

## Artigos

### Representação do discurso em narrativas factuais: Sexismo sistêmico e marginalização de mulheres em textos jornalísticos brasileiros

#### *Discourse Representation in Factual Narratives: Systemic Sexism and Marginalization of Women in Brazilian News Reports*

Daniel Alves\*

#### RESUMO

Este artigo discute padrões de representação do discurso em narrativas factuais publicadas por veículos brasileiros de comunicação entre 2012 e 2019, contrastando as formas como homens e mulheres são citados em textos jornalísticos. O artigo se afilia a um arcabouço teórico da Linguística Sistêmico-Funcional e da Análise Crítica do Discurso — como Fairclough (1998) e Halliday & Matthiessen (2014) —, em sua interface com métodos da Linguística de Corpus e utiliza dados coletados a partir do site *Corpus do Português* (Davies, 2016). A análise corrobora Caldas-Coulthard (1994) e aponta para uma sub-representação de mulheres em textos jornalísticos — tendo sido identificadas 37.199 ocorrências de ‘ele disse’ contra 14.583 de ‘ela disse’. Além disso, análise amostral revela um contraste nos contextos em que homens e mulheres são citados: 70,3% dos Dizentes Masculinos podem ser associados a contextos mais marcadamente públicos / institucionais, contra 36,3% de Dizentes Femininos para a mesma categoria. Por outro lado, em contextos mais marcadamente pessoais / privados, o número de Dizentes Femininos chega a 17,7%, enquanto o de Dizentes Masculinos fica em 4,1%. Os resultados revelam, portanto, indícios de um sexismo sistêmico em textos jornalísticos brasileiros.

**Palavras-chave:** *representação do discurso; linguística de corpus; sexismo sistêmico; textos jornalísticos brasileiros*

\* Universidade Federal da Paraíba (UFPB). Paraíba – Brasil. <https://orcid.org/0000-0002-3702-0895>. E-mail: [daniel@cchla.ufpb.br](mailto:daniel@cchla.ufpb.br)

## ABSTRACT

This paper discusses patterns of discourse representation in factual narratives published by Brazilian news organizations between 2012 and 2019, contrasting how the voices of men and women are quoted in news reports. This paper draws on theoretical works affiliated with Systemic Functional Linguistics and Critical Discourse Analysis — such as Fairclough (1998) and Halliday & Matthiessen (2014) — integrated with Corpus Linguistics methods, gathering data from the website *Corpus do Português* (Davies, 2016). The analysis corroborates Caldas-Coulthard's findings (1994) and points toward an underrepresentation of women in news reports — with 37,199 occurrences of 'ele disse' (he said) compared to 14,583 for 'ela disse' (she said). Moreover, a sample analysis reveals a contrast in the contexts in which the voices of men and women are quoted: 70.3% of Male Sayers can be associated with contexts more prominently public / institutional, compared to 36.3% of Female Sayers in the same category. Conversely, when it comes to contexts more prominently private / personal, the proportion of Female Sayers reaches 17.7%, compared to 4.1% for Male Sayers. The results reveal signs of systemic sexism in Brazilian news reports.

**Keywords:** *discourse representation; corpus linguistics; systemic sexism; brazilian news reports*

*How come you get the really interesting existential question, and I get the 'rabbit food' question?*

## 1. Introdução

Este artigo pretende discutir uma situação de sexismo sistêmico, identificada a partir da forma como veículos de comunicação representam as falas de homens e de mulheres, criando uma situação de marginalização destas últimas: enquanto homens têm espaço na mídia para falar sobre aspectos de suas vidas profissionais, às mulheres são relegados espaços relacionados a esferas pessoais. A epígrafe escolhida para enquadrar esta discussão — embora não esteja entre os dados que serão analisados no trabalho aqui apresentado — ilustra bem essa situação de marginalização. Em 2012, por ocasião da divulgação de um dos filmes da franquia Vingadores a atriz Scarlett Johansson e o seu colega de elenco, Robert Downey Jr. concediam entrevista coletiva<sup>1</sup>, quando ocorre a seguinte interação:

1. A entrevista coletiva em questão está disponível no endereço <https://youtu.be/wBbSNnGmPfw?t=190> (último acesso em 03/05/2024) e as perguntas a que me refiro podem ser acessadas a partir dos 3'10" do vídeo.

Em nome dos jornalistas japoneses, tenho uma pergunta para Robert e para Scarlett. Primeiramente para Robert: Ao longo dos filmes Ironman 1 e 2, Tony Stark começa como um personagem individualista, mas que aprende a fazer parte de um time. Como você abordou esse papel, tendo em mente o tipo de maturidade, como ser humano, quando se trata do personagem Tony Stark e se você aprendeu algo ao longo dos três filmes que você fez. E para Scarlett: Para entrar em forma para o papel de Viúva Negra, você teve que fazer algo especial em termos de dieta, como, por exemplo, você teve que comer alguma comida específica ou algo do tipo?<sup>2</sup>

Ao ouvir a pergunta, a atriz externa o seu desconforto com as perguntas, respondendo com a seguinte frase (que é apresentada como epígrafe deste artigo):

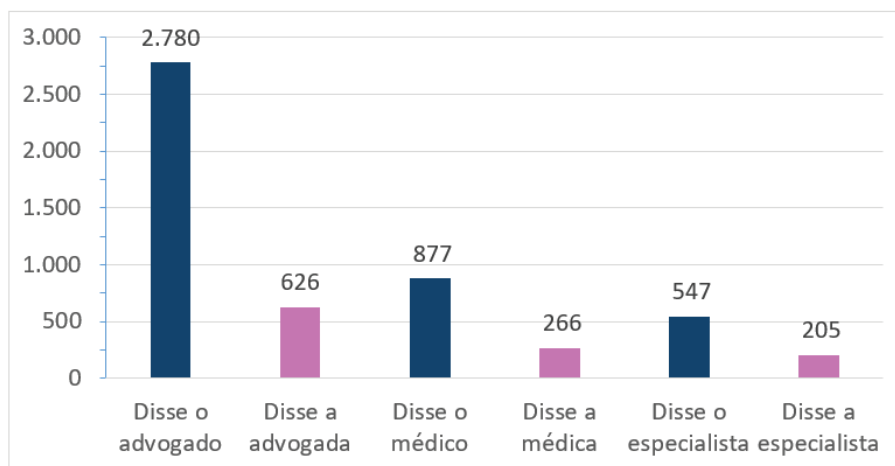
Por que as perguntas para você são interessantes e existenciais e as para mim são sobre ‘comida de coelho’?

Ao longo das diferentes etapas desta pesquisa — desde os momentos em que foram realizados levantamentos iniciais, para verificar a viabilidade da investigação —, foram identificados indícios da marginalização de mulheres, mencionada nesta introdução: havendo não apenas uma diferença quantitativa nos números de citações de homens e mulheres em narrativas factuais, como também uma frequência maior de Dizentes Masculinos em representações de fala associáveis a esferas profissionais, contrastada com uma frequência maior de Dizentes Femininos em representações associáveis a esferas familiares. Os Gráficos 1 e 2, a seguir, mostram resultados obtidos nos levantamentos iniciais desta pesquisa, realizados no site *Corpus do Português* (Davies, 2016) e ilustram as observações sobre o contraste entre os espaços, mostrando os números de ocorrências gerados a partir dos levantamentos pelos nós de busca (doravante ‘nódulos’) “Disse o advogado”, “Disse a advogada”, “Disse o médico”, “Disse a médica”, “Disse o especialista”, “Disse a especialista” (Gráfico 1) e “Disse o pai”, “Disse a mãe”, “Disse o tio”, “Disse a tia”, “Disse o avô”, “Disse a avó” (Gráfico 2):

---

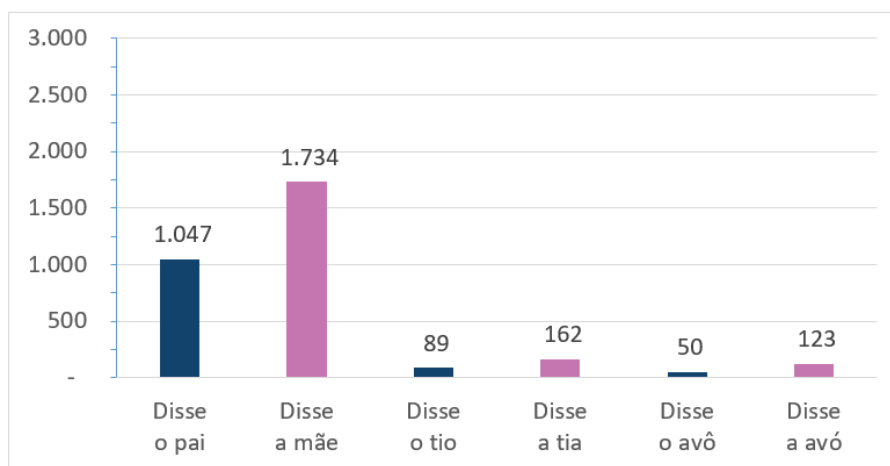
2. Minha tradução para: On behalf of Japanese journalists, I have a question to Robert and to Scarlett. Firstly to Robert: Throughout Ironman 1 and 2, Tony Stark started off as a very egotistic character, but learns how to play as a team, so how did you approach this role, bear in mind that kind of maturity as a human being when it comes to the Tony Stark character and did you learn anything throughout the three movies that you made. And to Scarlet, to get into the shape of Black Widow did you have anything special to do in terms of diet, like you have to eat some specific food or that sort of thing?

