



Mas... não é assim que se diz!

Convenções linguísticas e *corpora*

Stella E. O. Tagnin – USP

UFMG – Letras Debate

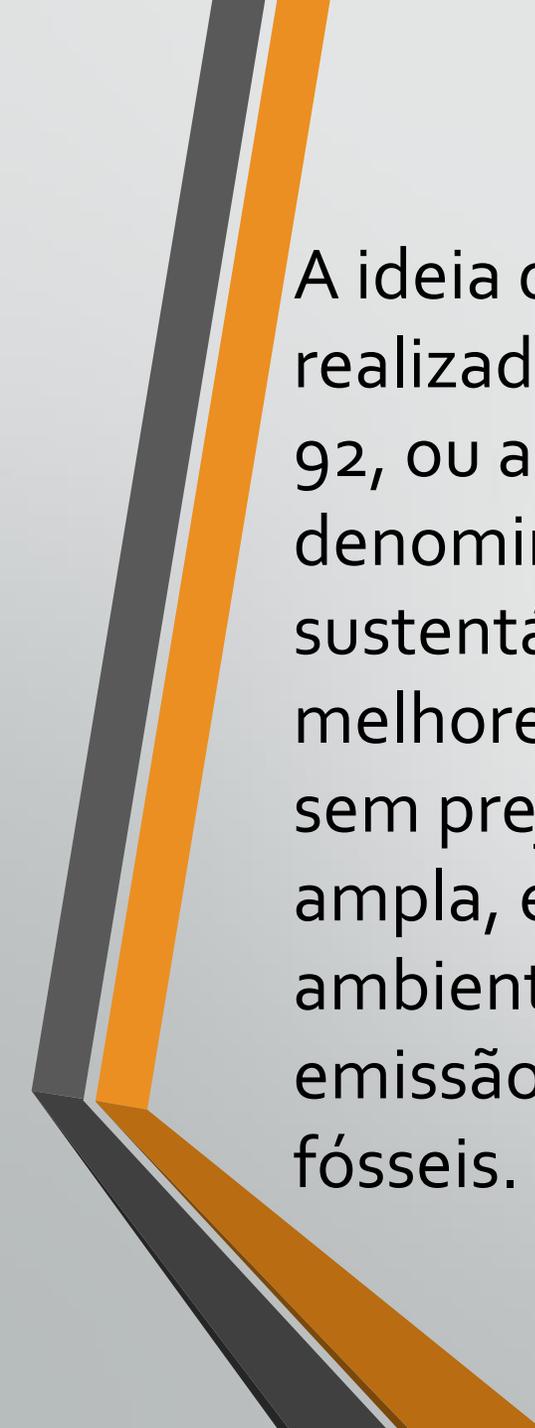
14 de maio de 2016

Resumo

- Convencionalidade e Fraseologia
 - Tipos de fraseologismos
- Linguística de *Corpus*
 - O que é
 - O que se faz com ela
- *Corpora on-line*
 - Exemplos



Convencionalidade e Frasesologia



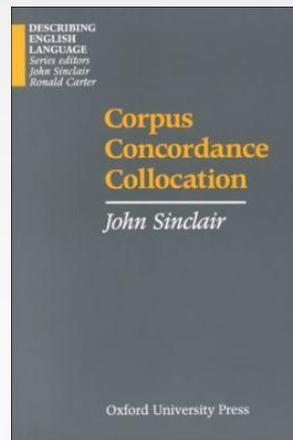
A ideia de desenvolvimento sustentável é associada à reunião realizada no Rio de Janeiro em 1992, a “Cúpula da Terra”, ou Eco-92, ou ainda Rio 92. Nela foi aprovado o compromisso denominado Agenda 21 que considera “desenvolvimento sustentável” o modo de progresso do mundo (global) em que melhorem (com equidade) as condições de vida da atual geração, sem prejuízo das gerações vindouras. Embora seja muito mais ampla, essa ideia é com muita frequência associada a questões ambientais que incluem o aquecimento global associado à emissão de gases resultantes da degradação de combustíveis fósseis.

A ideia de **desenvolvimento sustentável** é associada à reunião realizada no **Rio de Janeiro em 1992**, a “**Cúpula da Terra**”, ou **Eco-92**, ou ainda **Rio 92**. Nela **foi aprovado** o compromisso denominado **Agenda 21** que considera “**desenvolvimento sustentável**” o **modo de** progresso do mundo (global) em que melhorem (com equidade) as **condições de vida** da atual geração, **sem prejuízo das gerações vindouras**. Embora seja muito mais ampla, essa ideia é **com muita frequência associada a questões ambientais** que incluem o **aquecimento global** associado à **emissão de gases** resultantes da degradação de **combustíveis fósseis**.

