

O CORPUS DO PORTUGUÊS

1. Buscas básicas
2. Análise de variações
3. Navegação e busca das 40 mil palavras mais comuns
4. *Word sketches* (para as 40 mil palavras mais comuns)
5. Links para recursos externos
6. Corpora Virtuais
7. Análise de textos completos
8. Resumo

Os corpora do **Corpus do Português** oferecem bilhões de palavras provenientes de dados recentes de quatro países de língua portuguesa. Eles permitem que pesquisadores, estudantes e professores entendam melhor o português de formas que não são possíveis com nenhum outro recurso. Há três corpora diferentes no Corpus do Português:

	Corpus	Tamanho (n.º de palavras)	Conteúdo	Usos	Data de criação
1	Web / Dialeto *	1 bilhão	Páginas da internet de quatro países de língua portuguesa, aproximadamente 50% blogs e 50% gerais	Comparação de dialetos	2016
2	NOW (2012 - 2019)	1,1 bilhão	Páginas da internet dos mesmos quatro países, 2012 a 2019	Mudanças; dialetos	2018
3	Gênero / Histórico	45 milhões	Textos dos anos 1200 a 1900; equilibrado quanto ao gênero para os anos 1900 (oral, ficção, jornal, acadêmico)	Mudanças; gêneros	2006

Dos três corpora, o que admite a maior variedade de recursos é o corpus Web/Dialeto (o primeiro na tabela acima), e a maior parte da discussão a seguir se refere a esse corpus. O corpus Web/Dialeto tem mais de um bilhão de palavras de dados provenientes de quatro países de língua portuguesa:

País	Geral			Blogs			Total		
	Palavras	Páginas	Sites	Palavras	Páginas	Sites	Palavras	Páginas	Sites
Brasil	319.435.592	286.712	25.351	336.244.918	321.305	35.248	655.680.510	608.017	60.599
Portugal	136.144.529	184.512	12.082	190.503.822	221.338	9.005	326.648.351	405.850	21.087
Angola	17.877.399	19.178	1.240	17.255.595	21.233	418	35.132.994	40.411	1.658
Moçambique	16.936.743	19.236	1.065	15.070.829	17.910	404	32.007.572	37.146	1.469
TOTAL	490.394.263	509.638	39.738	559.075.164	581.786	45.075	1.049.469.427	1.091.424	84.813

Existem outros corpora de português que são maiores que o Corpus do Português, mas há [sérios problemas](#) com as anotações desses corpora, o que significa que são bastante problemáticos para muitos tipos de pesquisa linguística. O Corpus do Português tem os únicos corpora grandes de português que também são cuidadosamente anotados.

Nos exemplos a seguir, não oferecemos instruções sobre o que inserir exatamente no formulário de busca para obter os mesmos resultados. Porém, existem muitos arquivos de ajuda “sensíveis ao contexto”, dentro do próprio corpus, que mostram esse tipo de informação. Por exemplo, no formulário de busca principal, clique em qualquer tipo de busca (em vermelho) para ver instruções e exemplos de buscas (em verde) e os arquivos de ajuda relacionados (em amarelo).

List Chart Word Browse **Collocates** Compare KWIC -

ESPESSO Word/phrase [POS] ?

NOUN Collocates [POS]

+ 4 3 2 1 0 - 0 1 2 3 4 +

Find collocates Reset

☐ Sections ☐ Texts/Virtual ☐ Sort/Limit ☐ Options

(HIDE HELP)

COLLOCATES display

See what words occur near other words, which provides great insight into meaning and usage. For example, nouns before **espresso** or after **contar com**, verbs before **calma**, or any word near **garganta**, **brilhar**, **forte**, or **rapidamente**.

More information: [compare to LIST display](#), [types of collocates](#), [direction and distance of collocates](#).

1. Buscas básicas (voltar para o início)

Em seu nível mais básico, o Corpus do Português permite que os usuários façam buscas por palavra, expressão, parte do discurso, lema (formas de uma palavra), sinônimos e lista personalizada de palavras. Em quase todos os casos, a busca por mais de um bilhão de palavras de dados leva apenas 1 ou 2 segundos. Por exemplo:

NOUN SUAVE (NOUN = substantivo)

HELP	①	★	ALL FORMS (SAMPLE): 100 200 500 WORDS	FREQ	TOTAL 9,013 UNIQUE 1,693 +
1	①	★	FORMA SUAVE	381	
2	①	★	VOZ SUAVE	324	
3	①	★	MÚSICA SUAVE	215	
4	①	★	BRISA SUAVE	193	
5	①	★	MOVIMENTOS SUAVES	172	
6	①	★	TOQUE SUAVE	172	
7	①	★	CHEIRO SUAVE	168	
8	①	★	PALAVRAS SUAVES	160	
9	①	★	CORES SUAVES	156	
10	①	★	PELE SUAVE	154	

=MENINA =BELO (maiúsculas = todas as formas de cada palavra, como *bela*, *belas*)

HELP	①	★	ALL FORMS (SAMPLE): 100 200 500 WORDS	FREQ	TOTAL 2,812 UNIQUE 148 +
1	①	★	MENINA BONITA	367	
2	①	★	MENINA LINDA	364	
3	①	★	MENOR GRAÇA	234	
4	①	★	MOÇA BONITA	204	
5	①	★	MENINAS LINDAS	137	
6	①	★	MENINAS BONITAS	125	
7	①	★	GAROTA LINDA	112	
8	①	★	MOÇA LINDA	96	
9	①	★	GAROTA BONITA	90	
10	①	★	CRIANÇA LINDA	84	

PRON COLOCAR * @ROUPA (PRON = pronome)

HELP	①	★	ALL FORMS (SAMPLE): 100 200 500 WORDS	FREQ	TOTAL 165 UNIQUE 138 +
1	①	★	ELA COLOCOU O VESTIDO	4	
2	①	★	EU COLOCAR UM VESTIDO	4	
3	①	★	VOCÊ COLOCA UMA BLUSA	4	
4	①	★	EU COLOCO UMA CAMISA	3	
5	①	★	EU COLOCO UM VESTIDO	3	
6	①	★	ELE COLOCOU A BOLSA	3	
7	①	★	ELA COLOCAR O CINTO	2	
8	①	★	ELE COLOCA O COLAR	2	
9	①	★	EU COLOCO A BOLSA	2	
10	①	★	ELE COLOCOU UMA CARTEIRA	2	
11	①	★	EU COLOCO O CINTO	2	

Obviamente, em qualquer uma dessas buscas (ou das outras buscas mostradas neste documento) é possível clicar em qualquer palavra/expressão para visualizar as linhas de concordância:

CLICK FOR MORE CONTEXT			HELP		SAVE	TRANSLATE	ANALYZE
1	G BR	alexeiweb.org	🔍	🗨	📄	calçadas surpreendentemente limpas, nos jardins e canteiros bem cuidados, na brisa suave que passa por todas as ruas, na tranquilidade das caminh	
2	B BR	ancoradadama.blogspot.com	🔍	🗨	📄	"» Autor desconhecido "« A natureza humana quer velejar até o céu com uma brisa suave , em mar calmo e sem sobressaltos. A oração é um trabalho	
3	G PT	anuea.org	🔍	🗨	📄	, o cheiro quase imperceptível do ar e do rio e a brisa suave , transportavam-me em memórias duma realidade maior que a minha, fundindo em	
4	G BR	arionetorres.blogspot.com	🔍	🗨	📄	palavras nas minha visitas. Quero ser pelo menos o vento com sua brisa suave soprar no seu rosto. Quero de alguma forma dizer o quanto você é	
5	G PT	arpose.blogspot.com	🔍	🗨	📄	como: Manjo Fango, Marin Miéjour, ou o Damo que é uma brisa suave , mas persistente, que faz levantar as saias das senhoras... Em a	
6	G BR	artigos.gospelprime.com.br	🔍	🗨	📄	que pede, mas caso não veja, não terá. Foi quando uma brisa suave movimentou as vestes dos dois amigos, e o que era brisa se transformou	
7	G PT	asuldomundo.wordpress.com	🔍	🗨	📄	dias já não são dum calor sufocante. A a noite, uma brisa suave e morna tocava a pele afastando os tão indesejados mosquitos. Por duas noites tive	
9	B BR	auxilioebd.blogspot.com (1)	🔍	🗨	📄	o terremoto nem no fogo... Veio então um murmúrio duma brisa suave e Elias cobriu o rosto. Ele compreendeu que Deus tinha (e tem)	
10	G PT	blog-sic.blogspot.com	🔍	🗨	📄	a terminar. As temperaturas escaldantes dos últimos dias dão lugar a uma brisa suave , enquanto o sol vai desaparecendo. José Moutinho é o último a	

Além disso, é possível visualizar mais contexto, caso necessário:

Source information:

Source	http://asuldomundo.wordpress.com/
Country/genre	Portugal: General
Title	A Sul do Mundo Pequenos apontamentos do que se passa deste

