

## Análises em corpora linguística por meio de ferramentas digitais

### Corpus Linguistic analysis by means of digital tools

#### RESUMO

Esse artigo apresenta o percurso da iniciação científica ligado ao projeto de pesquisa Estudos em Linguística, Tecnologia e suas Interfaces, o qual contribui para a reflexão sobre como a tecnologia se relaciona com as línguas. Nesse sentido, a iniciação científica concluída teve como objetivo selecionar contextos em que aspectos do conhecimento linguístico, na sua relação com a produção de sentidos, fossem coletados e/ou analisados por meio de ferramentas tecnolinguísticas, tomando como base textos de estudantes no banco *Digitus*, projeto que visa coletar dados de textos escritos com o intuito de analisar características de escrita de jovens da cidade de Curitiba e região. Particularmente, esta iniciação científica se comprometeu a realizar três importantes tarefas para a execução do projeto de pesquisa, sendo elas: Contribuir para a continuidade do desenvolvimento do banco de textos *Digitus* (UTFPR/DALIC); Recortar contextos linguísticos com função semântica específica nos textos, sendo ela a ocorrência dos verbos impessoais no sentido de existência; Analisar esses contextos de produção de sentido por meio de ferramentas digitais. Como resultados principais, citamos a padronização dos documentos no projeto *Digitus* e a análise a respeito da baixa produtividade de verbos existenciais nos textos analisados.

**PALAVRAS-CHAVE:** Linguística. Tecnologia. Variação na escrita. Verbos existenciais.

#### ABSTRACT

This paper aims to present the main points regarding to the undergraduate research project which is part of the scientific research project called Studies in Linguistics, Technology and its Interfaces (UTFPR). This research project has the goal of bringing contributions regarding the reflection of how technology is related to languages. Thus, this undergraduate project aims to select contexts in which aspects of linguistic knowledge, in their relationship with the production of meanings, could be collected and/or analyzed by means of technolinguistic tools, based on students' writing essays in the *Digitus* database – a project that aims to collect data from written texts by students to analyze writing characteristics on them. More specifically, this research has undertaken to perform three important tasks for the execution of the research project, namely: to contribute to the further development of the *Digitus* text bank (UTFPR / DALIC); to specify linguistic contexts with typical semantic functions in these texts, being the occurrence of impersonal verbs in the sense of existence; to analyze these contexts of producing meaning for digital tools. The main results are the standardization of files on *Digitus* project and the analysis about the low occurrence of existential verbs in the analyzed texts.

**KEYWORDS:** Linguistics. Technology. Writing variation. Existential verbs.

Camila Eduarda Perez  
Kozarewicz  
[c.eduardaperez@hotmail.com](mailto:c.eduardaperez@hotmail.com)  
Universidade Tecnológica Federal  
do Paraná, Curitiba, Paraná, Brasil

Roberlei Alves Bertucci  
[bertucci@utfpr.edu.br](mailto:bertucci@utfpr.edu.br)  
Universidade Tecnológica Federal  
do Paraná, Curitiba, Paraná, Brasil

**Recebido:** 04 set. 2020.

**Aprovado:** 01 out. 2020.

**Direito autoral:** Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



## INTRODUÇÃO

Este artigo apresenta os pontos principais do percurso relacionado à iniciação científica “Análises em *corpora* linguística por meio de ferramentas digitais”. Este projeto é parte do projeto de pesquisa “Estudos em Linguística, Tecnologia e suas interfaces”. O principal objetivo do projeto é refletir sobre o modo como Linguística, Tecnologia e diferentes interfaces afetam a humanidade, em particular no contexto de escrita, já que um dos projetos em andamento é a formação de um banco de dados com textos de estudantes para análises linguísticas, em especial por meio de ferramentas tecnológicas.