John Sinclair

1933-2007

The idiom principle



The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.

(Sinclair, 1991:110)

O Falante Ingênuo → Tradutor Ingênuo

desconhece

- as convenções morfológicas:
jailer vs. prisoner
carcereiro vs. prisioneiro
- as convenções de ordem:
dead or alive vs. vivo ou morto
- as convenções de coocorrência:
have breakfast vs. tomar o café
have a meal vs. fazer uma refeição

O Falante/Tradutor Ingênuo

desconhece

- as convenções de aceitabilidade: by and large
o mais das vezes
- as convenções semânticas:
hard nut to crack
osso duro de roer
- as convenções pragmáticas:
How do you do?
Muito prazer

(Adaptado de Fillmore, C.J. 1979. *Innocence. A Second Idealization for Linguistics*, *Berkeley Linguistic Society* 5, 63-76)

Colocações Adjetivas

Japanese current - corrente do Japão

green light – sinal verde



red onion – cebola roxa



red wine – vinho tinto



red hair – cabelo ruivo



Colocações Nominais 1

N2 + N1 Labor Day — Dia do Trabalho



metal mouth
sorriso metálico

jug ears – orelhas de abano



N1 + Prep + N2

justice of **the** peace — juiz de paz

cost of **living** - custo de vida

Tree of Knowledge — Árvore **do** Conhecimento

House of Representatives – Câmara **dos** Deputados

Colocações Verbais

V + N _[obj]

muster courage — armar-se de coragem

pay a compliment — fazer um elogio

take a step — dar um passo / fare un paso

take vacation — tirar férias

fill a prescription — aviar uma receita / manipular uma receita

pit [a fruit] — tirar o caroço / descaroçar

Colocações Verbais

N_[subj] + V

a nose runs — um nariz escorre

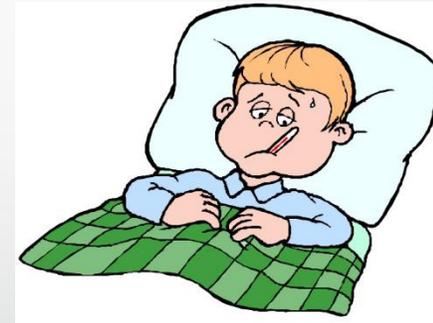
a volcano erupts — um vulcão **entra em erupção**



V + Adj

take ill — ficar doente

grow old — ficar velho



V + Adv

climb aboard — subir a bordo

come ashore — desembarcar

Colocações Adverbiais

Adv. + Adj.

lavishly illustrated — fartamente ilustrado

critically injured — gravemente ferido



Verb + Adverb

fully agree — concordar plenamente

cry loudly — chorar copiosamente

lie outright — mentir descaradamente

take seriously — levar a sério

pay dearly — pagar caro



Binômios e Multinômios

night and day* – dia e noite*

knife and fork – garfo e faca

dead or alive – vivo ou morto

cats and dogs – cães e gatos / gatos e cachorros



over and over and over

red, white and blue



cama, mesa e banho

casa, comida e roupa lavada

Expressões Convencionais e Idiomáticas

Expressões Convencionais

That's my fault — A culpa é minha

in the old days — nos velhos tempos

Will you marry me? – Quer casar comigo?



Expressões Idiomáticas

kick the bucket — bater as botas

pull sb's leg — tirar o sarro de alguém

blow one's top — perder a cabeça

raise hell — botar a boca no mundo



Don't cut off your nose
to spite your face



Do not let your overreaction
lead to self-harm.



Dar um tiro no
(próprio) pé



bedroom eyes



cara de paisagem

Fórmulas Discursivas

Citações

To be or not to be... — Ser ou não ser...

The die is cast — A sorte está lançada

Provérbios

Not all that glitters is gold.

Nem tudo que reluz é ouro.

Do as I say, not as I do

Faça o que eu digo, não faça o que eu faço.

Marcadores Conversacionais

Come to think of it,... — Por falar nisso,...
In other words, ... — Em outras palavras,
You see, ... - Sabe, ...
Well,... — Bem,...
May I interrupt? — Posso interromper?
What do you think? — O que você acha?
That's it. — É isso (aí).
I'm all ears. — Sou todo ouvidos.
Are you with me? — Está entendendo?
Mhm



Fórmulas Situacionais

Act your age — Não seja criança.

Have fun — Bom divertimento/Divirta-se

That's the pot calling the kettle black.
Olha o roto falando do esfarrapado.