Expanded context

estar mais errado! É tão difícil descrever o fascínio de este país. É pobre, é certo, tem muitas dificuldades, é certo, cria muitas chatices, também é verdade! Mas, lá em o fundo há algo que só cá quem vem pode experimentar. Um cocktail de história, gentes e culturas de sabor ímpar! Esta última semana foi passada a pensar que tudo o que fazia era " a última vez " ! É uma sensação estranhar pensar assim. Quase que me atrevo a dizer que em o meu subconsciente algo previa que este era de facto um " último adeus " (quem sabe definitivo) a muita coisa. Será que volto cá? Não sei porquê, mas esta semana o tempo esteve tão agradável! Os dias já não são de um calor sufocante. A a noite, uma brisa suave e morna tocava a pele afastando os tão indesejados mosquitos. Por duas noites tive a oportunidade de jantar ao ar livre junto a a barragem de Macarretane. Aqui, além de os sons habituais de a bicharada, a Natureza presenteou- nos com um céu estrelado impossível de esquecer e de descrever e um desfile constante de pequenos pirilampos por o mato. Fantástico! Tanta coisa que poderia dizer... ando há uma semana a pensar em o que escrever. Todos os dias tenho uma versão nova! Mas, quando foi preciso escrever este texto não me saia nada... Levo com mim uma bagagem quase vazia de roupa, mas recheada de lembranças inesquecíveis: umas boas, umas muito boas, outras más. Moçambique é assim, é preciso amar! Muitos acabam por abandonar o barco e perdem a essência de esta jovem nação. Um Estado em construção onde eu fui uma pequena peça de

Também é possível encontrar “colocados” (palavras vizinhas), que ajudam a entender o significado e o uso de uma palavra ou expressão. Por exemplo:

Substantivos colocados de *ESPESSO* (maiúsculas = todas as formas de *espesso*)

HELP	①	★	WORDS	FREQ	ALL	%	MI	
1	①	★	CAMADA	236	22451	1.05	7.71	
2	①	★	CREME	113	27521	0.41	6.36	
3	①	★	NUVENS	105	17248	0.61	6.93	
4	①	★	PELE	105	133321	0.08	3.98	
5	①	★	MOLHO	87	17532	0.50	6.63	
6	①	★	LÍQUIDO	82	27663	0.30	5.89	
7	①	★	CAMADAS	67	19460	0.34	6.10	
8	①	★	PAREDES	67	30473	0.22	5.46	
9	①	★	NUVEM	63	16102	0.39	6.29	
10	①	★	SANGUE	57	124502	0.05	3.19	
11	①	★	NEVOEIRO	55	3547	1.55	8.28	
12	①	★	MASSA	54	102251	0.05	3.40	

Colocados de *meio ambiente*

HELP	①	★	WORDS	FREQ	ALL	%	MI	
1	①	★	PRESERVAÇÃO	1736	19801	8.77	8.19	
2	①	★	MINISTÉRIO	1030	166607	0.62	4.36	
3	①	★	SECRETARIA	909	43902	2.07	6.11	
4	①	★	DESENVOLVIMENTO	722	287986	0.25	3.06	
5	①	★	RECURSOS	697	182618	0.38	3.67	
6	①	★	PROTEÇÃO	639	66012	0.97	5.01	
7	①	★	DEFESA	492	142944	0.34	3.52	
8	①	★	INSTITUTO	458	84195	0.54	4.18	
9	①	★	UNIDAS	434	31026	1.40	5.54	
10	①	★	ECONOMIA	416	151625	0.27	3.19	
11	①	★	MUNDIAL	387	157332	0.25	3.04	
12	①	★	ESTADUAL	385	45622	0.84	4.81	

Também é possível visualizar “linhas de concordância”, que mostram padrões nos quais a palavra ocorre. Por exemplo, os exemplos abaixo mostram linhas de concordância para *relação*:

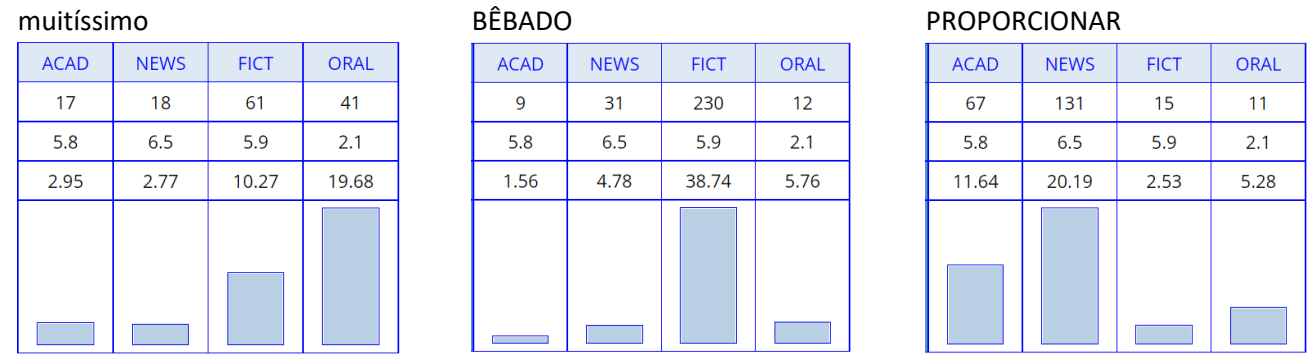
164	B PT	...urnbookbybrunocunha.blogspot.com	🔍🔍🔍	Q	os meus lindos e adorados progenitores souberam de a minha relação com o Marlboro Gold , ficaram chocados e não consegu
165	G BR	portaldodoamor.com.br	🔍🔍🔍	Q	muito injustiçada , pois ate ali fui sempre cuidados com nossa relação e sempre tratei a relação como se fosse um ser especia
166	B PT	ciadaniasc.blogspot.com	🔍🔍🔍	Q	projecto não é da minha autoria e não tenho qualquer relação contratua com a câmara ou com a Cascais Próxima "» ,
167	B PT	lei-natural.blogspot.com	🔍🔍🔍	Q	governantes (que, à partida , não têm qualquer relação familiar com si) . Naturalmente , sendo o Estado monár
168	B BR	psicologiadialetica.com	🔍🔍🔍	Q	de enfrentar o amadurecimento que a situação requeria já que relação de irmãos e casamento não combinam , ainda mais qua
169	B BR	aartedefazerbemfeito.blogspot.com	🔍🔍🔍	Q	e com o respeito aos limites individuais poderá levar sua relação com um novo patamar . Gosto de todas essas coisas , mas
170	G PT	ferias-cientificas.com	🔍🔍🔍	Q	os habitats de todos os seus animais e estudar a sua relação com as árvores ; Mas ainda mais interessante é que , co
171	B PT	resumosdatv.com	🔍🔍🔍	Q	de imaginar que as primas querem é acabar com a sua relação com Matilde ; Ivo regressa do funeral do irmão acompa
172	B BR	panoramadaaquicultura.com.br	🔍🔍🔍	Q	deter conhecimento prévio sobre a biologia de o patógeno , sua relação com o hospedeiro é o ambiente . Esses são os requisito:
173	G BR	hottopos.com	🔍🔍🔍	Q	educacional brasileira , verificando , em concreto , sua relação com o ponto de exclamação , e o que ele implica de
174	G PT	scielo.gpeari.mctes.pt	🔍🔍🔍	Q	artigo , enfim , compreendeu o sentido de pertença e sua relação com os códigos de a confiança e lealdade , conceitos aq
175	B BR	aposentadoinvocado1.blogspot.com	🔍🔍🔍	Q	, as finanças , o orçamento de a União , sua relação com os Estados ; A falta da Rede da Sustentabilidade en

2. Análise de variações (voltar para o início)

Os corpora do Corpus do Português permitem examinar de forma rápida e fácil variações no português – entre palavras, entre gêneros, ao longo do tempo e entre dialetos. Para exemplificar a **comparação de palavras**, as palavras a seguir são os colocados dos sinônimos *potente* e *poderoso*:

WORD 1 (W1): PODEROSO (6.50)						WORD 2 (W2): POTENTE (0.15)					
	WORD	W1	W2	W1/W2	SCORE		WORD	W2	W1	W2/W1	SCORE
1	ORAÇÃO	694	0	1,388.0	213.6	1	ANALGÉSICOS	14	0	28.0	181.9
2	TODO	421	0	842.0	129.6	2	DENTE	11	0	22.0	142.9
3	GENTE	238	0	476.0	73.3	3	CARROS	32	3	10.7	69.3
4	MULHER	171	0	342.0	52.6	4	MOTORES	27	3	9.0	58.5
5	GRUPOS	85	0	170.0	26.2	5	CARRO	27	4	6.8	43.9
6	INTERESSES	78	0	156.0	24.0	6	COMPUTADOR	13	2	6.5	42.2
7	MULHERES	136	1	136.0	20.9	7	HIDRATAÇÃO	10	3	3.3	21.7
8	FORÇAS	125	1	125.0	19.2	8	REMATE	19	6	3.2	20.6
9	HOMENS	125	1	125.0	19.2	9	MOTOR	48	21	2.3	14.9
10	MÃOS	109	1	109.0	16.8	10	PROCESSADORES	10	7	1.4	9.3
11	INIMIGOS	50	0	100.0	15.4	11	SOM	30	23	1.3	8.5
12	AMARRAÇÃO	50	0	100.0	15.4	12	GOLPES	10	8	1.3	8.1

