

RELATO DE EXPERIÊNCIA

CONSTITUIÇÃO DE **CORPUS** DE **GÊNEROS TEXTUAIS** **JORNALÍSTICOS** PARA INVESTIGAÇÃO DE VARIEDADE(S) CULTA(S)

Monique Débora Alves de Oliveira LIMA  
Universidade Federal do Rio de Janeiro (UFRJ)

RESUMO

Neste relato de experiência – parte de uma pesquisa de Doutorado em andamento –, descreve-se a etapa referente à composição de uma amostra de escrita culta, cuja constituição atende a um dos objetivos da investigação e é formada por gêneros textuais publicados no jornal O Globo (Rio de Janeiro). O estudo toma como base os pressupostos da Teoria da Variação e Mudança (WEINREICH; LABOV; HERZOG, 1968; LABOV, 1972) para composição de amostras que possibilitem investigar o comportamento das regras variáveis e seus condicionamentos. Consoante Vieira (2019), cuja proposta assume a conceituação de norma culta e norma-padrão (FARACO, 2008), buscou-se compor uma amostra de gêneros jornalísticos para estudo de fenômenos linguísticos na perspectiva do *continuum* fala-escrita (BORTONI-RICARDO, 2004; 2005; MARCUSCHI, 2001; 2008). O *corpus* organizado foi composto por sete gêneros textuais (artigo de opinião, carta de leitor, crônica, editorial, entrevista, notícia e tirinha) – cada um com cerca de 20.000 palavras – e editado de modo que fosse manipulável por computador, conforme preconizado pela Linguística de Corpus (SANCHES, 1995 *apud* SARDINHA, 2000).

 OPEN ACCESS

EDITORES

– Miguel Oliveira Jr. (UFAL)
– René Almeida (UFS)

AVALIADORES

– Rosane Berlinck (UNESP)
– Lorena Borges (UFAL)

DATAS

– Recebido: 20/08/2021
– Aceito: 19/09/2021
– Publicado: 16/11/2021

COMO CITAR

LIMA, M. D. A. O. (2021). Constituição de corpus de gêneros textuais jornalísticos para investigação de variedade(s) culta(s). *Cadernos de Linguística*, v. 2, n. 4, e564.

ABSTRACT

In this experience report – part of a PhD research in progress –, the stage related to the composition of a sample of cultured writing is described, whose constitution serves to one of the objectives of the investigation and is formed by textual genres published in the newspaper O Globo (Rio de Janeiro). The study is based on the assumptions of the Theory of Linguistic Variation and Change (WEINREICH; LABOV; HERZOG, 1968; LABOV, 1972) to compose samples that make it possible to investigate the behavior of variable rules and their conditioning. According to Vieira (2019), whose proposal assumes the concept of a cultured norm and a standard norm (FARACO, 2008), we sought to compose a sample of journalistic genres to study linguistic phenomena from the perspective of the speech-writing continuum. (BORTONI-RICARDO, 2004; 2005; MARCUSCHI, 2001; 2008). The organized corpus was composed of seven textual genres (opinion article, reader's letter, chronicle, editorial, interview, news and comic strip) – each with about 20,000 words – and edited so that it could be manipulated by computer, recommended accordingly by Corpus Linguistics (SANCHES, 1995 *apud* SARDINHA, 2000).

PALAVRAS-CHAVE

*Corpus de gêneros textuais jornalísticos; continuum
fala-escrita; sociolinguística.*

KEYWORDS

Corpus of journalistic textual genres; speech-writing
continuum; sociolinguistics.

INTRODUÇÃO

A constituição de *corpora* linguísticos – de modalidade falada ou escrita – há muito tempo é objeto de estudo e aprimoramento por parte dos linguistas. As amostras de dados linguísticos são fontes privilegiadas para a descrição da língua, seja em análises sincrônicas ou diacrônicas. Uma vez que o *locus* de observação da variação linguística é a própria língua, nas modalidades oral e escrita, justifica-se a importância da constituição de *corpora* linguísticos.

O trabalho de coleta/organização de amostras nem sempre é descrito por completo nos relatos das investigações realizadas. Freitag, Martins e Tavares (2012, p. 918) destacam que o texto reservado à descrição das amostras constituídas nos trabalhos (teses e dissertações), em geral ocupa um espaço exíguo, que não reflete “o real esforço e tempo dispendidos no processo de constituição de um banco de dados, desde a prospecção e seleção de informantes até a transcrição, armazenamento e disponibilização”.

Seguindo essa linha de pensamento dos autores, este texto tem por objetivo principal fazer um relato da experiência de constituição de uma amostra de escrita culta do Rio de Janeiro – formada por gêneros textuais publicados no Jornal O Globo –, a qual serve de base para uma pesquisa de Doutorado em andamento.

Para fins de organização da leitura, cumpre anunciar como se estrutura este texto. A esta introdução, seguem os pressupostos teóricos (Seção 2) para constituição de amostras; a justificativa para composição de bancos de dados representativos da escrita culta (Seção 3); a descrição do *corpus* de gêneros textuais do Jornal O Globo (Seção 4); a descrição dos gêneros escolhidos para composição do *corpus*; e, por fim, algumas considerações finais (Seção 5).

1. PRESSUPOSTOS PARA CONSTITUIÇÃO DE AMOSTRAS

No que tange aos estudos de variação e mudança, a Sociolinguística Laboviana (WEINREICH; LABOV; HERZOG, 1968; LABOV, 1972) contribuiu significativamente para a criação de um modelo teórico-metodológico, a partir do qual foram realizadas investigações sobre variedades linguísticas. Esse modelo, também conhecido como Sociolinguística Quantitativa, prevê análises estatísticas de dados linguísticos, nas quais se observa a atuação de condicionadores linguísticos ou extralinguísticos, que podem (des)favorecer o uso de uma ou outra forma. Nesse sentido, a pesquisa sociolinguística,

nos moldes labovianos, depende necessariamente de uma amostra representativa da fala ou da escrita, na qual são coletadas as ocorrências dos fenômenos variáveis.

Collischonn e Monaretto (2012), ao traçar características gerais do VARSUL – um *corpus* de variedades do Sul brasileiro –, afirmam que os dados quantitativos, obtidos na análise de amostras específicas,

lançam luz sobre muitas questões que não podem ser respondidas simplesmente pela intuição e permitem comparar estruturas hipotéticas com estruturas atestadas. Ou seja, estudos de *corpora* sociolinguísticos podem contribuir para uma sustentação mais sólida na realidade de análises de caráter mais teórico. (COLLISCHONN; MONARETTO, 2012, p. 838)

No que tange à modalidade falada do Português Brasileiro, os *corpora* constituídos desde a década 70 do século passado buscavam atender a um padrão, na medida do possível, que é o de preenchimento do envelope da variação, a partir de metodologia própria da Sociolinguística, para o estudo de fenômenos linguísticos diversos. A fim de constituir uma amostra que seja representativa da sociedade ou do grupo estudado e não seja demasiadamente grande, estratifica-se a amostra, dividindo-a em células, compostas por informantes com características sociais semelhantes.

Aplicando essa metodologia, uma pesquisa clássica – de primeira onda, segundo a classificação proposta por Eckert (2012) – busca preencher tais células a partir de fatores sociais como classe social, idade (faixa etária), grau de escolaridade e sexo/gênero (cf. SILVA, p. 121). Desse modo, determina-se um número específico de informantes com as mesmas características, como por exemplo, 5 homens de classe alta/média/baixa e 5 mulheres de classe alta/média/baixa. Nesse exemplo, foram combinadas as variáveis sociais sexo/gênero e classe social. Na amostra estratificada, a inserção de outras variáveis (como grau de escolaridade, por exemplo) significa também aumento no número de informantes, a fim de que todas as células sejam preenchidas.

