

Linguística de Corpus

Perspectivas

Organizadoras:

Maria José Bocorny Finatto

Rozane Rodrigues Rebechi

Simone Sarmiento

Ana Eliza Pereira Bocorny



INSTITUTO
DE LETRAS
UFRGS



UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL



UNIVERSIDADE
FEDERAL DO RIO
GRANDE DO SUL

Reitor

Rui Vicente Oppermann

Vice-Reitora e Pró-Reitora
de Coordenação Acadêmica

Jane Fraga Tutikian



INSTITUTO
DE LETRAS
UFRGS

Universidade Federal
do Rio Grande do Sul
Instituto de Letras

Diretor

Sérgio de Moura Menuzzi

Vice-diretora

Beatriz Cerisara Gil

Linguística de Corpus

Perspectivas

Organizadoras:

Maria José Bocorny Finatto

Rozane Rodrigues Rebechi

Simone Sarmento

Ana Eliza Pereira Bocorny

© dos autores
1ª edição: 2018

Direitos reservados desta edição:



Esta licença permite que outros distribuam, remixem, adaptem e criem a partir dos trabalhos aqui publicados, mesmo para fins comerciais, desde que lhes atribuam o devido crédito pela criação original.

Capa: Ethel Kawa
Preparação de originais: Carlos Batanoli Hallberg
Revisão: Lia Cremonese
Editoração eletrônica: Fernando Piccinini Schmitt

Esta coletânea foi publicada graças ao apoio recebido da FAPERGS, processo 17/0399-3, Edital 06/2016 EDITAL FAPERGS 06/2016 – AOE, que apoiou o XVI Encontro de Linguística de *Corpus* (ELC 2017) e IX Escola Brasileira de Linguística Computacional (EBRALC 2017). Esta coletânea é um livro derivado do evento, reúne uma seleção de trabalhos gerados a partir de diferentes atividades de ambos os eventos. Todos os trabalhos aqui publicados foram avaliados por Comissão Científica especialmente convidada. Os Anais do evento correspondem a uma outra publicação denominada “Caderno de Resumos do ELC-EBRALC 2017”, ISBN: 9788561424183.

O direito autoral dos textos deste livro foi liberado por seus autores e organizadores, visto que é proibida a sua comercialização, sendo seu acesso livre e gratuito através do *site* do PPG-LETRAS-UFRGS, na guia E-BOOKS. A edição é do Instituto de Letras da UFRGS.

Versão DIGITAL gratuita disponível em:
PPG-LETRAS-UFRGS:
<https://www.ufrgs.br/ppgletras/ebooks.html>

Site do evento:
<http://www.ufrgs.br/elc-ebralc2017>



L755 Linguística de *corpus* : perspectivas [recurso eletrônico] / Organizadoras: Maria José Bocorny Finatto, Rozane Rodrigues Rebechi, Simone Sarmento, Ana Eliza Pereira Bocorny. — Porto Alegre: Instituto de Letras - UFRGS, 2018.
575 p.

Requisitos do sistema: Adobe Reader.
Modo de acesso: World Wide Web

1. Linguística. 2. Linguística de *corpus*. I. Finatto, Maria José Bocorny. II. Rebechi, Rozane Rodrigues. III. Sarmento, Simone. IV. Bocorny, Ana Eliza Pereira.

CDD 410

Catálogo na publicação: Vladimir Luciano Pinto – CRB 10/1112

ISBN 978-85-64522-36-7

O discurso dos deputados na votação do *impeachment*: a LC combinada à ACD

Deputies' speeches in the impeachment drive:
a CL approach to CDA

Rozane Rodrigues Rebechi

Resumo: Em abril de 2016, o Congresso Nacional votou a admissibilidade do processo de *impeachment* da então presidente Dilma Rousseff. Antes mesmo do término da sessão, a mídia e as redes sociais publicaram comentários sobre a escolha lexical dos congressistas ao justificarem seus votos, em geral associando palavras relacionadas a Deus, família e nação com os discursos pró-*impeachment*. Utilizando a metodologia da Linguística de *Corpus*, combinada à Análise Crítica do Discurso, este capítulo visa à análise das transcrições das falas dos deputados a fim de se confirmar ou não a impressão do público e da imprensa. Os resultados apontaram para uma escolha lexical estatisticamente coincidente das palavras mencionadas, especialmente nos dois grupos com maior número de votantes.

