

Descrição de gêneros textuais/discursivos com apoio computacional

Com muita satisfação, apresentamos o volume 11, número 2, de 2017 da nossa querida revista Domínios de Linguagem, que nos tem trazido importantes trabalhos das mais variadas áreas dos estudos linguísticos. Desta vez, o volume temático gira em torno da descrição de gêneros discursivos e/ou textuais com algum tipo de suporte computacional. Instaure-se, assim, desde a chamada, alguma interdisciplinaridade, seja teórica ou metodológica, entre os Estudos do Texto e do Discurso e a Informática. Ao propor este volume, imaginamos poder reunir trabalhos que trouxessem, partindo da descrição e da caracterização – ou mesmo da problematização do estatuto de gênero, diferentes tipos de aplicações. A descrição do nosso objeto linguístico, tornado concreto via texto e práticas de discurso, conforme entendemos, dá suporte à análise. Dessa análise, feita sob diferentes pontos de vista teóricos, os quais iluminam determinadas facetas do nosso objeto, poderemos partir para a explicação e, posteriormente, para as generalizações e previsibilidades. Feito esse percurso, pode-se retornar ao ponto de partida teórico e verificar o quanto ele, efetivamente, nos dá suporte.

Desde muito, o papel do processamento, via computador, da língua natural tem já se mostrado um empreendimento científico e multidisciplinar tanto relevante quanto desafiador. Se hoje testemunhamos muitos avanços para línguas como o inglês, para o português, o avanço é ainda desigual, embora reconhecido como necessidade. Para se ter uma ideia, em 1999, já líamos em uma notícia do jornal português *Público* a chamada para a seguinte notícia (<https://www.publico.pt/ciencias/jornal/a-lingua-de-camoes-na-informatica-132347>): “A língua de Camões na Informática”. Nessa notícia, entrevistava-se a nossa colega pesquisadora Diana Santos, e assinalava-se “A adaptação dos computadores à língua portuguesa é considerada pelo Ministério da Ciência e da Tecnologia uma área prioritária”.

Anos mais tarde, em 2007, em livro organizado por Diana Santos (SANTOS, 2007), encontramos um exemplo do avanço disso no trabalho de Quaresma e Rodrigues (2007), ambos cientistas da Computação. Nele vemos uma pesquisa dedicada ao tema da Recuperação da Informação que reúne e explora uma base de texto 7 mil (!) pareceres da Procuradoria-Geral da República Portuguesa, produzidos tais documentos entre 1940, compondo um universo de cerca

de dez milhões de palavras. A proposta dos autores era a de um sistema interativo para recuperação de informações nesse universo de documentos. Para tanto, desde a unidade palavra até a unidade “tipo de texto” precisou ser enfrentada, e bem sabemos o quanto a linguagem jurídica é complexa.

Atendendo à proposta desses tipos de estudos, estão aqui artigos que nos trazem um rico panorama: da descrição da linguagem – falada ou escrita - para fins diversos – como a arquitetura de sistemas computacionais que lidam com textos e conseguem, de algum modo, representar seu conteúdo – à obtenção de recursos para o ensino/aprendizagem de línguas. A intersecção, entre os Estudos da Linguagem, o Processamento da Linguagem Natural e a Linguística de *Corpus*, de certo modo esperável nesse cenário, destaca mais pontos de diálogo possíveis, instigantes e produtivos, como o aporte da Linguística Sistêmico-Funcional.

Enfim, muitos linguistas e cientistas da Computação – entre outros cientistas – têm lidado com gêneros textuais e discursivos sob diferentes enfoques. Não raro, encontramos belos trabalhos sobre esse tema em eventos de áreas de conhecimento cujo diálogo com a área de Letras não tem sido muito frequente: não é difícil imaginar, por exemplo, em um evento de Farmacologia, vários estudiosos de Ciências da Saúde debruçando-se, entre outras coisas, sobre os problemas conceituais, linguísticos e terminológicos das bulas de medicamentos. Em paralelo, sabemos, se precisar supor, que esse mesmo objeto, a bula, já rendeu vários estudos na área de Letras, de artigos a teses de doutorado. Entendemos, assim, que o tópico gênero – seja textual ou discursivo – é um potente agregador, uma ponte que pode facilitar nosso encontro, cada um com a sua ciência em relação com o outro e com a sua ciência. Além disso, essa noção de gênero parece que precisa sempre ser retomada entre nós que pesquisamos sobre a linguagem em funcionamento, a linguagem que instaura diálogos e cumpre funções.