**Gráfico 1.** Levantamento inicial de representações do discurso em esferas profissionais



Fonte: elaboração própria, a partir de dados coletados do site *Corpus do Português*

**Gráfico 2.** Levantamento inicial de representações do discurso em esferas familiares



Fonte: elaboração própria, a partir de dados coletados do site *Corpus do Português*

Os Gráficos 1 e 2, acima, indicam algumas das observações feitas nos levantamentos iniciais realizados no decorrer desta pesquisa. No Gráfico 1, por exemplo, é possível observar indícios de que, quando se trata de contextos profissionais, os Dizentes Masculinos tendem a ser mais frequentes em textos jornalísticos: os resultados referentes às representações de médicos, advogados e especialistas (no masculino) são numericamente superiores aos de suas contrapartes femininas (respectivamente médicas, advogadas

e especialistas (no feminino)). Quando se trata de contextos familiares, no entanto, é possível observar, no Gráfico 2, indícios de uma tendência oposta: com Dizentes Femininos (mães, tias e avós) sendo mais frequentes do que suas contrapartes masculinas (respectivamente pais, tios e avôs).

Partindo dos contrastes observados nos levantamentos iniciais, esta pesquisa tenta responder às seguintes perguntas: i) Existem diferenças na forma como veículos brasileiros de comunicação representam as falas de homens e de mulheres? e ii) Existem, no contexto contemporâneo, indícios de uma marginalização das mulheres em termos de espaços públicos / institucionais de fala?

De modo a tentar responder a tais perguntas e, a partir de uma análise qualitativa, avaliar se há padrões de linguagem, em termos de acesso à voz de homens e mulheres, este trabalho levanta, a partir do site *Corpus do Português* (Davies, 2016), linhas de concordância que indicam a representação do discurso por meio do verbo de elocução neutro ('disse') e busca analisar padrões textuais no acesso a vozes de homens e mulheres em textos jornalísticos. Dessa forma, o artigo revisita o trabalho de Caldas-Coulthard (1994), que analisa textos jornalísticos em inglês e aponta diferenças quantitativas e qualitativas na forma como homens e mulheres são citados em narrativas factuais.

Cabe reconhecer, de antemão, que este artigo não estabelecerá uma comparação direta com os resultados de Caldas-Coulthard (1994). A principal razão para isso está no fato de os trabalhos lidarem com sistemas linguísticos e culturais diferentes — este trabalho levantando e analisando dados a partir de um corpus em português, e Caldas-Coulthard (1994), a partir de um corpus de língua inglesa. Ainda assim, o trabalho da autora será importante para subsidiar as investigações sobre a forma como textos jornalísticos dão voz a homens e a mulheres.

Para atingir seus objetivos, este artigo está estruturado em cinco seções, contando com esta introdução, a saber: Revisão teórica, que traz uma discussão enxuta<sup>3</sup> sobre representação de fala; Método, que descreve o uso

---

3. A decisão por uma revisão teórica mais enxuta se dá em função da extensão deste artigo e por ser um tema já discutido por diferentes autores (inclusive pelo autor deste trabalho, em outras oportunidades). Para uma discussão mais aprofundada sobre Representação do Discurso, recomendo a leitura de trabalhos como:

- Leech, G., & Short, M. (1981). *Style in fiction*. Longman.
- Short, M. (1988). Speech Presentation, the novel and the press. In Van Peer, W. (Ed.) *The taming of the text: explorations in language, literature and culture* (pp. 61-81). Routledge.

do *Corpus do Português* para o levantamento de dados, além dos procedimentos adotados para salvar e classificar os dados; Análise de dados e discussão, que traz os resultados quantitativos e qualitativos da pesquisa aqui empreendida; e Considerações finais, que retoma as perguntas de pesquisa e sugere propostas para investigações futuras.

## 2. Revisão teórica

Recorrer a representações da fala, como bem aponta Caldas-Coulthard (1994), é uma característica importante para muitos gêneros textuais: boletins de ocorrência, textos acadêmicos, notícias e textos literários, por exemplo, tendem a recorrer a recursos linguísticos para acessar as vozes e discursos de outrem. As formas como tais vozes são selecionadas para representação, e as formas como elas são efetivamente representadas, em função de sua complexidade, tendem a ser alvo de diferentes discussões, que abrangem desde as hierarquias sociais envolvidas até os mecanismos linguísticos acionados nesse processo.

Caldas-Coulthard (1994) trabalha com o conceito de voz acessada e analisa a forma como textos jornalísticos utilizam a representação da fala para dar voz a homens e a mulheres. A autora acessa dados linguísticos do *Bank of English*<sup>4</sup> e analisa linhas de concordâncias levantadas a partir de verbos de elocução em língua inglesa. Em sua amostra, a autora identifica que homens tendem a ser mais citados (497 citações, contra 62 de mulheres) e que mulheres tendem a ser citadas quando assuntos estão relacionados a esferas privadas (nos seus papéis como mães, filhas, mulheres, viúvas etc.). Segundo a análise da autora, essa diferença implica uma marginalização das mulheres em termos de espaços públicos / institucionais de fala e abre o questionamento sobre os veículos de comunicação não recorrerem às vozes de mulheres para discutir assuntos de interesse social.

Machin (2008), discutindo o papel do texto jornalístico na sociedade, ressalta a importância do jornalismo para a democracia, apontando a expectativa de que textos jornalísticos informem cidadãos e cidadãs sobre eventos significativos de forma tão acurada e verdadeira quanto possível. Reconhecendo a existência de vieses nesse processo, Machin (2008) se

---

4. A autora indica ter selecionado a base de dados referente ao ‘*The Times*’, a partir do *Cobuild* — (sigla em inglês para *Collins Birmingham University International Language Database*)

apoia em trabalhos afiliados à Análise Crítica do Discurso para defender que se conheçam os processos subjacentes à produção de textos — de modo a entender as práticas discursivas e as relações de poder que influenciam a construção de textos. Ao discutir sobre elementos que levam um evento a ser selecionado para ser noticiado, Machin (2008) elenca fatores que, no contexto da pesquisa aqui apresentada, podem ser associados aos processos de recurso a vozes de outrem — como, por exemplo, a Personalização (quando eventos complexos são apresentados sob o ponto de vista de um indivíduo) e a Dramatização (quando pedaços de informação são apresentados colocando personagens no centro dos eventos).

Dentro do arcabouço da Linguística Sistêmico-Funcional de Halliday, (Halliday e Matthiessen, 2014) a representação da fala é contemplada dentro do sistema de projeção, ocorrendo por meio de complexos oracionais. No sistema de transitividade dessa teoria linguística, a representação da fala ocorre em Orações Verbais, realizadas por um Dizente. A terminologia ‘Dizente’ é adotada ao longo deste trabalho e indica o sujeito de uma Oração Verbal.

Quanto a ‘discurso’, este trabalho reconhece, como discute Fairclough (1993) a complexidade do termo. Dentre as muitas acepções discutidas pelo autor, interessa aqui aquela que se refere aos “diferentes tipos de linguagem usada em diferentes tipos de situação social” (Fairclough, 1993, p.3), podendo o termo ser empregado em situações diversas, como ‘discurso de jornal’, ‘discurso publicitário’ e outros, exemplificados pelo autor. Dentro da lógica da argumentação do autor, mais do que apenas refletir entidades sociais e relações, os discursos “as constroem ou as ‘constituem’ (...) [posicionando] as pessoas de diversas maneiras como sujeitos sociais” (Fairclough, 1993, p. 3) <sup>5</sup>.

Quanto à escolha pelo termo ‘representação do discurso’, este trabalho segue a decisão, já discutida em trabalhos anteriores, como Alves e Pagano (2016), de entendê-lo como mais abrangente e mais compreensivo, tanto por ele contemplar diferentes tipos de discurso quanto por ele abranger a ideia de que não é neutra a decisão de representar um discurso, sempre havendo, como aponta Fairclough (1988, p.125) uma “decisão de se interpretá-lo e representá-lo de uma maneira e não de outra”.

5. As traduções das citações de Fairclough apresentadas neste parágrafo são de Izabel Magalhães (2001, p.23-24). Publicada em:

- Fairclough, N. (2001). *Discurso e mudança social*. Tradução de Izabel Magalhães. Editora Universidade de Brasília, 2001.