It takes two to tango.

Quando um não quer, dois não brigam.



Fórmulas de Rotina

Cumprimentos

Good morning, Hello, How are you?

Bom dia, Oi, Como vai?

Desculpas

Sorry, Excuse me, Pardon me

Desculpe, Sinto muito, Perdão

Votos



Feliz Ano Novo



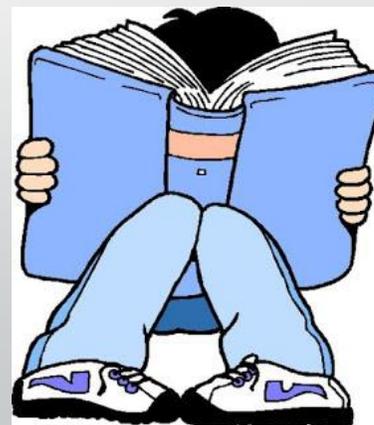
Estimo as melhoras / Sare logo

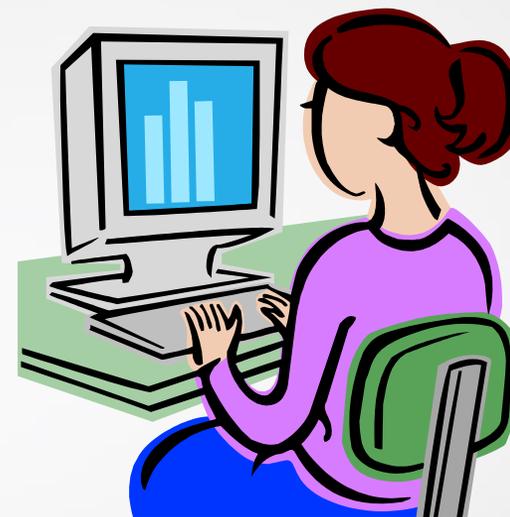


Feliz Natal



Como aprender tudo isso?





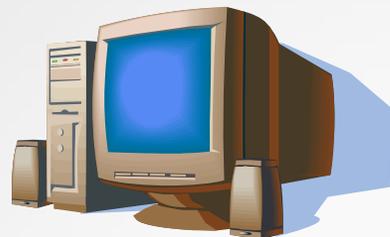
A Linguística de *Corpus*

O que é um *corpus*?

Uma coletânea de textos em formato eletrônico, compilada segundo critérios específicos, considerada representativa de uma língua (ou da parte que se pretende estudar), destinada à pesquisa

A História dos *Corpora*

- Década de 1960:



- Década de 1970: consolidação, corpora diversos, línguas diversas; corpora maiores, corpora anotados (Escandinávia) – tudo digitalizado



- Década de 1980



- Década de 1990: disponibilidade de textos já digitalizados
- Novo milênio: textos da internet
- Web como corpus





Qual a grande vantagem dos
corpora?

**Observação da
linguagem em uso**

- natural -

probabilidade

contrapõe-se à introspecção Chomskyana

- gramatical -

possibilidade

Para que servem os *corpora*?

- Ensino e aprendizagem de línguas
- Estudos terminológicos
- Estudos da Tradução
- PLN: Processamento Automático de Língua Natural
 - tradução automática
 - treinamento de ferramentas



now is a time to [look](#) at the other side of
the style I want?" Well, [look](#) no further. I, Adam Polselli
on in and take a " [look](#) ," literally! Be on the lookout
update is a brand new [look](#) that I call "Chic Simplicity
call "Chic Simplicity." This new [look](#) was inspired by the outdated
Centre is opposite Coral Reef - [look](#) out for the brown signs
homepage [look](#) trn: How not to [look](#) like an idiot on Usenet
you want to do is [look](#) like an idiot or be
group.name or take a [look](#) at news.answers. If you
careful of advertising. Take a [look](#) at `trn: How to advertise
Service Pack 2 Beta first [look](#) by Kurt Hutchinson A

harder [look](#) at after-school help By Lisa
in the take a closer [look](#) page for several weeks. My
for several weeks. My students [look](#) forward to the picture each
Friday March 5, 2004 Hey [look](#) , it's a new comic. That's
seen it yet, take a [look](#) . The new aqua appearance is

O que é uma concordância?

