

EAD | Letras  
Libras



# ESTUDOS DA TRADUÇÃO II

**Prof. Carlos Rodrigues**  
***11 de abril de 2015***

**O que são corpora eletrônicos?**

**Como os corpora eletrônicos contribuem com a atividade dos tradutores e intérpretes?**

# Corpus/ Corpora? Corpora eletrônicos?

Corpus pode ser entendido como “coleção de algum material”. No contexto dos Estudos da Tradução, a definição de corpus possui conotações mais específicas: *formato digital, textos completos, auto configuráveis e representativos.*

“Uma coleção autoconfigurável de textos completos digitalizados, analisáveis automática ou semi-automaticamente e, coletados a fim de serem representativos ao máximo do fenômeno tradutório sendo examinado.”

Baker, 1995

“Corpora eletrônicos são coletâneas de textos escritos ou transcrições de fala reunidos em formato de arquivo legível por computador.”

Sardinha, 2003

# Tipos de Corpora para pesquisa e ensino de Tradução (BAKER, 1995):

- **Corpora comparáveis**

“consistem em duas compilações separadas de textos na mesma língua: um corpus consiste de textos originais na língua em questão e o outro consiste de traduções naquela língua a partir de uma dada língua fonte ou línguas”.

# Tipos de Corpora para pesquisa e ensino de Tradução (BAKER, 1995):

- **Corpora Paralelos**

“consistem de textos originais da língua fonte A e suas versões traduzidas na língua B”.

# Tipos de Corpora para pesquisa e ensino de Tradução (BAKER, 1995):

- **Corpora Multilíngues**

“conjuntos de dois ou mais corpora monolíngues em línguas diferentes, construídos ou pelas mesmas, ou diferentes instituições, tendo como base critérios de desenho semelhantes”.

# Os Corpora segundo Fernandes (2006):

Segundo o autor, o termo multilíngue não possui nenhuma característica contrastiva que possa distingui-lo dos outros dois tipos principais de corpora. Portanto, os corpora podem ser de dois tipos: **comparáveis** e **paralelos**.

**TRADUTÓRIO OU LINGUÍSTICO**  
(Área)

**GERAL OU RESTRITO**  
(Domínio)

**ESCRITO e/ou FALADO**  
(Modo)

**COMPARÁVEL e/ou PARALELO**  
(Tipo de Relação entre os Textos)

**DIACRÔNICO ou SINCRÔNICO**  
(Restrição Temporal)

**MONOLÍNGUE, BILÍNGUE ou**  
**MULTILÍNGUE**  
(Número de Línguas)

**UNIDIRECIONAL, BIDIRECIONAL ou MULTIDIRECIONAL**  
(Direcionalidade)

# TIPO DE TEXTO:

comparável

Em um corpus comparável os textos são selecionados tendo como base uma **relação textual** entre eles.

Por exemplo, os *textos são selecionados de acordo com suas semelhanças em termos de assunto, tipo de texto, função comunicativa, etc.*

# TIPO DE TEXTO:

## paralelo

Em um corpus paralelo os textos são agrupados tendo como base uma **relação tradutória**.

Por exemplo, os *textos são selecionados de acordo com algum tipo de relação de tradução* entre eles.

# ÁREA: linguístico ou tradutório?

- **Linguístico:**

estudos baseados em corpora desenhados para o estudo de línguas.

- **Tradutório:**

estudos baseados em corpora desenhados com vistas a investigar produtos e processos da tradução.

# **DOMÍNIO:** geral ou restrito?

- **Geral:**

possui um escopo mais amplo, por ser construído para estudar a linguagem do material traduzido como um todo.

- **Restrito:**

investiga a tradução da linguagem de gêneros e tipos de textos específicos.

# **MODO:** escrito ou falado?

- **Falado:**

texto transcrito a partir de um fonte de vídeo ou áudio.

- **Escrito:**

texto escaneado a partir de um livro e convertido em formato eletrônico.

# RESTRIÇÃO TEMPORAL:

sincrônico ou diacrônico?

- **Sincrônico:**

quando centra sua atenção em um objeto de estudo em um período temporal específico.

- **Diacrônico:**

quando se preocupa com o desenvolvimento histórico deste objeto de estudo através do tempo.

# NÚMERO DE LÍNGUAS:

- **Monolíngue:**

Envolve uma língua.

- **Bilíngue/ Trilíngue:**

Envolve duas línguas/ três línguas

- **Multilíngue:**

Envolve mais de três línguas.

# DIRECIONALIDADE:

- **Unidirecional:**

formado por textos originalmente escritos em L1 e suas respectivas traduções em L2.

- **Bidirecional:**

formado de textos originalmente escritos em L1 e suas traduções em L2, mais originais em L2 e suas respectivas traduções em L1.

# DIRECIONALIDADE:

- **Multidirecional:**

mais de duas línguas estão envolvidas e a direção das traduções não está centrada na L1, mas na interação de todas as línguas que integram o corpus.

# Processamento de um Corpus

- Existem técnicas de processamento de corpus que permitem que o usuário acesse, manipule e exiba as informações contidas em um corpus de várias formas úteis.

# Processamento de um Corpus

- **Lista de Palavras:** possibilidade de informações estatísticas sobre número de tipos e de ocorrências, sendo que a relação entre tipos e ocorrências indica a amplitude do vocabulário usado;
- **Concordâncias:** possibilidade de listar a ocorrência de todas as palavras (caracteres curingas \* ?) de forma monolíngue ou bilíngue.

# Compilação de um Corpus

- (i) **desenho** do corpus – discussão das questões teóricas gerais associadas com o planejamento do corpus;
- (ii) **construção** do corpus – descrição das decisões técnicas feitas durante a compilação do corpus;
- (iii) **processamento** do corpus – especificação dos equipamentos, dos programas e do conjunto de ferramentas computacionais utilizados para o processamento do corpus.

# Alguns Programas de Processamento de Corpus

- WordSmith Tools  
(<http://www.lexically.net/wordsmith/>)
- ParaConc  
(<http://www.athel.com/para.html>)
- Multiconcord  
(<http://web.bham.ac.uk/johnstf/lingua.htm>)

# Corpora de Línguas de Sinais

- British Sign Language Corpus Project  
(<http://www.bslcorpusproject.org/>)
- Corpus de Libras  
([http://nals.cce.ufsc.br/?page\\_id=71](http://nals.cce.ufsc.br/?page_id=71))
- ASL Corpus  
(<http://www.bu.edu/asllrp/>)

# Considerações

- Os corpora eletrônicos podem contribuir significativamente com a prática de tradução e com a pesquisa.
- É muito importante observar que as ferramentas de análise de corpora não interpretam os dados – é de responsabilidade do tradutor ou pesquisador analisar as informações encontradas em um corpus