



Evento	Salão UFRGS 2022: SIC - XXXIV SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2022
Local	Campus Centro - UFRGS
Título	Como dar forma a um corpus? Construção de corpus especializado na área do Patrimônio Cultural Imaterial
Autor	MARIANA DA TRINDADE CARPI NEJAR
Orientador	CLECI REGINA BEVILACQUA

Esta apresentação faz parte do projeto *A terminologia do patrimônio cultural imaterial*, desenvolvido de agosto de 2021 a julho de 2024 pelo grupo de pesquisa Termisul. O objetivo geral deste projeto é identificar e representar, em uma base de dados *on-line* e gratuita, os termos e UFEs da área do Patrimônio Cultural Imaterial em português e seus equivalentes nas línguas estrangeiras (espanhol, francês, inglês, italiano e russo). Entre os objetivos específicos, está a coleta e preparação de *corpora* comparáveis (língua portuguesa e línguas estrangeiras), conformados por textos de diferentes gêneros (acadêmicos, normativos, etc.). Essa etapa é fundamental, pois, a partir desses *corpora*, os termos e fraseologias usados pelos profissionais da área serão selecionados na etapa seguinte do projeto. Nesta apresentação, objetivamos descrever o processo de construção do *corpus* em espanhol sobre o Patrimônio Cultural Imaterial. Justificamos tal pesquisa e relato de experiência pela importância de construir *corpora* comparáveis com critérios bem estabelecidos, a fim de extrair deles as informações linguísticas adequadas aos usuários previstos para a base (tradutores, redatores, estudantes, etc.). Serão apresentados os critérios de seleção, catalogação e limpeza, bem como as dificuldades e as soluções encontradas na busca dos textos que compõem o *corpus*. Uma das dificuldades, por exemplo, foi estabelecer subcategorias (*subcorpora*). Primeiramente, dividimos o *corpus* do espanhol em subcategorias de cada país – Uruguai, Argentina, Paraguai, etc. –, mas tal disposição se mostrou insuficiente, pois não abarcava todos os textos coletados. Foi necessário incluir três subcategorias adicionais: a) Mercosul (textos sobre o bloco econômico); b) Textos em castelhano (textos com coautorias de diferentes nacionalidades e/ou com temáticas de países que não são a terra natal dos autores); c) Textos traduzidos (do português ao espanhol). Até o momento, coletamos 4 textos para Mercosul e 24 para Textos em castelhano; a subcategoria Textos traduzidos está em vias de execução.