- Lista de todas as ocorrências da **palavra de busca** em contexto
- **Formato: KWIC:** “key word in context” –
 - em geral centralizada
- **Leitura vertical:** permite identificar padrões gramaticais e lexicais - repetição
- **Leitura horizontal:** permite identificar colocados e diferentes sentidos



now is a time to look at the other side of
the style I want?" Well, look no further. I, Adam Polselli
on in and take a "look ," literally! Be on the lookout
update is a brand new look that I call "Chic Simplicity
call "Chic Simplicity." This new look was inspired by the outdated
Centre is opposite Coral Reef - look out for the brown signs
homepage trn: How not to look like an idiot on Usenet
you want to do is look like an idiot or be
group.name or take a look at news.answers. If you
careful of advertising. Take a look at `trn: How to advertise
Service Pack 2 Beta first look by Kurt Hutchinson A

harder look at after-school help By Lisa
in the take a closer look page for several weeks. My
for several weeks. My students look forward to the picture each
Friday March 5, 2004 Hey look , it's a new comic. That's
seen it yet, take a look . The new aqua appearance is



now is a time **to** look at the other side of
the style I want?" Well, look no further. I, Adam Polselli
on in and **take a** " look ," literally! Be on the lookout
update is a brand **new** look that I call "Chic Simplicity
call "Chic Simplicity." This **new** look was inspired by the outdated
Centre is opposite Coral Reef - look out for the brown signs
homepage trn: How not **to** look **like** an idiot on Usenet
you want to do is look **like** an idiot or be
group.name or **take a** look **at** news.answers. If you
careful of advertising. **Take a** look **at** `trn: How to advertise
Service Pack 2 Beta first look by Kurt Hutchinson A

harder look **at** after-school help By Lisa
in the **take a** closer look page for several weeks. My
for several weeks. My students look forward to the picture each
Friday March 5, 2004 Hey look , it's a new comic. That's
seen it yet, **take a** look . The new aqua appearance is



now is a time **to** look at the other side of
the style I want?" Well, look no further. I, Adam Polselli
on in and **take a** " look ," literally! Be on the lookout
update is a brand **new** look that I call "Chic Simplicity
call "Chic Simplicity." This **new** look was inspired by the outdated
Centre is opposite Coral Reef - look **out for** the brown signs
homepage trn: How not **to** look **like** an idiot on Usenet
you want to do is look **like** an idiot or be
group.name or **take a** look **at** news.answers. If you
careful of advertising. **Take a** look **at** `trn: How to advertise
Service Pack 2 Beta first look by Kurt Hutchinson A

harder look **at** after-school help By Lisa
in the **take a closer** look page for several weeks. My
for several weeks. My students look **forward to** the picture each
Friday March 5, 2004 Hey look , it's a new comic. That's
seen it yet, **take a** look . The new aqua appearance is

Como se diz, em inglês,

Ela causou uma boa impressão?





ventos mais fortes do oeste e aumento de tempestades violentas, **causou danos** sem precedentes nas propriedades situadas na costa

Por isso, **causou** grande **impacto** a descoberta do oceanógrafo americano Dudley Chelton

membro do conselho executivo deve fazer é "A nossa empresa **causou** uma primeira **boa impressão** em consumidores potenciais suficientes para garantir

de 1995, quando já assumira a presidência do Banco Central, **causou repercussão** na imprensa o fato de ter passado o carnaval

fato teve não teve a repercussão equivalente ao **prejuízo** que **causou** a centenas de empresas que hoje dependem prioritariamente da rede



family thinks Vicki Rock's former job at the plant **caused** her son's **cancer**.

but he denies that the plant **caused leukemia** in families that have brought suits against the company

One of two things could have **caused** this **error**: The Pobox.com user whose web site you

Ok: BM **caused** TONS of **problems** over last few weeks...

Holes that **caused** fuel **leak** on shuttle located in nozzle July 28: Columbia

reduce **deaths, injuries, property damage, economic losses** and human **suffering** **caused** by natural disasters.

Cause vs. causar

- Cause: prosódia semântica negativa
- Causar: prosódia semântica neutra

Portanto:

*cause a good impression →

make a good impression

V+or+V

to protest Japan's continued whaling, call or write to the Japanese Embassy, 101-104
the police know of anything suspicious you see or hear, you are helping to reduce
in breast tissue long before a woman can feel or notice anything unusual or before
Neither does Midlands Electricity plc confirm or deny any of the claims made in
Uvistat Sun Cream spf15), so the sun won't burn or damage her skin. Unfortunately,
aircraft are a scourge to those who live, work or relax in the country. [p] For several
which you lay brick or stone paths can shorten or increase their apparent distance.
was pressed by mps of all parties to confirm or deny reports that the cost of
in town. Except for Ella, no one would know or care where Sue Anne Norton
humankind or the goodwill of God can do to avert or transform the effects of
was my experience that a patient's will to live (or die) made a substantial difference
be inadequate because it was unable to predict or suggest how managers'
credit cards are often useless. You can't bribe or barter with plastic, sometimes
alcohol or other sedatives. 4. Not drive or operate machinery whilst under the
should use any opportunity that he has to visit or contact friends or relations he has
two major companies and their decisions can make or break small publications. [p]
years are very important and whether I stay or go remains to be seen. [p] But I'