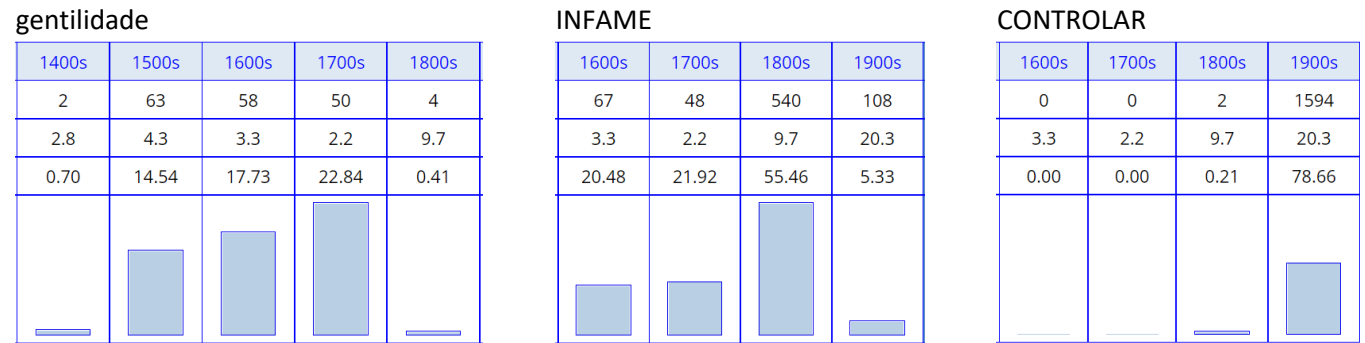
O corpus Gênero/Histórico permite comparar a frequência de palavras ou expressões nos quatro **gêneros** (oral, ficção, jornal e acadêmico). Por exemplo, eis as tabelas para três palavras diferentes:



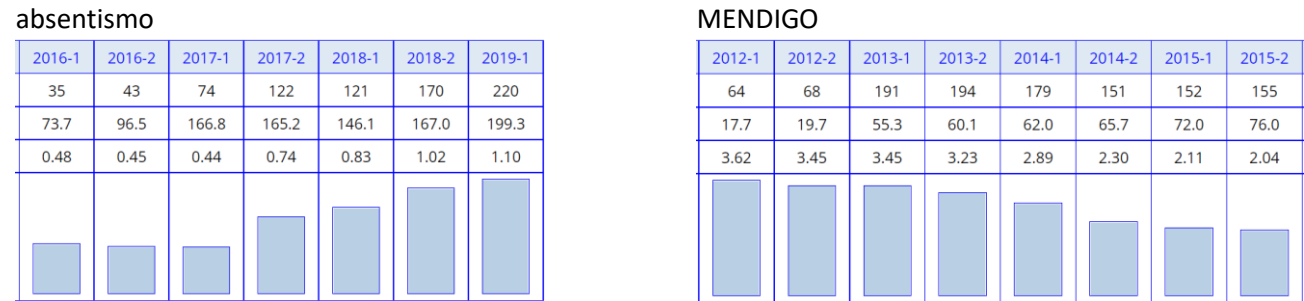
Também é possível comparar palavras, expressões ou colocados nos diferentes gêneros. Por exemplo, a tabela a seguir mostra uma comparação dos verbos que começam com *des** em textos de ficção (esquerda) e acadêmicos (direita):

SEC 1 (FICT): 5,937,484 WORDS							SEC 2 (ACAD): 5,754,506 WORDS						
	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	DESABAFAR	47	0	7.9	0.0	791.6	1	DESCONSIDERAR	12	0	2.1	0.0	208.5
2	DESATAR	27	0	4.5	0.0	454.7	2	DESCONSTRUIR	6	0	1.0	0.0	104.3
3	DESPEJAR	27	0	4.5	0.0	454.7	3	DESCENTRALIZAR	6	0	1.0	0.0	104.3
4	DESMAIAR	25	0	4.2	0.0	421.1	4	DESTRONAR	6	0	1.0	0.0	104.3
5	DESPEGAR	16	0	2.7	0.0	269.5	5	DESOBRIGAR	4	0	0.7	0.0	69.5
6	DESDENHAR	14	0	2.4	0.0	235.8	6	DESBLOQUEAR	3	0	0.5	0.0	52.1
7	DESCAIR	13	0	2.2	0.0	218.9	7	DESACTIVAR	3	0	0.5	0.0	52.1
8	DESCASCAR	13	0	2.2	0.0	218.9	8	DESESTIMULAR	3	0	0.5	0.0	52.1
9	DESGRAÇAR	12	0	2.0	0.0	202.1	9	DESINTEGRAR	3	0	0.5	0.0	52.1
10	DESFALECER	12	0	2.0	0.0	202.1	10	DESINFECTAR	3	0	0.5	0.0	52.1

Também é possível comparar por **intervalo de tempo**. Como exemplo, veja as frequências das palavras a seguir ao longo do tempo:



O corpus NOW-Portuguese, com 1,1 bilhão de palavras, permite comparar a frequência no curto período entre 2012 e 2019 para observar mudanças recentes:



Por último, o corpus Web/Dialetos (bem como o corpus NOW-Portuguese, que inclui textos dos mesmos países) permite comparar a frequência das palavras e expressões em **quatro países de língua portuguesa**. As tabelas a seguir são apenas alguns exemplos:



trem

Brasil	Portugal	Angola	Moçambique
17391	1931	240	168
558.4	270.2	38.4	27.9
31.15	7.15	6.26	6.03

comboio

Brasil	Portugal	Angola	Moçambique
1263	10858	1229	434
558.4	270.2	38.4	27.9
2.26	40.18	32.05	15.57

machicombo (Moçambique)

Brasil	Portugal	Angola	Moçambique
8	35	16	69
558.4	270.2	38.4	27.9
0.01	0.13	0.42	2.48

mata-bicho (Moçambique e Angola)

Brasil	Portugal	Angola	Moçambique
5	63	30	36
558.4	270.2	38.4	27.9
0.01	0.23	0.78	1.29


E buscas por qualquer uma das palavras e expressões a seguir mostrarão diferenças semelhantes por país (a lista a seguir é principalmente para o português brasileiro e europeu):

Inglês	Brasil	Portugal	Inglês	Brasil	Portugal
train	trem	comboio	truck	caminhão	camião
bus	ônibus	autocarro	jacket	jaqueta	blusão
ice cream	sorvete	gelado	breakfast	café da manhã	pequeno-almoço
cup	xícara	chávena	cancer	câncer	cancro
sports	esporte	desporto	slum	favela	bairro de lata
goal (futebol)	gol	golo	"cool"	(muito) legal	(muito) fixe
cell/mobile (phone)	celular	telemóvel	swimsuit	maiô	fato de banho
bathroom	banheiro	sala de banhos	(electronic) screen	tela	escrã

Os exemplos acima são apenas palavras isoladas ou expressões curtas. O corpus também pode ser usado para comparar características sintáticas e morfológicas entre dialetos. Por exemplo:

Brasil	Explicação	Portugal	Explicação
eu iria VERB	condicional “analítico”		
estava* [vpp*]	construção progressiva	estava* a [vr*]	construção progressiva
eu [po*] VERB	clítico pré-verbal (com verbo finito)	eu *-_v [po*]	clítico pós-verbal (com verbo finito)
e me lhe disse	clítico pré-verbal (com verbo finito)	eu *-_v [po*]	clítico pré-verbal (com verbo finito)
para pra DAR lhe	clítico pós-verbal (com verbo não finito)	para pra lhe DAR	clítico pré-verbal (com verbo não finito)
para pra fazer- lo	clítico pós-verbal (com verbo não finito)	para pra o fazer	clítico pré-verbal (com verbo não finito)
eu VERB você	você em vez do clítico		
eu VERB para pra ela ele	ênfase no pronome em vez do clítico		
DAR para pra mim	ênfase no pronome em vez do clítico		
		me os as a o [vis*]	pronomes contraídos *

A título de exemplo, a tabela a seguir mostra a frequência no corpus da sequência de caracteres *e me | lhe disse* (com um pronome objeto pré-verbal após *e*), que é mais comum no Brasil:

SECTION	ALL	Brasil	Portugal	Angola	Moçambique
FREQ	2151	1810	280	37	24
WORDS (M)	1000	558.4	270.2	38.4	27.9
PER MIL	2.15	3.24	1.04	0.96	0.86
					