Palavras-chave: *Impeachment*. Discurso político. Linguística de *Corpus*. Análise Crítica do Discurso.

Rozane Rodrigues Rebechi – Professora adjunta na Universidade Federal do Rio Grande do Sul, doutora pela Universidade de São Paulo – rozanereb@gmail.com.

Abstract: In April 2016, Brazil's Lower House of Congress voted the impeachment drive of then president Dilma Rousseff. The session was not even over when internet users and mass media started commenting on the vocabulary the legislators used, in most cases associating words related to God, family and nation to the pro-impeachment speeches. By combining *Corpus Linguistics* and *Critical Discourse Analysis*, this chapter aims at analyzing the transcripts of the deputies' talks in order to confirm or not the impression of the general public and the media. Results pointed to a statistically similar lexical choice of the aforementioned words, especially in the two groups with the largest number of voters.

Keywords: Impeachment. Political Discourse. *Corpus Linguistics*. *Critical Discourse Analysis*.

1 Introdução

No dia 17 de abril de 2016, a Câmara dos Deputados votou a admissibilidade do processo de *impeachment* da então presidente do Brasil, Dilma Rousseff. Dos 513 representantes do Poder Legislativo, 367 votaram a favor do processo, 137 contra, sete se abstiveram e dois faltaram à sessão, resultando no prosseguimento do caso para o Senado. Além de escolher entre uma das três possibilidades de voto – sim, não e abstenção –, a grande maioria dos votantes optou por justificar suas escolhas, utilizando os dez segundos ao microfone a que tinham direito¹. Antes mesmo do término da votação, que durou mais de cinco horas, as redes sociais e a mídia passaram a postar comentários de historiadores, cientistas políticos, jornalistas e público em geral, que calcularam e discutiram as palavras mais recorrentes nessas falas.

Deus, palavras relacionadas à família, à nação, ao combate à corrupção, entre outras, também deram vazão a inúmeros *memes* que viralizaram nas redes sociais. A fim de legitimar a escolha do voto, diversos congressistas iniciaram as falas com o vocábulo “pelo(a)(s)” (contração da preposição “per” + artigo definido), seguido de justificativa, como em “Pela minha família”, “Por Deus”, “Pelo meu país” etc. Com o intuito de ridicularizar essa recorrência, os internautas criaram *memes* com o mesmo início, “defendendo” interesses diversos, o que resultou em apelos jocosos como “Pelo Wi-Fi grátis”, “Pelo emagrecimento fácil, sem dieta e academia”, “Pela volta da Caverna do Dragão” etc., sempre finalizados com “eu voto sim”.

A mídia também se pronunciou a respeito dos discursos dos deputados. O *site* UOL Notícias publicou uma matéria com o título “*Minha*”, “*meus*”, “*família*”...: *a lista das palavras mais citadas na sessão do impeachment*², em que é feita

¹ Muitos deputados extrapolaram esse tempo e continuavam se pronunciando mesmo mediante a tentativa de interrupção pelo presidente da sessão.

² Disponível em: <https://noticias.uol.com.br/politica/ultimas-noticias/2016/04/20/familia-e-democracia-sao-citadas-mais-de-100-vezes-por-deputados-veja-outras.htm>. Acesso em: 10 out. 2017.

uma análise dos discursos dos parlamentares, elencando os vocábulos mais citados. Os trechos reproduzidos a fim de ilustrar a recorrência de palavras relacionadas à família e a Deus foram extraídos dos discursos pró-*impeachment*, sendo que apenas um excerto transcreve uma fala de votante contrário ao processo, na qual critica a menção a Deus por deputados que votaram a favor da continuidade do processo.

Outras matérias foram ainda mais incisivas. Em artigo intitulado *O que os discursos dos deputados pró-impeachment revelam sobre a construção da nossa democracia*, publicado em *Carta Capital*³, o historiador Luan Aiuá denomina “*show de horrores*” os discursos favoráveis ao processo, chamando especial atenção à “*triade família, Deus e nação*”, que teria permeado as falas desses “*conservadores*”. Já *Congresso em Foco* publicou matéria com a chamada *Deputados citaram “Deus” 59 vezes na votação do impeachment*^{4,5}. O artigo reproduz uma passagem do *blog* do teólogo Leonardo Boff, na qual afirma: “*Dezenas de parlamentares da bancada evangélica fizeram claramente discursos de tom religioso e invocando o nome de Deus. E todos, sem exceção, votaram pelo impedimento*”. Com a manchete *Deus derruba a presidenta do Brasil*, seguida do *lead* *Deputados justificam seus votos em Deus, na moralidade e a família: o motivo real da votação é esquecido*, a matéria publicada no jornal *on-line El País*, assinada por María Martín, compara as justificativas dos votos pelo *impeachment* a um programa de auditório⁶.