A variável dependente, no modelo laboviano, é a regra variável linguística que está sendo analisada – por exemplo, a representação do acusativo anafórico de terceira pessoa. Essa variável de análise é cruzada com outras variáveis – de ordem linguística ou extralinguística –, a fim de verificar a correlação desses fatores com o emprego de determinada variante. Dentre os fatores extralinguísticos, há as chamadas variáveis clássicas – mencionadas anteriormente –, usadas desde os primeiros estudos, como sexo/gênero, faixa etária e grau de escolaridade.

Uma vez determinado o fenômeno linguístico a ser estudado, o *corpus* constituído para análise da atuação de sexo/gênero, faixa etária e grau de escolaridade na escolha por uma ou outra forma desse fenômeno precisa necessariamente ser composto considerando-se essas variáveis. O conceito de “envelope da variação” é aplicado desta forma na constituição da amostra: seleciona-se o mesmo número de informantes para preencher cada uma das células das variáveis pré-estabelecidas como hipóteses de pes-

quisa. Essa metodologia proveniente da Sociolinguística Variacionista orientou muitas das amostras já constituídas.

Consoante Silva (2015), os bancos de dados linguísticos são fonte privilegiada para a descrição da língua. A fim de contribuir com a divulgação dessas fontes de estudo, a pesquisadora fez um levantamento dos *corpora* linguísticos de variedades do Português e reuniu informações sobre importantes amostras representativas do Português Brasileiro. Nesse levantamento, Silva (2015) cita as seguintes amostras nacionais, de fala ou de escrita: Projeto NURC, *Corpus* Censo, Projeto APERJ, Projeto Discurso e Gramática, PHPB-Rio, Projeto ALiB, Projeto VARPORT, Projeto AVAL-RJ, Projeto do *Corpus* do Português clássico e moderno, Projeto *Corpus* Brasileiro, *Corpus* histórico do Português TYCHO BRAHE. Dessas amostras, a pioneira foi a do Projeto NURC, cuja composição se iniciou na década de 70 e abrangeu cinco capitais brasileiras.

Há ainda outros trabalhos de constituição de banco de dados não elencados pela autora, como o VALPB, o VARSUL e o Falares Sergipanos. Outra iniciativa, de cooperação internacional, é o “Projeto 21: Estudo comparativo de variedades africanas, brasileiras e europeias do Português (COMPARAPORT)”, coordenado por Sílvia Rodrigues Vieira e Sílvia Figueiredo Brandão, no âmbito da ALFAL, que contempla quaisquer fenômenos morfossintáticos ou fonético-fonológicos em variedades do Português.

Salomão (2011), ao fazer um panorama dos grupos de pesquisa brasileiros que estudam a língua por meio dos pressupostos da Teoria da Variação e Mudança, elenca 48 projetos, muitos dos quais são responsáveis por composição de amostras. Diante desses dados levantados, tem-se um retrato do esforço da pesquisa brasileira, no sentido de prover tanto banco de dados quanto investigações acerca do Português e suas variedades.

Diante do exposto, percebe-se que já há uma tradição consolidada no que diz respeito à constituição de amostras por parte de pesquisadores brasileiros. A composição do *corpus* ora descrito fundamenta-se no que já foi estabelecido. Além disso, toma-se ainda como pressuposto a definição de *corpus* advinda da Linguística de Corpus, área que também provê aparato metodológico para composição de amostras. Os estudiosos dessa área entendem um *corpus* linguístico como

um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade de uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise. (SANCHES, 1995, p. 8-9, *apud* SARDINHA, 2000, p. 338)

Como se pode observar, é necessário estabelecer/seguir critérios específicos ao se organizar uma amostra para estudo de dados linguísticos. Pode-se dizer que um dos grandes desafios desta pesquisa de Doutorado é o de composição de um *corpus* que

atendesse ao princípio da Linguística de Corpus, quanto à incumbência de torná-lo processável por computador. De modo geral, entende-se que a constituição de *corpora* é uma tarefa árdua, mas necessária para o mapeamento dos fenômenos linguísticos.

2. JUSTIFICATIVA PARA COMPOSIÇÃO DE BANCO DE DADOS DE ESCRITA CULTA

Nesta investigação, acredita-se na urgência, a partir de Freitag, Martins e Tavares (2012) e outros pesquisadores, de constituição de *corpora* nacionais homogêneos em termos de parâmetros de constituição. Os autores defendem a padronização dos procedimentos metodológicos para que sejam realizados estudos contrastivos entre as variedades, possibilitando uma descrição mais acurada do Português Brasileiro.

Como evidenciado na última seção, é inegável o trabalho árduo e produtivo de pesquisadores brasileiros na constituição de *corpora* do Português em suas variedades, haja vista a vastidão de amostras compostas, brevemente citadas. No que tange à modalidade escrita, entretanto, segundo Lima, Vieira (2019, p. 10), “a sistematização do comportamento das variedades em função do *continuum* fala-escrita ainda está por ser feita”.

Consoante Faraco (2008, p. 73), norma culta representa um “conjunto de fenômenos linguísticos que ocorrem habitualmente no uso dos falantes letrados em situações mais monitoradas de fala e escrita”. Essa conceituação é confundida, em diversos âmbitos, com norma-padrão, “uma codificação relativamente abstrata, uma baliza extraída do uso real para servir de referência, em sociedades marcadas por acentuada dialeção, a projetos políticos de uniformização linguística” (FARACO, 2008, p. 75).

Partindo da conceituação de Faraco (2008), Vieira (2019) assume que a delimitação da norma-padrão – para fins de atualização, uma vez que sua construção se deu no século XIX (cf. PAGOTTO, 1998; FARACO, 2008) – deve ocorrer a partir do conhecimento detalhado das variedades cultas. A pesquisadora propõe que a descrição das variedades cultas de escrita se dê a partir da observação do comportamento de regras variáveis em textos de diferentes gêneros textuais, produzidos por falantes com alto grau de escolarização. Nos textos jornalísticos e acadêmicos, portanto, seriam encontradas essas expressões máximas de norma(s) culta(s).

A proposta de Vieira toma como base os *continua* de variação linguística de Bortoni-Ricardo (2004; 2005), que percebe a variação linguística em três linhas imaginárias, nas quais as variantes se distribuem de um polo a outro (rural-urbano, oralidade-letramento e monitoração estilística). Vieira ainda se fundamenta no *continuum* fala-escrita proposto por Marcuschi (2001; 2008), em que os usos linguísticos são colocados numa organização

escalar, em função da caracterização dos gêneros textuais, com base nos princípios de meio e concepção.

Vieira (2019, p. 104) postula uma hipótese central de que “consoante o maior grau de letramento e/ou formalidade atribuído(s) ao gênero em questão, maior seria o uso da variante considerada de prestígio, identificada muitas das vezes como variante padrão”.

O *corpus* de gêneros textuais para análise desta pesquisa de Doutorado em andamento inspira-se em uma experiência de constituição de outra amostra, realizada em 2018, no âmbito de uma disciplina de Pós-Graduação de Letras Vernáculas (UFRJ), ministrada pela professora Silvia Rodrigues Vieira. Em Vieira, Lima (2019, p. 6), livro organizado com as investigações realizadas no curso em questão, defende-se a ideia de que “embora haja vasta literatura sobre os usos relativos a fenômenos diversos na fala brasileira, percebe-se que o mapeamento de dados em meios escritos segundo estilos variados e contemplando diversos tópicos gramaticais ainda está por ser construído”. O trabalho desenvolvido na disciplina buscou contribuir com esse mapeamento.