Outro ponto importante para este trabalho diz respeito à decisão em empregar um método de levantamento de dados associado à Linguística de Corpus. Para Semino & Short (2004), o uso de ferramentas eletrônicas (como é o caso da Linguística de Corpus) para o levantamento de dados é uma forma de promover análises mais amplas (em oposição a modelos de análise que selecionam apenas exemplos que condizem com os modelos de análise previamente decididos pelo/a analista). Além disso, como método de trabalho, a Linguística de Corpus permite a identificação de padrões de linguagem não perceptíveis por outros meios de análise — como discutem diversos autores, entre eles Berber Sardinha (2000), McEnery & Hardie (2012), Alves e Assis (2016) e Assunção & Araújo (2019). A próxima seção desenvolve a discussão sobre a Linguística de Corpus, apresentando tanto o corpus utilizado para levantamento de dados, quanto os passos adotados para a construção da pesquisa.

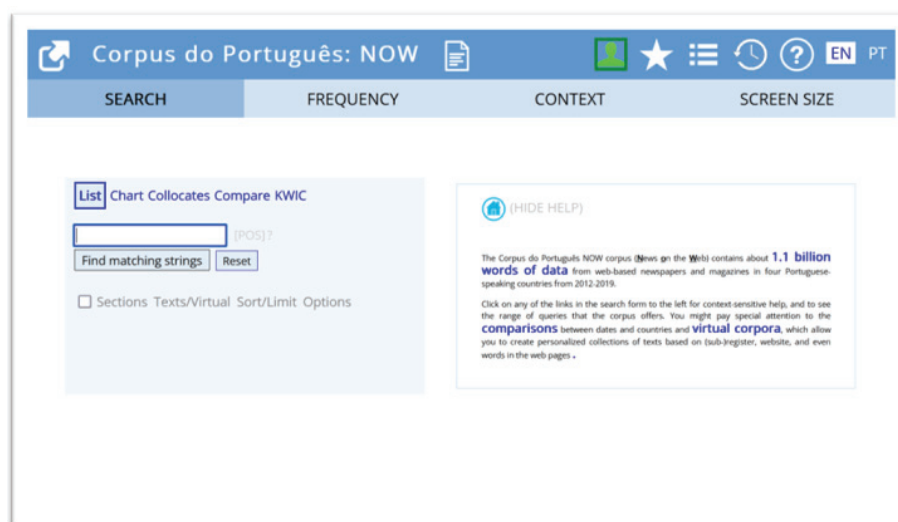
### 3. Método

#### *Sobre o corpus*

Como apontado no final da seção anterior, a investigação aqui proposta utiliza recursos da Linguística de Corpus, buscando identificar padrões linguísticos que talvez não pudessem ser percebidos por outros meios de análise. Os dados utilizados para a análise aqui apresentada são autênticos, no sentido de terem sido produzidos com fins comunicativos reais — que é o caso das notícias jornalísticas aqui analisadas —, levantados partir do site *Corpus do Português*.

De modo a analisar apenas narrativas factuais, dentro do *Corpus do Português* (Davies, 2016), é selecionada a seção **NOW (News on the web)** — disponível a partir do endereço <https://www.corpusdoportugues.org/now/> — cujos resultados são referentes a notícias e dados publicados em veículos jornalísticos, entre os anos de 2012 e 2019. A Figura 1, a seguir, ilustra a tela inicial da seção NOW:



**Figura 1.** Tela de busca do site Corpus do Português

Fonte: Captura de tela feita pelo autor do artigo

Na Figura 1, acima, é possível ver algumas das informações sobre a seção NOW do *Corpus do Português*, incluindo a sua apresentação, em que é revelado o fato de a seção contar com cerca de 1,1 bilhão de palavras. Criado por Mark Davies, o *Corpus do Português* oferece subcorpora com bilhões de palavras provenientes de dados recentes de quatro países de língua portuguesa, a saber: Angola, Brasil, Moçambique e Portugal. A ferramenta (online) oferece recursos que permitem visualizar linhas de concordância, acessar os textos fonte e observar contextos que permitam a identificação de padrões linguísticos. Mais informações sobre o projeto podem ser encontradas em <https://www.corpusdoportugues.org/>.

### *Levantamento e organização dos dados*

Com o objetivo de analisar se existem diferenças na forma como veículos brasileiros de comunicação representam as falas de homens e de mulheres, levantam-se linhas de concordância a partir de buscas por ‘Ele disse’ e por ‘Ela disse’. Como já mencionado, tais levantamentos são feitos a partir da seção NOW. A Figura 2, a seguir, ilustra as linhas de concordância levantadas a partir de ‘Ela disse’:

**Figura 2.** Linhas de concordância geradas pelo site Corpus do Português a partir da busca por ‘ela disse’

|                                                                                                                         | SEARCH      | FREQUENCY             | CONTEXT                                                                                                           | LOG IN |
|-------------------------------------------------------------------------------------------------------------------------|-------------|-----------------------|-------------------------------------------------------------------------------------------------------------------|--------|
| FIND SAMPLE: 100 200 500 1000<br>AGE: 1 / 146<br>14583 ENTRIES: 13189 TEXTS<br>LIMITS: NONE<br>SORTING: DATE ↑, COUNTRY |             |                       |                                                                                                                   |        |
| CLICK FOR MORE CONTEXT<br>SAVE TRANSLATE ANALYZE HELP                                                                   |             |                       |                                                                                                                   |        |
| 1                                                                                                                       | 18-04-27 MZ | @Verdade Online       | é preciso reforçar as medidas de segurança e fiscalização. # Todavia, o que <b>ela disse</b> tem sido amplamente  |        |
| 2                                                                                                                       | 18-01-15 PT | Revista VIP           | e decidi responder de forma privada, depois de algum tempo a processar o que <b>ela disse</b> », lê-se. « Foi a   |        |
| 3                                                                                                                       | 18-01-15 PT | Revista VIP           | não tinha sido assim para ela, fiquei surpreendido e confuso. Li o que <b>ela disse</b> e decidi responder de fo  |        |
| 4                                                                                                                       | 18-05-10 PT | SAPO Lifestyle        | casei com a minha primeira mulher, ela era bailarina em Boston, e quando <b>ela disse</b> a os pais que estav     |        |
| 5                                                                                                                       | 18-01-18 BR | 24Horas New           | " meu bebê, meu bebê ". Perguntei se ela estava grávida, mas <b>ela disse</b> que não. De repente, a avó cheg     |        |
| 6                                                                                                                       | 18-01-20 BR | Globo.com             | eu não fazia vistoria de local. " # Seguindo o mesmo raciocínio, <b>ela disse</b> ainda que foi orientada por sua |        |
| 7                                                                                                                       | 18-01-22 BR | Internacional Estadão | antes de a eleição para comprar seu silêncio em relação a o caso extramatrimonial que <b>ela disse</b> ter tido   |        |
| 8                                                                                                                       | 18-03-09 BR | Globo.com             | crimes contra a administração pública e por improbidade administrativa. Procurada por o G1, <b>ela disse</b> e    |        |
| 9                                                                                                                       | 18-03-11 BR | Blasting News         | fato de ela ter ameaçado Ayrton em um determinado momento e contra Paula, que <b>ela disse</b> que não ge         |        |

Fonte: Captura de tela feita pelo autor do artigo

A Figura 2, acima, ilustra o processo de levantamento de dados a partir da busca por ‘ela disse’, mostrando as primeiras linhas de concordância geradas pelo site *Corpus do Português*. Na imagem, é possível ver, da esquerda para a direita: 1) os números das linhas de concordância; 2) seguidos pela data de publicação da notícia (no formato AA-MM-DD) e da variedade linguística do português —, havendo as opções Angolano (AO), Brasileiro (BR), Europeu (PT) ou Moçambicano (MZ) —; 3) os veículos de informação nos quais as notícias foram publicadas; 4) algumas opções para leitura e tradução do texto (em integração com outras ferramentas baseadas na Internet); e, por fim, 5) as linhas de concordância, com os nódulos de busca destacados em negrito e realçados em verde.