Sucedendo à publicação deste volume, vale já anunciar aos interessados no tema aqui tratado o XIV Encontro de Linguística de *Corpus* (ELC) e a IX Escola Brasileira de Linguística Computacional (EBRALC). São dois eventos associados, que devem ocorrer em agosto de 2017. Na edição de 2017, O ELC, justamente, visa provocar reflexões sobre as contribuições da Linguística de *Corpus* para os estudos sobre gêneros textuais e discursivos. E a EBRALC quer mostrar como podemos nos aproximar das tecnologias computacionais para o estudo da linguagem. Fica, portanto, reforçada a temática dos gêneros como algo importante, sem esquecer de mencionar aqui o próprio SIGET - IX Simpósio Internacional de Estudos dos Gêneros Textuais, maior evento da área, que ocorre também em agosto de 2017.

Feita essa breve introdução passamos agora a comentar um pouco sobre cada um dos trabalhos que perfazem este volume, esperando que esse comentário possa contribuir para a entusiasmar a leitura de cada um.

Inicia o volume o artigo intitulado “**O gênero entrevista radiofônica em comunidades hispânicas: um aporte da análise textual automática**” da autoria de Leandro Silveira de Araujo. Nesse trabalho, é descrito o gênero *entrevista radiofônica* em comunidades de língua espanhola. Para tanto, recorre-se a uma retomada dos principais aspectos envolvidos na caracterização dos gêneros discursivos conforme a abordagem sociointeracionista. O trabalho nos apresenta o *Tropes 7.23*, um *software* de processamento automático de textos, útil para a descrição dos âmbitos temático, estilístico, estrutural e funcional dos enunciados que compõem a *entrevista radiofônica*. Foram estudadas entrevistas produzidas em Madri, Buenos Aires e San Miguel de Tucumán – cidade da Argentina – difundidas por rádio e pela *web*. Esse trabalho nos traz ótimas ideias para novos estudos que se ocupem de questões relacionadas à hispanofonia, à variação diatópica e apresenta-nos um suporte metodológico muito atraente. A ferramenta computacional, a serviço da descrição desse gênero, conforme vemos no trabalho, permite perceber algumas nuances que talvez não fossem tão evidentes sem esse aporte tecnológico.

O segundo artigo é “**Interação em rede social: das reações às características do gênero comentário**” de Roberlei Alves Bertucci e Paula Ávila Nunes. O objetivo do trabalho é analisar a relação entre o gênero comentário em uma página no *Facebook* e reações de usuários nesse contexto. Nada mais atual, considerando o papel das redes sociais no cenário atual. Com o aplicativo *Netvizz*, os autores examinam o “odiar” manifestado pelos comentaristas. Conforme os autores mesmos apontam, estudar a linguagem em contextos da chamada era digital “é imperativo em nossa constante tentativa, como linguistas, de entender o modo pelo qual a linguagem se coloca como fundamento antropológico principal, sobretudo em sua relação com a tecnologia”. Outro ponto a destacar neste artigo é a verificação das condições e características de um gênero textual emergente. Esse tipo de objeto, o teor de comentário de usuários em redes sociais, bem sabemos, tem rendido toda uma série de pesquisas, do Marketing à prospecção de tendências políticas. Em Processamento de Linguagem Natural, esse campo tem sido denominado de *Análise de Sentimentos* e reúne várias áreas, como Linguística, Psicologia e Computação. Para os linguistas interessados nessa inter- e multidisciplinaridade, sugerimos consultar a tese recente de Silva (2016), em Ciência da

Computação, cujo título é, justamente, *Análise de sentimentos em textos curtos provenientes de redes sociais*. Veja a referência ao final deste texto.