CLICK FOR MORE CONTEXT			HELP		SAVE	TRANSLATE	ANALYZE
1	B BR	escribacafe.com	🔍	🗨️	🔍	ela veio até mim e me apontou uma banca cheia de peças, e me disse da seguinte forma: -- Essas aqui são as da promoção! Imediatamente	
2	G BR	ocandoble.wordpress.com	🔍	🗨️	🔍	! A mãe duma colega minha há 5 anos leu minha mão e me disse coisas a meu respeito que quase ninguém sabia e não tinha como ela saber.	
3	G BR	blogdomarcone.com.br	🔍	🗨️	🔍	beijos, nada de mais. Só que ele sabe que eu namoro e me disse para não criar expectativas quanto ao que havia acontecido. Falou que continua	
4	G BR	aguasvivas.ws	🔍	🗨️	🔍	. Só a síntese. E lhe perguntou pelo que sabia. E lhe disse: ' Você deveria ir lá onde está o seu pai; mas Deus lhe	
5	B BR	nao2nao1.com.br	🔍	🗨️	🔍	soltei um "« você está linda hoje "». Ela fechou a cara e me disse: não sou bonita nos demais dias? Não sou interessante nos demais	
6	G BR	ocandoble.wordpress.com	🔍	🗨️	🔍	seria minha mãe, (A de lemanjá) virou onde moro, e me disse que lemanjá está esperando que eu vá até ela, para resolver com ela o	
7	B BR	sonhodetres.blogspot.com	🔍	🗨️	🔍	eu a colocasse, logo em seguida minha GO entrou no quarto e me disse que iria introduzir um comprimidinho que ajudaria a dilatar ainda mais o	
8	G BR	verboeterno.wordpress.com	🔍	🗨️	🔍	ele ainda nem tinha e que ele já estava casando pensando em separar e me disse claramente que não poderia me garantir que iria ficar com min	
9	B BR	pitacosdelua.blogspot.com	🔍	🗨️	🔍	o Submarino, entrou novamente em contato por telefone hoje, 02/03, e me disse que o Submarino já providenciou o envio de outro livro. Aguard	
10	G BR	literaturabrasileira.ufsc.br	🔍	🗨️	🔍	puero tuo et salvum fac filium ancillae tuae, apareceu- lhe o encoberto e lhe disse: Deixai- me rezar, não me inquieteis; e se me ordenais que repit	
11	G BR	oblogodeplastico.com	🔍	🗨️	🔍	usou um remédio para aliviar os quelóides de acne no meu rosto e me disse que tenho uma tendência a quelóides, caso um dia eu venha fazer a	
12	B BR	vivoblog.com.br	🔍	🗨️	🔍	DIA. Mas a primeira atendente não parece se lembrar do treinamento e me disse que minha linha é pré paga, por isso não tem plano. Ok,	

Os exemplos acima mostram uma busca por uma palavra ou expressão específicas, já selecionada anteriormente. No entanto, o corpus também pode mostrar **todas as palavras, expressões ou colocados correspondentes que sejam mais comuns em um país (ou grupo de países) que em outro**. Por exemplo, a tabela a seguir mostra as palavras com **ismo* que são mais comuns no Brasil (esquerda) que em Portugal (direita):

SEC 1 (Brazil): 558,357,341 WORDS

SEC 2 (Portugal): 270,234,249 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	PETISMO	840	2	1.5	0.0	203.3	1	SPORTINGUISMO	515	1	1.9	0.0	1,064.1
2	TRANSFEMINISMO	344	2	0.6	0.0	83.2	2	BENFIQUISMO	570	2	2.1	0.0	588.9
3	TUNGUISMO	444	0	0.8	0.0	79.5	3	DESPESISMO	565	5	2.1	0.0	233.5
4	EMPREGUISMO	314	2	0.6	0.0	76.0	4	PROFISSIONALISMO	195	2	0.7	0.0	201.5
5	NEOPENTECOSTALISMO	139	1	0.2	0.0	67.3	5	GUIONISMO	61	1	0.2	0.0	126.0
6	DENUNCISMO	126	1	0.2	0.0	61.0	6	CIRCUNSTANCIALISMO	60	1	0.2	0.0	124.0
7	ERGOTISMO	117	1	0.2	0.0	56.6	7	ESCUTISMO	266	5	1.0	0.0	109.9
8	PRÓ-HOMOSSEXUALISMO	114	1	0.2	0.0	55.2	8	FOSFENISMO	271	0	1.0	0.0	100.3
9	BOM-MOCISMO	97	1	0.2	0.0	46.9	9	PSICADELISMO	138	3	0.5	0.0	95.0
10	MASCULINISMO	246	0	0.4	0.0	44.1	10	MEDIATISMO	495	11	1.8	0.0	93.0
11	LULO-PETISMO	89	1	0.2	0.0	43.1	11	CHICO-PERTISMO	134	3	0.5	0.0	92.3
12	BOLIVARIANISMO	242	3	0.4	0.0	39.0	12	SIDONISMO	44	1	0.2	0.0	90.9

E a tabela a seguir mostra os colocados de *placa* que são mais comuns no Brasil (esquerda) ou em Portugal (direita), o que ajuda a entender o significado e o uso de *placa* nos dois países:

SEC 1 (Brazil): 558,357,341 WORDS

SEC 2 (Portugal): 270,234,249 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	VEÍCULO	222	1	0.4	0.0	107.4	1	ELIMINAÇÃO	27	1	0.1	0.0	55.8
2	VERMELHA	164	1	0.3	0.0	79.4	2	FORNO	26	1	0.1	0.0	53.7
3	IGREJA	172	2	0.3	0.0	41.6	3	GIRATORIA	45	2	0.2	0.0	46.5
4	PRETA	167	2	0.3	0.0	40.4	4	EPIFISÁRIA	20	1	0.1	0.0	41.3
5	COR	297	4	0.5	0.0	35.9	5	ACUMULAÇÃO	35	2	0.1	0.0	36.2
6	CAMINHÃO	61	1	0.1	0.0	29.5	6	MASSA	65	4	0.2	0.0	33.6
7	CARRO	263	5	0.5	0.0	25.5	7	GAMA	16	1	0.1	0.0	33.1
8	DENOMINAÇÃO	45	1	0.1	0.0	21.8	8	INDUÇÃO	45	3	0.2	0.0	31.0
9	FIAT	41	1	0.1	0.0	19.8	9	Á	14	1	0.1	0.0	28.9
10	ONBOARD	41	1	0.1	0.0	19.8	10	ELIMINA	27	2	0.1	0.0	27.9
11	HONDA	110	0	0.2	0.0	19.7	11	CABELO	13	1	0.0	0.0	26.9
12	MOTO	104	0	0.2	0.0	18.6	12	CONTROLO	25	2	0.1	0.0	25.8

3. Navegação e busca das 40 mil palavras mais comuns (voltar para o início)

No corpus Web/Dialetos, criamos uma lista cuidadosamente revisada das 40 mil palavras mais comuns no corpus de um bilhão de palavras, e é possível fazer buscas dessas palavras – por forma da palavra, parte do discurso, frequência da palavra e tradução (básica) para o inglês. Por exemplo, a tabela a seguir apresenta algumas palavras em três faixas de **frequência** diferentes: perto de 4.200 (ou seja, a 4.200^a palavra (lema) mais frequente no corpus), perto de 16.000 e perto de 35.000:

	RANK #	PoS	WORD	MORE DEFIN	IMAGE	VIDEO	WEB	ACAD	NEWS	FIC	SPOK
1	4209	V	COLETAR	Info Prib Mich Dic Ling WR Rev Coll Oxf			13113	30	9	0	1
2	4210	J	EMERGENTE	Info Prib Mich Dic Ling WR Rev Coll Oxf			13109	41	67	14	8
3	4211	J	SILENCIOSO	Info Prib Mich Dic Ling WR Rev Coll Oxf			13103	10	23	222	7
4	4212	V	COZER	Info Prib Mich Dic Ling WR Rev Coll Oxf			13103	1	9	3	2
5	4213	N	RODADA	Info Prib Mich Dic Ling WR Rev Coll Oxf			13102	29	2	16	3
6	4214	N	DECEPÇÃO	Info Prib Mich Dic Ling WR Rev Coll Oxf			13099	5	24	55	13
7	4215	N	PROCESSAMENTO	Info Prib Mich Dic Ling WR Rev Coll Oxf			13097	594	40	0	11

	RANK #	PoS	WORD	MORE DEFIN	IMAGE	VIDEO	WEB	ACAD	NEWS	FIC	SPOK
1	16020	J	FULGURANTE	Info Prib Mich Dic Ling WR Rev Coll Oxf			1070	6	4	45	1
2	16021	V	IMOLAR	Info Prib Mich Dic Ling WR Rev Coll Oxf			1069	2	25	3	0
3	16022	N	ADVERSÁRIA	Info Prib Mich Dic Ling WR Rev Coll Oxf			1069	3	32	2	2
4	16023	N	DEPOENTE	Info Prib Mich Dic Ling WR Rev Coll Oxf			1069	61	1	1	1
5	16024	V	DESONERAR	Info Prib Mich Dic Ling WR Rev Coll Oxf			1069	1	2	2	1
6	16025	N	FIGURAÇÃO	Info Prib Mich Dic Ling WR Rev Coll Oxf			1068	23	3	16	1
7	16026	V	APEAR	Info Prib Mich Dic Ling WR Rev Coll Oxf			1068	0	1	47	0