Nesta iniciação, trabalhou com um objetivo principal: utilizar os meios tecnológicos, em especial o espaço em nuvem, para a padronização dos documentos, com vias de torná-los materiais mais acessíveis e claros para os pesquisadores. Assim, a fim de verificar se a padronização estava adequada, optou-se por uma análise breve a respeito da variação escrita de verbos existenciais nos textos dos alunos. O resultado foi um baixo número de verbos, mas a facilidade com que se procurou o fenômeno mostrou que o objetivo anterior fora bem cumprido.

Neste artigo, apresentamos a metodologia que utilizamos para o trabalho, os resultados e discussões, e finalizamos com a conclusão.

## METODOLOGIA

A divisão do Projeto de Iniciação Científica ocorreu em três etapas.

**Fase 1. Organização do bando.** Foi necessário nessa primeira fase a padronização e alimentação do banco de dados *Digitus*, sendo destinados os primeiros meses, de agosto a dezembro de 2019, para finalizar a proposta. Essa fase foi necessária para que houvesse textos suficientes para pesquisas posteriores, por esse motivo, deu-se prioridade aos textos dos alunos do Cursinho Solidário de 2018. Depois da familiarização com o banco de dados e os textos, passamos a organizar os textos já prontos, separando-os nas pastas e conferindo a repetição ou falta de arquivos. Os arquivos ficaram em duplas (PDF e Word), sendo o primeiro um documento (PDF) com o texto original do aluno, contendo a correção e nota já escritas pelo corretor. O outro documento (Word), sem correções, apenas com o texto correspondente ao original, facilita o trabalho de pesquisa no interior do documento, inclusive quando se trabalha com analisadores digitais. No mês seguinte, foram escritos todos os textos em Word faltantes, bem como foram retirados os nomes dos autores que estavam presentes nos PDFs. Isso é importante para se garantir o anonimato, conforme rege o Termo de Consentimento Livre e Esclarecido que os participantes assinam. Um ponto importante para se facilitar a construção do bando de dados foi a organização dos textos e das propostas em uma planilha. Percebemos que seria melhor renomear as pastas e arquivos, tanto dos textos quanto das propostas, para que a tabela de pesquisa ficasse mais intuitiva e acessível (e que pudesse ser passível de uma programação, no futuro). Por fim, em dezembro, passei a organizar alguns textos dos alunos do curso de Letras Português da Universidade Tecnológica Federal do Paraná, ingressados no primeiro semestre de 2018. As provas foram feitas e

enviadas já em documento Word, sendo assim, um mês foi o suficiente para retirar todas as identificações, renomear pastas e arquivos, organizar e fazer a planilha.

**Fase 2. Pesquisa teórica e ferramenta digital.** A partir de Janeiro/2020, estudamos algumas partes do livro "Por que a escola não ensina gramática assim?" (Parábola, 2014), em especial o capítulo 6, "Verbos impessoais: variação no uso de haver, ter e fazer", escrito por Freitas e Maria Alice Fernandes de Souza, assim como outras pesquisas sobre o assunto (VITÓRIO, 2015; 2018). O livro seria a base para a pesquisa que iniciáramos na segunda metade de março de 2020, tentando colocar em prática o uso do banco por meio de uma pesquisa linguística. Esse capítulo em questão foi decisivo para o entendimento da pesquisa e contribuiu para a busca de como os textos da base se comportavam em relação aos verbos existenciais. Após as leituras, no início de março, foi necessário marcamos uma reunião por vídeo chamada para discutirmos com precisão o tema da pesquisa, pois a instituição teve que ser fechada devido à situação de pandemia. Nessa reunião chegamos à conclusão de que o objeto de pesquisa seria a presença dos verbos impessoais *ter* e *haver* no sentido de existência, presentes nos textos escritos pelos alunos do Cursinho Solidário encontrados no banco de dados *Digitus*. Neste momento, também utilizamos a plataforma *Linguakit* para separar os dados de que necessitávamos nos documentos escolhidos.

