



Glossair: Metodologia da Compilação de um Corpus Especializado

Pedro Garbinatto Schoenardie ¹, Ana Eliza Pereira Bocorny ²

¹ Pedro Garbinatto Schoenardie, Faculdade de Ciências Aeronáuticas, PUCRS

² Ana Eliza Pereira Bocorny, PUCRS



XXV SIC
Salão Iniciação Científica

LLA - Linguística, Letras e Artes

CONTEXTO

Corpus especializado para criação de um glossário bilíngue online baseado nas necessidades dos usuários (aprendizes da área de segurança de voo).

Interface: Aviação **X** Língua

OBJETIVO

Compilar um corpus especializado com textos sobre segurança de voo, utilizados por alunos do curso de Ciências Aeronáuticas da PUCRS.

METODOLOGIA

Para a compilação do corpus foram estabelecidas as seguintes etapas metodológicas:

Seleção de textos a partir dos seguintes critérios de seleção:

- Quanto ao assunto: corpus temático;
- Quanto à origem: ser acadêmico (artigos, livros, teses, dissertações). O corpus de estudo foi compilado a partir de material estudado na graduação do curso de Ciências Aeronáuticas;
- Quanto ao propósito/finalidade: elaboração do glossário bilíngue de segurança de voo.

Tais textos servirão como fontes de extração de termos, contextos e definições (base aproximada de 40 artigos acadêmicos provenientes de material estudado na graduação).

- Alimentação dos referidos textos para o Google Drive: Nomeação dos arquivos de textos organizados de forma cronológica:

Legenda Para Entendimento do Nome Padrão dos Textos do Corpus

Ex: **SG AR 01 SD**



- Manipulação (limpeza): Transformação de todos os textos para o formato “.txt” (arquivo de texto: legível e sem formatação);

- Inserção do cabeçalho:

```
<link></link>
<title></title>
<author></author>
<year></year>
```

- Inserção dos textos na plataforma de hospedagem – sistema online para gestão dos termos selecionados: “Antconc”.



RESULTADOS

Um corpus composto de livros, artigos, teses e dissertações no âmbito da segurança de voo, escrito por especialistas e direcionado para especialistas e semi-especialistas: “CAVI - Flight Safety”.

Tal corpus possui 40 textos que resultaram em 481888 tokens e 18090 types gerando um “type/token ratio” de 3,75. Esse resultado demonstra uma variedade lexical relativamente baixa para um corpus especializado..

FUNDAMENTAÇÃO TEÓRICA

- Linguística de Corpus (BERBER SARDINHA);
- Teoria Comunicativa de Terminologia – TCT (CABRÉ);
- Function Theory – User Needs (FUERTES-OLIVEIRA, TARP).

CONCLUSÃO

A compilação de um corpus especializado é uma importante tarefa relativa à constituição do produto terminográfico. Este corpus servirá como base para a criação do glossário bilíngue para aprendizes da área de segurança de voo na medida em que dele serão extraídos contextos e definições.

REFERÊNCIAS

- BERBER SARDINHA, T. Linguística de corpus. São Paulo: Manole, 2004.
- BERBER SARDINHA, T. Linguística de Corpus: histórico e problemática. DELTA, São Paulo, v. 16, n. 2, 2000.
- BERBER SARDINHA, T. Tamanho de corpus. The ESpecialist, São Paulo, vol. 23, nº 2. p. 103-122, 2003.

INFORMAÇÃO PARA CONTATO

- Pedro Garbinatto Schoenardie – garbinatto@live.com



MODALIDADE DE BOLSA

Iniciação Científica