine

 **Ask a question**
Let our support team help you

 **Interface introduction**
Show 1-minute introduction to the interface

 **Get notified**
About new corpora and tools

 **Bibliography**
Sketch Engine references



Leituras:

Guião do SK

- <https://www.sketchengine.eu/guide/word-sketch-collocations-and-word-combinations/>

Avaliação



Próxima aula

- Avaliação dos resumos submetidos



EN	The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
PT	O apoio da Comissão Europeia à produção desta publicação não constitui um aval do seu conteúdo, que reflete unicamente o ponto de vista dos autores, e a Comissão não pode ser considerada responsável por eventuais utilizações que possam ser feitas com as informações nela contidas.