

Ferramenta linguístico-computacional como facilitadora para o ensino de gramática na escola

Lívia Vicente Dutra¹, Natália Sathler Sigiliano¹

¹FrameNet Brasil - Faculdade de Letras

Universidade Federal de Juiz de Fora (UFJF)

Rua José Lourenço Kelmer, s/nº, Campus Universitário

36036-900 – Juiz de Fora – Minas Gerais – Brasil

livia.vdutrael@gmail.com, natalia.sigiliano@ufjf.br

Abstract. *The research "Genres, textual typologies and linguistic analysis: constitution of didactic resources for the contextualized work of linguistic knowledge in an approach guided by textual genres" aims to develop a linguistic-computational tool to assist professors in the approach to content of grammar based on textual genres, given that the relationship between textual genres and their linguistic-discursive construction still constitutes a challenge for the teaching of Portuguese language. This tool is built based on annotations from FrameNet Brasil. Through this resource, it is expected to make it possible for the user to search for textual genres that are more conducive to teaching certain grammatical topics, given their prominence in these genres.*

Resumo. *A pesquisa "Gêneros, tipologias textuais e análise linguística: constituição de recursos didáticos para o trabalho contextualizado dos conhecimentos linguísticos em uma abordagem orientada pelos gêneros textuais" objetiva desenvolver uma ferramenta linguístico-computacional de auxílio ao professor para a abordagem de conteúdos de gramática pautados em gêneros textuais, já que a relação entre gêneros textuais e sua construção linguístico-discursiva ainda se constitui em um desafio para o ensino de língua. Esta ferramenta é construída com base em anotações da FrameNet Brasil. Por meio dela, espera-se tornar possível ao usuário a busca por gêneros textuais mais propícios para o ensino de determinados tópicos gramaticais, tendo em vista sua proeminência nesses gêneros.*

1- Introdução

A pesquisa "Gêneros, tipologias textuais e análise linguística: constituição de recursos didáticos para o trabalho contextualizado dos conhecimentos linguísticos em uma abordagem orientada pelos gêneros textuais" foi motivada a partir de uma demanda frequente no Brasil quanto ao ensino de gramática na educação básica de forma contextualizada e pautado no uso de gêneros textuais. Tal necessidade se revela em práticas de ensino de gramática ainda arraigadas no nível da análise do período e da oração, especialmente no que diz respeito ao ensino de sintaxe e morfologia. Como forma de reaver tal relação, os professores de Língua Portuguesa (LP) buscam no nível da textualidade alternativas para a exploração dos efeitos de sentido no uso de determinadas categorias gramaticais disponíveis no texto. Entretanto,

concebe-se que “as escolhas linguístico-discursivas presentes num dado gênero não são aleatórias, mas ali estão para permitirem que um gênero funcione socialmente” (MENDONÇA, 2007). Dessa forma, a análise das estruturas micro textuais características de determinados gêneros textuais pode auxiliar nas escolhas de conteúdos gramaticais a serem desenvolvidos nas aulas de LP. Essa escolha, por vezes, tem ocorrido com base no repertório do professor ou do autor de material didático com relação aos seus conhecimentos anteriores com o gênero textual em estudo. Sendo assim, a pesquisa tem sido desenvolvida na FrameNet Brasil (FN-Br) com o objetivo de elaborar uma ferramenta linguístico-computacional que auxilie o professor da educação básica no ensino de LP, associando tipos e gêneros textuais a seus itens gramaticais mais proeminentes.

A prática de Análise Linguística (AL) prevê o relacionamento estreito com outras práticas de ensino aprendizagem de LP, ou seja, à leitura e à produção de textos (GERALDI, 1984; MENDONÇA, 2006; BRASIL, 1997; BRASIL, 2017). Contudo, diversas pesquisas revelam que os materiais didáticos e os professores de língua materna ainda encaram o ensino de AL como independente de outras práticas, travando uma relação de ensino de LP descontextualizado e afastado de práticas sociais. Ao mesmo tempo, muitos manuais de ensino e docentes já inovam suas práticas ao preverem aproximações bastante claras, como aquelas relativas ao estudo do modo imperativo a partir de textos instrucionais. Não há, no entanto, muitas destas associações reconhecidas e desenvolvidas no que tange aos elementos sintáticos e às categorias morfológicas (SIGILIANO e SILVA, 2019).

Esta pesquisa assume o propósito de auxiliar o professor da escola básica no cumprimento de abordagens mais recentes para o ensino de LP, a partir de um trabalho conjunto entre Linguística Computacional, Ciência da computação e Ensino de Língua Materna.

2- Metodologia

A pesquisa foi subdividida em quatro etapas, sendo elas: seleção de *corpus*, anotação de texto corrido, extração de padrões linguísticos dos textos e, por fim, elaboração da ferramenta.