confirm or deny

any way. Neither does Midlands Electricity plc **confirm or deny** any of the claims
Wakeham was pressed by mps of all parties to **confirm or deny** reports that the cost
and flown to America. Dr Swire has refused to **confirm or deny** reports that the
political prisoners. He has always refused to **confirm or deny** whether he's a
J. Dore which the Princess refused either to **confirm or deny** she wore during the
Household had seen them, it was impossible to **confirm or deny** that they were any
of the reason and no one is prepared to **confirm or deny** that the Princess forced
to show on his face that would either **confirm or deny**. When's the next time
merely smiled and nodded and refused to **confirm or deny** anything. Of the three
were, ended there and then. Neither side will **confirm or deny** that they took place
[p] The Prison Service yesterday refused to **confirm or deny** whether women were
the Hong Kong police were last night able to **confirm or deny** the report, which
by Lord Stevens of Ludgate, has refused to **confirm or deny** claims that it is
soon. [p] The Agriculture Minister refused to **confirm or deny** reports that he had
s Playboy magazine Elizabeth Ward refuses to **confirm or deny** whether she has
[p] But Downing Street later refused to **confirm or deny** whether Maxwell's

confirm or deny

any way. Neither does Midlands Electricity plc confirm or deny any of the claims

Wakeham was pressed by mps of all parties to confirm or deny reports that the cost and flown to America. Dr Swire has refused to confirm or deny reports that the political prisoners. He has always refused to confirm or deny whether he's a

J. Dore which the Princess refused either to confirm or deny she wore during the Household had seen them, it was impossible to confirm or deny that they were any of the reason and no one is prepared to confirm or deny that the Princess forced to show on his face that would either confirm or deny. When's the next time merely smiled and nodded and refused to confirm or deny anything. Of the

were, ended there and then. Neither side will confirm or deny that they took place

[p] The Prison Service yesterday refused to confirm or deny whether women were the Hong Kong police were last night able to confirm or deny the report, which by Lord Stevens of Ludgate, has refused to confirm or deny claims that it is

soon. [p] The Agriculture Minister refused to confirm or deny reports that he had s Playboy magazine Elizabeth Ward refuses to confirm or deny whether she has

[p] But Downing Street later refused to confirm or deny whether Maxwell's

confirm or deny

any way. **Neither** does Midlands Electricity plc confirm or deny any of the claims
Wakeham was pressed by mps of all parties to confirm or deny reports that the cost
and flown to America. Dr Swire has **refused to** confirm or deny reports that the
political prisoners. He has always **refused to** confirm or deny whether he's a
J. Dore which the Princess **refused** either **to** confirm or deny she wore during the
Household had seen them, it was **impossible** to confirm or deny that they were any
of the reason and **no one** is prepared to confirm or deny that the Princess forced
to show on his face that would either confirm or deny. When's the next time
merely smiled and nodded and **refused to** confirm or deny anything. Of the
were, ended there and then. **Neither** side will confirm or deny that they took place
[p] The Prison Service yesterday **refused to** confirm or deny whether women were
the Hong Kong police were last night able to confirm or deny the report, which
by Lord Stevens of Ludgate, has **refused to** confirm or deny claims that it is
soon. [p] The Agriculture Minister **refused to** confirm or deny reports that he had
s Playboy magazine Elizabeth Ward **refuses to** confirm or deny whether she has
[p] But Downing Street later **refused to** confirm or deny whether Maxwell's

V+or+V

to protest Japan's continued whaling, call or write to the Japanese Embassy, 101-104
the police know of anything suspicious you **see or hear**, you are helping to reduce
in breast tissue long before a woman can feel or notice anything unusual or before
Neither does Midlands Electricity plc **confirm or deny** any of the claims made in
Uvistat Sun Cream spf15), so the sun won't burn or damage her skin. Unfortunately,
aircraft are a scourge to those who live, work or relax in the country. [p] For several
which you lay brick or stone paths can shorten or increase their apparent distance.
was pressed by mps of all parties to **confirm or deny** reports that the cost of
in town. Except for Ella, no one would know or care where Sue Anne Norton
humankind or the goodwill of God can do to avert or transform the effects of
was my experience that a patient's will to **live (or die)** made a substantial difference
be inadequate because it was unable to predict or suggest how managers'
credit cards are often useless. You can't **bribe or barter** with plastic, sometimes
alcohol or other sedatives. 4. Not **drive or operate** machinery whilst under the
should use any opportunity that he has to visit or contact friends or relations he has
two major companies and their decisions can make or break small publications. [p]
years are very important and whether I **stay or go remains to be seen**. [p] But I'