De maneira geral, foram tomadas as seguintes providências no âmbito da disciplina:

(i) construção de um *corpus* com diversidade de gêneros textuais principalmente dos domínios jornalístico e acadêmico, tomados como expressões de usos cultos; (ii) levantamento de dados de oito fenômenos variáveis; e (iii) sistematização dos resultados buscando uma distribuição dos dados em função do referido *continuum*. (LIMA; VIEIRA, 2019, p. 8-9)

A amostra em questão foi constituída a partir da contribuição coletiva dos discentes do curso e composta por dez gêneros textuais: entrevistas sociolinguísticas (provenientes do *corpus* Concordância); tirinhas; anúncios; entrevistas impressas (em jornais ou revistas); cartas de leitor; crônicas jornalísticas; notícias; editoriais; teses/dissertações; artigos em revistas científicas. Os gêneros acadêmicos foram retirados de periódicos e páginas de programas de Pós-graduação da área de Comunicação Social

Uma vez organizado o *corpus*, os discentes, divididos em grupos, procederam à coleta de dados de regras variáveis de oito fenômenos específicos, quatro dos quais foram publicados na obra de 2019. Os resultados foram discutidos com o objetivo de se elaborar um *continuum* de fala-escrita que hipoteticamente se aliará ao de monitoração estilística.

Cumpra esclarecer que, embora o *corpus* não tenha seguido rigorosas condições de constituição – em função do trabalho coletivo, em curto tempo –, as análises desenvolvidas com os dados coletados revelaram “notável sistematicidade” (cf. LIMA; VIEIRA, 2019, p. 16) na distribuição das formas analisadas. A partir dessa experiência, julgou-se relevante constituir uma nova amostra – a que se descreve na próxima seção – , em que se adotassem critérios mais rígidos de organização.

A constituição da amostra de gêneros textuais proposta nesta pesquisa de Doutorado integra o projeto “Do *continuum* fala-escrita para a norma-padrão: limites e

possibilidades”, coordenado pela professora Sílvia Rodrigues Vieira (UFRJ), com a participação, na condição de colaboradores, desta pesquisadora, dos professores Flávio Brandão Silva (Universidade Estadual de Londrina – UEL), Juliana Bertucci Barbosa (Universidade Federal do Triângulo Mineiro – UFTM) e do professor Carlos Alberto Faraco (Universidade Federal do Paraná – UFPR), no papel de consultor.

O projeto foi recentemente contemplado em um edital da FAPERJ (PROGRAMA Cientista do Nosso Estado – 2020) e tem por objetivo constituir amostras de textos jornalísticos e acadêmicos, de quatro capitais distintas, a saber: Recife/PE, Rio de Janeiro/RJ, Salvador/BA e São Paulo/SP. No âmbito deste projeto nacional, o *corpus* descrito adiante representa a amostra de textos jornalísticos do Rio de Janeiro. Seus parâmetros de constituição servirão de modelo para a composição das demais amostras estabelecidas para figurar a experiência-piloto do projeto: sete gêneros, com cerca de 20.000 palavras cada.

Quanto à constituição da amostra de gêneros acadêmicos, parte-se da amostra organizada por Ana Carolina Alves Caetano. No trabalho de Mestrado – ainda em andamento –, a pesquisadora coletou 42 artigos científicos, publicados em revistas de *qualis* A1 e A2, e 42 teses de doutorado – textos representativos do Rio de Janeiro. Seguindo o critério estabelecido para o tamanho da amostra de gêneros jornalísticos, a fim de compor a amostra acadêmica, selecionaram-se 500 palavras de cada artigo e de cada tese, para que se chegasse ao número de 20.000 palavras por gênero, considerando as três áreas escolhidas: Linguística, Comunicação Social e Economia. Para o *corpus* do projeto nacional, os textos serão apenas da área de Comunicação Social, para ficar na mesma área de estudo dos jornalistas.

Na próxima seção, descreve-se o *corpus* de gêneros textuais jornalísticos do Rio de Janeiro.

3. O CORPUS DE ESCRITA CULTA DO JORNAL O GLOBO: METODOLOGIA DE CONSTITUIÇÃO

A constituição desse *corpus* atende a um dos objetivos da pesquisa de Doutorado em andamento, que é o de propiciar um material no qual se possa observar a heterogeneidade da chamada norma culta. Acredita-se, a partir de Vieira (2019), que, com base em um *continuum* de gêneros textuais, é possível aferir o comportamento altamente variável das variedades cultas e elaborar orientações normativas para o contexto escolar. Com base na amostra constituída em 2018, o *corpus* organizado para análise desta pesquisa privilegiou o domínio jornalístico e foi composto por sete gêneros textuais,

quais sejam: (1) *artigo de opinião*, (2) *carta de leitor*, (3) *crônica*, (4) *editorial*, (5) *entrevista*, (6) *notícia* e (7) *tirinha*.

Os gêneros selecionados para constituição da amostra pertencem à modalidade escrita do Português Brasileiro e foram publicados em versão impressa do jornal. A escolha dos textos foi baseada na proposta de Marcuschi (2001; 2008) acerca da concepção envolta em cada gênero. O autor propõe que a caracterização dos gêneros textuais seja determinada pelos princípios de meio (*sonoro versus gráfico*) e de concepção (*oral versus escrita*). Na proposta de Marcuschi (2001; 2008), há gêneros que são mais prototípicos da fala (*meio sonoro e concepção oral*), aqueles mais representativos da escrita (*meio escrito e concepção escrita*) e os híbridos (*meio escrito e concepção oral e vice-versa*).

No *corpus* organizado, as entrevistas e as tirinhas, por exemplo, poderiam ser alocadas supostamente entre os gêneros mais próximos da concepção de oralidade, uma vez que, apesar de pertencerem ao meio escrito, representariam diálogos. Já as crônicas, as cartas de leitor e as notícias seriam gêneros mais intermediários quanto à concepção de oralidade/letramento em função de diversos fatores, como grau de autoria (alguns são textos assinados), uso de discurso direto (por vezes presente), temáticas diversas, dentre outros fatores. Por fim, o *corpus* conta, ainda, com textos pertencentes a gêneros que, no domínio jornalístico, poderiam ser considerados mais prototípicos da modalidade escrita, tanto em *meio* quanto em *concepção*: os artigos e os editoriais.

Na presente investigação, decidiu-se por trabalhar com os textos publicados no Jornal O Globo. Tal escolha se deve ao fato de que esse é o jornal voltado para os leitores mais escolarizados, e por consequência, mais elitizados. Esse jornal, produzido pelo Grupo Globo, da Fundação Roberto Marinho, foi criado em 1925 e possui um acervo digital que abrange textos publicados desde então. A Figura 1, a seguir, mostra parte da capa da primeira edição (29 de julho de 1925). Essa foi uma edição vespertina¹.



Figura 1. Capa da primeira edição do jornal O Globo, publicada em 29/07/1925. Fonte: Jornal O Globo (29/07/1925).

1 Até 1962, as edições foram vespertinas, a partir de quando passaram a ser matutinas.

De acordo com os dados do Instituto Verificador de Comunicação (IVC), referentes a maio/2018 e divulgados no site Infoglobo, o jornal conta com 1.194.000 leitores. Em relação à classe social, mais de 60% dos leitores pertencem às classes A (15%) e B (50%), dos indivíduos que recebem acima de 10 salários-mínimos. No que tange ao fator escolaridade, 39% dos leitores completaram o nível superior e 22%, o nível médio.

