



Universidade Federal  
de Santa Catarina

# Exemplos de *Corpora* Online

Aprendendo a classificar um *corpus*

Equipe de Estudos da Tradução I



- Coleção de textos selecionados e agrupados de acordo com critérios claramente definidos (Dayrell, 2012, p.1);
- Textos em formato eletrônico, analisados de forma automática ou semi-automaticamente.



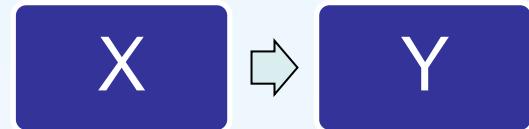
Universidade Federal  
de Santa Catarina

# Tipos de Corpus

Um corpus é **paralelo** quando alinha texto fonte em uma língua “x” com e sua respectiva tradução na língua “y”, “h”, “j”, etc.



Universidade Federal  
de Santa Catarina



Texto Fonte	Texto Alvo
You've forgotten the <b>magic</b> word,' said Harry irritably.	Você esqueceu a palavra mágica – "disse Harry irritado.
Improper Use of <b>Magic</b> Office	ESCRITÓRIO DE CONTROLE DO USO INDEVIDO DE MÁGICA



Universidade Federal  
de Santa Catarina

# Tipos de Corpus

É **comparável** se traz textos originalmente escritos na língua “x” e outros textos similares traduzidos para a língua “x” a partir de várias línguas diferentes.

## Sobre os Corpora do CorTec

O CorTec é constituído de 5 corpora comparáveis, ou seja, com textos semelhantes, em inglês e português originais, nas seguintes áreas:

1. **Culinária** - Receitas Culinárias
2. **Ecoturismo** - Meio ambiente
3. **Hipertensão** - Cardiologia
4. **Informática** - Geral
5. **Instrumentos Contratuais** - Direito Contratual
6. **Astronomia** - Geral
7. **Insuficiência Renal** - Geral
8. **Linguística** - Geral
9. **Medidores Eletromagnéticos de Vazão** - Geral
10. **Suplementos Nutricionais** - Geral
11. **Computação** - Hardware
12. **Futebol** - Geral
13. **Café** - Colheita e Processamento
14. **Turismo Cultural** - Geral
15. **Culinária 2** - Geral
16. **Prostodontia** -
17. **Fotografia** -
18. **Autoclaves** -
19. **Moda** -
20. **Turismo - Hotelaria** - Turismo - Hotelaria

Cada corpus técnico é composto por aproximadamente 200.000 palavras em cada língua.



Um *corpus* é:

- **Monolíngue**: se traz textos em apenas um idioma;
- **Bilíngue**: se traz textos em dois idiomas;
- **Trilíngue**: se traz textos em três idiomas;
- **Multilíngue**: quando traz textos em muitos idiomas.



Um *corpus* é:

- **Unidirecional**: os textos estão alinhados do texto fonte para o texto-alvo;
- **Bidirecional**: os textos podem ser alinhados tanto do texto fonte para o texto-alvo quanto do texto-alvo para o texto fonte.





Universidade Federal  
de Santa Catarina

# Tipos de texto

- Que gêneros textuais são contemplados pelo *corpus*?

Ex.: Jornalísticos, Literários, Culinária, Resumos Acadêmicos, Etc.





Universidade Federal  
de Santa Catarina

# Períodos

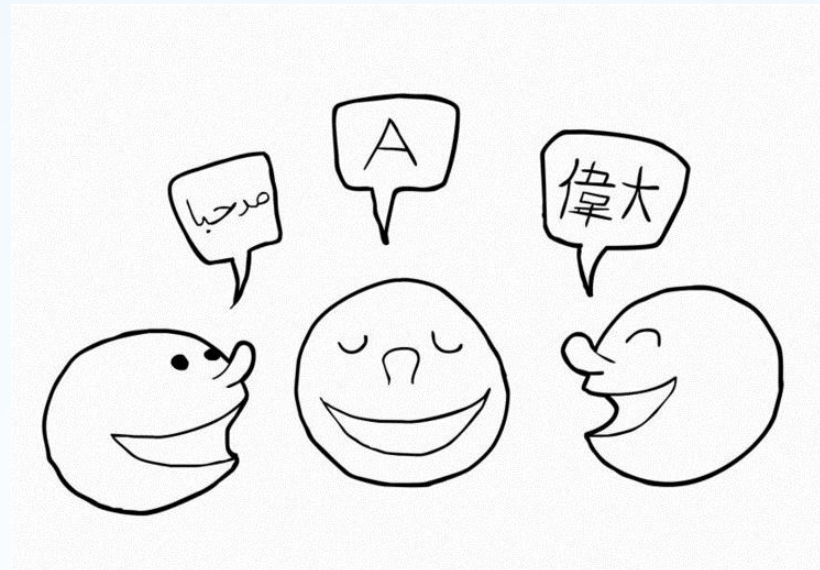
Um corpus é **sincrônico** quando apresenta textos de apenas um determinado período histórico e **diacrônico** quando traz textos de períodos históricos diferentes.