	RANK #	PoS	WORD	MORE DEFIN	IMAGE	VIDEO	WEB	ACAD	NEWS	FIC	SPOK
1	35000	V	FLIRTAR	Info Prib Mich Dic Ling WR Rev Coll Oxf			79	0	1	0	0
2	35001	J	ALUADO	Info Prib Mich Dic Ling WR Rev Coll Oxf			79	0	0	6	4
3	35002	J	ESOTÉRICO	Info Prib Mich Dic Ling WR Rev Coll Oxf			79	9	8	10	6
4	35003	N	AFERIDOR	Info Prib Mich Dic Ling WR Rev Coll Oxf			79	1	0	1	0
5	35004	J	AVINAGRADO	Info Prib Mich Dic Ling WR Rev Coll Oxf			79	0	0	1	0
6	35005	R	ENCANTADORAMENTE	Info Prib Mich Dic Ling WR Rev Coll Oxf			79	0	1	2	0
7	35006	N	NEURALGIA	Info Prib Mich Dic Ling WR Rev Coll Oxf			79	1	0	0	0

As palavras a seguir são palavras com **ista* menos frequentes (abaixo do número 20.000 da lista):

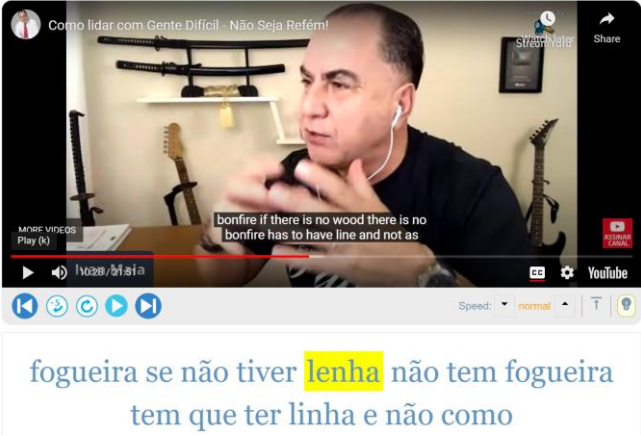
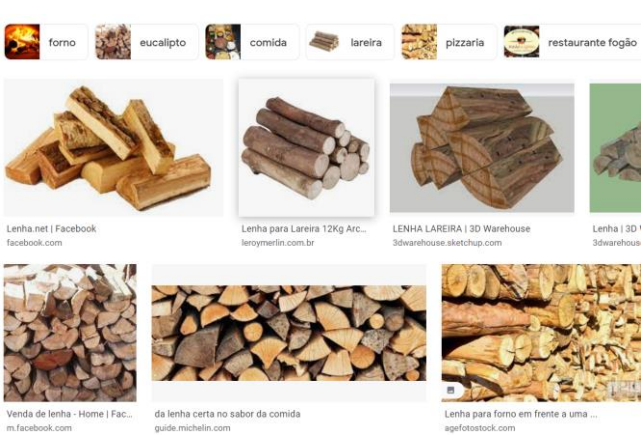
	RANK #	PoS	WORD	MORE DEFIN	IMAGE	VIDEO	WEB	ACAD	NEWS	FIC	SPOK
1	20007	J	UTILITARISTA	Info Prib Mich Dic Ling WR Rev Coll Oxf			583	18	0	4	2
2	20057	J	NEOCOLONIALISTA	Info Prib Mich Dic Ling WR Rev Coll Oxf			579	2	0	0	1
3	20066	N	PACIFISTA	Info Prib Mich Dic Ling WR Rev Coll Oxf			578	5	4	3	1
4	20102	J	REVISIONISTA	Info Prib Mich Dic Ling WR Rev Coll Oxf			574	4	2	0	4
5	20131	J	FANTASISTA	Info Prib Mich Dic Ling WR Rev Coll Oxf			571	1	1	1	1
6	20190	N	PUGILISTA	Info Prib Mich Dic Ling WR Rev Coll Oxf			567	13	4	2	2
7	20217	N	LINGUISTA	Info Prib Mich Dic Ling WR Rev Coll Oxf			565	14	3	1	2

Já as palavras a seguir têm *wood* em sua glosa básica em inglês:

	RANK #	PoS	WORD	DEFINITION	MORE DEFIN	IMAGE	VIDEO	WEB	ACAD	NEWS	FIC	SPOK
1	2019	N	MADEIRA	<u>wood</u> , (timber)	Info Prib Mich Dic Ling WR Rev Coll Oxf			39293	171	409	354	329
2	3100	N	PAU	<u>wood</u> , stick, cock, (bar)	Info Prib Mich Dic Ling WR Rev Coll Oxf			21753	9	74	233	32
3	4150	N	MATA	forest, <u>wood</u> , (thumb)	Info Prib Mich Dic Ling WR Rev Coll Oxf			13418	23	112	135	63
4	6309	N	LENHA	fire <u>wood</u> , (fuel)	Info Prib Mich Dic Ling WR Rev Coll Oxf			6764	6	11	74	5
5	7740	N	BOSQUE	forest, <u>wood</u> , grove, (<u>wood</u> land)	Info Prib Mich Dic Ling WR Rev Coll Oxf			4758	84	16	55	1
6	9410	N	CARIMBO	stamp, imprint, postmark, (<u>wood</u>)	Info Prib Mich Dic Ling WR Rev Coll Oxf			3299	1	16	9	3
7	10524	N	CARPINTEIRO	carpenter, (<u>wood</u> worker)	Info Prib Mich Dic Ling WR Rev Coll Oxf			2669	4	4	48	1

Para cada palavra da lista, é possível ver a posição de classificação (1 a 40.000), a parte do discurso, a glosa básica em inglês, links para vários **dicionários** on-line, **imagens** do Google, **vídeos** do Youglish.com (em português), a

frequência geral e por gênero no corpus Gênero/Histórico. E, claro, também há um link para o *word sketch* da palavra (veja a seção a seguir). Por exemplo, a seguir são mostrados imagens, vídeos e definições para *lenha*, uma das palavras na lista acima:



lenha /ɛ/

le.nha • 'le(j)ɲɐ, 'leɲɐ

nome feminino

1. madeira utilizada para alimentar a combustão; madeira
2. popular surra; pancada; espancamento
3. *Brasil coloquial* dificuldade; complicação; problema
4. *Brasil gíria* prova automobilística

arranjar lenha para se queimar

trabalhar em seu próprio prejuízo

Brasil popular **baixar lenha em**

1. bater; surrar
2. criticar algo ou alguém com dureza

deitar lenha na fogueira

exaltar os ânimos; avivar uma discórdia, uma disputa, etc.; acicatar

Brasil popular **entrar na lenha**

apanhar pancada

i Do latim *ligna*, plural de *lignu-*, «lenha»

Traduções principais		
Português	Inglês	
lenha <i>sf</i>	(madeira para fazer fogo)	firewood <i>n</i> kindling <i>n</i>
Está faltando alguma coisa importante ? Notifique-nos a respeito de erros ou sugestões para que possamos aprimorar o nosso sistema.		
WordReference English-Portuguese Dictionary © 2022:		
Entradas correspondentes do outro lado do dicionário		
Traduções principais		
Inglês	Português	
firewood <i>n</i>	(wood for fuel) Matt keeps a stack of firewood in his garage during winter.	lenha <i>sf</i>
deadwood <i>n</i>	(dead parts of a tree)	lenha <i>sf</i>
log <i>n</i>	(firewood) He grabbed four logs to throw on the fire. (1) Ele pegou lenha para jogar na fogueira.	(sing) lenha <i>sf</i>

4. Word sketches (voltar para o início)

Para cada uma das 40 mil palavras (lemas) mais comuns do corpus, é possível obter uma grande variedade de informações, que incluem: [1] palavra (lema) [2] informação de frequência (posição de classificação, número 1 a 40.000) [3] links para imagens e vídeos (com pronúncia) [4] glosas simples em inglês, [5] links para registros em dicionários monolíngues detalhados (inclusive vários com etimologias) [6] frequência por gênero (do corpus Gênero/Histórico) [7] adição da palavra à lista de “favoritos” (onde as palavras podem ser ordenadas e agrupadas) e Histórico (para ver palavras recentes) [8] sinônimos (é possível clicar em qualquer sinônimo, colocado ou tópico para ver a página correspondente da palavra) [9] colocados (palavras vizinhas; 4 à esquerda a 4 à direita) [10] tópicos (palavras que coocorrem em qualquer parte dos mais de 1 bilhão de textos do corpus) [11] frequência por dialeto [12] frequência das formas da palavra (possivelmente mais útil para verbos) [13] 200 exemplos de linhas de concordância reordenáveis.

1

CARTÃO *n* (RANK 1121, FREQ 76536)

2

DEFINITIONS (click for more)

card. cardboard. (board)

4

MORE: I P M D L W R C O

3

SPOK

FIC

NEWS

ACAD

5

COLLOCATES

(NOUN) crédito, visita, débito, pagamento, cartão, número, cidadão, dado, caixa, compra

(ADJ) amarelo, postal, vermelho, internacional, bancário, magnético, válido, corporativo, automático, especial

(MISC) pagar, aceitar, creditar, já, só, pré-pagar, até, imprimir, emitir, cancelar

7

☆

🔄

✖

cartão, bilhete, boletim, bilhete, bilhete

8

SYNONYMS (click to see word)

BEST: bilhete, carta, carteira, convite, crachá, credencial, distintivo, ficha, identidade, identificação, passagem, passe, postal, senha, OTHERS: alvará, autorização, boletim, botão, caixa, caixote, canhoto, cartolina, central, chave, correio, correspondência, denominação, desfiladeiro, determinante, embalagem, essencial, formulário, fundamental, id, identificador, importante, ingresso, insígnia, legenda, letra, licença, ofício, papelão, pedido, permissão, porta, principal, recibo, símbolo, talão,

9

TOPICS (click to see)

crédito *n* compra *n* pagamento *n* cliente *n* loja *n* banco *n* desconto *n* conta *n* taxa *n* caixa *n* preço *n* telefone *n* pedido *n* débito *n* viagem *n* comprar *v* aparelho *n* bancário *j* juro *n* dado *n* consumidor *n* pagar *v* enviar *v* correio *n* aplicativo *n* celular *j* hotel *n* gasto *n* gratuito *j* bilhete *n*

11

BRAZIL

PORTUGAL

ANGOLA

MOZAMBIQUE

12

FORMS

60,625 cartão 19,663 cartões 93 cartao

13

CONCORDANCE LINES PROBLEMS?

