

XII Semana da Filologia na USP

PARTE 2: CORPORA ON-LINE E FERRAMENTAS DA LC: WL,
KW E CONCORDÂNCIAS

Objetivos

- *Corpora online*
 - Corpus do Português e *Corpora do Projeto CoMET*
- Ferramentas para *corpora offline*
 - AntConc e análise de *corpora offline*

Projeto CoMET

- <http://comet.fflch.usp.br/>



Projeto

Equipe

CorTec

CorTrad

CoMAprend

Publicações

Corpora Online

Bem-Vindos

PROJETO **CO** rpus

M ultilíngue para

E nsino e

T radução



Projeto Equipe CorTec CorTrad CoMAprend Publicações Corpa

Início

Corpora Online

LINKS ÚTEIS DE LINGUÍSTICA

[Corpora de Alemão](#)

[Corpora de Espanhol](#)

[Corpora de Francês](#)

[Corpora de Inglês](#)

[Corpora de Italiano](#)

[Corpora de Português](#)

[Corpora Multilingue](#)

[Ferramentas de Exploração de Corpora](#)

[Informações sobre Linguística de Corpus](#)

[Miscelânea](#)

Quer sugerir um link? Encontrou um link não operante?

Por favor, escreva para: projetocomet@usp.br



Projeto Equipe CorTec CorTrad CoMAprend Publicações Corpora Online

Início

Corpora de Português

Banco do Português: site brasileiro com um *corpus* de português contemporâneo, atualizado constantemente (em 2004, possuía 223 milhões de palavras). Apenas uma amostra está disponível para consulta e uso com as próprias ferramentas do site.

CetenFOLHA (*Corpus* de Extractos de Textos Electrónicos NILC/Folha de São Paulo): *corpus* de cerca de 24 milhões de palavras em português brasileiro retirados do jornal "A Folha de São Paulo" - os textos podem ser baixados via FTP / HTTP ou consultados no [Projeto AC/DC](#).

CetemPUBLICO (*Corpus* de Extractos de Textos Electrónicos MCT/Público): *corpus* de aproximadamente 180 milhões de palavras em português europeu retirados do jornal português "Público" - disponibiliza as versões anotada e sem anotação para consulta online e envia cd com o *corpus* mediante solicitação.

COMPARA: site português com um *corpus* paralelo - originais e respectivas traduções - de/para as diversas variantes do português e do inglês.

Corpus do Português: *corpus* diacrônico com 45 milhões de palavras, composto por textos dos séculos XV ao XX, nas variantes portuguesa e brasileira. Criado por Michael Ferreira (Universidade de Georgetown) e Mark Davies (Universidade Brigham Young). Apresenta as linhas de concordância em formato KWIC, ou seja, a palavra de busca é centralizada e apresentada em negrito. Permite buscas complexas por categoria gramatical ou tipo de texto.

Corpus Brasileiro: *corpus* de português brasileiro contemporâneo, língua geral, ainda em desenvolvimento junto ao LAEL da PUC-SP, é etiquetado morfossintaticamente e permite pesquisas por gênero textual / registro.

Lácio-Web: site brasileiro com um *corpus* contemporâneo de língua geral, subdividido em *subcorpora* que representam vários gêneros e tipos textuais - textos disponíveis para download e para serem usados com as próprias ferramentas do site.

Linguateca: projeto português que disponibiliza diversos recursos para os quais foi desenvolvida uma interface online. A interface permite consultar um *corpus* de cada vez (inclusive versão anotada) - são vários os *corpora* disponíveis, a maioria em português de Portugal. Seu objetivo é melhorar significativamente as condições para o processamento do português, ao facilitar a disponibilização dos recursos existentes, ajudar ao desenvolvimento de recursos públicos, fornecer programas para obter *corpora* através da internet, criar *corpora* suficientemente grandes que possam ser usados como referência, facilitar o processamento de *corpora* portuguesas em geral e criar ferramentas que possam ser reutilizadas por outros investigadores ou desenvolvedores. Desde 2012 conta com a colaboração de investigadores de instituições em vários países.