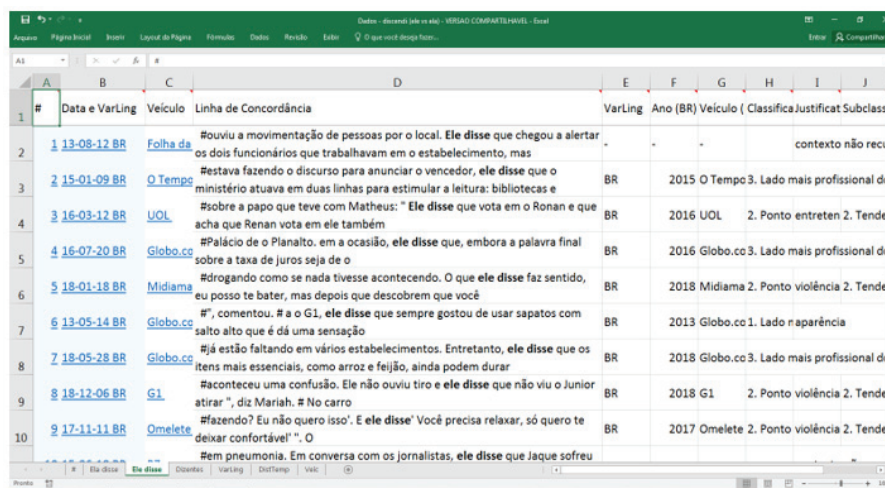
As linhas de concordância resultantes da busca são copiadas e salvas em uma planilha eletrônica<sup>6</sup>, seguindo um método de trabalho discutido em Alves e Assis (2016). As informações são dispostas em colunas cuja organização acompanha a forma como as informações são apresentadas pelo corpus. Para fins de análise, outras colunas são trabalhadas na planilha,

6. De modo a promover a verificabilidade e a falseabilidade dos dados, a planilha está disponível online no endereço [https://docs.google.com/spreadsheets/d/1V9X\\_VONyBdgnFQJ2dyr-TK2KZpjb7Ciez/edit?usp=sharing&ouid=111659250017905934562&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1V9X_VONyBdgnFQJ2dyr-TK2KZpjb7Ciez/edit?usp=sharing&ouid=111659250017905934562&rtpof=true&sd=true).

seja para fazer classificações e fórmulas, seja para desmembrar informações fornecidas pelo *Corpus do Português*, de modo a facilitar análises.

Um exemplo de informação desmembrada para facilitar análises é a separação da data de publicação e da variedade linguística em que a notícia foi escrita: enquanto o site *Corpus do Português* apresenta ambas informações em um mesmo espaço, para a organização da planilha as informações são separadas, com uma coluna sendo reservada para organizar as datas de publicação da notícia (no formato DD/MM/AAAA), outra coluna sendo reservada para organizar as informações sobre variedade linguística e uma terceira coluna, para colocar, em duplicidade, as informações sobre o ano de publicação (de modo a contabilizar a distribuição temporal dos resultados). A Figura 3, a seguir, ilustra a organização da planilha:

**Figura 3.** Organização dos dados em uma planilha eletrônica



| #  | Data e VarLing | Veículo  | Linha de Concordância                                                                                                                         | VarLing | Ano (BR) | Veículo  | Classifica | Justificat                | Subclass |
|----|----------------|----------|-----------------------------------------------------------------------------------------------------------------------------------------------|---------|----------|----------|------------|---------------------------|----------|
| 1  | 13-08-12 BR    | Folha da | #Ouvii a movimentação de pessoas por o local. Ele disse que chegou a alertar os dois funcionários que trabalhavam em o estabelecimento, mas   | -       | -        | -        |            | contexto não recu         |          |
| 2  | 15-01-09 BR    | O Tempo  | #Estava fazendo o discurso para anunciar o vencedor, ele disse que o ministério atuava em duas linhas para estimular a leitura: bibliotecas e | BR      | 2015     | O Tempo  | 3.         | Lado mais profissional de |          |
| 3  | 16-03-12 BR    | UOL      | #Sobre a papo que teve com Matheus: "Ele disse que vota em o Ronan e que acha que Renan vota em ele também                                    | BR      | 2016     | UOL      | 2.         | Ponto entreten 2. Tende   |          |
| 4  | 16-07-20 BR    | Globo.co | #Palácio de o Planalto, em a ocasião, ele disse que, embora a palavra final sobre a taxa de juros seja de o                                   | BR      | 2016     | Globo.co | 3.         | Lado mais profissional de |          |
| 5  | 18-01-18 BR    | Midiana  | #drogando como se nada tivesse acontecendo. O que ele disse faz sentido, eu posso te bater, mas depois que descobrem que você                 | BR      | 2018     | Midiana  | 2.         | Ponto violência 2. Tende  |          |
| 6  | 13-05-14 BR    | Globo.co | "", comentou. #a o G1, ele disse que sempre gostou de usar sapatos com salto alto que é dá uma sensação                                       | BR      | 2013     | Globo.co | 1.         | Lado n aparência          |          |
| 7  | 18-05-28 BR    | Globo.co | #já estão faltando em vários estabelecimentos. Entretanto, ele disse que os itens mais essenciais, como arroz e feijão, ainda podem durar     | BR      | 2018     | Globo.co | 3.         | Lado mais profissional de |          |
| 8  | 18-12-06 BR    | G1       | #aconteceu uma confusão. Ele não ouviu tiro e ele disse que não viu o Junior atirar ", diz Mariah. # No carro                                 | BR      | 2018     | G1       | 2.         | Ponto violência 2. Tende  |          |
| 9  | 17-11-11 BR    | Omelete  | #fazendo? Eu não quero isso'. E ele disse' Você precisa relaxar, só quero te deixar confortável! ". O                                         | BR      | 2017     | Omelete  | 2.         | Ponto violência 2. Tende  |          |
| 10 |                |          | #em pneumonia. Em conversa com os jornalistas, ele disse que Jaque sofreu                                                                     |         |          |          |            |                           |          |

Fonte: Captura de tela feita pelo autor do artigo

A Figura 3 ilustra a organização da planilha para a análise de dados. Na parte superior da planilha, é possível ver a seguinte distribuição de colunas:

- Coluna A – Indica o número de cada linha de concordância;
- Coluna B – Data de publicação e variedade linguística de cada linha;
- Coluna C – Fonte de dados de cada notícia, com link para acesso ao site;

- Coluna D – Linha de concordância (com o nódulo de busca destacado em negrito);
- Coluna E – Duplica a informação de da variação linguística, separando-a da data da notícia;
- Coluna F – Duplica a informação sobre o ano de publicação da notícia, separando-a das demais fornecidas pelo site;
- Coluna G – Duplica a informação sobre a fonte de dados de cada notícia, sem link de acesso, além de permitir a indicação de linhas a não serem analisadas em etapas posteriores da análise;
- Coluna H – Permite a classificação de dados, segundo a esfera da vida em que a representação se dá, havendo três possibilidades de classificação, a saber: ‘1. Lado mais pessoal do espectro’; ‘2. Ponto intermediário (pessoal x profissional)’ e ‘3. Lado mais profissional do espectro’
- Coluna I – Permite registrar interpretações (subjetivas) que levaram à classificação registrada na Coluna H;
- Coluna J – Permite a subclassificação dos casos que, na Coluna H, foram classificados como ‘2. Ponto intermediário (pessoal x profissional)’, havendo duas possibilidades de classificação, a saber ‘1. Tendendo para pessoal’ e ‘2. Tendendo para profissional’.

Como se pode observar a partir da lista acima, as colunas A, B, C e D apresentam as informações tal como elas são apresentadas pelo site *Corpus do Português*. As demais colunas são utilizadas para realizar os procedimentos de classificação de dados. A razão para utilizar as Colunas E, F e G para duplicar dados é permitir que os resultados gerados pelo Corpus do Português sejam inalterados, permitindo conferências futuras.

Outro detalhe que pode ser observado na parte superior da Figura 3 são pequenos triângulos vermelhos na parte superior direita de cada célula da linha de número 1 (um). Tais triângulos indicam a inclusão de um comentário na planilha eletrônica. Esses comentários são utilizados como uma forma simplificada de caderneta de categorias, explicando a organização de cada coluna, as possibilidades de valores inseridos e outras meta-informações voltadas para o esclarecimento da base de dados. A construção de cadernetas de categorias (ou ‘*code books*’, no termo em inglês) tende a ser uma boa prática

na organização de bancos de dados, por permitir a identificação dos dados, facilitando consultas e correções posteriores, como aponta Lobo (1993).

Na parte inferior da Figura 3, também é possível ver as diferentes abas da planilha, que organizam os dados referentes a cada um dos levantamentos ('Ele disse'; e 'Ela disse'), além de planilhas de consolidação de dados, que contabilizam os resultados da identificação de Dizentes, da variação linguística, da distribuição temporal e da distribuição de veículos de comunicação. Cada um desses resultados está apresentado a partir da seção 4 deste trabalho.

### *Definição de amostra, exclusão de linhas e classificação de linhas*

Em função do elevado número de resultados (mais de 51 mil linhas de concordância), delimita-se uma amostra para análise, utilizando-se a fórmula de Slovin. A escolha pela fórmula se justifica não apenas por sua simplicidade, mas por ser ela permitir a definição de uma amostragem quando não se tem informações sobre o comportamento da população. Para uma margem de erro inferior a 5%, a fórmula indica uma amostra de 389 linhas para o levantamento por 'Ela disse' e 396 linhas para o levantamento por 'Ele disse'. Considerando as delimitações e exclusões de linhas (a serem explicadas na sequência deste parágrafo), a opção, para esta pesquisa, é analisar as primeiras 500 linhas de concordância de cada um dos levantamentos — o que, mesmo com as delimitações, garante uma margem de erro inferior a 4,8% para a construção das duas amostras.