Diante de afirmações tão contundentes em relação à escolha lexical dos deputados durante a votação, surgiu o interesse em realizar uma pesquisa linguística mais aprofundada, não respaldada apenas pela frequência de palavras, com o intuito de confirmar – ou não – a representação dessas falas nas publicações em redes sociais e na mídia.

Por meio da combinação entre a metodologia subjacente à Linguística de *Corpus* (LC) e à Análise Crítica do Discurso (ACD), este estudo visa à análise das falas dos deputados na votação do processo de *impeachment*, a partir do levantamento quantitativo de palavras-chave e colocados, possibilitado por ferramentas computacionais.

³ Disponível em: <http://justificando.cartacapital.com.br/2016/04/19/o-que-os-discursos-dos-deputados-pro-impeachment-revelam-sobre-a-construcao-de-nossa-democracia/>. Acesso em: 10 out. 2017.

⁴ Disponível em: <http://congressoemfoco.uol.com.br/noticias/deputados-citaram-%E2%80%9Cdeus-%E2%80%9D-59-vezes-na-votacao-do-impeachment/>. Acesso em: 10 out. 2017.

⁵ Segundo lista de palavras do *corpus*, a palavra “Deus” ocorre 56 vezes.

⁶ Disponível em: https://brasil.elpais.com/brasil/2016/04/18/politica/1460935957_433496.html. Acesso em: 17 out. 2017.

2 Linguística de *Corpus* aplicada à análise de discurso

Antes de tudo, faz-se relevante explicitar o uso da palavra “discurso” neste estudo. Se, em inglês, a polissemia de *discourse* pode ser problemática, uma vez que pode suscitar diferentes acepções – tipos de linguagens (por exemplo, discurso midiático, político etc.), linguagem em uso, unidade de língua além da sentença, entre outras (ver BAKER, 2006, para uma análise dos conceitos subjacentes a essa palavra) –, em português o problema se agrava, pois a palavra pode ainda se referir à exposição oral feita em público, ou seja, poderia ser usada para se referir às falas dos deputados na sessão da câmara. Contudo, a fim de evitar mal-entendidos, neste estudo, como em Baker (2006, p. 2), procuramos restringir “discurso” à acepção Foucaultiana, qual seja, a de um sistema de sentenças utilizadas na construção de um objeto.

Na presente pesquisa, dois tipos de discursos são analisados: (i) o discurso da mídia, que relata as falas dos deputados durante sessão de votação do *impeachment*, e (ii) as falas propriamente ditas. Em seguida, apresenta-se uma breve explanação dessas duas formas de discurso, conforme entendidas neste trabalho.

2.1 Análise Crítica do Discurso e mídia

Para Fairclough (1989), a linguagem deve ser entendida como meio de ação, uma vez que, quer seja falada ou escrita, constitui atos de fala, como promessa, declaração, advertência etc., e envolve relações de poder, nem sempre explícitas. Sobre o poder no discurso da mídia, afirma:

O discurso dos meios de comunicação de massa [televisão, rádio, cinema, jornais] é interessante porque a natureza das relações de poder que ele estabelece muitas vezes não é clara, e há razões para entendê-lo como envolvendo relações *escusas* de poder⁷ (FAIRCLOUGH, 1989, p. 49, grifo do autor).

Enfatiza, ainda, que esse poder é construído a partir de sistematizações, ou seja, de repetições de informação nas atividades midiáticas:

Um único texto por si só é bastante insignificante: os efeitos do poder das mídias são cumulativos, funcionando por meio da repetição de formas específicas de se lidar com causalidade e agência, formas específicas de posicionamento do leitor, e assim por diante⁸ (FAIRCLOUGH, 1989, p. 54).

⁷ “[...] *mass media discourse* [television, radio, film, newspapers] is interesting because the nature of the power relations enacted in it is often not clear, and there are reasons for seeing it as involving hidden relations of power.”