remains to be seen

to fulfil its promise in education. It **remains to be seen** whether this book will help will translate onto their forthcoming album **remains to be seen**, but for now length of time they have been in Britain. It **remains to be seen** whether the most to enfranchise its 120 women members. [p] It **remains to be seen** how many other Whether that's a desirable development **remains to be seen**, but at this protests against the president's visit But it **remains to be seen** whether the emperor not on a normal trans-Atlantic service. So it **remains to be seen** if the trophy itself and the animosity of the Cold War. It **remains to be seen** what kind of aid have been arguing about over a year but it **remains to be seen** whether the voters, something the current constitution forbids. It **remains to be seen** though, whether the outright defeat for the fledgling opposition **remains to be seen**. Ivorians should on the island of Brioni. THE GUARDIAN says it **remains to be seen** whether the Actuality - demonstration/chanting. BEARD: It **remains to be seen** whether the on them and with them enthusiastically. It **remains to be seen** how far Third house. Whether The Eighth Wonder will travel **remains to be seen**. [p] There are of British food over subsequent decades **remains to be seen**. [p] There was current year," he said. [p] Just how robust **remains to be seen**. In a

remains to be seen

to fulfil its promise in education. It remains to be seen **whether** this book will help m
will translate onto their forthcoming album remains to be seen, but for now
length of time they have been in Britain. It remains to be seen **whether** the most
to enfranchise its 120 women members. [p] It remains to be seen how many other
whether that's a desirable development remains to be seen, but at this
protests against the president's visit But it remains to be seen **whether** the emperor
not on a normal trans-Atlantic service. So it remains to be seen if the trophy itself
and the animosity of the Cold War. It remains to be seen what kind of aid
have been arguing about over a year but it remains to be seen **whether** the voters,
something the current constitution forbids. It remains to be seen though, **whether**
outright defeat for the fledgling opposition remains to be seen. Ivorians should
on the island of Brioni. THE GUARDIAN says It remains to be seen **whether** the
Actuality - demonstration/chanting. BEARD: It remains to be seen **whether** the
on them and with them enthusiastically. It remains to be seen how far Third
house. **whether** The Eighth Wonder will travel remains to be seen. [p] There are
of British food over subsequent decades remains to be seen. [p] There was
current year," he said. [p] Just how robust remains to be seen. In a

remains to be seen

to fulfil its promise in education. It remains to be seen **whether** this book will help m
will translate onto their forthcoming album remains to be seen, but for now
length of time they have been in Britain. It remains to be seen **whether** the most
to enfranchise its 120 women members. [p] It remains to be seen **how many** other
whether that's a desirable development remains to be seen, but at this
protests against the president's visit But it remains to be seen **whether** the emperor
not on a normal trans-Atlantic service. So it remains to be seen **if** the trophy itself
and the animosity of the Cold War. It remains to be seen **what kind** of aid
have been arguing about over a year but it remains to be seen **whether** the voters,
something the current constitution forbids. It remains to be seen though, **whether**
outright defeat for the fledgling opposition remains to be seen. Ivorians should have
on the island of Brioni. THE GUARDIAN says It remains to be seen **whether** the
Actuality - demonstration/chanting. BEARD: It remains to be seen **whether** the
on them and with them enthusiastically. It remains to be seen **how far** Third
house. **whether** The Eighth Wonder will travel remains to be seen. [p] There are
of British food over subsequent decades remains to be seen. [p] There was
current year," he said. [p] Just **how robust** remains to be seen. In a

WebCorp
Live



Alguns corpora
para vocês se divertirem

Linguateca

CORPUS OF CONTEMPORARY
AMERICAN ENGLISH

Corpora on-line

- COCA – Corpus of Contemporary American English -
<http://corpus.byu.edu/coca/>
 - 520 milhões de palavras
 - 1990 – 2015
- BNC-BYU = BNC completo
<http://corpus.byu.edu/bnc/>
 - 100 milhões de palavras
 - 1980s - 1993

corpus.byu.edu

corpora, size, queries = better resources, more insight

[Contribute](#) [Contributors](#)

[Academic site license](#)

[Overview](#)

[Corpora](#)

[Size, speed, queries](#)

[Insight into variation](#)

[Upcoming \(May 2016\)](#)

[History / updates](#)

[FAQ / questions](#)

[Researchers](#)

[Log in / password](#)

[Profile / register](#)

[Related resources](#)

[Full-text data](#)

[Word frequency](#)

[Collocates](#)

[N-grams](#)

[WordAndPhrase](#)

[Academic vocabulary](#)

[Problems](#)

[Contact us](#)

Created by [Mark Davies](#), [BYU](#). [Overview](#), [search types](#), [looking at variation](#), [corpus-based resources](#), [upcoming](#).