Quanto à delimitação de dados, mencionada no parágrafo anterior, a primeira iniciativa nesse sentido se dá pela variedade linguística. De modo a delimitar a análise ao contexto brasileiro, toma-se a decisão de pesquisa de considerar, para etapas posteriores da classificação, apenas dados referentes a variedade brasileira do português.

A segunda delimitação implica a não inclusão, em etapas posteriores da análise, de dados que, embora sejam condizentes com os nódulos utilizados, NÃO estão diretamente relacionados ao fenômeno investigado nesta pesquisa. Incluem-se aí, por exemplo, resultados que apontam para títulos de filmes e para representações do discurso em narrativas ficcionais, como mostrado nos três exemplos a seguir:

- Ex. 1. #se encontrar foi um pulo. # iG: “ Ela Disse, Ele Disse: O Namoro “  
resgata os conflitos entre meninos e  
Ex. 2. #Grupo Globo, que concluiu recentemente o longa “ Ela Disse Ele Disse  
“, vai jogar a campanha de lançamento em cima de a  
Ex. 3. #carta em a manga. No último episódio, **ele disse** para Sansa desconfiar  
de todos - se ele mantém essa filosofia, Sansa e

Nos Exemplos 1 e 2, acima, nota-se a referência ao título de filme *Ela Disse, Ele Disse*, de 2019, com direção de Claudia Castro e roteiro de Tati Ingrid Adão e Thalita Rebouças de acordo com o site IMDb<sup>7</sup>. A exclusão das linhas de concordância levantadas em função do título do filme não implica qualquer crítica ou demérito à obra. Trata-se apenas de um caso que não condiz com a proposta desta investigação. Já no Exemplo 3, é possível notar que se trata de uma linha que faz referência a um contexto ficcional (no caso específico, da série *Game of Thrones*, produzida e exibida pelo canal *HBO* entre os anos de 2011 e 2019, também de acordo com o site IMDb<sup>8</sup>). Como anteriormente dito, trata-se de dados não intrinsecamente ligados ao fenômeno aqui investigado e, portanto, não reveladores para uma análise sobre representações do discurso em narrativas factuais publicadas por veículos brasileiros de comunicação.

Além disso, há casos em que a exclusão das linhas se justifica pelo fato de o site que originalmente publica a notícia estar fora do ar, impedindo o acesso a um contexto mais amplo da notícia (algumas vezes necessário para a classificação dos dados). Os exemplos a seguir se encaixam nessa categoria:

- Ex. 4. #o motivo de elas não estarem se identificando “, **disse ele**. # A boliviana Zailin Miranda, de 27 anos, e seus  
Ex. 5. #porque viu a notícia em a televisão “, **disse ela**. # A gravação de tela que me pediram... obrigada a todas as mensagens

As linhas exibidas nos Exemplos 4 e 5, acima, são exemplos de casos excluídos em função de o contexto não ser recuperável (e/ou o link estar fora do ar). A linha exibida no Exemplo 4 aponta para o link <http://www.correio24horas.com.br/detalhe/categoria/noticia/sobe-para-quatro-numero-de-mortos-em-incendio-em-cortico-em-sao-paulo/>, ao passo que a linha

7. *Ela Disse, Ele Disse* (2019) – IMDb, disponível em <https://www.imdb.com/title/tt9660006/> último acesso em 11 de setembro de 2023.

8. *Game of Thrones* (série de TV 2011-2019) – IMDb, disponível em <https://www.imdb.com/title/tt0944947/> último acesso em 11 de setembro de 2023.

exibida no Exemplo 5 aponta para o link <https://manausalerta.com.br/apos-grave-acusacao-de-aliciar-menores-mc-mirella-se-pronuncia/>. A primeira, quando acessada em outubro de 2023, gera um erro 404, esclarecendo que a página pode ter sido movida ou não existe, ao passo que a segunda abre apenas a página inicial do portal *Manausalerta*, não tendo sido encontrada forma de acessar o texto em específico.

### *Classificação dos dados*

Como mencionado na seção 3.2 deste texto, as linhas de concordância são classificadas em função da esfera da vida (profissional/institucional vs pessoal/privada), de modo a entender os contextos em que as vozes de homens e de mulheres são acessadas por veículos brasileiros de comunicação e, eventualmente, interpretações e decisões de pesquisa são registradas nas colunas adequadas da planilha.

Essa classificação se mostra mais desafiadora do que inicialmente esperado. Uma das razões para isso diz respeito aos motivos que levam textos jornalísticos a acessar as vozes de outrem, como discutido anteriormente neste texto, fazendo referência a Machin (2008). Tais motivos podem variar — incluindo a simples abordagem de questões pessoais de um/a entrevistado, o recurso a experiências de uma pessoa para ilustrar casos/notícias de interesse público, ou mesmo para dar voz a autoridades —, podendo haver muitas zonas cinzentas entre o totalmente profissional/público e o totalmente pessoal/privado. Opta-se, nesta pesquisa, por construir uma escala, organizada da seguinte maneira:

- 1) Lado mais pessoal do espectro
- 2) Ponto intermediário (pessoal x profissional)
  - 2.1. Tendendo para pessoal
  - 2.2. Tendendo para profissional
- 3) Lado mais profissional do espectro

A escala acima não estabelece limites rígidos para as classificações e permite reconhecer as muitas zonas cinzentas possíveis entre os extremos da escala aqui trabalhada. Um ponto importante, para este artigo, é o reconhecimento de que a classificação é fundamentalmente baseada em uma percepção (individual e subjetiva) do pesquisador. Ao reconhecer, de antemão, a natureza epistemológica desse trabalho de classificação, busca-se



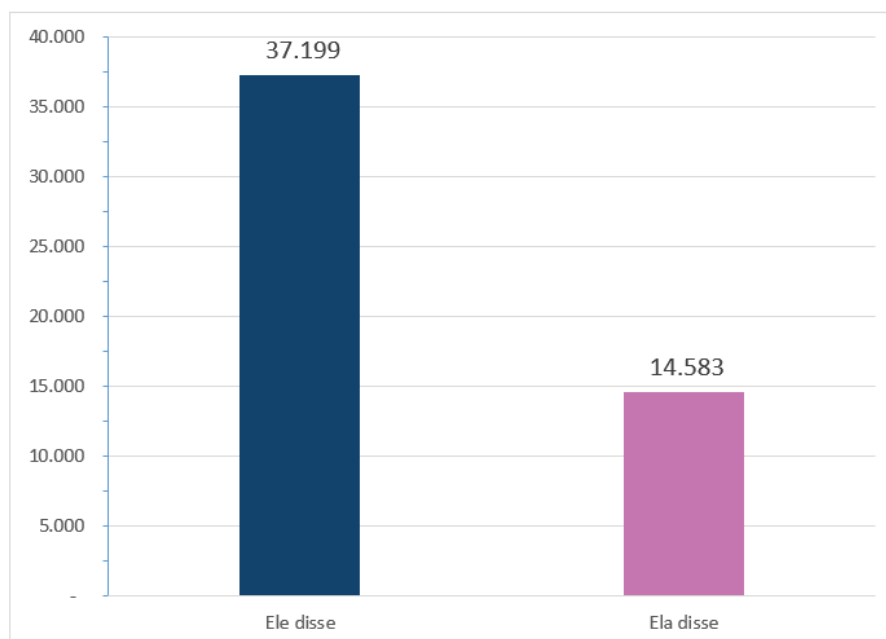
deixar explícito que este trabalho não pretende estabelecer limites rígidos para um objeto que se constrói em muitas zonas cinzentas, tampouco atingir um resultado inquestionável. O objetivo deste trabalho é contribuir para as reflexões e debates sobre assimetrias na sociedade, inclusive convidando outros pesquisadores e outras pesquisadoras a se debruçarem sobre estes mesmos dados, trazendo múltiplos pontos de vista e jogando mais luz à discussão.

A próxima seção traz os resultados (tanto quantitativos quanto qualitativos) da análise de dados, além de algumas possíveis leituras.