Em termos comparativos com os demais jornais veiculados pelo Grupo Globo, o jornal Extra, intermediário, possui 1.300.000 leitores (mais do que os do O Globo), dos quais 17% possuem nível superior. Já o jornal Expresso – tabloide voltado para “os leitores das classes C/D”, segundo a descrição do site Infoglobo – possui 133.000 leitores, dos quais apenas 6% possuem nível superior completo. Note-se que o Grupo Globo não possui um jornal destinado àqueles que pertencem à classe E (recebem até dois salários-mínimos).

Para a constituição do *corpus* de gêneros textuais jornalísticos desta pesquisa em andamento, estabeleceu-se que seriam extraídos textos do jornal O Globo. Na medida do possível, os textos são referentes ao Rio de Janeiro: as cartas de leitor, metade das entrevistas e as notícias pertencem à editoria “Rio”. Com relação às crônicas, apesar de serem publicadas fora dessa editoria – no segundo caderno –, todas são assinadas por autores do Rio. Não foi possível seguir esse padrão em relação aos editoriais – uma vez que representam “a voz” do jornal, são publicados na editoria “País” e não são, portanto, vinculados necessariamente ao Rio – aos artigos – pois ficaria um número reduzido de colonistas – e às tirinhas – pois algumas são traduzidas.

Todos os textos selecionados foram publicados na versão impressa do jornal e se encontram disponíveis no acervo online, cujo acesso é permitido apenas para assinantes. Para a constituição do *corpus*, foi estabelecido o período de agosto de 2018 a julho de 2020, exceto para as tirinhas, como se verificará adiante.

O critério adotado para delimitação do presente *corpus* foi o de número de palavras por texto selecionado, de modo a viabilizar uma amostra equilibrada quanto ao tamanho. A adoção desse critério se deu a partir do trabalho de Biazolli (2016), que constituiu uma amostra de quatro gêneros textuais (dois de modalidade oral e dois escritos), cada um com 35.000 palavras. No que diz respeito ao *corpus* dessa pesquisa, determinou-se um número menor de palavras, visto que há mais gêneros textuais na amostra.

Na primeira etapa da pesquisa, a de qualificação, o *corpus* foi constituído por gêneros que variavam entre 8 e 10 mil palavras. No entanto, análises prévias sobre o acusativo anafórico de terceira pessoa, mostraram que esse número poderia restringir a ocorrência dos dados em determinados contextos. Uma vez que o *corpus* deve servir para outras análises além do fenômeno do acusativo, decidiu-se, então, aumentar esse número estabelecido previamente, a fim de evitar enviesamento. Optou-se, assim, por coletar cerca de 20.000 palavras/gênero. Cabe esclarecer que não são exatamente 20.000

palavras, pois nenhum texto foi cortado ao se atingir esse número; no processo de seleção, foi preservada a unidade textual.

Uma vez determinados os gêneros, a próxima etapa de organização do *corpus* foi a de seleção dos textos que o comporiam, no acervo online. Objetiva-se deixá-los editados de tal modo que viabilizasse sua leitura por programas computacionais. Se os textos estão em formato .txt, por exemplo, podem ser submetidos a programas de processamento de dados. Ocorre que os textos são disponibilizados no acervo do jornal O Globo em extensão PDF – cada página do jornal corresponde a um arquivo de PDF –, mas sendo uma imagem sem caracteres reconhecidos pelo Word.

Com o propósito de contornar esse obstáculo, foi necessário submeter cada texto a um programa de reconhecimento óptico de caracteres (OCR), a fim de que o texto fosse transformado de imagem para caracteres. O programa utilizado foi o aplicativo “Office”, da Microsoft², disponível na loja de aplicativos do sistema Android. O caminho para uso é: Office > Ações > Imagem em texto. A seguir, a Figura 2 mostra o aplicativo na loja virtual e a Figura 3 ilustra a interface desse aplicativo.



Figura 2. Aplicativo “Office”.
Fonte: Play store.



Figura 3. Interface do aplicativo “Office”.
Fonte: Aplicativo Office

2 O aplicativo “Office” está disponível apenas para quem possui conta da Microsoft[®]. Optou-se por usar esse aplicativo em vez de outras opções gratuitas, com o Google Drive, uma vez que o resultado da leitura era mais acurado do que as demais ferramentas ofereciam.

A primeira etapa de edição consistiu em recortar os textos baixados do site do acervo O Globo nas dimensões do texto gênero em questão – para que o programa não lesse os demais textos publicados na mesma página – e submetê-los ao leitor de OCR. Em diversas ocasiões, foi necessário recortar o texto que se pretendia recolher em várias partes, já que as configurações de editoração do jornal – com textos em colunas – não seguem o padrão de texto contínuo do Word – programa no qual foram organizados, editados e revisados os textos. Na Figura 4, pode-se observar uma entrevista, da editoria de cultura, recortada.

ENTREVISTA
Karol Conka/ CANTORA E APRESENTADORA
Cinco anos após seu primeiro álbum, a curitibana volta com 'Ambulante', no qual passeia por vários estilos musicais sem perder a língua afiada
SILVIO ESSINGER silvio.essinger@oglobo.com.br

‘NÃO ME ACHO EMPODERADA, EU ME ACHO É PODEROSA’

Figura 4. Recorte da Entrevista 10 para submissão ao programa de OCR. Fonte: Jornal O Globo (data).

Uma vez recortados, na segunda etapa, os textos (ou suas partes) eram submetidos ao processo de extração de imagem em texto, no aplicativo Office, e o resultado dessa ação, copiado para um arquivo de Word, no qual era possível, além de editar, contar o número de palavras. Dessa maneira, organizaram-se sete arquivos principais, nos quais foram colados os textos pertencentes a cada gênero textual, até que se atingisse cerca de 20.000 palavras/gênero.

Na terceira etapa, os textos foram revisados, comparando-se com os originais, para que fossem corrigidos possíveis erros. Os mais comuns eram de ortografia e de falta de trechos. A tecnologia de OCR contribui bastante para esse processo de extração de textos, mas não é perfeita. Dois dos erros mais comuns ocorridos foram o de acentuação de algumas palavras com acento agudo em vez de grave e de troca de “s” por “c” em algumas palavras. Após a revisão, na quarta e última etapa, criou-se um arquivo de Word para cada texto e seu correspondente em .txt, a fim de tornar o *corpus* legível/manipulável por computador.

Após as ações de coleta/edição/formatação, elaborou-se uma planilha do Excel, na qual foram inseridos dados específicos de cada um dos textos selecionados: número do arquivo original, data, título, manchete, autor(es), temática, a depender dos gêneros.

Dessa maneira, chegou-se à seguinte distribuição do quantitativo de textos coletados por gêneros textuais, sistematizado no Quadro 1. Observe-se que esse valor varia em função dos gêneros. A título de exemplo, em termos de volume textual, no jornal O Globo, 36 artigos correspondem a 261 cartas de leitor ou a 45 editoriais.

Composição do *corpus* de gêneros textuais do jornal O Globo

	GÊNERO TEXTUAL	TOTAL
1	<i>Artigo</i>	36
2	<i>Carta de leitor</i>	261
3	<i>Crônica</i>	36
4	<i>Editorial</i>	45
5	<i>Entrevista</i>	22
6	<i>Notícia</i>	45
7	<i>Tirinha</i>	480
		925 textos

Quadro 1. Composição do corpus de gêneros textuais do jornal O Globo. Fonte: Elaboração da autora.

4. OS GÊNEROS TEXTUAIS DO CORPUS COONSTITUÍDO

Nesta seção, descrevem-se os gêneros do *corpus* de gêneros textuais do jornal O Globo.