O terceiro artigo, que também trata do tópico *Análise de Sentimentos*, é “**Um estudo sobre referenciação e a construção da opinião a partir de um corpus textual extraído do YouTube**”, de Alexandre Ribeiro e Afonso Jordão Té. Os dois autores não são linguistas; são formados e atuantes nas áreas da Ciência/Gestão da Informação e Computação e empreendem um belo trabalho de diálogo com os estudos da Linguagem, especialmente com a Linguística do Texto. Esse trabalho analisa um *corpus* de opiniões extraído das postagens do *YouTube*. Os autores verificam como o fenômeno discursivo da referenciação ocorre em tais postagens, isto é, buscam compreender como um dado objeto de interesse em discussão na mídia é renomeado. Foram examinadas as postagens opinativas favoráveis ao *impeachment* da então presidente do Brasil, Dilma Rousseff, e as postagens contrárias ao processo. Os dados foram gerados de três vídeos que relatam ou comentam sobre o tema. A ferramenta computacional para tratamento do texto foi o nosso conhecido *software* AntConc, recurso de acesso gratuito bastante utilizado por quem lida com Linguística de *Corpus* no Brasil.

O quarto trabalho do volume, “**Gêneros textuais datasheet e artigos científico em aulas de ESP: levantamentos léxico-estatísticos para fins educacionais**” de Sabrina Bonqueves Fadanelli e Andrea Jessica Borges Monzón, ambas professoras de inglês em cursos tecnológicos no Sul do Brasil, objetivou extrair dados relevantes, de um *corpus* textual composto por textos técnicos e científicos, para a preparação de material para o ensino de terminologia técnica e científica em aulas de ESP – *English for Specific Purposes*. As autoras examinaram artigos acadêmicos da área de Computação e um gênero textual bem específico da área de Eletrotécnica: os *datasheets*. As ferramentas computacionais empregadas foram o já antes citado AntConc e o extrator automático de candidatos a terminologias, também de acesso gratuito, o *TermoStat*. Interessaram às autoras especialmente as colocações mais frequentes em torno das terminologias empregadas nos textos.

O quinto artigo é “**Português para Fins Acadêmicos sob o Aporte da Linguística de Corpus e do Processamento de Linguagem Natural**”, que reúne, como autores Cristina Becker Lopes Perna, professora e pesquisadora da área de inglês e Linguística, Lucelene Lopes, graduada em Matemática e pesquisadora de Ciência da Computação/PLN, e Lucas Zambrano Rollsing também da área de Letras/Linguística. O artigo traz resultados iniciais de estudo sendo desenvolvido entre dois programas de Pós-Graduação da Pontifícia Universidade Católica do

Rio Grande do Sul (PUCRS), o Programa de Pós-Graduação em Letras (PPGL), na sua área de concentração em Linguística, através do Grupo de Pesquisa sobre o Uso e Processamento de Língua Adicional (UPLA), e o Programa de Pós-Graduação em Ciência da Computação (PPGCC), através do Grupo de Processamento em Linguagem Natural. O tópico de interesse do trabalho conjunto é o Português para Fins Acadêmicos, representando por um *corpus* de 200 dissertações e teses sobre Linguística produzidas na PUCRS. Com apoio da ferramenta EXATO, esse acervo de pesquisas é tratado como *corpus* e são reconhecidos, automaticamente, os principais temas dos trabalhos, nas diferentes subáreas de Linguística abrigada pelo PPGL. Com o estudo, demonstra-se como se poderia, por exemplo, representar os principais temas de pesquisa ao longo de todo um quadriênio de um programa de pós-graduação. Em tempos de avaliação e prospecção da produtividade de nossos cursos, esse sistema poderia ser muito útil para gestores e para os próprios pesquisadores-produtores dos trabalhos.