#### 4. Análise de dados e discussão

Seguindo os passos delineados na seção anterior, para esta pesquisa, levantam-se os casos de representação do discurso a partir de duas buscas, a saber: ‘ele disse’ e ‘ela disse’. A partir desses levantamentos, é possível observar que os Dizentes Masculinos são numericamente superiores aos Dizentes Femininos. O gráfico a seguir mostra os resultados (quantitativos) desse levantamento:

**Gráfico 3.** Levantamento inicial de representações do discurso em esferas profissionais



Fonte: elaboração própria



No Gráfico 3, acima, é possível ver que os resultados referentes aos Dizentes Masculinos são numericamente superiores aos resultados de Dizentes Femininos. O levantamento de casos de ‘ele disse’ gera um total de 37.199 resultados, 22.616 a mais do que os 14.583 gerados no levantamento por ‘ela disse’. Ainda que seja apenas um primeiro resultado — antes que sejam feitas maiores explorações — a diferença numérica já pode ser vista como um indício daquilo que Caldas-Coulthard (1994) aponta como sendo uma tendência de narrativas factuais em citar homens com mais frequência do que mulheres.

Como explicado no método de trabalho, desse total, nas etapas seguintes da pesquisa analisam-se as primeiras 500 linhas de cada levantamento — o que garante um erro inferior a 5% na construção das duas amostragens (mesmo com as delimitações e exclusões de dados). Essas 500 primeiras linhas de concordância são filtradas em função da variedade linguística, buscando, como anteriormente dito, selecionar para as etapas subsequentes da investigação apenas os dados na variedade brasileira da língua portuguesa. A Tabela 1, a seguir, apresenta os resultados referentes a esta etapa da exploração dos dados.

**Tabela 1.** Variedade Linguística das linhas de concordância da amostra analisada

| Variedade        | Ela disse  |               | Ele disse  |               |
|------------------|------------|---------------|------------|---------------|
| Angolano (AO)    | -          | 0,0%          | -          | 0,0%          |
| Brasileiro (BR)  | 446        | 89,2%         | 437        | 87,4%         |
| Europeu (PT)     | 26         | 5,2%          | 27         | 5,4%          |
| Moçambicano (MZ) | -          | 0,0%          | 2          | 0,4%          |
| Linhas excluídas | 28         | 5,6%          | 34         | 6,8%          |
| <b>TOTAL</b>     | <b>500</b> | <b>100,0%</b> | <b>500</b> | <b>100,0%</b> |

Fonte: elaboração própria

Os dados apresentados na Tabela 1, acima, mostram que a variedade brasileira da língua portuguesa tende a ser predominante nos resultados gerados pelo site *Corpus do Português*. Em todos os casos levantados, os resultados na variedade brasileira ficam próximos da marca de 90% dos dados. A decisão por limitar as etapas futuras da análise à variedade brasileira se justifica por uma tentativa de limitar o número de variáveis culturais, como exposto na seção 3.3 deste texto. Os resultados apresentados na Tabela 1 reforçam essa decisão, ao indicar que as demais variedades da língua portuguesa têm representatividade menor na construção do corpus.

Na sequência, os dados são analisados em termos de distribuição temporal. A Tabela 2, a seguir, organiza tais resultados:

**Tabela 2.** Distribuição temporal das linhas de concordância da amostra analisada

|              | Ela disse  |               | Ele disse  |               |
|--------------|------------|---------------|------------|---------------|
| Ano 2012     | 27         | 6,1%          | 18         | 4,1%          |
| Ano 2013     | 57         | 12,8%         | 53         | 12,1%         |
| Ano 2014     | 43         | 9,6%          | 30         | 6,9%          |
| Ano 2015     | 42         | 9,4%          | 35         | 8,0%          |
| Ano 2016     | 53         | 11,9%         | 57         | 13,0%         |
| Ano 2017     | 83         | 18,6%         | 93         | 21,3%         |
| Ano 2018     | 79         | 17,7%         | 88         | 20,1%         |
| Ano 2019     | 62         | 13,9%         | 63         | 14,4%         |
| <b>TOTAL</b> | <b>446</b> | <b>100,0%</b> | <b>437</b> | <b>100,0%</b> |

Fonte: elaboração própria

A Tabela 2, acima, apresenta a distribuição temporal das linhas de concordância da amostra (em português brasileiro) analisada nesta pesquisa. Os resultados apontam que o ano de 2012 é sistematicamente o ano com menor número percentual, o que pode ser um indício<sup>9</sup> de que a compilação do *Corpus do Português* não contempla aquele ano integralmente. Além disso, é possível ver na Tabela 2 que a maior parte dos dados, em termos percentuais, se concentra nos anos de 2017 e 2018. Mais do que a simples organização de dados ao longo dos anos, é possível ver, a partir dos dados da Tabela 2, uma distribuição ao longo do período estudado — algo que reforça a importância de projetos de compilação de grandes corpora (como é o caso do *Corpus do Português*).

Outro ponto a ser destacado na Tabela 2 é que os valores da linha que apresenta o total de dados analisados correspondem àqueles que, na Tabela 1, estão apresentados como sendo dados referentes ao Português Brasileiro. Esses números (446 para ‘Ela disse’ e 437 para ‘Ele disse’) poderão ser vistos ao longo dos próximos resultados e nas próximas tabelas.

9. Ao falar em indício, reforço que as análises aqui apresentadas são limitadas pelas permissões de acesso que tenho para o site *Corpus do Português*. Como usuário, sem permissão de administrador ou acesso aos registros de compilação do corpus, não é possível fazer análises mais categóricas sobre flutuações temporais na publicação de notícias, tampouco sobre eventuais variações na produtividade dos nódulos escolhidos para esta pesquisa.

Na sequência, a amostra de dados é analisada em função dos veículos de comunicação que os publicaram. A Tabela 3, a seguir, organiza tais resultados:

**Tabela 3.** Principais veículos de comunicação da amostra de linhas de concordância referente aos dados levantados por ‘Ela disse’ e ‘Ele disse’

| Ela disse                   |            |               | Ele disse                   |            |               |
|-----------------------------|------------|---------------|-----------------------------|------------|---------------|
| Fonte                       | Resultados | (%)           | Fonte                       | Resultados | (%)           |
| Globo.com                   | 123        | 27,6%         | Globo.com                   | 84         | 19,2%         |
| G1                          | 30         | 6,7%          | G1                          | 27         | 6,2%          |
| UOL                         | 14         | 3,1%          | UOL                         | 17         | 3,9%          |
| Purepeople.com.br           | 13         | 2,9%          | Folha de S.Paulo            | 17         | 3,9%          |
| Folha de S.Paulo            | 11         | 2,5%          | Estado de Minas             | 11         | 2,5%          |
| R7                          | 10         | 2,2%          | Jornal Extra                | 11         | 2,5%          |
| Terra Brasil                | 9          | 2,0%          | Jornal O Globo              | 11         | 2,5%          |
| Jornal Extra                | 8          | 1,8%          | Correio da Bahia            | 9          | 2,1%          |
| Correio da Bahia            | 8          | 1,8%          | EBC                         | 8          | 1,8%          |
| Jornal O Globo              | 7          | 1,6%          | Terra Brasil                |            | 1,8%          |
| Cidadeverde.com             | 6          | 1,3%          | R7                          | 8          | 1,8%          |
| Consultor Jurídico          | 6          | 1,3%          | Último Segundo - iG         | 7          | 1,6%          |
| Blasting News               | 6          | 1,3%          | O POVO Online               | 7          | 1,6%          |
| O Dia Online                | 5          | 1,1%          | EXAME.com                   | 7          | 1,6%          |
| Folhama                     | 5          | 1,1%          | Terra                       | 6          | 1,4%          |
| Outros (131 veículos)       | 185        | 41,5%         | Outros (116 veículos)       | 199        | 45,5%         |
| <b>TOTAL (146 veículos)</b> | <b>446</b> | <b>100,0%</b> | <b>TOTAL (131 veículos)</b> | <b>437</b> | <b>100,0%</b> |

Fonte: elaboração própria

Ao todo, 201 veículos de informação são identificados na composição do corpus (considerando a amostra analisada), sendo 146 veículos na consulta por ‘Ela disse’ e 131 na consulta por ‘Ele disse’ (com 76 veículos repetidos nas duas listas). Do total de veículos identificados, 44,2% dos dados estão associados a 184 veículos de comunicação, havendo 101 veículos que são fonte de dados para apenas UMA linha de concordância (considerando as duas planilhas). Como contraste, 16 veículos de comunicação respondem por mais de 55,8% dos dados — isso significa uma concentração dos dados a partir de um número reduzido de veículos.

Nesse grupo reduzido que concentra a maior parte dos dados, destacam-se os veículos ligados ao grupo *Globo* e ao *Grupo Folha* (como *globo.com*, *G1*, *UOL* e *Folha de S.Paulo*, por exemplo). Uma possível explicação para essa representatividade pode estar relacionada ao tamanho dos grupos e ao fato de que tais grupos têm forte presença online, publicando muitos de seus textos nesse meio. O grupo *Folha*, por exemplo, destaca, desde a campanha comemorativa dos 80 anos<sup>10</sup> de sua fundação, que não é um grupo exclusivamente voltado para a mídia impressa e que a presença online (por meio do portal *UOL*) é um de seus principais investimentos estratégicos. Já o grupo *Globo*<sup>11</sup> vem investindo, ao longo dos últimos anos, na proposta de se constituir como uma *media tech*, buscando desenvolver forte presença online e investir em (novos) produtos digitais.