Este corpus contém cerca de 45 milhões de palavras de **quase 57.000 textos em português** entre os anos de 1300 a 1900. As 20 milhões de palavras no texto correspondentes a 1900 são uma mistura entre os gêneros de conversação, ficção, jornais e academia. Este corpus permite **comparar todos os gêneros e períodos de tempo**.

guês

Este corpus contém aproximadamente **um bilhão de palavras em Português**, retiradas de mais ou menos um milhão de páginas de web de quatro países que falam Português (Brasil, Portugal, Angola e Moçambique). O corpus **está a ser**

<http://www.corpusdoportugues.org/xp.asp>

o corpus do português

English **Português**

Criado pelo Professor [Mark Davies, BYU](#). Financiada pelo [National Endowment for the Humanities](#) (2004, 2015). Faz parte da coleção [corpora da BYU](#).

	Corpus	Tamanho	Criado	Mais informação
1	Gênero / Histórico	45 milhões de palavras	2004-06	Info
2	Web / Dialectos	1 <i>bilhão</i> de palavras	<u>2015-16</u>	Info

Os corpora
Nova interface
Tamanho dos corpus
Comparar com outros corpora
Recursos
Pesquisadores (inglês)

Seja voluntário!

Problemas (inglês)
Contatar

A maioria das pessoas é

X







A maioria das pessoas são

Com a expressão «a maioria de», recomenda-se a concordância verbal no singular, com a palavra **maioria**, adotando um critério estritamente sintático. Mas, como muitas vezes a construção «a maioria de...» tem por complemento uma expressão nominal no plural, aceita-se que o verbo ocorra também no plural concordando com essa expressão nominal. Por outras palavras, não se pode dizer que estejam incorretas as sequências «a maioria das pessoas vivem» ou «a maioria das pessoas são».

A	B	C	. A enxaqueca é um problema de saúde que acomete muitas pessoas, na maioria das vezes as mulheres, principalmente quando as mesmas se encontram nos dias
A	B	C	irritabilidade, desejos por alimentos como chocolate e sonolência. Como se adquire Em a maioria das vezes a enxaqueca é hereditária, mas acredita-se que ela seja multifatorial
A	B	C	o tipo de dor de cabeça. Depois de identificada a causa, na maioria das vezes são realizados tratamentos medicamentosos, os quais visam acabar com as crises
A	B	C	do que saber que foi traída por algo que não tem importância. -- A maioria das mulheres não gosta de sair com homens de boné. Geralmente, os
A	B	C	585965 Como perdoar uma traição e continuar o relacionamento a dois? Em a maioria das vezes não conseguimos perdoar uma traição, mas há aquelas pessoas que preferem
A	B	C	sempre acompanhados dum adulto. Benefícios das atividades físicas em geral A grande maioria das atividades físicas garante inúmeros benefícios a nossa saúde, podemos citar entre eles
A	B	C	foi traída e não poder fazer nada para reverter essa situação. Em a grande maioria das vezes os homens acabam sempre piorando essa situação. A pior coisa para