**Fase 3. Pesquisa no banco *Digitus*.** Nessa última etapa, passamos a utilizar o banco *Digitus* para a pesquisa. A primeira tarefa foi selecionar as propostas a serem analisadas de acordo com os critérios discutidos entre orientada e orientador. Foram selecionadas, inicialmente, duas delas. Decidimos utilizar, primeiramente, dez textos, cinco de cada proposta, para análise. Percebemos, no entanto, que as ocorrências eram muito baixas e resolvemos separar mais textos, sendo nesse momento dez de cada tema. Ainda assim, poucas ocorrências apareceram. Então, decidimos selecionar todos os textos produzidos das duas propostas e acrescentar mais propostas que possuísem grande quantidade de textos produzidos. Entretanto, devido à pandemia citada anteriormente, houve um corte significativo na minha renda familiar, causando assim o cancelamento dos meus dados de internet. Fiquei impossibilitada de realizar as devidas pesquisas e me concentrei nos seguintes meses em iniciar este relatório. Só consegui acesso à internet novamente no início do mês de junho, quando conversei com meu orientador e resolvemos manter a pesquisa apesar do inconveniente enfrentado, possibilitando assim a finalização dela. Finalmente, em junho, pude acrescentar mais textos para verificar as ocorrências e fechar a pesquisa.

## RESULTADOS E DISCUSSÕES

Mesmo após compilar 49 textos, os dados foram baixos. Pode-se observar no Quadro 1 que foram utilizadas 4 (quatro) diferentes propostas, possuindo entre 9 a 15 textos cada, e, apesar disso, encontrou-se apenas 6 resultados com a ocorrência dos verbos impessoais *ter* e *haver* no sentido de existir. Dentre os resultados, um fato a se observar é uma única ocorrência do verbo *ter* considerada um erro na gramática escrita normativa.

Quadro 1 - Número de textos, análises e ocorrências

Propostas		Número de análises	Ocorrências significativas	
			TER	HAYER
1	Tirinha	13	-	1
2	Resumo	15	-	2
5	Tabela	12	1	-
6	Carta	9	-	2
Total		49	1	5

Fonte - A autoria própria (2020).

A seguir apresentamos os trechos escritos pelos alunos com as ocorrências apresentadas no Quadro 1, todos com os verbos impessoais *ter* e *haver*, no sentido de existência, destacados:

1. "...é demonstrado esse pensamento através disso e mostram que ainda **há** pessoas com coração de carne.";
2. "Tal sentido é trazido ao texto pois **há** dificuldades de amadurecimento intelectual e de percepção do próximo entre os jovens.";
3. "... Thomaz Wood Jr. defende que **há** uma diferença entre alunos e estudantes...";
4. "...35% deles foram em bares, mesmo **tendo** uma lei que proíbe a venda e consumo nesses lugares e por menores";
5. "...pois não **houve** questionamento algum sobre o problema do 'incentivo indireto' do consumo do álcool...";
6. "Mas, o que me incomoda nessa reportagem, é o fato de não **haver** nenhuma menção ao consumo de álcool entre adolescentes...".

Além de uma única ocorrência equivocada (de acordo com a gramática prescritiva) do verbo *ter* no sentido de existir (4), já comentada, destacamos o fato de que todos os usos de *haver* seguem a mesma prescrição gramatical, já que lá se lê que "de acordo com o português padrão, o emprego de *haver* como verbo impessoal, no sentido de existir, ficará sempre na terceira pessoa do singular, em qualquer que seja o tempo" (FREITAS; SOUZA, 2014, p.166). Assim, observamos que cerca de 85% das ocorrências com verbos impessoais localizadas estão de acordo com a gramática normativa.

Quais as possíveis causas para o número de acertos ser tão alto? Podemos destacar primeiramente que, além de ser um gênero escrito, que naturalmente exige uma maior rigidez na produção de sentenças, os textos foram produzidos no contexto de pré-vestibular, ou seja, há, por parte dos redatores, um maior cuidado nas questões gramaticais. Outra possibilidade é que, com um tempo prolongado para a produção do texto, junto com a possibilidade de recursos digitais para correção automática ou para pesquisas, possa haver essa interferência nas escolhas. Por fim e com grande ênfase, não podemos esquecer que, justamente por serem estudantes concorrendo à uma vaga nas universidades, os alunos são



expostos há um maior fluxo de leituras que, se antes não submetidos, podem agregar, e muito, no repertório linguístico.