Na sequência, procede-se à análise qualitativa dos dados, visando a analisar se existem indícios do que Caldas-Coulthard (1994) discute como marginalização das mulheres em termos de espaços profissionais / públicos / institucionais de fala, segundo a qual homens tendem a ser mais citados, de uma forma geral, e mulheres tendem a ser relegadas a assuntos relacionados a esferas privadas de suas vidas (como mães, filhas, mulheres, viúvas, etc.).

Como discutido na seção 3.4 deste texto, a classificação dos dados em relação à contraposição ‘público vs privado’ se mostra mais desafiadora do que inicialmente esperado, havendo desde casos em que a distinção é relativamente simples até casos em que a classificação é complexa, podendo até mesmo ser controversa.

Tratando primeiramente de casos de classificação relativamente simples, quando a voz é acessada para tratar estritamente de aspectos pessoais, familiares, afetivos, sexuais e outros, por exemplo, não há grandes dificuldades em justificar a classificação da linha de concordância como fundamentalmente relacionada a esferas da vida privada. Os exemplos a seguir mostram casos em que as vozes acessadas tratam de aspectos privados, relativos a esferas pessoais de suas vidas:

10. Fundado em 1921, o grupo completou 80 anos em 2001 e o texto *Jornal cresce e se torna grupo de mídia*, no qual o veículo discute sua presença na internet e novos produtos de informação está disponível em [https://www1.folha.uol.com.br/folha/80anos/grupo\\_folha.shtml](https://www1.folha.uol.com.br/folha/80anos/grupo_folha.shtml) (último acesso em 26/09/2023).

11. A discussão sobre o modelo de reorganização do grupo *Globo*, incluindo os esforços de unificação das companhias que compõem o grupo e a ênfase em investir em conteúdo e tecnologia está disponível em <https://valor.globo.com/empresas/noticia/2021/07/30/unificada-globo-estuda-novos-modelos-de-negocio-sembarreira.ghtml> (último acesso em 26/09/2023).

Ex. 6. #Questionada por QUEM se gostou de o novo visual, **ela disse** que sim: “Amei! Estou me acostumando ainda, mas gostei

Ex. 7. #aconteceu“, desabafou Loreto em o Instagram. **Ele disse** ainda que sente muito por ter “manchado a imagem“ de a ex-mulher

Os Exemplos 6 e 7, acima, ilustram casos em que os Dizentes dão declarações sobre aspectos fundamentalmente pessoais de suas vidas, incluindo comentários sobre aparência física, relacionamentos afetivos.

Por outro lado, mas ainda tratando de casos de classificação menos controversa, quando a voz é acessada para trazer a fala de um/a especialista, comentando sobre sua área de especialidade ou sobre questões função de um cargo público e outros aspectos claramente profissionais, tampouco há dificuldades em justificar a classificação como fundamentalmente relacionada a esferas profissionais da vida. Os exemplos a seguir trazem casos em que a representação do discurso pode ser associada a esferas profissionais dos/das Dizentes:

Ex. 8. #) com o ministro de a Cidadania Osmar Terra. **Ela disse** que conversou com o ministro sobre a possibilidade de intensificar a valorização de

Ex. 9. #a Comissão Mista de Orçamento (CMO), **ele disse** que a reformulação de expectativas diante de a demora em a aprovação de a

Os Exemplos 8 e 9, acima, ilustram casos em que os Dizentes dão declarações sobre aspectos de suas vidas profissionais. O Exemplo 8, publicado pelo jornal *Folha de São Paulo*<sup>12</sup>, dá voz à então vice-ministra italiana Lucia Borgonzoni em um contexto de diálogo sobre promoção cultural; o Exemplo 12, publicado pelo *D24am*<sup>13</sup>, dá voz ao então ministro Paulo Guedes, em uma situação de audiência na Comissão Mista de Orçamento (CMO).

No entanto, entre os dois extremos da contraposição ‘pessoal vs profissional’, há uma série de pontos intermediários, cuja classificação é complexa e pode fomentar debates epistemológicos. Dois exemplos, que efetivamente aparecem no corpus, são quando uma ministra de estado concede uma en-

12. Texto ‘Itália oferece peças ao Museu Nacional, mas diretor diz que a prioridade é ter verba’, publicado em 19/06/2018, em <https://www1.folha.uol.com.br/ciencia/2019/06/italia-anuncia-intencao-de-emprestar-pecas-e-know-how-para-recuperacao-do-museu-nacional.shtml> (último acesso em 26/09/2023).

13. Texto ‘Guedes diz que governo já trabalha com crescimento de 1,5% em 2019’, publicado em 14/05/2019, em <https://d24am.com/politica/guedes-diz-que-governo-ja-trabalha-com-crescimento-de-15-em-2019/> (último acesso em 26/09/2023).

trevista, em função de sua posição como ministra de estado, mas é citada em seu relato sobre um trauma de infância; ou quando uma pessoa que é convidada a comentar sobre uma doença de que sofre, no contexto de matéria que trata sobre tal tema como algo de interesse público. Ambos os casos podem ser interpretados como situações de relatos pessoais, mas colocados em contextos não privados, como discursos públicos em função da posição da entrevistada ou do interesse público que o texto jornalístico constrói em torno do tema. Os exemplos a seguir trazem alguns desses casos pessoais, apresentados em contexto de interesse público:

Ex. 10. #para me mostrar a cicatriz em seu seio direito. **Ela disse** que, depois de curada, teve três filhos, todos afillhados de

Ex. 11. #há 11 anos e pai de quatro filhos, **ele disse** que educa com bons exemplos e com respeito. “ Não é necessário agredir

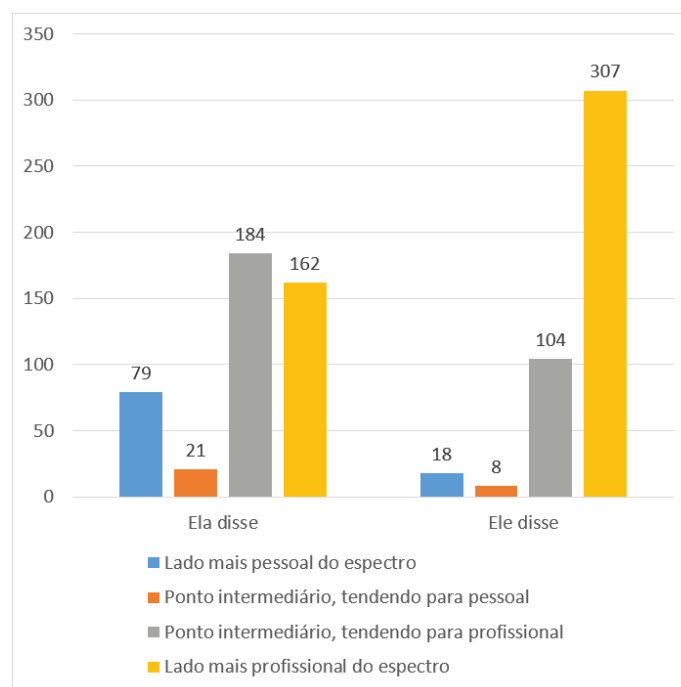
Os Exemplos 10 e 11, acima, ilustram casos em que a distinção ‘público vs privado’ é de classificação complexa. O Exemplo 10, publicado pela *Época*<sup>14</sup>, traz o depoimento de uma senhora (não nomeada no texto) para ilustrar uma discussão sobre rituais de cura pela fé, em um contexto de escândalo, por acusações de abuso; o Exemplo 11, publicado pelo *Correio do Estado*<sup>15</sup>, traz a fala de um policial (também não nomeado) que, após prender um pai por violência contra criança, dá seu depoimento como pai e sua visão como educador. Nos Exemplos 10 e 11, é possível notar a complexidade da distinção ‘público vs privado’, havendo elementos que apontam para aspectos da vida privada dos Dizentes (questões de fé, doença, abuso e depoimentos familiares) e elementos que apontam para aspectos públicos (seja pelo interesse público em torno de escândalos e casos de abuso, seja pela situação profissional do Dizente).

A escala trabalhada, descrita na seção 3.4, tenta abraçar essa diversidade, reconhecendo as diversas zonas cinzentas entre os dois extremos da contraposição ‘público vs privado’. Os resultados da classificação estão representados no Gráfico 4, a seguir:

14. Texto ‘Dez dias em Abadiânia, a cidade que vive em torno de João de Deus’, publicado em 08/12/2018, em <https://epoca.globo.com/dez-dias-em-abadiania-cidade-que-vive-em-torno-de-joao-de-deus-23286309> (último acesso em 26/09/2023).