Artigos

No jornal O Globo, os artigos são publicados diariamente, por autores diferentes, escalados semanalmente. Esses textos figuram nas primeiras seções do Primeiro Caderno, logo após os editoriais e versam sobre assuntos relacionados ao governo, a políticas públicas, à economia, a reformas, dentre outros assuntos atuais interessantes à sociedade. Como se espera nos gêneros opinativos, em todos, um ponto de vista é defendido a partir de argumentos diversos: de autoridade, de causa e consequência, de analogia, entre outros.

Os textos selecionados têm autoria de treze colunistas: Ana Maria Machado, Ascânio Seleme, Cacá Diegues, Carlos Alberto Sardenberg, Carlos Andreazza, Elio Gaspari, Fabio Giambiagi, Fernando Gabeira, José Casado, Luis Fernando Verissimo, Merval Pereira, Roberto Damatta e Zuenir Ventura. No Quadro 2, reproduz-se um dos artigos do *corpus*.

<ARTIGO 19

ELIO GASPARI

Crivella quer o Porto Jogatina>

Num mesmo dia, o prefeito Marcelo Crivella disse que “O Rio de Janeiro é o epicentro da corrupção” e anunciou um futuro radiante para o projeto do Porto Maravilha. Prometeu R\$ 10 bilhões em investimentos com a construção de duas torres de hotéis, um centro de convenções e... um cassino.

O prefeito do “epicentro da corrupção” defende a legalização da jogatina para salvar um projeto megalomaniaco atolado na Zona Portuária da cidade. Isso num estado que tem dois governadores presos, e foram apanhados em roubalheiras dezenas de deputados, secretários do governo e conselheiros do Tribunal de Contas. Dois cardeais da sacrossanta Arquidiocese de Dom Eugenio Sales viram suas atividades tisanadas por malfeitos de pessoas que lhes eram próximas. Tudo isso sem que o jogo seja legalizado.

Um policial militar que trabalhou com a família Bolsonaro e orgulhou-se de “fazer dinheiro” ainda não ofereceu uma versão consistente para explicar suas movimentações financeiras. Um capitão da tropa de elite da PM teve a mãe e a mulher empregadas no gabinete do filho do presidente. Alcunhado “Caveira”, o oficial foi expulso da corporação e está foragido. Ele era donatário de uma milícia da cidade.

O Rio de Janeiro elegeu um juiz para o governo do estado. Outro policial, que se apresentava como seu consultor para assuntos de segurança, está na cadeia, acusado de extorsão. Na última eleição, esse policial foi candidato a deputado federal pelo partido do governador.

Sem cassinos, o Rio já está assim. Nenhuma pessoa de boa-fé pode acreditar que alguma coisa melhorará com estímulos à jogatina e à abertura de lavanderias de dinheiro. Ao crime organizado, Crivella quer juntar o jogo legalizado.

O prefeito não joga, não fuma, não bebe e sabe que está apenas criando uma nova miragem para uma cidade ludibriada por fantasias como as da Copa do Mundo e a da Olimpíada. De miragem em miragem, o Rio vive uma eterna Quarta-Feira de Cinzas. Crivella sabe que a reabertura dos cassinos depende da aprovação de uma lei pelo Congresso. Conhecendo a tessitura do crime organizado na cidade, dificilmente Jair Bolsonaro perfilhará a legalização do jogo.

No mundo real, a única pessoa tenuemente interessada nas torres e no cassino prometidos por Crivella é o bilionário americano Sheldon Adelson, que tem complexos de turismo e jogo em Las Vegas, Macau e Cingapura. Ele começou a trabalhar aos 12 anos (tem 85) e já juntou US\$ 33,3 bilhões. É um campeão da causa de Israel e a ele se atribui a abertura dos cofres de muitos republicanos para Donald Trump. É também um protetor do primeiro-ministro Benjamin Netanyahu. Dele viriam os R\$ 10 bilhões imaginados por Crivella.

O Porto Maravilha de Eduardo Paes atolou porque era um projeto demóforo. O Rio da Zona Portuária nunca poderia ter sido o que é o Puerto Madero argentino, como a Barra da Tijuca nunca será uma Miami.

Aquela área está num bairro popular e centenário. Quem quiser conferir, que ande pelas ruas da Gamboa e de São Cristóvão. A megalomania imobiliária encalhou porque foram poucos os interessados em levar suas empresas para lá. Ali, o povo do Rio sempre viveu em casas modestas. Miami é em outro lugar.

Na região do Porto Maravilha construíram-se dois novos museus. A poucos quilômetros dali, pegou fogo o Museu Nacional. (Quarenta anos antes, incendiou-se o Museu de Arte Moderna. Ganha um fim de semana em Caracas quem souber de outra cidade com semelhante desempenho.)

Quadro 2. Exemplo de texto: Artigo 19. Fonte: Jornal O Globo (03/03/2019).

Cartas de leitor

No jornal O Globo, as cartas dos leitores são publicadas na editoria “Rio” e versam sobre as diversas temáticas que apareceram nos gêneros publicados nas editorias do dia anterior ou da semana (notícias, artigos, editoriais, crônicas etc.). Apesar de se tratar de publicação na editoria específica da cidade (“Rio”), cabe esclarecer que há alguns textos de leitores de outras cidades e estados. Considerando isso, a fim de coletar apenas textos de leitores do Rio de Janeiro, não foram selecionadas as cartas cujos leitores fossem de outro estado.

<CARTA DE LEITOR 23

Artistas no metrô

ANGELA BRANT, RIO>

Antes de sancionar a lei permitindo a apresentação de artistas no metrô (“Proibição à arte que embala as viagens está prestes a cair”, 3/9), os vereadores deveriam saber o quanto podem ser perturbadoras essas manifestações para quem usa o metrô para dormir, ler, estudar ou trabalhar. É permitido tocar nas estações, mas, segundo alguns artistas, as regras são perversas, razão pela qual a maioria se arrisca nos vagões. Que se flexibilizem as regras, criem-se pequenos palcos nas estações, sem que seja invadida a privacidade dos passageiros.

Quadro 3. Exemplo de texto: Carta de leitor 23. Fonte: Jornal O Globo (04/09/2018).

Crônicas

As crônicas, por sua vez, são publicadas todos os dias, no Segundo Caderno do jornal O Globo, em uma das últimas páginas. De acordo com informações dispostas na coluna em que as crônicas são exibidas, essas publicações obedecem a um rodízio de escritores, de acordo com o dia da semana.

Para a constituição do *corpus*, foram selecionados textos de todos os dias da semana, a fim de ter mais de uma crônica de cada autor. Os autores cujas crônicas compõem o *corpus* publicam nos seguintes dias: segunda – Joaquim Ferreira dos Santos; terça – Arnaldo Bloch; quarta – Ana Paula Lisboa e Maria Ribeiro (quinzenal); quinta – Cora Rónai; sexta – Leo Aversa e Ruth de Aquino (quinzenal); sábado – Geovani Martins (quinzenal); domingo – Artur Xexéo. Um dos cronistas não selecionados foi José Eduardo Agualusa, que é angolano, de ascendência portuguesa e brasileira, o que o exclui do rol de cronistas cariocas.

No Quadro 4, a seguir, expõe-se a Crônica 30.