	GENRE		SORT	SORT	SORT
1	PT G	carris.pt	Zapping O Zapping consiste em o carregamento de dinheiro em o	cartão	7 Colinas Viva Viagem ou Lisboa Viva , que vai sendo
2	BR G	ipaddicas.com	Grato desde já , Fernando . Eu venho comprando os meus	cartões	a mais de 1 ano , com o pessoal de a http://www.
3	BR B	pix.mbzhost.com	o prazo de 15 dias uteis tenham se passado e o	cartão	ainda não tenha sido recebido , o procedimento é ligar para o
4	PT B	briosapress.blogspot.com	respeito a o clube com o qual tem contrato ! 24'	CARTÃO	AMARELO para Fabrice Fokobo (Sporting) 24' PENALTY ASSINALADO
5	BR B	sportv.globo.com	Em esse caso , o Rogério Ceni deveria ter tomado o	cartão	amarelo porque estava com uma bola em a mão e a jogou
6	PT B	ad-ninense.blogspot.com	disciplina e respeito mutuo de os jogadores sem mostrar um	cartão	amarelo é algo de louvável e muito raro de acontecer hoje em
7	PT B	ruitavares.net	de uma confusão inaudita , e o Presidente já mostrou vários	cartões	amarelos Em as bancadas especula-se que a intenção é mesmo
8	MZ B	noainacio.blogspot.com	visto que o povo atento de a Beira os daria um	cartão	bastante vermelho em outros pleitos eleitorais para o
9	PT B	lucrandoleal.com	sobre a promoção de a BB e a utilização de o	cartão	BB em lessas promoções . Utilizem a página BB Watch (disponível
26	PT G	destinia.com	Destinia.com , por seguridade a os seus passageiros e a os	cartões	de crédito de os mesmos , pode solicitar-se cópia de a
27	PT G	help.eventbrite.pt	, dado que o portal de pagamentos não guarda informações de	cartões	de crédito durante mais de seis meses . Nota : Por predefinição
28	BR B	minimalizo.blogspot.com	de paciência , para ligar para o atendimento de o meu	cartão	de crédito e negociar a anuidade . Que perda de tempo ...
29	BR B	acertodecontas.blog.br	31209 Super dívida em o	cartão	de crédito e nome em o Serasa . O que fazer ?
30	MZ G	sautulisslam.com	a conta bancária pode usufruir de as facilidades de o seu	cartão	de crédito e ter o valor gasto disponível em a conta antes
31	PT B	comunit.com	Pedimos- lhe então que nos envie um scan de o seu	cartão	de crédito e de o cartão de estudante de a sua universidade
32	BR G	amato.com.br	ficou abaixo de setores como serviços de assistência técnica ,	cartões	de crédito e operadoras de telefonia celular , com nota geral de
33	BR G	ipaddicas.com	é em torno de R\$ 4 por mês . Ter um	cartão	de crédito e economizar dinheiro em a AppStore é uma opção .

O item [3], imagens e vídeos, leva para recursos externos:

aniversário

desenho

agradecimento

visita

empresa

Nubank

negativados disponibilizado ...
jornalcontabil.com.br

Cartão Nubank: cartão de crédito sem ...
blog.nubank.com.br

Nubank: 3 dicas para aumentar o limit...
seucreditodigital.com.br

Com canal

Caixa Econômica Federal

Cartões Nacionais MasterCard / VL...
caixa.gov.br

Cartão Platinum - Cartões de Crédi...
caixa.gov.br

Cartões CAIXA | CAIXA
caixa.gov.br

Cartão CAIXA Leve | CAIXA
caixa.gov.br

Cartão de crédito: você sabe utilizá-lo a seu favor?

Watch later

Share

MORE VIDEOS

0:29 / 6:08

cc

Settings

YouTube

Speed: normal

T

Light

faço meu controle semanal da minha fatura do

cartão de crédito.

É possível encontrar palavras “aleatórias” (usando o ícone de “dado” no topo do corpus) e definir a faixa de palavras aleatórias, como palavras nas posições 15 mil a 20 mil da lista de frequência (útil para aprendizes do idioma).

TRIBUTO *n* (RANK 4681, FREQ 11169)

DEFINITIONS (click for more)

tribute, (tax)

MORE: I P M D L W R C O

SPOK

FIC

NEWS

ACAD

0k 1k 3k 6k 10k 15k 20k

tributo, besteira, sintético, soquete, adorador, micronutriente

SYNONYMS (click to see word)

BEST: contribuição, elogio, fiscal, fiscalidade, homenagem, honra, imposto, louve

OTHERS: comemoração, cotação, custo, desconto, discurso, elegia, finança, glóri

orgulho, pagamento, panegírico, portagem, preço, prêmio, privilégio, rendimento

COLLOCATES

(NOUN) pagamento, imposto, contribuição, arrecadação, recolhimento, espécie, cobrança, tributo, taxa, isenção

(ADI) federal, relativo, incidente, estadual, legal, devido, competente, efetivo,

TOPICS (click to see)

tributário j imposto n contribuinte r

tributação n alíquota n contribuiçã

n crédito n recolhimento n tributar

Também é possível ver “Palavras do dia” em diferentes faixas de frequência:

Click below to see detailed information on word

Frequency band	Rank #	Word	PoS
1 1-999	214	produto	n
	314	música	n
2 1000-2999	1626	avisar	v
	2305	tabela	n
3 3000-5999	3015	muro	n
	4767	sombrio	j
4 6000-9999	9753	belga	j
	9880	desocupar	v
5 10000-14999	14433	mitigação	n
	14997	body	n
6 15000-19999	15744	haitiano	j
	16082	muçulmana	n
7 20000-29999	21007	autoclismo	n
	27527	patelar	j

22-05-21 ◀ ▶ 22-05-23

"Words of the Day" for [22-05-22]

The 20 words to the left were selected at random by the computer from the top 40,000 words in PORT, with two words in each of the seven frequency ranges.

You might want to come back to this page every day and see how many of the words you know. You can also expand your vocabulary by focusing on words that are unfamiliar, and then "save" these words for later review via on the "word" page. .

You can also click below to see the words for past days, from 21-04-01 through 22-06-22 (one month from today).

Search by date

To get to this page in the future, click on and then .

You can also find many random words for each of the ten frequency ranges (e.g. words 3000-6000) via the "word page".

Or [browse](#) to focus on an even more narrow frequency band, or to search by word form, pronunciation, meaning, synonyms, etc.

5. Links para recursos externos (inclusive traduções) (voltar para o início)

Conforme mostrado nas seções anteriores de “navegação” e “word sketches”, há muitos links para recursos externos. Isso também ocorrem em outras páginas do corpus. Por exemplo, na página básica de “resultados” que aparece após a busca, a ação predefinida é a visualização das linhas de concordância para a palavra, expressão ou colocado. Mas também é possível optar pela visualização de recursos externos para o registro: tradução (mais de 100 idiomas no Google Translate), busca no Google, vídeo/pronúncia ou busca no Google Books

ON CLICK:		CONTEXT	TRANSLATE (ES)	ENTIRE PAGE	GOOGLE	IMAGE	PRON/VIDEO	BOOK	(HELP)
HELP	1	★	ALL FORMS (SAMPLE): 100 200 500 WORDS			FREQ	TOTAL 9,013 UNIQUE 1,693		
1	1	★	FORMA SUAVE			381			
2	1	★	VOZ SUAVE			324			
3	1	★	MÚSICA SUAVE			215			
4	1	★	BRISA SUAVE			193			
5	1	★	MOVIMENTOS SUAVES			172			
6	1	★	TOQUE SUAVE			172			
7	1	★	CHEIRO SUAVE			168			
8	1	★	PALAVRAS SUAVES			160			
9	1	★	CORES SUAVES			156			
10	1	★	PELE SUAVE			154			

Imagens

Vídeos

Livros

Também é possível ver a tradução de todos os registros da página:

ON CLICK:

CONTEXT

TRANSLATE (ES)

ENTIRE PAGE

HELP	1	★	ALL FORMS (SAMPLE): 100 200 500 V		
1	1	★	FORMA SUAVE		
2	1	★	VOZ SUAVE		
3	1	★	MÚSICA SUAVE		
4	1	★	BRISA SUAVE		
5	1	★	MOVIMENTOS SUAVES		
6	1	★	TOQUE SUAVE		
7	1	★	CHEIRO SUAVE		

