

SALÃO DE
INICIAÇÃO CIENTÍFICA
XXIX SIC

UFRGS
PROPESQ



múltipla 
UNIVERSIDADE
inovadora  inspiradora

Evento	Salão UFRGS 2017: SIC - XXIX SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2017
Local	Campus do Vale
Título	Conservação dos Bens Culturais Móveis em papel: a construção de corpora em português e espanhol
Autor	VITÓRIA SIGNORI ROSO
Orientador	CLECI REGINA BEVILACQUA

Apresentadoras: Sabrina Alexsandra Pacheco Gomes – alexsandra718@gmail.com,
Vitória Signori Roso – vitoriasignori@gmail.com (UFRGS).

Resumo: O presente trabalho tem como objetivo 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. Esta pesquisa insere-se em um projeto maior que intitulado *A Linguagem do Patrimônio Cultural Brasileiro: Conservação dos Bens Culturais Móveis*, do grupo Termisul. Seu objetivo é 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. Os fundamentos teóricos seguem os princípios da Teoria Comunicativa da Terminologia e da Linguística de Corpus. Para a seleção dos textos que constituem os *corpora* em ambas as línguas, foram aplicados os seguintes critérios: a) conterem as palavras-chave “documento, documentação, conservação, papel, patrimônio, preservação, restauração e restauro”; b) pertencerem aos gêneros acadêmicos - livros, manuais, revistas científicas e boletins informativos de associações da área; c) estarem incluídos em fontes confiáveis - *sites* de universidades, instituições de pesquisa, dentre outros, cuja língua original fosse o português e o espanhol. Seguindo tais parâmetros, foram selecionados cerca de 177 em português e 88 em espanhol. Durante a pesquisa, uma das dificuldades encontradas foi a obtenção de textos de alguns países hispanofalantes como o Chile, Paraguai e Uruguai. Por essa razão, atingimos, até o presente momento, um total de 07 textos do Chile; 02 textos do Paraguai; e 08 textos do Uruguai; em contrapartida, obtivemos maior sucesso na coleta de textos da Argentina e Espanha, alcançando um total de 47 e 24 textos, respectivamente. Após a coleta, iniciamos o processo de limpeza e armazenamento. No processo de limpeza são retiradas todas as informações não significativas para a pesquisa, tais como agradecimentos, referências bibliográficas, figuras e fotos. Em seguida, os textos são convertidos de PDF ou Word para .txt, atendendo aos pré-requisitos de processamento no programa *AntConc*. Este programa é utilizado para extração de informações linguísticas, entre elas, os candidatos a termos. Após essa etapa, os textos são nomeados com um código estabelecido pelo grupo. Por exemplo, arPP001, indica o primeiro texto (001) em língua espanhola-Argentina (ar), do projeto PaPel (PP). Por fim, organiza-se um catálogo que contém os códigos e as referências completas dos textos selecionados. Esse catálogo será disponibilizado juntamente com a base de dados terminológica para dar maior confiabilidade aos usuários em relação às fontes de coleta dos termos. Todos esses procedimentos fazem parte de uma metodologia para a construção de *corpus*, garantindo, 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.