<p><CRÔNICA 30 O que quer dizer o 'foda-se' Joaquim Ferreira Dos Santos></p> <p>Uma tendência mundial, ligar o foda-se para ver o que acontece, nem aí para essas racionalidades antigas de se preocupar com regras civilizatórias. O Brasil está dentro. É o país da tecla F, todo na base do fetiche pelo furo e pela fake news.</p> <p>Moderníssimo, o novo governo segue esse fervor desde o início, quando mandou imediatamente desligar os pardais das estradas e tirar o pé do freio. Ligou o foda-se rodoviário. O resultado já apareceu no número recorde de mortos deste carnaval. Um foda-se fúnebre.</p> <p>A lista dos livros de autoajuda que o GLOBO publicou no sábado deixa o governo à vontade para seguir em frente com o frisson fuderoso. Em primeiro lugar está o invencível, não sai dali faz mais de dois anos, "A sutil arte de ligar o foda-se", de Mark Manson. É um mantra desse tempo chinfrim. A sutil arte de não se responsabilizar pelas consequências.</p> <p>Bolsonaro não leu esse livro, nem parece ter lido qualquer outro que não a biografia do Ustra, mas tem antenas afiadas para as novas conspirações que flanam pelo planeta. O furor pelo Führer. Nada mais natural então que, ao ter como plataforma de governo o foco na ditadura da tecla F e falar à nação vestindo camiseta de clube de futebol (foda-se a liturgia do cargo), ele convoque fãs da força para o Dia Nacional do Foda-se. Foi um general quem deu o brado retumbante, agora contra o congresso e a necessidade de boas relações entre os poderes. Foda-se a federação, é como se pode traduzir o militar falastrão. Ele reclamou que deputados e senadores chantageavam o presidente. Em seu fricote de repugnância pela ordem, lançou como senha para o golpe um palavrão.</p> <p>Desde o início o governo se anuncia como liberal e, regido por um filósofo fissurado no fálico, dele não se podia esperar outra coisa – uma liberalidade de clube de swing. O ministro da Educação já havia assaltado a gramática e, imprecionante!, liberado a ortografia. Agora chegou a vez de o general liberar carnavalescamente a outra banda do vernáculo. Ninguém precisa mais usar o pudico (*), o asterisco que a velha imprensa inventou para esconder as palavras ofensivas à família. É o foda-se ao vocabulário, um furibundo 'Chupa, Houaiss'.</p> <p>A civilização do foda-se tem avançado todo dia, como se houvesse uma vontade institucional de mudar a bandeira e cunhar "Desordem e atraso" entre as estrelas do pavilhão. Nem aí. É o descaso de baixo calão contra a cultura, os índios, a preservação das florestas, a limpeza dos rios e tudo mais que até 400 dias atrás parecia ser cláusula pétrea do orgulho nacional. O importante é aumentar de 20 para 40 a pontuação que um infrator precisa atingir para perder a carteira de habilitação. Pé em Deus, fé na tábua e – o Brasil tem pressa – saiam da frente as disposições em contrário de toda a humanidade ao redor.</p> <p>Por algum resquício de pundonor, o verbo lamentável só foi dito duas semanas atrás, mas ele já administrava o país desde o ano passado. Nenhuma novidade. É a modernização cafajuste para a frase do ministro Jarbas Passarinho, aquele que em dezembro de 1968 disse "às favas com todos os escrúpulos de consciência" e deu de ombros, nem aí, oferecendo seu sim ao AI-5 que radicalizou a ditadura militar, prendeu, matou, fechou o Congresso e censurou os jornais. O foda-se quer dizer a mesma coisa.</p>

Quadro 4. Exemplo de texto: Crônica 30. Fonte: Jornal O Globo (02/03/2020).

Editoriais

Os editoriais do jornal O Globo também são publicados em todos os dias. Uma breve consulta aos jornais de 2018 revelou uma mudança na maneira como esses gêneros são organizados no jornal. Até julho de 2018, os editoriais apareciam ao final da editoria “País”, antes da editoria “Rio”. Além disso, eram publicados dois textos: um representando a opinião do jornal, sem assinatura e outro, com a opinião de algum convidado. A partir de agosto de 2018, os editoriais passaram a figurar na seção intitulada “Opinião O GLOBO”, na qual são publicados dois textos sem assinatura, ambos representando a opinião do corpo editorial.

Esses textos são publicados em pares diariamente e versam sobre temáticas atuais distintas, relacionadas aos cadernos/seções do jornal: País, Economia, Política, Cidade, entre outros. No Quadro 5, há o Editorial 32 do *corpus*:

<EDITORIAL 32>
 Recessão na indústria mostra o desafio da prolongada estagnação
É preciso chegar a um consenso sobre a retomada de investimentos, além da reforma da Previdência

A indústria está em recessão, alerta o Instituto de Estudos para o Desenvolvimento Industrial (Iedi), que reúne meia centena das maiores empresas do país. A queda na produção é constante desde meados do ano passado. Em março, o declínio já atingia dois terços do parque industrial brasileiro, com tendência a se espalhar por toda a base produtiva dos 15 estados mais industrializados.

Nos 26 estados e no Distrito Federal, confirma o IBGE, a retração no produto industrial, que havia sido de 0,3% no último trimestre do ano passado, avançou para 0,7% nos primeiros três meses deste ano.

Se excluídos os ramos de construção civil e de extrativismo — este fortemente influenciado pela tragédia da Vale em Brumadinho (MG) —, tem-se um retrato mais preciso da indústria de transformação: a queda foi de 1,5% entre outubro e dezembro do ano passado e passou a 1,7% de janeiro a março deste ano. Em São Paulo, o retrocesso completou nove meses seguidos, com 72% das fábricas afetadas neste ano.

O país está estagnado — para alguns economistas, à beira de uma depressão. De 1981 até 2018, cresceu muito abaixo da média dos países em desenvolvimento.

Há quase oito anos o Produto Interno Bruto brasileiro patina na faixa de 0,5% ao ano. Por 84 meses, até dezembro, a renda per capita evoluiu 0,3% anuais enquanto a população aumentava 0,8% ao ano. Significa empobrecimento numa economia com 13,2 milhões de desempregados. O Fundo Monetário Internacional lembra que, ao final da década, o Brasil terá registrado crescimento de 0,9%, inferior ao de 90% dos países.

O panorama não é animador. Não há estímulo ao investimento em escala suficiente para uma recuperação, por exemplo, acima de 2%. E sem o motor industrial, obviamente a retomada permanece distante.

A indústria tem papel essencial na produção de inovações tecnológicas, ajuda na geração de superávit na balança comercial e é vital no aumento da produtividade da economia. Reformas do Estado são vitais, sobretudo num quadro de atrofia no qual um servidor aposentado custa à União quase 15 vezes mais do que um aposentado da iniciativa privada.

É necessário, porém, que o foco governamental nas necessárias reformas para melhoria do desempenho futuro do setor público não eclipse o desafio da superação do cenário atual, de estagnação ou de pré-depressão. O governo precisa, com urgência, chegar a um consenso como setor privado sobre os meios para se estimular a retomada dos investimentos no curto prazo. Além da reforma da Previdência.

Quadro 5. Exemplo de texto: Editorial 32. Fonte: Jornal O Globo (03/06/2019).

Entrevistas

Nas edições do Jornal O Globo, em todos os meses do ano estabelecido, houve publicações de entrevistas no Primeiro ou no Segundo Cadernos. As do primeiro grupo faziam parte das diversas editorias que figuram nesse espaço do jornal: “País”, “Mundo”, “Sociedade”, “Economia” e “Rio”. Para composição do *corpus*, selecionaram-se

entrevistas desta última editoria, por representarem, de forma majoritária, falantes do Rio de Janeiro.