Choose a language to translate to:

English

Submit

Danish

Dutch

English

Esperanto

Estonian

Filipino

Finnish

Google Translate

Portuguese → English

381 forma suave	smooth shape
324 voz suave	soft voice
215 música suave	soft music
193 brisa suave	gentle breeze
172 movimentos suaves	smooth movements
172 toque suave	smooth touch
168 cheiro suave	soft smell
160 palavras suaves	soft words
156 cores suaves	soft colors
154 pele suave	smooth skin
140 sabor suave	mild flavor
133 maneira suave	smooth way
118 declives suaves	gentle slopes

Na página de contexto/concordância/KWIC, é possível ver a tradução de qualquer linha de texto [1] e escutar a pronúncia do texto em português [2]:

CLICK FOR MORE CONTEXTHELP

1	B PT	aondadarita.com	o rosto, Piz Buin Allergy conta o Creme de Rosto que mantém a pele suave, hidratada e protegida
2	G PT	beleza-natural2012.blogspot.com	e F.=Recomendado para o tratamento das mãos secas e envelhecidas; Limpa a pele suave e eficaz
3	G BR	beleza.umcomo.com.br	, é imprescindível prestar atenção ao cuidado do rosto para brilhar com uma pele suave e radiante
4	B PT	blog.antonioiuliscampos.com	que hoje abundam. Depois, banho de mangueirada, água gelida na pele suave de criança, bem lá
5	G PT	braun.braun.com	corporais são igualmente uma boa forma de remover as células mortas e manter a pele suave. Ap
6	G PT	braun.braun.com	utilize o aparelho uma vez de dois em dois meses, para manter a pele suave, dia após dia. Até um,
7	G PT	braun.braun.com	fácil. As 20 pinças removem os pêlos pela raiz, deixando a sua pele suave. Ao mesmo tempo que a
8	G PT	braun.braun.com	mesmo os pelos mais curtos pela raiz. Para que possa ter uma pele suave sempre que quiser. Sen
9	G PT	braun.braun.com	corporal enriquecida nas pernas das modelos. É a base para uma pele suave e brilhante e cria aqu

DETECT LANGUAGEPORTUGUESEENGLISHSPANISH

SPANISHENGLISHPORTUGUESE

vital e a saúde do corpo. Os lípidios ajudam a manter a pele suave e a fortalecer a barreira cutânea natural que a protege de as agressões externas.

vitality and the health of the body. Lipids help keep the skin smooth and strengthen the skin's natural barrier that protects it from external aggressions.

148 / 5,000

Por último, pode-se obter a tradução a partir da página de “Contexto expandido”:

1278213 Como eliminar a celulite A celulite é um mal comum e uma de as grandes inimigas de as mulheres. Apesar de também ocorrer em alguns corpos masculinos, a celulite assombra o mundo feminino. Estima-se que cerca de 90 % de as mulheres seja afectado por qualquer tipo de celulite, a partir de a adolescência. Não é doença e muito menos é algo grave, mas o facto é que as pequenas ondulações em a pele, a o qual damos o nome de celulite, são causa para uma auto-tima inferior e para a procura incessante de ter o corpo perfeito. O aspecto de casca de laranja em a pele, causado por o celulite, é totalmente diferente de a pele suave e definida que as mulheres desejam ter. O que é a celulite O que caracteriza a celulite é a chamada degeneração de o tecido gorduroso, que tem algumas deficiências em a circulação. As covinhas são criadas através de as saliências de a gordura hipodérmica em a pele. Ou seja, a gordura localizada. Como o tecido adiposo de a hipoderme se deposita em grandes feixes verticais, todos esses feixes separam-se através de os septos fibrosos, que existem em a pele com celulite, e que separam as células gordurosas em grupos. Este processo, localizado em a hipoderme, provoca o temido efeito de “a casca de laranja”. Apesar de falarmos em gordura localizada, este processo não acontece, necessariamente, a pessoas gordas. A celulite afecta pessoas gordas e magras, não se restringindo a um espectro de a população. Desengane-se quem acha que uma pessoa gorda tem celulite

1278213 How to eliminate cellulite Cellulite is a common evil and one of the great enemies of women. Although it also occurs in some male bodies, cellulite haunts the female world. It is estimated that around 90% of women are affected by any type of cellulite, starting in adolescence. It's not a disease and even less is it something serious, but the fact is that the small ripples in the skin, which we call cellulite, are the cause of an inferior self-esteem and the incessant search for the perfect body. The appearance of orange peel on the skin, caused by cellulite, is totally different from smooth skin, and defined that women want to have. What is cellulite What characterizes cellulite is the so-called degeneration of fatty tissue, which has some deficiencies in circulation. The dimples are created through the pits of hypodermic fat in the skin. In other words, localized fat. As the adipose tissue of the hypodermis is deposited in large vertical bundles, all these bundles are separated through the fibrous septa, which exist in skin with cellulite, and which separate the fat cells into groups. This process, located in the hypodermis, causes the dreaded effect of “orange peel”. Although we talk about localized fat, this process does not necessarily happen to fat people. Cellulite affects both fat and thin people and is not restricted to a spectrum of the population.

6. Corpora Virtuais (voltar para o início)

É possível criar “Corpora Virtuais”, uma espécie de “corpus dentro do corpus”, de forma rápida e simples. Por exemplo, pode-se buscar artigos com a expressão meio ambiente ou buscar artigos do Brasil na parte “Geral” (não blog) do corpus Web/Dialetos que tenham nuclear no título:

ListChartWordBrowse +

meio ambiente [POS]?
Find matching stringsReset

☐ SectionsTexts/VirtualSort/Lin

FIND TEXTS
MY CORPORA

Create corpus
Edit corpora

Web domain	<input type="text"/> (can use substring, e.g. *futbol*, *.mx)	Find sources
Article title	nuclear	
Country	<div>-IGNORE- Brasil Portugal Angola Mozambique</div>	
Genre	General	
Words in text	<input type="text"/>	
SubmitReset		

Um ou dois segundos depois, aparece uma lista dos textos com a melhor correspondência. Basta dar um nome para o corpus, e o Corpus Virtual terá sido criado.

(Optional) SAVE AS: <input type="text" value="meio_ambiente"/>		OR ADD TO: <input type="text" value="--SELECT--"/>	<input type="button" value="SUBMIT"/>	<input type="button" value="RESET"/>	HELP		
HELP	<input type="checkbox"/> 100	TEXT	SEE PAGE	# WORDS	# HITS ↓	RELEVANCE ↓	PER MILLION WORDS
1	<input checked="" type="checkbox"/>	G BR: MONTESCLAROS.COM		273739	101	369.0	<div></div>
2	<input checked="" type="checkbox"/>	B BR: AMBIENTALSUSTENTAVEL.ORG		16997	80	4,706.7	<div></div>
3	<input checked="" type="checkbox"/>	B BR: BLOGS.UNIGRANRIO.COM.BR		20912	74	3,538.6	<div></div>
4	<input checked="" type="checkbox"/>	G PT: SLIDESHARE.NET		22959	68	2,961.8	<div></div>
5	<input checked="" type="checkbox"/>	G BR: ECCLESIA.COM.BR		7887	66	8,368.2	<div></div>
6	<input checked="" type="checkbox"/>	G BR: AMBITO-JURIDICO.COM.BR		6089	63	10,346.5	<div></div>
7	<input checked="" type="checkbox"/>	G BR: ARTEESOCIEDADE.COM		30681	49	1,597.1	<div></div>
8	<input checked="" type="checkbox"/>	G BR: MMA.GOV.BR		4447	49	11,018.7	<div></div>
9	<input checked="" type="checkbox"/>	G BR: AMBITO-JURIDICO.COM.BR		11480	46	4,007.0	<div></div>
10	<input checked="" type="checkbox"/>	B BR: BLOGS.UNIGRANRIO.COM.BR		19666	44	2,237.4	<div></div>
11	<input checked="" type="checkbox"/>	G BR: SURJOURNAL.ORG		9902	42	4,241.6	<div></div>
12	<input checked="" type="checkbox"/>	G BR: MPRS.MP.BR		16836	42	2,494.7	<div></div>

Após a criação do Corpus Virtual, é possível limitar as futuras buscas a esse corpus (para encontrar correspondências de expressões ou colocados, por exemplo) e organizar os corpora de várias maneiras. E bastam um ou dois segundos a mais para encontrar “palavras-chave” do Corpus Virtual, conforme mostrado abaixo para substantivos nos corpora [meio ambiente] e [nuclear] que acabamos de criar. Também é possível ver a tradução de todas as palavras-chave em qualquer idioma.