Na primeira etapa, foram selecionados 25 textos de diferentes gêneros textuais, de caráter modelar, dentre os tipos argumentativo, descritivo, expositivo, injuntivo e narrativo. Buscou-se equilibrar a extensão dos textos. Estes textos compuseram o *corpus* de análise da pesquisa, que foi incorporado à base de dados da FN-Br para que as anotações de texto corrido pudessem ser realizadas.

As anotações computacionais foram feitas na plataforma da FN-Br, denominada *Web Annotation Tool (WebTool)*¹, seguindo as diretrizes de anotação estabelecidas por Ruppenhofer et al (2016). O grupo de anotadores envolvidos foi composto por dez alunos da graduação que cursaram a oficina de anotação computacional, somando um total de 30 horas de dedicação no ano de 2019. Posteriormente, três bolsistas de Iniciação Científica (IC) envolvidos na pesquisa e treinados nas diretrizes de anotação fizeram as revisões e deram continuidade à tarefa de anotação de texto corrido, constante durante todas as etapas do trabalho. Uma anotação de texto corrido consiste na análise de cada uma das Unidades

¹ Link para acesso: <http://webtool.framenetbr.ufjf.br/index.php/webtool/main>

Lexicais (UL) constantes no texto, as quais evocam um determinado frame² cada. Essa análise é realizada em três níveis: o nível dos elementos de frame, o nível sintagmático e o nível das funções gramaticais, ou seja, ela é capaz de extrair e sinalizar dados sintáticos-semânticos. Um exemplo deste tipo de anotação pode ser observado na Figura 1, em que a UL *dedicar* foi anotada no frame *Causar_fazer_progresso*. Na primeira camada de anotação, foram marcados os elementos de frame *projeto* e *duração* de forma explícita e o *agente* a partir de uma Instanciação Nula Construcional (CNI), que indica que o sujeito é marcado pela desinência do verbo. Nas segunda e terceira, observam-se as funções e classificações gramaticais correspondentes. Esses dados são armazenados no sistema, possibilitando, por meio da elaboração de queries, a extração de padrões linguísticos, suas respectivas frequências e contextos de ocorrências relacionados aos gêneros textuais. Os itens gramaticais anotados foram nomes, verbos, adjetivos e advérbios.

	FE	GF	PT
Causar_fazer_progresso.dedicar.v	CNI		
Projeto		ObjD	NP
Duração		ObjD	NP

Figura 1 - Anotação da FrameNet Brasil em três camadas para a Unidade Lexical *dedicar* no frame *Causar_fazer_progresso*

A última fase envolve a elaboração da ferramenta linguístico-computacional, uma Interface de Consulta *online*. Ela ainda está em fase de teste e aprimoramento, por apresentar, por exemplo, dados com as nomenclaturas específicas à anotação na base de dados da FrameNet, mas já é possível fazer buscas e identificar alguns padrões linguísticos extraídos das anotações.

3- Resultados

A versão disponível da Interface de Consulta permite realizar buscas por *corpus*, tipos e gêneros textuais, abrangendo resultados semânticos e sintáticos, que possibilitam encontrar dados que evidenciam a relação dos gêneros com determinadas classes gramaticais ou padrões sintáticos (Marcuschi, 2002). A ocorrência, por exemplo, do modo imperativo em textos instrucionais aparece nos dados comprovando uma metodologia de ensino, em um primeiro momento, empírica. Ao escolher o gênero receita, observa-se que, por ter predominância de verbos no imperativo, os sujeitos são marcados como CNIs, ou seja, licenciados pela construção gramatical do imperativo. O mesmo não ocorre quando a busca é realizada pelo gênero notícia em que há domínio de sujeitos oracionais, simples ou compostos, ou seja, expressos através de sintagmas nominais ou orações.

São estes modelos de resultados que se espera aplicar para todos os gêneros e tipos textuais disponíveis na Interface. No entanto, resultados gerados a partir da nomenclatura usada para as anotações não são funcionais para o professor da escola básica que, por vezes, desconhece tais termos por serem de uso muito específico da Semântica de Frames

² De acordo com Fillmore (1982) frames são “qualquer sistema de conceitos relacionados de tal forma que para entender qualquer um deles é preciso entender toda a estrutura em que ele se encaixa; quando uma das coisas em tal estrutura é introduzida em um texto ou em uma conversa, todas as outras estão automaticamente disponíveis.”