A	B	C	que devem ser mantidos sempre com uma aparência legal. No entanto, a grande maioria das mulheres não está contente com as madeixas, onde optam por uma série
A	B	C	de tratamentos visando melhorar a aparência de seus cabelos e aumentar a autoestima. A maioria hoje opta pelos alisamentos, que deixam os cabelos mais bonitos e lisos,
A	B	C	a escova progressiva, causa muita polêmica e desconforto entre suas clientes, pois a maioria deste método usa formol. Ela permanece no cabelo por mais ou menos
A	B	C	vaginal normal. Gonorreia e Clamídia A Gonorreia infecta o colo uterino e na maioria dos casos é difícil identificar os sintomas. Algumas mulheres apresentam corrimento, mas
A	B	C	o time não teve um bom desempenho, no entanto Liedson se destacou na maioria das partidas e marcou 15 gols em 29 jogos disputados. Em o Corinthians
A	B	C	. No entanto, há sempre um que acaba os agradando mais. A grande maioria dos adolescentes se identifica mais com o rock ou pop rock, as músicas
A	B	C	exame recebeu mais de 4,6 milhões de inscrições de todo o Brasil, sendo a maioria do Sudeste. Boa sorte! Erica você deverá imprimir-la diretamente de a
A	B	C	mulher para mulher, elas que sempre deixam um mistério para os homens. A maioria dos itens podem ser considerados inúteis, mas elas garantem que usam quase tudo
A	B	C	tratamentos para os calos sumirem ou amenizarem! Esse é um problema que perturba a maioria das pessoas, sejam mulheres ou homens, e até crianças. Mas,
A	B	C	têm dúvidas na hora de fazer a mala de viagem, onde na maioria das vezes acabam optando por levar um monte de roupas, das quais
A	B	C	você acredite, vale a pena apenas conhecer o que cada hora significa. A maioria dos significados é em relação ao amado, porém nem sempre são coisas
A	B	C	permitir uma maior rapidez nas transferências de arquivos. * Slots de expansão A maioria das placas-mãe de hoje oferecem slots de expansão PCI-EXPRESS, é a partir de
A	B	C	para o problema, mas ele pode ser controlado através de tratamentos. Em a maioria das vezes, seus sintomas são confundidos com sinais de estresse. Porém,
A	B	C	para estarem sempre na moda. Com certeza, a maior preocupação para a maioria são os cabelos que devem ser mantidos sempre hidratados e com boa aparência. Por
A	B	C	. Rosto oval As mulheres que têm rosto oval são bastante privilegiadas, pois a maioria dos cortes de cabelo fica bom para elas. Por isso, ouse e
A	B	C	são os principais de eles e tome cuidado para não afastar-lo. Em a maioria das vezes tudo não passa de coisas da nossa cabeça, mas é
A	B	C	em aquele espaço de tempo. E o mais incrível é que na suprema maioria das vezes, as coisas parecem-se bem parecidas com seu jeito de ser
A	B	C	balada e ainda contam, achando tudo bonito... As moças sagitarianas, na maioria das vezes, moram sozinhas, demoram-se para casar, são independentes e