MEIO_AMBIENTE [3,836,623 WORDS, 100 TEXTS] (0.4% OF TOTAL)		NOUN	VERB	ADJ	ADV	N+N	ADJ+N	[ALL CORPORA] SAVE LIST	
HELP	SAVE	WORD (CLICK FOR CONTEXT) TRANSLATE ALL ENTRIES	FREQ	# TEXTS	FREQ	SPECIFIC	TEXTS	ENTIRE CORPUS	EXPECTED
					90	10			
1	★	RESÍDUO	678	20	72.8			2,428	9.3
2	★	MILHAR	280	10	41.5			1,759	6.7
3	★	HABITANTE	323	11	24.5			3,430	13.2
4	★	TONELADA	173	10	17.4			2,589	9.9
5	★	MORADOR	386	16	13.9			7,216	27.7
6	★	ATERRO	117	10	13.2			2,316	8.9
7	★	POLUIÇÃO	544	35	10.1			14,085	54.0
8	★	TRABALHADOR	1312	11	9.2			37,024	142.0
9	★	ECOSSISTEMA	155	13	8.9			4,553	17.5

NUCLEAR [107,793 WORDS, 100 TEXTS] (0.0% OF TOTAL)		NOUN	VERB	ADJ	ADV	N+N	ADJ+N	[ALL CORPORA] SAVE LIST	
HELP	SAVE	WORD (CLICK FOR CONTEXT) TRANSLATE ALL ENTRIES	FREQ	# TEXTS	FREQ	SPECIFIC	TEXTS	ENTIRE CORPUS	EXPECTED
					20	5			
1	★	REATOR	279	20	1,758.4			1,472	0.2
2	★	FISSÃO	88	9	1,607.0			508	0.1
3	★	URÂNIO	114	10	407.1			2,598	0.3
4	★	USINA	240	17	388.7			5,728	0.6
5	★	ELETRICIDADE	46	5	51.1			8,357	0.9
6	★	FUSÃO	65	7	50.8			11,880	1.3
7	★	NÚCLEO	107	9	46.0			21,556	2.3
8	★	COMBUSTÍVEL	79	5	35.1			20,864	2.2
9	★	ARMA	124	10	32.8			35,043	3.8
10	★	ENERGIA	551	30	28.6			178,595	19.3
11	★	REAÇÃO	91	8	23.4			36,053	3.9

7. Análise de textos (voltar para o início)

Também é possível analisar textos completos usando os dados do corpus. Por exemplo, você pode copiar um artigo de um jornal on-line e colar no formulário de “Análise de texto” (usando o ícone “Analyze”, no topo do corpus)

Esper Kallás
Médico infectologista, é professor titular do departamento de moléstias infecciosas e parasitárias da Faculdade de Medicina da USP e pesquisador na mesma universidade.

SEGUIR

Varíola de macacos e a ameaça de nova pandemia
O que já sabemos sobre o vírus e a doença

Varíola de macacos e a ameaça de nova pandemia
O que já sabemos sobre o vírus e a doença

SEARCH CLEAR HELP

From this page, you can find keywords word in the text to synonyms, collocates (that word). You can also find phrases in the Corpus.

There are several ways

1. Just copy and paste (Word file) into the box.
2. Click on SAMPLES, as you choose. You can then find that country.

Após clicar em Submit, a página seguinte mostra [1] o número de palavras nas faixas de frequência alta/média/baixa (para verificar a “dificuldade” do texto) [2] as palavras mais frequentes em cada uma dessas faixas de frequência [3] o texto, com as palavras menos frequentes (no corpus todo) destacadas em verde e principalmente amarelo.

EDIT TEXT SAVE TEXT

WORD PHRASE

FREQ RANGE	1-500	501-3000	> 3000
415 WORDS	69 %	18 %	7 %

Varíola de macacos e a ameaça de nova pandemia
O que já sabemos sobre o vírus e a doença

Aconteceu com um brasileiro de 26 anos, que chegou à Bavária, na Alemanha, depois de passar por Portugal e Espanha. A presença de vesículas na pele levantou a suspeita de ser um caso da varíola do macaco. Autoridades alemãs confirmaram se tratar do primeiro caso da doença naquele país. Com este, já são 15 os países que registram ao menos um caso, até a conclusão deste texto.

Ainda vivendo os rescaldos da pandemia de Covid-19, surge uma nova ameaça de doença transmissível, que vem chamando a atenção da mídia internacional. Em inglês: "monkeypox".

(CLICK TO SEE WORD SKETCH)

LOW FREQ	MID FREQ	HIGH FREQ
4: varíola 3: vesículas 1: acometidas, antivirais, branda, calafrios, combatem, crostas, denominadas, detectam, imprecisos, linfáticos, medicações, naquele, negligenciadas, nódulos, pandemia, purulentas, registram, roedores, transmissível, vacinados	5: vírus 3: macacos 2: ameaça, direto, macaco, ocorrer, testes 1: absolutamente, alemãs, aprovação, aproximadamente, assusta, bolhas, causador, chamando, cicatrizes, comprometimento, comumente, concentrados, confere, confirmar, confirmaram, cooperação, décadas, defeitos, deste, dores, durar, erradicação, espalhadas, febre, felizmente, genético, grávidas, identificado, importados, incomum, inesperada, inflamação, lesões, levantou, limitado, músculos,	20: de 15: a 11: e, que 8: com, o 7: da, em, para 5: caso, do, doença 4: anos, os, pele, se, uma 3: as, foi, mais, na, no, podem, por, um 2: ao, casos, contato, embora, internacional, menos, nova, ou, países, parte, pelo, pode, primeiro, são, ser, seu, sistema, também, tornar 1: abaixo, acesso, aconteceu, ainda, além, alguns, apenas, apresentação, atenção, autoridades, bastante, boa,

É possível então clicar em qualquer uma das palavras [2] nas faixas de frequência ou [3] no texto em si para ver o word sketch:

PANDEMIA n (RANK 11491, FREQ 2238)

DEFINITIONS (click for more)
(pandemic)

MORE: I P M D L W R C O

SYNONYMS (click to see word)

BEST: doença, epidemia, infecção, praga, surto,
OTHERS: contágio, contaminação, crise, deflagração, peste, vírus,

FIC NEWS ACAD

COLLOCATES

(NOUN) gripe, epidemia, democracia, vírus, doença, luta, pandemia, influenza, combate, ameaça
(ADJ) global, mundial, suíno, eventual, falso, bubônico, devastador, considerado, viral, espanhol
(MISC) gerar, causar, afetar, provocar, já, declarar, só, espalhar, travar, derrubar

TOPICS (click to see)

vírus n gripe n vacina n infectar v infecção n epidemia n doença n suíno j ave n vacinação n influenza n vacinar v sintoma n transmissão n respiratório j sazonal j contaminar v mutação n porco n surto n célula n febre n prevenção n subtipo n transmitir v imunidade n adjuvante j complicação n clínico j dose n

Também podem-se encontrar expressões relacionadas no corpus de um bilhão de palavras. Por exemplo, pode-se realçar a expressão *ameaça de nova pandemia* no texto [1] para encontrar expressões semelhantes no corpus. Isso insere as palavras em [2], e é possível clicar nos botões abaixo para [3] adicionar ou deletar palavras, encontrar

todas as formas de uma palavra, palavras da mesma parte do discurso, sinônimos de uma palavra da expressão, etc. Por último, são localizadas sequências de caracteres correspondentes no corpus [4]:

EDIT TEXT

SAVE TEXT

WORD

PHRASE

FREQ RANGE

1-500

501-3000

> 3000

415 WORDS

69 %

18 %

7 %

ameaça

de

NOUN

+ ADD

- DELETE

*

FORMS

SYNONYMS

POS:

NOUN

VERB

ADJ

ADV

PREP

Other:

SUBMIT

RESET

HELP

1

2

3

4

Resumo (voltar para o início)

O Corpus do Português contém os maiores corpora disponíveis da língua portuguesa de qualquer fonte, cuidadosamente anotados. O corpus permite a realização de buscas por palavra, expressão, lema (todas as formas de uma palavra), parte do discurso, sinônimo e lista personalizada de palavras, além da visualização de palavras, sequências de caracteres ou colocados correspondentes em apenas um ou dois segundos. Os corpora podem ser utilizados para observar variações por gênero, período histórico e dialeto. Também é possível criar “Corpora Virtuais” relacionados a qualquer tópico (ou com base nos metadados dos textos) de forma rápida e fácil, e em seguida limitar as buscas a esse Corpus Virtual ou visualizar as palavras-chave do Corpus Virtual.

Os corpora permitem fazer buscas “orientadas por palavras” de formas que são impossíveis com qualquer outro corpus. Pode-se navegar e buscar as 40 mil palavras (lemas) mais comuns do corpus por forma da palavra, parte do discurso, faixa de frequência e glosa básica em inglês. Para cada uma das 40 mil palavras mais comuns, é possível visualizar um *word sketch*, o qual oferece uma quantidade incrível de informações, como informação de frequência, definição, sinônimos, colocados, tópicos relacionados, linhas de concordância e links para imagens e vídeos externos. Também é possível analisar textos completos e em seguida clicar em qualquer palavra do texto para visualizar o *word sketch* ou encontrar expressões relacionados no corpus. E, em todas essas buscas, há uma grande variedade de links para dados externos muito úteis (traduções, definições, imagens, vídeos, etc.).

Em resumo, os corpora do Corpus do Português permitem que pesquisadores, professores e estudantes entendam melhor a língua portuguesa, o que não está disponível com nenhum outro recurso.

Nossos agradecimentos ao [Beatriz Baker Méio](#), que traduziu este documento do inglês para o português.