Destarte, os dados (ainda que poucos), mostram que os textos analisados em sua grande maioria respeitam o uso normativo dos verbos impessoais, em especial pela ausência de ter no lugar de haver, fenômeno observado mais amplamente no português brasileiro, em especial na fala.

## CONCLUSÕES

O estudo dos verbos impessoais no sentido de existência já estão sendo alvo de muitas pesquisas pois, apesar da normatização padrão do português, a língua é variável, está em constante transformação, e entre elas está esse grupo. Diversas pesquisas apontam que, principalmente no português brasileiro, a fala apresenta uma grande utilização do verbo ter para substituir o verbo existir, já a escrita percebe-se uma maior utilização de haver mantendo a norma padrão da língua. Entretanto, essa última percebe-se uma variação maior, dependendo da região do Brasil e da escolaridade dos cidadãos. Para a pesquisa proposta nessa iniciação científica, analisamos dados de alunos de classe baixa, participantes do curso pré-vestibular Formação Solidária, alunos de Curitiba e região que possuem como objetivo a ingresso no ensino superior.

Para recolhimento de dados utilizamos a plataforma Digitus em construção com a supervisão do doutor e professor da Universidade Tecnológica Federal do Paraná, Roberlei Alves Bertucci. Nos dados recolhidos, notamos pouca presença da utilização dos verbos impessoais no sentido de existência e, mesmo com poucos dados para análise, encontramos um total de seis dados dentre quarenta e nove textos, sendo apenas um com a utilização do verbo ter.

Temos que destacar que consideramos erro apenas se observarmos pela perspectiva da gramática normativa, pois não podemos esquecer que a língua portuguesa, assim como todas as línguas vivas, estão em constante mudanças e ressignificações, e percebemos que se a língua escrita apresenta essas variações é porque na língua falada essa variação já está presente no cotidiano.

Finalmente, concluímos que o presente projeto contribuiu muito para o desenvolvimento do projeto de pesquisa a ele relacionado, tanto pela organização do banco de textos, parte fundamental para que outras pesquisas sejam realizadas, quanto pela análise de textos em contextos de tecnologia, o que reforça o caráter amplo e necessário do estudo da linguagem em contextos tecnológicos.

## AGRADECIMENTOS

Em primeiro lugar, agradecemos à Fundação Araucária pelo apoio financeiro para o desenvolvimento desta pesquisa. Depois, aos estudantes do Cursinho Solidário e do curso de Letras que cederam seus textos para a composição do banco *Digitus*. E aos corretores, pela aplicação na ajuda que oferecem aos estudantes.

## REFERÊNCIAS

FREITAS, V. A. L.; SOUZA, M. A. F. Verbos impessoais: variação no uso de haver, ter e fazer. *In*: BORTONI-RICARDO, S. M.; SOUZA, R. M.; FREITAS, V. A. L.; MACHADO, V. R. (Orgs.). **Por que a escola não ensina gramática assim?** São Paulo: Parábola Editorial, 2014. p.161-179.

VITÓRIO, E. G. S. L. A. A Competição Ter, Haver e Existir na Escrita Escolar.

**SIGNUM**: Estudos Linguísticos, Londrina, n. 18/1, p. 365-391, jun. 2015.

Disponível em:

<<http://www.uel.br/revistas/uel/index.php/signum/article/view/20334>> Acesso em: 03 set. 2020.

VITÓRIO, E. G. S. L. A. As construções existenciais e o problema da avaliação linguística / **Domínios de Lingu@gem**, Uberlândia, vol. 12, n. 3, p. 1825-1858, jul./set. 2018. Disponível em:

<<http://www.seer.ufu.br/index.php/dominiosdelinguagem/article/view/40617>> Acesso em 03 set. 2020.