- <file:///C:/Users/joacy/Desktop/Corpus%20do%20Portugu%C3%AAs.html>

A meu ver x do meu ver

Corpus do Português: Web/Dialetos       EN **PT**

PESQUISAR FREQUÊNCIA CONTEXTO AJUDA

VER CONTEXTO: CLIQUE NA PALAVRA (TODAS AS SEÇÕES), NO NÚMERO (UMA SEÇÃO), OU [CONTEXTO] (VÁRIAS) [\[AJUDA...\]](#) [COMPARAR](#)

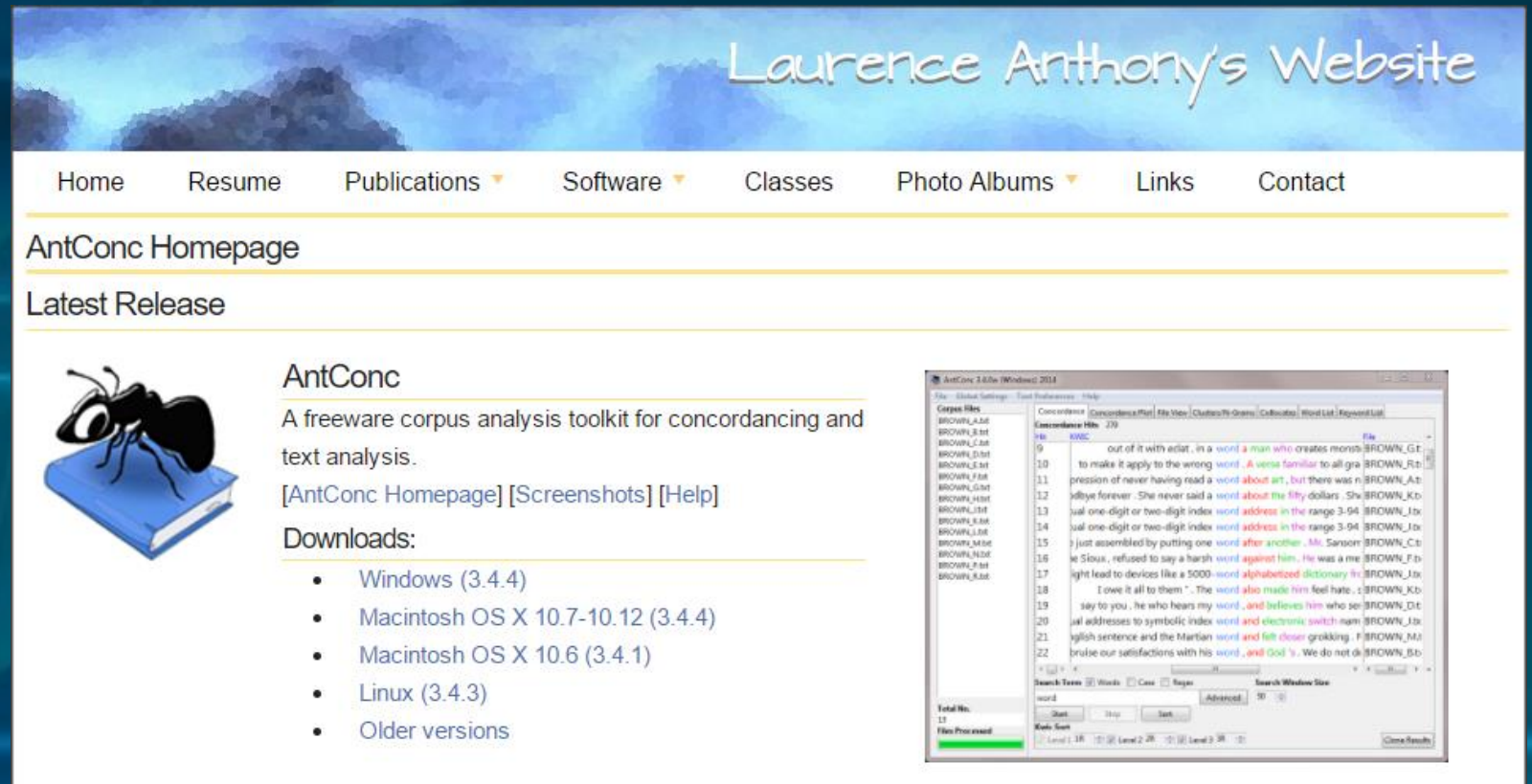
	<input type="checkbox"/>	CONTEXTO	FREQ	
1	<input type="checkbox"/>	A MEU VER	10862	<div style="width: 100%; height: 10px; background-color: #007bff;"></div>