Já para as entrevistas do Segundo Caderno, de cultura e arte, não foi possível fazer esse refinamento seguindo o critério de região, visto que há uma heterogeneidade maior da origem do entrevistado. O Quadro 6, a seguir, reproduz a Entrevista 01, retirada do caderno de cultura:

<ENTREVISTA 01
Renata Carvalho / ATRIZ
Atriz defende espetáculo que traz de novo ao Rio. Polêmica em Garanhuns entrará no livro da autora Jo Clifford sobre a trajetória do monólogo no país.
GUSTAVO CUNHA
"A PEÇA É UMA PARÁBOLA BÍBLICA, É PRÓ-CRISTÃ"
A intrincada trajetória da peça "Evangelho segundo Jesus, Rainha do Céu" no Brasil vai virar livro, escrito pela própria dramaturga, a britânica Jo Clifford. Após enfrentar resistência para encená-la em Jundiá, Salvador, Porto Alegre e Rio, a atriz Renata Carvalho passou por mais um teste de fogo na última semana, dessa vez em Garanhuns (PE). "Se desisto da peça, desisto de viver", diz a artista, que volta à Sede das Cias. hoje para uma microtemporada.>

Você traz a peça ao Rio após mais um episódio difícil em Garanhuns (PE). O que houve?
Foi mais um caso de censura. Depois de sermos retirados da programação oficial do Festival de Garanhuns, fizemos uma vaquinha virtual em que conseguimos o dobro do valor solicitado para montar a peça, paralelamente ao evento, num espaço alugado. A primeira sessão foi um sucesso. Mas assim que acabou, explodiram uma bomba na porta do lugar. Os policiais que faziam a nossa segurança precisaram fechar uma área de dois quarteirões. Logo depois, chegaram oficiais de Justiça e um desembargador, que cancelaram a segunda sessão. Fiquei revoltada, abri o portão e mandei o público de 400 pessoas entrar. Quando subi no palco e comecei o espetáculo, eles me interromperam e arrancaram o toldo do local. Continuei debaixo da chuva. Mas não tinha como prosseguir. Fui recebida por uma verdadeira tropa de choque. Os oficiais queriam a prova do contrato do aluguel que havíamos feito por aquele espaço. Foi uma violência desonesta.

Depois desta temporada no Rio, haverá mais apresentações pelo país?
Vamos fazer mais algumas em São Paulo. Depois, seguirei com uma turnê do espetáculo "Domínio público" por algumas cidades, com Wagner Schwartz e Maikon K, e direção de Guilherme Weber. Mas em novembro participo de um festival de teatro em Cabo Verde, na África, onde volto com "Jesus Rainha do Céu". Jo Clifford vai me acompanhar por lá. Depois, ela viaja comigo para o Brasil, onde seguirá minha rotina por mais de um mês. O resultado será um livro, escrito por ela, sobre toda a trajetória da peça no Brasil.

Em nenhum momento pensou em desistir do projeto?
Se desisto da peça, desisto de viver. Na estreia do projeto, no ano passado, quando nos censuraram logo em nossa primeira vez, falei para a minha diretora: "Bem-vinda à transfobia". Vivo isso diariamente. Se saio de casa, a violência fica baforando no meu cangote durante as 24 horas do dia, pois as pessoas acham que é normal tirar sarro de uma travesti. Por isso não trabalho mais para mim. O que faço é lutar para que as pessoas trans que estão nascendo agora sofram menos do que eu. Carregamos um corpo endemoniado pelos outros. Tem gente que fala como vai me matar, que vai me dar um tiro no meio da rua.

E você não denuncia?
É difícil. Se resolvesse denunciar, passaria todos os dias da minha vida lidando com isso. Mas a corrente do bem é infinitamente maior.

Já houve manifestações de apoio por parte de pessoas vinculadas a igrejas?
Claro. Apresentei o espetáculo, inclusive, dentro de três igrejas, em São Paulo e em Curitiba. Numa dessas vezes, depois que encerrei a peça, um bispo me chamou para conversar e brincou que era para eu voltar no domingo, para fazer a missa. Ele disse que o que eu fazia, com o meu discurso, atingia os fiéis de uma forma muito mais rápida. É isso. A peça é pró-cristã. Falo sobre amor, solidão, tolerância, feminismo... É uma parábola bíblica, em que levanto a seguinte questão: o que faria uma pessoa ser atacada e crucificada hoje, da mesma forma que aconteceu com Jesus?

A questão de gênero continuará a ser um tema central em seus espetáculos?
Costumo dizer que meu corpo já nasceu com marcas e estigmas que foram introjetados em mim. Preciso lutar por representatividade. Em março de 2017, fundei o Movimento Nacional de Artistas Trans, o Monat. Vamos lançar um acordo com artistas cisgêneros, solicitando que eles parem de interpretar personagens trans nos próximos 30 anos.

Quadro 6. Exemplo de texto: Entrevista 01. Fonte: Jornal O Globo (01/08/2018).

Notícias

As notícias recolhidas para a constituição do *corpus* foram retiradas da editoria “Rio”, como já mencionado, e versam sobre temáticas diversificadas relacionadas à cidade, tais quais: infraestrutura, políticas públicas, violência, chuvas, turismo, entre outras.

No jornal O Globo, os textos desse gênero são assinados pelos jornalistas envolvidos no processo de pesquisa e escrita da matéria. Sendo assim, há notícias assinadas por um ou mais jornalistas. No Quadro 7, em que se reproduz a Notícia 24 do *corpus*, três jornalistas participaram da elaboração do texto. Quando não há envolvimento total na produção da notícia, o nome aparece citado como colaboração (por exemplo: “LUANA SANTIAGO, Colaborou Priscilla Aguiar Litwak” – Notícia 22).

<NOTÍCIA 24
GISELLE OUCHANA, HELLEN GUIMARÃES E SAULO PEREIRA GUIMARÃES-
 Avenida Niemeyer é tomada por lama durante temporal
Situação da via, interditada desde terça-feira, volta a preocupar moradores da região; no Humaitá, queda de palmeira centenária fechou rua por três horas

A chuva que atingiu a cidade na manhã de ontem provocou a formação de bolsões d'água e queda de árvores em diferentes pontos do Rio. Parte da Avenida Niemeyer, fechada desde terça-feira por ordem judicial devido ao risco de novos deslizamentos, foi tomada por lama que descia de encostas. Equipes da Geo-Rio tentaram conter a enxurrada desobstruindo canaletas das redes de esgoto e águas pluviais, mas a sujeira acabou indo para o mar. Moradores do Vidigal, que já tiveram a rotina impactada por causa das restrições de mobilidade, voltaram a ficar aflitos, com medo de deslizamentos.

– Se tem chuva, começa a correria. Alguns saem de casa e tentam arrumar lugar para ficar – disse Edivaldo Corrêa, que vive na comunidade.

Ontem, a Niemeyer também serviu de área de treino para esportistas, alheios ao perigo. O analista de sistemas Mauro Aquino, de 59 anos, morador do Leblon, subiu a via de bicicleta pelo acesso de São Conrado:

– Sabendo que não teria carro, vim para ver como está o local.

Presidente da Associação de Moradores e Amigos de São Conrado (Amasco), José Britz acredita que ocupação do Vidigal sem controle e o consequente lançamento de esgoto nas encostas da região configuram o maior problema da área hoje.

No Humaitá, uma palmeira imperial caiu e deixou a Rua Humaitá interditada por três horas, na altura do Largo dos Leões. O problema foi visto como um alerta por moradores, que reclamam da falta de poda regular e de conservação das árvores.

– Antigamente, o serviço não era perfeito, mas era recorrente. O problema aumentou muito – reclamou Alexander Polasek, de 47 anos, temendo que vítimas sejam atingidas por galhos e árvores.

A palmeira, maior que um prédio de oito andares, caiu por volta das 9h15m. Ninguém foi atingido.

A vereadora Teresa Bergher (PSDB) vai enviar um ofício à prefeitura solicitando que a poda de árvores volte a ser realizada pela Fundação Parques e Jardins, não mais pela Comlurb.

