

CapaDeRevista2017: construção de um corpus de reportagens de capa

Diulia Justin Deon¹, Leonardo Pereira dos Santos¹, Kathlen Luana de Oliveira¹,
Alexandre Ricardo Lobo de Sousa¹, Larissa Astrogildo de Freitas^{1*}
*Orientadora

¹Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) -
Campus Osório, Osório, RS

A mídia constrói uma narrativa e uma representação de fenômenos que é reconhecida como um retrato da realidade. As revistas, em seus editoriais e reportagens, retratam posicionamentos frente aos fatos histórico-político-culturais, defendendo cosmovisões, formando uma interpretação e uma posição de um grupo, indo além da informação. Suspeita-se que as premissas jornalísticas de narração dos fatos estejam comprometidas com os interesses de empresas, impossibilitando interpretações críticas sobre a realidade, violando exigências da concessão constitucional e até os direitos humanos. Considerando a liberdade e a influência da mídia, esta pesquisa tem como objetivo identificar e analisar posicionamentos nas reportagens de capa do primeiro semestre de 2017 de quatro revistas semanais brasileiras: CartaCapital, Época, IstoÉ e Veja. O critério de escolha foi estipulado de acordo com o IVC (Instituto Verificador de Circulação), o qual considera essas revistas como amplamente utilizadas para informação geral. Utilizando a linguística computacional, analisar-se-á como a imprensa pretende consolidar uma visão de mundo, defendendo discursos e posicionamentos. A coleta de dados distinguirá criticamente posicionamentos de revistas, avaliando se estas são neutras, imparciais e informativas. Os métodos utilizados perfazem três etapas: pesquisa bibliográfica acerca do texto jornalístico; construção de um corpus contendo os objetos de análise e construção e/ou adaptação de recursos computacionais para realização de análises linguísticas das reportagens de capa. Um corpus é um conjunto de documentos sobre um tema, elaborados a partir do estudo da estrutura dos elementos. Esses elementos irão compor a ferramenta, os documentos de uma determinada língua, os quais são armazenados em formato eletrônico, permanecendo disponíveis a consultas. A construção do corpus intitulado CapaDeRevista2017 deu-se em três fases: análise da estrutura dos acervos digitais das revistas, coleta e adaptação do material e população do banco de dados. As revistas Época e IstoÉ disponibilizam páginas web, logo, os textos foram extraídos dos HTMLs. Já, as revistas Veja e CartaCapital disponibilizam PDFs, o que exigiu a conversão para texto editável através de um processo semi-automático. Ainda, foi construída uma base de dados na qual foram inseridas as reportagens de capa, seus títulos e suas edições. O corpus CapaDeRevista2017 é composto por 26 edições das revistas Veja, Época e IstoÉ, e 18 edições da revista CartaCapital, totalizando 96 textos. Nessa perspectiva, a pesquisa construiu um corpus de aproximadamente 258 mil palavras. Como trabalhos futuros, planeja-se a disponibilização do corpus em plataforma online, e ainda, serão realizadas análises linguísticas dos textos, buscando a identificação dos posicionamentos.

Palavras-chave: Linguística de corpus. Texto jornalístico. Reportagem. Imprensa.