



## **BIG DATA MEDIA ANALYSIS AND THE REPRESENTATION OF URBAN VIOLENCE IN BRAZIL - FORMAÇÃO DO CORPUS**

Marta Joanita Cassânego (PROBIC-FAPERGS), Heloisa Pedroso De Moraes Feltes (Orientador(a))

Nesta comunicação descrevem-se os procedimentos utilizados para a construção do *corpus* jornalístico da pesquisa DATA-UV, como parte da cooperação internacional de pesquisa UCS-CASS/Lancaster University/UK. A partir desse *corpus* visa-se analisar e problematizar, com a aplicação de métodos, técnicas e procedimentos da Linguística de *Corpus*, a veiculação dos casos de violência urbana na mídia brasileira. O processo de formação do *corpus* inicia com a definição da amostra de jornais brasileiros. Nesse caso, foram eleitos os jornais Pioneiro, Zero Hora, Folha de São Paulo e Estado de São Paulo, representando dois estados brasileiros. Com o apoio financeiro do CASS/Lancaster University foi acessada a base de dados Factiva. Essa base de dados é restrita e proporciona o acesso a mais de 36 mil fontes de conteúdo, em diferentes línguas, de jornais, revistas e publicações de economia. Nessa base de dados foram selecionados os jornais, nas edições de 01/01 a 31/12/2014, a partir de termos de busca como 'violência urbana', 'segurança pública' e 'criminalidade'. Com base nos critérios estabelecidos pelos termos de busca foi feito o *download* de matérias jornalísticas. Utilizando-se o *software* SarAnt, foram eliminados elementos não textuais irrelevantes para a análise posterior com o *software* AntConc. Como resultado, partindo de um *corpus* de 5.127 artigos, obteve-se um *corpus* final composto por 1.778.282 palavras, o qual foi submetido aos procedimentos rigorosos da Linguística de *Corpus* para a consecução dos objetivos propostos na cooperação internacional. Tais procedimentos possibilitaram a construção preliminar de um corpo de dados a ser mais profundamente analisado com relação à representação da violência urbana na mídia nacional brasileira; um conhecimento técnico mais avançado a respeito do uso de *softwares* para análise de dados linguísticos e, também, uma experiência prática bastante efetiva no campo da Linguística de *Corpus*.

Palavras-chave: Linguística de Corpus, Violência urbana, Mídia jornalística, Metodologia

Apoio: FAPERGS