O penúltimo artigo, “**Ideologia em gêneros discursivos pela análise associada da Linguística Sistêmico-Funcional e Linguística de *Corpus*: cantigas galego-portuguesas**” de Ulisses de Oliveira, logo de início, situa a recorrente discordância, entre linguistas, na definição de gêneros do discurso. Razão pela qual esse volume emprega, justamente, gêneros discursivos/textuais. Além disso, o artigo propõe tratar do tópico ideologia, uma dimensão contextual superior, que considera ser geralmente desprestigiada nas análises de gênero, dado seu caráter amplo e multinível, de complexa manipulação para o pesquisador. O trabalho traz uma proposta de análise cruzada de marcadores ideológicos postos nos textos, com aporte da Linguística Sistêmico-Funcional e da Linguística de *Corpus*. Enfim, temos um belo exemplo de enfrentamento de um tema bastante complexo, a Ideologia.

O último trabalho do volume intitula-se “**Modelagem topológica do perfil sistêmico-funcional de manuais de instrução no espaço gramatical**”. Esse trabalho, de autoria de Francieli Silvéria Oliveira, Giacomo Patrocínio Figueredo e Thaís Guerra Campolina, todos egressos da área de Letras/Linguística, investiga o modo pelo qual os sistemas gramaticais da oração contribuem para a produção de significados do texto, apoiando-se também em pressupostos da Linguística Sistêmico-Funcional e da Linguística de *Corpus*. O objeto sob estudo, tomado como *corpus*, é o gênero *manual de instrução* em português brasileiro. São examinados diferentes tipos de manuais, que abrangem desde tema como a criação de peixes até procedimentos de Enfermagem. O estudo, apoiado pelo *software UAM Corpus Tool*, demonstra, com grande detalhamento, como a análise gramatical é capaz de revelar como se

organizam aspectos fundamentais dos gêneros textuais. O tratamento quantitativo dos dados, fartamente ilustrado com gráficos multidimensionais, é um rico diferencial do trabalho. Conforme os autores, é possível prever, com confiança, o comportamento da gramática oracional do português brasileiro para quaisquer textos caracterizados como manual.

Assim, conforme o nosso leitor poderá conferir, com a leitura dos artigos deste volume, as pesquisas recentes com os *corpora* envolvidos em cada um dos estudos não trouxe mera contagens de palavras em textos com a ajuda de algum *software*. O apoio computacional não é um fim, ele é um meio, e ele não precisa nos intimidar.

Os artigos deste volume, no seu conjunto, mostram e validam métodos, sim, mas também apontam e reiteram correlações entre gêneros e tipologias textuais, cenários comunicativos, elementos lexicais, gramaticais, discursivos, semânticos e temáticos, sejam feitos por linguistas, por cientistas de Computação/Informação ou por ambos, associados. Esses trabalhos também se colocam epistemologicamente, filiam-se a teorias e a diferentes ciências. Esse é o diferencial desses trabalhos, ousar em direção ao diálogo interdisciplinar e instigar novas pesquisas, cada um a seu modo. Dos exemplos citados no início desta apresentação, da notícia sobre o processamento computacional do português de 1999, à pesquisa de 2007 com textos jurídicos, até o todo representado pela pequena amostra que vemos aqui reunida neste ano de 2017, pode ser percebido um percurso evolutivo em prol do tratamento da linguagem. Assim pensando, espero que cada vez mais o diálogo entre nós linguistas e os colegas de Computação e de áreas afins possa ser fortalecido. Desejo uma boa leitura a todos e que os trabalhos deste volume possam nos inspirar a prosseguir outros tantos.

Maria José Bocorny Finatto (UFRGS/CNPq/CAPES).
Abril de 2017.

Referências

QUARESMA, P.; RODRIGUES, I. Avaliação de sistemas interactivos de recuperação de informação em bases de texto jurídicas. p. 257-265. In: SANTOS, D. (Ed.) **Avaliação conjunta. Um novo paradigma no processamento computacional da língua portuguesa**. Ist Press: Lisboa, 2007. p. 257-265.

SANTOS, D. (Ed.) **Avaliação conjunta. Um novo paradigma no processamento computacional da língua portuguesa**. Ist Press: Lisboa, 2007.

SILVA, N. F. F. da. **Análise de sentimentos em textos curtos provenientes de redes sociais** [online]. Tese de Doutorado em Ciências de Computação e Matemática Computacional. São Carlos: Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2016. Disponível em: <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-27092016-143947/> Acesso em: 7 de abril de 2017.