15. Texto ‘Policial interferiu depois de ver pai agredir criança de 2 anos com tapa no rosto’, publicado em 03/10/2015, em <http://www.correiodoestado.com.br/cidades/campo-grande/policial-interferiu-depois-de-ver-pai-agredir-crianca-de-2-anos-com/259457/> (último acesso em 26/09/2023).

**Gráfico 4.** Levantamento inicial de representações do discurso na escala entre as esferas profissionais e privadas



Fonte: elaboração própria

A partir do Gráfico 4, acima, é possível ver o resultado da classificação das linhas de concordância em discursos profissionais / públicos. No gráfico, é possível notar que, em ambos os conjuntos de dados, os resultados referentes a vozes acessadas em esferas profissionais superam o de vozes acessadas em esferas privadas. Todavia, observando-se cada conjunto de dados, é possível notar indícios de uma associação entre discursos de homens e contextos profissionais / públicos e de uma associação de discursos de mulheres a contextos privados / pessoais.

Os dados referentes a Dizentes Masculinos ('ele disse') são mais marcadamente referentes a esferas profissionais: 307 linhas de concordância (ou 70,3% do total analisado) se encaixam no lado mais profissional do espectro, além de 104 linhas (ou 23,8%) que se encontram em ponto intermediário, mas tendem para o profissional. Como contraste, nos dados referentes a Dizentes Femininos ('ela disse') o número de linhas que se encaixa no lado mais profissional do espectro cai para 162 linhas (ou 36,3% do total

analisado) e o que se encontra em um ponto intermediário sobe para 184 linhas (ou 41,3%). Já os dados referentes a Dizentes Masculinos em esferas mais marcadamente pessoais ficam em 18 linhas (o equivalente a 4,1% da amostra), contra os dados referentes a Dizentes Femininos na mesma categoria chegando a 79 linhas (ou 17,7% da amostra).

Os dados da pesquisa parecem, portanto, corroborar as observações de Caldas-Coulthard (1994), indicando haver indícios de que textos jornalísticos brasileiros tendem utilizar a representação da fala para dar voz a homens e a mulheres de modo diferente: Além de homens tenderem a ser mais citados (quando se consideram números absolutos), mulheres tendem a ser mais citadas quando assuntos estão relacionados a esferas privadas. E, mesmo considerando as diferentes zonas cinzentas dentro dos contextos públicos / profissionais, os dados ainda parecem apontar para uma predileção por acessar vozes masculinas nos contextos que são mais marcadamente profissionais.

É possível, portanto, apontar que existem indícios de um sexismo sistêmico na forma como narrativas factuais são construídas — ou, utilizando os termos discutidos por Caldas-Coulthard (1994), de que existe uma marginalização das mulheres em termos de espaços públicos / institucionais de fala. Mesmo passados quase 30 anos da pesquisa da autora e mesmo considerando as diferenças culturais implícitas à construção dos corpora (da autora, de língua inglesa, e desta pesquisa, em português brasileiro).

## 5. Considerações finais

Este artigo se propôs a revisitar o trabalho de Caldas-Coulthard (1994) para discutir o fenômeno de marginalização das mulheres de espaços públicos de fala, utilizando dados levantados a partir do site *Corpus do Português*, para analisar a forma como homens e mulheres são citados em narrativas factuais. Tentando responder às perguntas de pesquisa, explicitadas no começo deste texto, é possível dizer que existem indícios de que os veículos brasileiros de comunicação representam as falas de homens e de mulheres de maneira diferente, com as citações de homens sendo mais frequentes do que as de mulheres, além de também haver indícios de que existem, no contexto contemporâneo, indícios de uma marginalização das mulheres em termos de espaços públicos / institucionais de fala.



Os dados analisados nesta pesquisa não são totalmente comparáveis com os de Caldas-Coulthard (1994) — por se tratar de dados de épocas diferentes, escritos em idiomas diferentes e publicados em países diferentes. Seria interessante se outras pesquisas realizassem uma análise diacrônica (tanto em português quanto em inglês), para ver se é possível identificar padrões de mudança social a partir das formas com as vozes de homens e de mulheres são acessadas em narrativas factuais. A realização dessa análise diacrônica poderia revelar mais aspectos históricos sobre a construção da fala feminina em narrativas factuais e, por extensão, na sociedade. Outra sugestão de pesquisa futura seria a realização de análises estratificadas por veículo de comunicação, ou levando em consideração os diferentes perfis de publicações — algo que não foi realizado nesta pesquisa.

Por fim, reitero aqui um ponto colocado na seção 3.4: os resultados aqui apresentados respondem, necessariamente, à percepção (individual e subjetiva) deste pesquisador sobre os dados. Trata-se de algo intrínseco a trabalhos de classificação, dada a sua natureza epistemológica. Esta pesquisa reconhece que outras pessoas podem ter perspectivas e compreensões diferentes sobre os mesmos dados e que não se pretende chegar, aqui, a resultados inquestionáveis, mas, sim, abrir um convite ao debate saudável sobre o tema — especialmente por ser algo que pode afetar uma parcela significativa da sociedade, seria interessante que outras pessoas se debruçassem sobre estes mesmos dados e trouxessem diferentes pontos de vista e dessem mais luz à discussão sobre as assimetrias sociais, sobre sexismo sistêmico e sobre a marginalização de mulheres na mídia.

### Conflito de Interesses

*Declaro não ter qualquer conflito de interesse, em potencial, neste estudo e assumo responsabilidade total pelo conteúdo do artigo.*

### Disponibilidade de dados

*Todo o conjunto de dados que dá suporte aos resultados deste estudo foi disponibilizado via Google Drive e pode ser acessado em [https://docs.google.com/spreadsheets/d/1V9X\\_VONyBdgnFQJ2dyrTK2KZpjb7Ciez/edit?usp=sharing&oid=111659250017905934562&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1V9X_VONyBdgnFQJ2dyrTK2KZpjb7Ciez/edit?usp=sharing&oid=111659250017905934562&rtpof=true&sd=true). O conjunto de dados também pode ser disponibilizado mediante solicitação ao autor.*

## Referências

- Alves, D., & Assis, R.C. (2016). Métodos de investigação em corpora: Ferramentas para classificação de dados extraídos de corpora de pequenas dimensões para análises discursivas, p. 1-17. In: *Anais do EBRALC 2015 & ELC 2015 [=Blucher Social Science Proceedings, n.3 v.2]*. São Paulo. [http://doi.org/10.5151/sosci-viiiiblc-xiii-elc-04\\_artigo\\_01](http://doi.org/10.5151/sosci-viiiiblc-xiii-elc-04_artigo_01)
- Alves, D., & Pagano, A. (2016). Com palavras minhas: A tradução de verbos de elocução neutros no corpus paralelo The Adventures Of Huckleberry Finn || As Aventuras de Huck. *Cadernos De Tradução*, 36(1), 34–61. <https://doi.org/10.5007/2175-7968.2016v36n1p34>.
- Assunção, C., & Araújo, C. (2019). Linguística de corpus: teoria, perspectivas metodológicas e ensino das línguas. *Filologia E Linguística Portuguesa*, 21(2), 271-288. <https://doi.org/10.11606/issn.2176-9419.v21i2p271-288>.
- Berber Sardinha, T. (2000). Linguística de Corpus: histórico e problemática. *DELTA: Documentação De Estudos Em Linguística Teórica E Aplicada*, 16(2), 323–367. <https://doi.org/10.1590/S0102-44502000000200005>
- Caldas-Coulthard, C. (1994). On Reporting Reporting: the representation of speech infactual and factional narratives. In M. Coulthard. *Advances in Written Text Analysis*. Routledge.
- Davies, M. (2016). *Corpus do Português*. Corpus do Português. Disponível em: <https://www.corpusdoportugues.org>. Acesso em 5 de maio de 2024.
- Fairclough, N. (1993). *Discourse and Social Change*. Polity Press.
- Fairclough, N. (1998). Discourse Representation in Media Discourse. *Sociolinguistics*, n. 17, 125-139.
- Halliday, M.A.K., & Matthiessen, C. (2014). *Halliday's Introduction to Functional Grammar*. Routledge.
- Lobo, M.C. (1993). Books-A Critical Link in the Research Process. *Western Journal of Nursing Research*, 15(3), 377-385. <https://doi.org/10.1177/019394599301500310>
- Machin, D. (2008). News Discourse I: Understanding the social goings-on behind news texts. In A. Mayr (ed.). *Language and Power: An Introduction to Institutional Discourse*. Continuum.
- McEnery, T., Hardie, A. (2012). *Corpus Linguistics Method, Theory and Practice*. Cambridge University Press.
- Semino, E., & Short, M. (2004). *Corpus Stylistics: Speech, writing and thought presentation in a corpus of English writing*. Routledge.

Recebido em: 07.11.2023

Aprovado em: 12.05.2024