(FILLMORE, 1982) e da Gramática de Construções (KAY & FILLMORE, 1999). Dessa forma, foi adicionada à tarefa de anotação uma quarta camada, denominada *sent* (sentença), que permite apresentar os padrões linguísticos encontrados em terminologias comuns ao contexto da escola, referentes à análise linguística, como é ilustrado na Figura 2, em que a UL *dedicar* é marcada como transitivo direto e indireto e seus complementos identificados. A inclusão da camada *sent* ocorreu na atual fase do projeto e, posteriormente, será adicionada à metodologia. A anotação da camada foi iniciada no ano de 2021 e está sendo realizada pelos bolsistas de IC da FN-Br. Estes dados serão disponibilizados na plataforma e, assim, o professor terá acesso a nomenclaturas que possam ser apresentadas a ele com mais clareza no que diz respeito à relação delas com os gêneros textuais com maior proeminência.

Layer	Annotation	Target
FE	CNI	dediquei
GF	Projeto	dediquei
PT	ObjInd	dediquei
Sent	Objeto_indire	dediquei
Frame	Causar_fazer_progresso.dedicar.v	dediquei
Frame	Duração	os melhores anos de minha vida
Frame	ObjD	dediquei

Figura 2 - Anotação da FrameNet Brasil para a Unidade Lexical *dedicar* no frame *Causar_fazer_Progresso* com a camada *sent*

Numa fase posterior, as anotações das quatro camadas realizadas sobre o *corpus* piloto serão usadas para treinamento de um algoritmo de aprendizagem de máquina de modo que, ao final do trabalho, todo e qualquer texto anotado no padrão FrameNet possa ser analisado automaticamente quanto aos elementos gramaticais tipicamente explorados no contexto escolar e, assim, possa passar a compor a Interface de Consulta.

Para além destes resultados, a pesquisa tem beneficiado não somente a elaboração de uma ferramenta *online* de busca de textos exemplares para que professores possam explorar competências de Análise Linguística de forma mais relevante para cada gênero textual, mas também enriquecido a base de dados da FN-Br. As anotações no *corpus* piloto geraram até o momento mais de 3500 UL anotadas em frames já existentes na base de dados da WebTool, além de muitas outras Unidades Lexicais que têm alavancado discussões sobre a criação de novos frames mais adequados para seus contextos.

4- Conclusões

Com base nos resultados preliminares expostos, já é possível comprovar a hipótese inicial da pesquisa que defende a associação de tipos e gêneros textuais a seus itens gramaticais mais predominantes, através de uma análise textual semântico-sintática e abstração de padrões linguísticos, aprimorando a metodologia de ensino de análise linguística. Dessa maneira, fica evidente que a Linguística Computacional pode ser uma grande aliada no ensino de língua, auxiliando professores em suas práticas pedagógicas.

Os próximos passos da pesquisa envolvem a finalização da anotação da camada *sent*, o aumento da base de dados, incluindo novos gêneros e mais exemplares de textos, o desenvolvimento do algoritmo de aprendizagem de máquina e o aprimoramento da interface de consulta para que esta se torne mais acessível e funcional para o seu usuário, o professor de Língua Portuguesa da educação básica.

Referências

- Brasil.(2017). Base Nacional Comum Curricular. Brasília: MEC/Secretaria de Educação Básica. Disponível em: <http://basbroenacionalcomum.mec.gov.br/>. Acesso em: agosto de 2021.
- Brasil.(1997). Parâmetros curriculares nacionais: introdução aos parâmetros curriculares nacionais. Brasília: MEC/SEF .
- Fillmore, C.J. (1982). Frame semantics. In *Linguistics in the Morning Calm*, (ed. by The Linguistic Society of Korea), Seoul: Hanshin
- Geraldi, J. W. (1984). *O texto na sala de aula*. Cascavel: Assoeste.
- Marcuschi, L. A. (2002). Gêneros textuais: definição e funcionalidade. *Gêneros textuais e ensino*, 2, 19-36.
- Mendonça, M.(2007) Análise linguística: refletindo sobre o que há de especial nos gêneros. In: Santos, C. F., Mendonça, M., Cavalcanti, M. C. B. (Orgs.). *Diversidade textual: os gêneros na sala de aula* (pp. 73-87). Belo Horizonte: Autêntica.
- Mendonça, M.(2006). Análise linguística no ensino médio: um novo olhar, um outro objeto. In: Bunzen, C.; Mendonça, M. (Orgs.). *Português no ensino médio e formação de professor*(pp. 199-226). São Paulo: Parábola Editorial.
- Ruppenhofer, J, Ellsworth, M. Petruck, M. Johnson, C. , Baker, C. And Scheffczyk, J. (2016). *FrameNet II: Extended Theory and Practice*. Disponível em <https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf>. Acesso em outubro de 2020
- Sigiliano, N. S.; Silva, W. R.(2017) Diagnóstico de propostas de análise linguística em livros didáticos aprovados em programa oficial. In: Magalhães, T. G., Reis, A. R. G.; Ferreira, H. M. (Orgs.). *Concepção Discursiva da Linguagem, ensino e formação docente*. (pp. 19-40). Campinas: Pontes.