Universidade Federal  
de Santa Catarina

# Línguas

- Os textos que compõe o *corpus* estão escritos em que idioma?  
Ex.: Português, Espanhol, Inglês, Alemão, etc.





Universidade Federal  
de Santa Catarina

# Ferramentas

## Concordanciador

é um programa de computador que constrói automaticamente uma lista de ocorrências de uma determinada palavra ou frase.

## :: Busca Simples ::

Termo:

**olhe**

Corpus ou Subcorpus:

Todo corpus (2 texto(s) indexad

Língua 1:

Português (1 texto(s) indexados

Língua 2:

Inglês (1 texto(s) indexados).

Realizar Busca no COPA-TRAD »

## Concordância do COPA-TRAD

Língua 1: Português

Língua 2: Inglês

Total em Exibição: 19  
Total Processado: 19  
Total Encontrado: 19  
Tempo Decorrido: 0.008s

Resultados por palavra:

**olhe**: Entradas » 19. Total de Ocorrências » 19.

[Traduzir](#)



" Agora vamos subir muito quietinhos - recomendou Fred – e esperar mamãe nos chamar para tomar o café da manhã. Então Rony, você desce correndo e diz: "Mamãe, **olhe** só quem apareceu durante a noite!" e ela vai ficar contente de ver o Harry e ninguém vai precisar saber que saímos voando no carro."

'Now, we'll go upstairs really quietly,' said Fred, 'and wait for Mum to call us for breakfast. Then Ron, you come bounding downstairs going, 'Mum, look who turned up in the night!' and she'll be all pleased to see Harry and no one need ever know we flew the car.'

[Traduzir](#)



Mas vocês são trouxas! – exclamou o Sr Weasley encantado. – Precisamos tomar um drinque! Que é que têm aí? Ah, estão trocando dinheiro de trouxas. Molly, **olhe!** – Ele apontou excitado para as notas de dez libras na mão do Sr Granger.

'But you're Muggles!' said Mr Weasley delightedly. 'We must have a drink! What's that you've got there? Oh, you're changing Muggle money. Molly, look!' He pointed excitedly at the ten-pound notes in Mr Granger's hand.

[Traduzir](#)



Minha varinha – respondeu Rony com a voz trêmula. – **Olhe** a minha varinha.

'My wand,' said Ron, in a shaky voice. 'Look at my wand.'



Universidade Federal  
de Santa Catarina

# Ferramentas

## Lista de frequência

é uma lista que contém o número de vezes que cada palavra do *corpus* aparece dentro de seus textos.



Universidade Federal  
de Santa Catarina



Bem-vindo **Carlos** [Visitar o site](#) | [Sair](#) ✕

[Painel Inicial](#)

[COPA-TRAD](#)

[Gerência](#)

[Equipe](#)

[Contato](#)

**COPA-TRAD TOOLS**

Busca Simples

Letra:

A

Língua:

Português

**Buscar Entradas »**

## COPA Wordlist

Palavra	Frequência
a	2989
as	518
ao	287
ate	168
agora	130
ainda	125



Universidade Federal  
de Santa Catarina

# Ferramentas

## Palavras-chave no contexto (KWIC- Key-words in context)

É um formato de apresentação de dados no qual o termo procurado é alinhado no centro da entrada mostrada.



**:: Concordância ::**

Termo:

Língua:

Formato:

Subcorpus:

[Realizar Busca no COPA-TRAD »](#)**Concordância do COPA-TRAD****Língua Português**É » **óbvio** »

» que não – concluiu o Sr Malfoy. – Ora veja, de que serve ser uma vergonha de bruxo se nem ao menos lhe pagam bem para isso?

Isto – tornou a professora – é » **óbvio** »Não era hábito de Hagrid criticar professores de Hogwarts, e Harry olhou-o surpreso. Mione, porém, disse num tom mais alto do que de costume: – Acho que você está sendo injusto. É » **óbvio** »

» que o Prof Dumbledore achou que ele era o melhor candidato para a vaga...

Isso esclarece tudo – disse Rony em tom triunfante. – Lúcio Malfoy deve ter aberto a Câmara quando esteve aqui na escola e agora ensinou ao nosso querido Draco como fazer o mesmo. É » **óbvio** »

» Mas eu bem gostaria que Dobby tivesse lhe dito que tipo de monstro tem lá dentro. Quero saber como é que ninguém reparou nele rondando a escola.



Universidade Federal  
de Santa Catarina

# Possíveis usos

- Ensino, prática e pesquisa em tradução;
- Ensino de línguas estrangeiras;
- Entre outros.



Universidade Federal  
de Santa Catarina

# Referencias

Dayrell, Carmen. **O uso de Corpora para o estudo da tradução**: objetivos e pressupostos. 2012.

Projeto Comet. Corpus Multilíngüe para Ensino e Tradução – USP. Disponível em:

<http://www.fflch.usp.br/dlm/comet/>

CopaTrad. Corpus Paralelo de Tradução - UFSC.

Disponível em: <http://www.copa-trad.ufsc.br>