0.953 seconds

que Mark A.Z. Dippe entregou aquele primeiro filme do Spawn que, a meu ver é bem.....e só valeu mais a pena pelo Violador do John Leguizamo acho que o "« Muito Além do Peso "» é uma denúncia. A meu ver, ele é um retrato. Se em pronto-socorros de hospitais infantis existe refrigerante e o governo sobre a propaganda direcionada ao público infantil? ER: A meu ver, permitir que a publicidade fale com nossos filhos é o mesmo que per a produção crítica e cinematográfica em comparação com os anos 60. "« A meu ver, a questão da crítica vai junto com a questão do cinema, e de conquista de aperfeiçoamento espiritual e intelectual. As várias interpretações, a meu ver, surgiram da convivência dos próprios maçons com o "» em nosso lar maçônico. O argumento é interessante, mas, a meu ver, não colhe, por várias razões: 1) O Dicionário Priberam de a Mestre, naturalmente um dos Mestres, eleito pelos restantes. A meu ver, esta reestruturação das Lojas Azuis completa a transformação da Maçonaria por dificultar a sua evolução e o acompanhamento das mudanças sociais. A meu ver, este sistema de voto unânime quanto à admissão, tendo as ól "» em nosso lar maçônico. O argumento é interessante, mas, a meu ver, não colhe, por várias razões: 1) O Dicionário Priberam de a festividade social levada a espaço no espaço dum templo maçônico. A meu ver, deve-se reservar o espaço e o tempo e o modo de reunião de Maçonaria foi abolida há já perto de quarenta anos. Não é, a meu ver, argumento válido o facto de ainda haver muito preconceito contra a Maçonaria entre a Tolerância e a consideração de existência de limites á mesma. A meu ver, esta é uma falsa questão, que um pouco de reflexão facilmente re a simplicidade da referência direta a uma ferramenta operativa que, e a meu ver com muito peso, suporta o argumento anterior. Com efeito, entre e assim abstenho-me de dar opinião definitiva sobre esta segunda situação. A meu ver, a melhor explanação sobre o assunto (ainda assim não con que justificasse essa improvável mudança. O que forçou então esta mudança? A meu ver, um segundo catalisador (considerando o primeiro, e long (a evolução da Maçonaria Operativa para a Maçonaria Especulativa)? A meu ver, um segundo catalisador (considerando o primeiro, e longínquo, a ir uma dimensão e influência superiores às que efetivamente detém. Resulta, a meu ver, esta situação, do facto da Maçonaria, pelas circunstâncias er que justificasse essa improvável mudança. O que forçou então esta mudança? A meu ver, um segundo catalisador (considerando o primeiro, e long ocorrido que tenha funcionado como propiciador e ou acelerador dessa transformação. A meu ver, houve, não um, mas dois fatores catalisadores. desejo e libido, pois e condição de perpetuação da especie. A meu ver o sexo e fonte de muitos problemas e de distanciamento entre pessoas alem só os dois tipos ser for leva ao pé da letra A o meu ver, considero que pertença a categoria dos assexuais aromânticos, por não sentir atração) que seja ser muito estranho, passa do quesito amizade ao meu ver... meio pedófilo... Agora o último filme que vi foi... ` Abril Despedaçado de o ser. Não sou passiva de classificações, ser assexuada ao meu ver está mais para uma característica da minha identidade como pessoa do que , mas posso estar outra vez errado, motivos para alarde. A o meu ver, o que houve foi um benefício. Vejamos: Dizia o Art. 9º, a apresentar as palavras de George Vaillant que definem o foco, que a meu ver é o mais adequado diante da experiência que acumulei ao longo de mas considero inócua tal discussão. O verdadeiro e necessário alvo de debates a meu ver é a necessidade da modernização das metodologias utiliz Digo isso porque a educação física militar, praticada nos quartéis ao meu ver, pouco ajuda em esse preparo. Pelo menos aqui no DF. A Mas para eles eu não vou deixar. Idnorar a CONSTITUIÇÃO, ao meu ver gera IMPROBIDADE ADMINISTRATIVA, Governador o nosso Estatdo é livre as por vários aspectos até louváveis da parte dos senhores. Mas ao meu ver a ONU quando fala em extinção da PM, eles também, acho eu

AntConc para corpora offline

- Ferramenta gratuita para análise de textos e linhas de concordância.
- Busque no Google “AntConc” e localize o site: <http://laurenceanthony.net/software/antconc/>




Laurence Anthony's Website

Home Resume Publications Software Classes Photo Albums Links Contact

AntConc Homepage

Latest Release



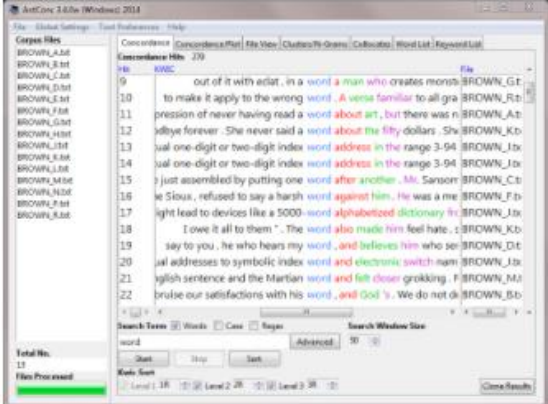
AntConc

A freeware corpus analysis toolkit for concordancing and text analysis.

[AntConc Homepage] [Screenshots] [Help]

Downloads:

- Windows (3.4.4)
- Macintosh OS X 10.7-10.12 (3.4.4)
- Macintosh OS X 10.6 (3.4.1)
- Linux (3.4.3)
- Older versions



Análise do *corpus* Assassinato de Mércia Nakashima

- Análise do *corpus* Assassinato de Mércia Nakashima no AntConc.



**MUITO
OBRIGADO!!!**

- Joacyr Oliveira – Joacyr@cel.unicamp.br