

Contextualização

Este trabalho insere-se no projeto *A Linguagem do Patrimônio Cultural Brasileiro: Conservação dos Bens Culturais Móveis*, desenvolvido pelo grupo TERMISUL. O projeto visa identificar e organizar a terminologia da área em uma base de dados terminológica *on-line*, oferecendo recursos a todos os profissionais que atuam na produção de textos: assessores linguísticos, tradutores, redatores técnicos, entre outros.

Objetivos

Apresentar a construção de *corpora* textuais em língua portuguesa e espanhola e seu processo de limpeza e armazenamento, tendo como foco a terminologia da *conservação dos bens culturais móveis em suporte de papel*. Destacamos a metodologia aplicada e os resultados obtidos.

Fundamentação Teórica

- Teoria Comunicativa da Terminologia (CABRÉ, 1999).
- Linguística de Corpus (BERBER SARDINHA, 2003).
- Tradução (Hurtado Albir, 2001).

Metodologia

1) Seleção de textos, a partir dos seguintes critérios:

- a) conter as palavras-chave: *documento, documentação, conservação, papel, patrimônio, preservação, restauração e restauro* e seus respectivos equivalentes em espanhol;
- b) pertencer aos gêneros acadêmico e científico: livros, manuais, revistas científicas, teses, dissertações e boletins informativos de associações da área;
- c) estar incluídos em fontes confiáveis: *sites* de universidades, instituições de pesquisa, dentre outros, cujas línguas originais fossem o português e o espanhol.

2) Conversão dos textos de <.pdf> ou <.doc> para <.txt> para atender aos pré-requisitos de processamento no programa *AntConc*, utilizado para a extração de candidatos a termos.

3) Limpeza dos textos para a retirada de informações não pertinentes à pesquisa: agradecimentos, referências bibliográficas, figuras e fotos.

4) Armazenamento:

- a) nomeação dos textos com um código estabelecido pela equipe. Por exemplo: ptPP001, indica o primeiro texto (001) em língua portuguesa (pt), do projeto PaPel (PP).
- b) catalogação dos textos: inserção dos códigos e das referências completas de cada texto selecionado em uma planilha *Excel*.

Código:	Fonte:	Disponível em:	Acesso em:
ptPP107	FRONER, Yacy-Ara. Conservação e Restauração - A legitimação da ciência. Revista Acervo, Rio de Janeiro, V. 23, n. 2, jul-dez, 2010.	http://revista.arquivonacional.gov.br	12 mar. 2017
arPP001	SCALA, Maria Ximenna. Insectos Bibliófagos, parte II. Boletín: n. x, Septiembre 2010.	https://www.inti.gov.ar/	21 mar. 2017

Justificativa

A necessidade de realização dos procedimentos aqui relatados garantem, na próxima fase do trabalho, a qualidade na extração dos termos. Além disso, o catálogo de textos auxiliará os usuários a encontrarem suas respectivas fontes.

Resultados

Português

Total de textos	Total de palavras
157	1.016.034

País	Total de Textos	Total de palavras
Argentina	35	231.061
Chile	13	66.205
Espanha	25	308.625
Paraguai	1	4.712
Uruguai	6	19.062
Total	80	629.665

Considerações finais

1) Durante a pesquisa para a construção de ambos os *corpora*, uma das dificuldades encontradas foi a conversão de textos em formato <.pdf> para <.txt>, pois alguns estavam bloqueados. A solução encontrada foi utilizar o OneNote para resolver o problema.

2) Especificamente em relação ao *corpus* em espanhol uma das dificuldades encontradas foi a obtenção de textos em alguns países, sobretudo no Paraguai e Uruguai. Tal fato pode indicar que nesses países há uma restrição maior ao acesso a esse tipo de material. A solução encontrada foi procurar textos em outros países como Chile e Espanha.

Referências

BERBER SARDINHA, Antonio P. Uso de corpora na formação de tradutores. *DELTA* [online]. 2003, vol.19, n.spe, pp.43-70. ISSN 0102-4450. Disponível em: <http://dx.doi.org/10.1590/S0102-44502003000300005>.

CABRÉ, M. T. *La terminología, representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona: IULA, 1999.

HURTADO ALBIR, Amparo. *Traducción y Traductología. Introducción a la Traductología*. Madri: Ediciones Cátedra, 2001.

ANTHONY, L. *AntConc* (3.2.14w). Tokyo: Waseda University, 2008. Disponível em <http://www.antlab.sci.waseda.ac.jp/software.html>. Acesso em: 20 de ago 2017.