Quedas de árvores causaram ainda interdição na Grajaú-Jacarepaguá e interrupção no fornecimento de energia no Rio Comprido e em Santa Teresa.

Hoje, a temperatura deve variar entre 19 e 30 graus. O céu ficará nublado, com possibilidade de pancadas de chuva a qualquer momento.

Quadro 7. Exemplo de texto: Notícia 24. Fonte: Jornal O Globo (03/06/2019).

Tirinhas

No jornal O Globo, os mesmos títulos de tirinhas são publicados de segunda a sábado; já no domingo, são novos títulos, alguns compostos por seis quadrinhos, diferentemente das tirinhas tradicionais, de três quadrinhos. Inicialmente, no processo de seleção, seriam coletadas apenas as produções nacionais; entretanto, para diversificar os títulos/autores, optou-se por reunir todos os títulos publicados, isto é, textos criados originalmente em português ou traduzidos.

Sendo assim, foram coletados textos de oito títulos distintos: “A arte de zoar”, de Reinaldo; “Bichinhos de jardim”, de Clara Gomes; “O corpo é o porto”, de André Dahmer; “Urbano, o aposentado”, de Antônio Silvério; “Valente”, de Vitor Cafaggi; “Dustin”, de Steve Kelley e Jeff Parker; “Zits”, de Scott e Borgman; e “Zoé e Zezé”, Jerry Scoott e Rick Kirkman. Os últimos três títulos são tirinhas traduzidas para o português. O Quadro 8, a seguir, exibe a Tirinha 224.



Quadro 8. Exemplo de texto: Tirinha 24. Fonte: Jornal O Globo (05/03/2019).

5. CONSIDERAÇÕES FINAIS

Neste texto, buscou-se descrever as etapas de constituição de um *corpus* de escrita culta representativo do Rio de Janeiro. A etapa da pesquisa ora relatada procurou atingir um dos objetivos da investigação em andamento, que é o de organizar uma amostra heterogênea no que tange aos gêneros textuais e autores selecionados, mas homogênea quanto aos critérios de composição. Além do estabelecimento de critérios para composição do *corpus*, no trabalho realizado, empenhou-se para tornar a amostra manipulável por computador, a fim de viabilizar buscas automáticas por dados linguísticos – como as realizadas com aplicação de expressões regulares.

Acredita-se que, como toda etapa de uma pesquisa científica, o *corpus* organizado é passível de críticas e de reestruturação, se assim for julgado necessário. Considera-se, contudo, que o produto gerado pode contribuir para a descrição de variedades cultas, ao propiciar a observação de formas variantes, no âmbito do *continuum* de fala-escrita (concepção). Nas próximas etapas da pesquisa de Doutorado, pretende-se realizar análises sobre variáveis linguísticas com base no *corpus* ora composto.

REFERÊNCIAS

- BIAZOLLI, Caroline Carnielli. *Posição de clíticos pronominais em duas variedades do português: inter-relações de estilo, gênero, modalidade e norma*. Araraquara: Faculdade de Ciências e Letras, Universidade Estadual Paulista, 2016. (Tese de Doutorado)
- BORTONI-RICARDO, Stella Maris. "Um modelo para análise sociolinguística do português brasileiro". In: *Nós chegemos na escola, e agora? Sociolinguística e educação*. São Paulo: Parábola editorial, 2005. p. 45-52.
- BORTONI-RICARDO, Stella Maris. O Português brasileiro. In: BORTONI-RICARDO, Stella Maris. *Educação em língua materna: a sociolinguística na sala de aula*. São Paulo: Parábola Editorial, 2004. p. 51-70.
- COLLISCHONN, Gisela; MONARETTO, Valéria de Oliveira. Banco de dados VARSUL: a relevância de suas características e a abrangência de seus resultados. *Alfa: Revista de Linguística*, São Paulo, v. 56, n. 3, p. 835-853. 2012.
- ECKERT, Penelope. Variation, convention, and social meaning. Plenary talking. *Annual meeting of the Linguistic variation*. San Francisco, 2005.
- FARACO, Carlos Alberto. *Norma culta brasileira – desatando alguns nós*. São Paulo: Parábola Editorial, 2008.
- FREITAG, Raquel Meister Ko; MARTINS, Marco; TAVARES, Maria Alice. Bancos de dados sociolinguísticos do Português Brasileiro e os estudos de terceira onda: potencialidades e limitações. *Alfa: Revista de Linguística*, São Paulo, v. 56, n.3, p. 917-944, 2012.
- LABOV, William. *Padrões sociolinguísticos*. Trad.: Marcos Bagno; Marta Scherre e Caroline Cardoso. São Paulo: Parábola, 2008 [1972].
- LIMA, Monique Débora Alves de Oliveira; VIEIRA, Sílvia Rodrigues. Para uma abordagem da norma no continuum fala-escrita. In: VIEIRA, S. R.; LIMA, M. D. A. de. (Orgs.). *Variação, gêneros textuais e ensino de Português: da norma culta à norma-padrão*. 1ed. Rio de Janeiro: Letras UFRJ, 2019. p. 8-16.
- MARCUSCHI, Luiz Antônio. Oralidade e letramento. *Da fala para a escrita: atividades de retextualização*. São Paulo: Cortez, 2001. p. 15-44.
- MARCUSCHI, Luiz Antônio. *Produção textual, análise de gêneros e compreensão*. São Paulo: Parábola Editorial, 2008.
- PAGOTTO, Emílio. Norma e condescendência; ciência e pureza. *Línguas e instrumentos linguísticos*. Campinas: Pontes, n. 2, 1998. p. 49-68.
- SALOMÃO, Ana Cristina Biondo. Variação e mudança linguística: panorama e perspectivas da sociolinguística variacionista no Brasil. *Fórum Linguístico*, Florianópolis, v. 8, n. 2, p. 187-207, jul./dez. 2011. <http://dx.doi.org/10.5007/1984-8412.2011v8n2p187>
- SARDINHA, Tony Berber. Linguística de corpus: histórico e problemática. *Revista D.E.L.T.A.*, v. 16, n. 2, p. 323-367, 2000.
- SILVA, Edila Vianna da. Bancos de dados sociolinguísticos em português. *IDIOMA*, Rio de Janeiro, n. 29, v.2, p. 168-180. 2015.
- SILVA, Giselle Machline de Oliveira e Silva. Coleta de dados. IN: MOLLICA, Maria Cecília; BRAGA, Maria Luiza. (Orgs.). *Introdução à Sociolinguística: o tratamento da variação*. 4. ed. São Paulo: Contexto, 2015. p. 117-134.
- VIEIRA, Sílvia Rodrigues; LIMA, Monique Débora Alves de Oliveira Lima. (Orgs.). *Variação, gêneros textuais e ensino de Português: da norma culta à norma-padrão*. 1. ed. Rio de Janeiro: Letras UFRJ, 2019. Disponível em: <https://figshare.com/articles/book/Varia_o_g_neros_textuais_e_ensino_de_Portugu_s_da_norma_culta_norma_padr_o/11857965>
- VIEIRA, Sílvia Rodrigues; BRANDÃO, Sílvia Figueiredo. *CORPORAPORT: Variedades do Português em análise*. Rio de Janeiro: Faculdade de Letras-UFRJ. Disponível em: <www.corporaport.letras.ufrj.br>

WEINREICH, Uriel; LABOV, William; HERZOG, Marvin. *Fundamentos empíricos para uma teoria da mudança linguística*. Tradução: Marcos Bagno. Revisão técnica: Carlos Alberto Faraco. São Paulo: Parábola Editorial, 2006 [1968].

Site Infoglobo: <https://www.infoglobo.com.br/Anuncie/ProdutosDetalhe.aspx?IdProduto=91>