The most widely used online corpora -- more than [130,000](#) distinct [researchers](#), teachers, and students each month.

English	# words	language/dialect	time period	compare
NEW Hansard Corpus (British Parliament)	1.6 billion	British	1803-2005	Info
Wikipedia Corpus (with virtual corpora)	1.9 billion	English	-2014	Info
Global Web-Based English (GloWbE)	1.9 billion	20 countries	2012-13	
Corpus of Contemporary American English (COCA)	520 million	American	1990-2015	* * * * *
Corpus of Historical American English (COHA)	400 million	American	1810-2009	* *
TIME Magazine Corpus	100 million	American	1923-2006	
Corpus of American Soap Operas	100 million	American	2001-2012	*
British National Corpus (BYU-BNC)*	100 million	British	1980s-1993	* *
Strathy Corpus (Canada)	50 million	Canadian	1970s-2000s	
Other languages				
Corpus del Español (see also...)	100 million	Spanish	1200s-1900s	*
Corpus do Português (see also...)	45 million	Portuguese	1300s-1900s	
N-grams				
Google Books: American English	155 billion	American	1500s-2000s	*

Corpora on-line

- Corpus do Português

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

- 45 milhões de palavras
- 1300s – 1900s

- Corpus del Español

<http://www.corpusdelespanol.org/>

- 100 milhões de palavras
- 1200s-1900s

- Corpus Brasileiro

www.linguateca.pt → Acesso a recursos → [AC/DC](#) Acesso a corpora
/Disponibilização de corpora

- 1 bilhão de palavras - provavelmente até 2010

Linguatca

[Estrutura](#)

[Equipa](#)

[Apresentação](#)

[Acesso a recursos](#)

▪ AC/DC

○ - [Procura](#)

○ - [Frequência](#)

○ - [PJR e exemplos](#)

○ - [Corte-e-costura](#)

○ - [Ensinador](#)

○ - [VARRA](#)

▪ [CETEMPúblico](#)

▪ [CETENFolha](#)

▪ [CHAVE](#)

▪ [COMPARA](#)

▪ [Corpógrafo](#)

▪ [CorTrad](#)

▪ [Esfinge](#)

▪ [Floresta Sintá\(c\)tica](#)

▪ [GeoNetPT](#)

▪ [PAPEL](#)

▪ [PoNTE](#)

▪ [ReEscreve](#)

▪ [REMBRANDT](#)

▪ [REPENTINO](#)

▪ [Repositório](#)

▪ [WebJspell](#)

▪ [WPI](#)

[Catálogo de recursos](#)

[Catálogo de ferramentas](#)

[Catálogo de atores](#)

[Catálogo de publicações](#)

[Informação importante](#)

Acesso a corpos de português: Projecto AC/DC

[Linguatca](#)

[Information](#)

O projecto AC/DC (Acesso a corpos/Disponibilização de corpos), iniciado em 1999, surgiu da necessidade de juntar os poucos recursos disponíveis num único ponto na rede e dessa forma facilitar a comparação e reutilização do material, permitindo ao mesmo tempo acesso a uma ferramenta poderosa de interrogação de corpos, o sistema [CWB](#) (versão nova do *IMS corpus workbench*), para o qual desenvolvemos esta interface.

Desde 2000, a anotação dos corpos tem sido feita automaticamente pelo [PALAVRAS](#) de Eckhard Bick, e convertida para o "formato AC/DC", descrito pormenorizadamente na página de [Anotação](#).

Uma descrição quantitativa inicial dos corpos servidos presentemente pelo AC/DC encontra-se na tabela abaixo. **Clique num dos corpos para o interrogar.** Para cada corpo, pode pedir concordâncias, distribuições de frequências simples e complexas, veja os [exemplos](#). Se é a primeira vez que visita o AC/DC e quer apenas experimentar, [procure no corpo Vercial](#). Veja também a nossa [PJR](#): lista de perguntas já respondidas.

Breve descrição dos corpos

Corpo	Tamanho (unidades)	Tamanho (palavras)	Tamanho (frases)	Variante(s)	Breve descrição
AmostrA-NILC	128.324	98.720	4.964	BR	AmostrA-NILC
ANCIB	1.707.799	1.257.221	83.509	BR	Correio electrónico correspondente ao tráfego na lista ANCIB
Avantel	8.189.915	6.762.307	212.183	PT	Semanário político Avantel, 1997-2002
Corpus Brasileiro	792.765.372	666.509.591	29.919.794	BR	Corpus Brasileiro, um bilhão de palavras de português do Brasil (por enquanto apenas uma fração)
CD HAREM	290.001	225.766	12.558	PT BR	Colecção dourada do HAREM
CETEMPúblico	240.425.222	191.277.678	8.128.620	PT	Jornal PÚBLICO, dividido em extractos, 1991-1998
CHAVE	124.070.878	97.886.617	4.682.363	PT BR	Jornais PÚBLICO e Folha de São Paulo, 1994-1995
CONDIVport	7.156.558	5.576.474	318.776	PT BR	Jornais desportivos e revistas de moda e saúde
CoNE	923.002	675.520	31.361	PT BR	Mensagens de correio electrónico não-endereçadas
C-Oral-Brasil	121.082	65.303	9.329	BR	C-Oral-Brasil, português brasileiro oral informal
DiaCLAV	7.721.767	6.550.966	210.373	PT	Diário de Coimbra, Diário de Leiria, Diário de Aveiro, Viseu Diário
Diáspora TL-PT	27.348	21.754	1.035	TL	Diáspora TL-PT, entrevistas a timorenses em Portugal
ECI-EBR	921.977	723.995	44.381	BR	Texto do corpo Borba-Ramsey, compilado pelo ECI



Corpora on-line
Inglês e outras línguas

- WebCorp: <http://www.webcorp.org.uk>

WebCorp Live lets you access the Web as a corpus - a large collection of texts from which examples of real language use can be extracted. [More...](#)

Search:



Case



Insensitive:

Span:

50 characters ▼



Search API:

Google ▼

Language:

Not specified ▼



Advanced Options

Redefinir

Search

Not specified
Not specified
Afrikaans
Arabic
Armenian
Belarusian
Bulgarian
Catalan
Chinese (Simplified)
Chinese (Traditional)
Croatian
Czech
Danish
Dutch
English
Esperanto
Estonian
Filipino
Finnish
French
German

WebCorp Linguist's Search Engine

Have you tried WebCorp LSE?

Our large-scale search engine with more search options, part-of-speech tags and quantitative analyses.

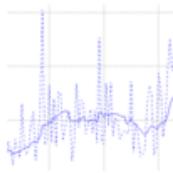
[More details...](#)

The WebCorp Linguist's Search Engine is a tool for the study of language on the web. The corpora below were built by crawling the web and extracting textual content from web pages. Searches can be performed to find words or phrases, including pattern matching, wildcards and part-of-speech. Results are given as concordance lines in KWIC format. Post-search analyses are possible including time series, collocation tables, sorting and summaries of meta-data from the matched web pages.



Synchronic English Web Corpus

470 million word corpus built from web-extracted texts. Including a randomly selected 'mini-web' and high-level subject classification.

[Search ▶](#)[About](#)

Diachronic English Web Corpus

130 million word corpus randomly selected from a larger collection and balanced to contain the same number of words per month.

[Search ▶](#)[About](#)

lovely	2057
nice	1846
another	1633
beautiful	1407
wonderful	1444
started	1210
discovered	1191
enjoy	1052
see	924

Birmingham Blog Corpus

630 million word corpus built from blogging websites. Including a 180 million word sub-section separated into posts and comments.

[Search ▶](#)[About](#)

Anglo-Norman Correspondence Corpus

A corpus of approximately 150 personal letters written by users of Anglo-Norman. Including bespoke part-of-speech annotation.

[Search ▶](#)[About](#)

Novels of Charles Dickens

A searchable collection of the novels of Charles Dickens. Results can be visualised across chapters and novels.

[Search ▶](#)[About](#)

Corpora on-line com Português

Monolíngue:

- Lácio-Web: <http://www.nilc.icmc.usp.br/lacioweb> (fora do ar)

Bilíngue:

- COMPARA: <http://www.linguateca.pt/COMPARA>
- CoMET: comet.fflch.usp.br/
 - CorTec: <http://comet.fflch.usp.br/cortec>
 - CorTrad: comet.fflch.usp.br/cortrad



Muitíssimo obrigada!

Stella