⁸ “A single text on its own is quite insignificant: the effects of media power are cumulative, working through the repetition of particular ways of handling causality and agency, particular ways of positioning the reader, and so forth.”

Em análises de discurso realizadas em jornais para a identificação de ideologias implícitas, Fairclough (1995) defende o conceito de “representação de discurso”, uma vez que, em geral, nas publicações não se reporta de forma transparente o que foi falado ou escrito; o que se observa é uma tomada de decisão a partir de uma interpretação, e posterior representação, da informação que se deseja transmitir. Distingue, portanto, o discurso primário [*primary discourse*], ou seja, o relato propriamente dito, e o discurso secundário [*secondary discourse*], ou a representação do discurso, que é permeada por interpretação. Nesse sentido, Fairclough (1995) chama a atenção para dois “mitos”: o primeiro seria a crença de que a mídia é um “espelho” da realidade, e o outro, que a própria realidade seria “transparente”, de forma que poderia ser “lida” sem mediação ou interpretação. Sendo assim, entende a representação do discurso na mídia como um processo de grande importância social.

Apesar de o público não ser passivo, uma vez que o significado é criado a partir da interação entre o texto e seus leitores/ouvintes, o jornalismo os influencia com a produção de novos discursos ou com a reformulação de outros, já existentes (BAKER, 2006, p. 72). Assim, é umas das funções da ACD evidenciar (i) de que forma o grupo que detém o poder controla o discurso e (ii) como esse discurso influencia os grupos menos poderosos (VAN DIJK, 2001, apud BAKER 2006, p. 73-74). Nesta pesquisa, o discurso primário é constituído pelas falas transcritas dos deputados, e o secundário, pela representação dessas falas pela mídia, esta que, nessa relação, seria a detentora do poder, e, portanto, capaz de influenciar o leitor.

2.2 Discurso político

Para Chilton (2004, p. 3), “política” pode ser definida como “[...] luta pelo poder entre aqueles que buscam reivindicá-lo e mantê-lo e aqueles que resistem a ele”⁹ e “[...] cooperação, como as práticas e instituições das quais a sociedade se utiliza para resolver conflitos de interesse em relação a questões financeiras, influência, liberdade etc.”¹⁰. Linguagem e política estão intimamente ligadas, pois é por meio da linguagem que os atores políticos produzem efeitos. Durante a comunicação, os interlocutores sempre supõem receber tanto informações verdadeiras quanto falsas, e é justamente a expectativa do recebimento de informações verdadeiras que possibilita que o agente do discurso engane ou distorça a verdade.

A partir da análise de entrevistas com políticos, além de discursos políticos diversos, Chilton (2004) conclui que, para imprimir veracidade aos seus

⁹ “[...] *a struggle for power, between those who seek to assert and maintain their power and those who seek to resist it.*”

¹⁰ “[...] *cooperation, as the practices and institutions that a society has for resolving clashes of interest over money, influence, liberty, and the like.*”

enunciados, os emissores fazem uso de “evidências” como forma de legitimar seu discurso. Citando Fetzer (2002), conclui: “[...] do ponto de vista político, o que interessa é se o falante tem ‘credibilidade’”¹¹ (CHILTON, 2004, p. 32).

Várias estratégias linguísticas são utilizadas pelos políticos com a finalidade de se “aproximar” do interlocutor e convencê-lo da “veracidade” de suas afirmações. Apelar para o patriotismo, para a causa dos menos favorecidos e para a união são algumas delas. Também usual é o político usar repetidamente o pronome em primeira pessoa do plural – “nós” –, assim como o pronome adjetivo possessivo relacionado a ele – “nosso(a)”. Entre o uso estratégico da língua pelos políticos, Chilton (2004) identifica, ainda, duas vertentes: legitimação e deslegitimação. A primeira envolve pressupostos sobre a vontade dos eleitores, princípios ideológicos em geral, atitudes carismáticas e autorrepresentação positiva. Já a segunda está relacionada à representação dos opositores, em geral retratados com características negativas, por meio de acusações e ofensas.

O controle da informação pressupõe o controle do discurso, este que pode ser qualitativo ou quantitativo. O controle qualitativo pode ser realizado por meio do uso de informações falsas, omissões e negações. Já o quantitativo costuma se favorecer das generalizações. Tais estratégias puderam ser observadas nas falas dos deputados aqui analisadas, conforme veremos adiante.

2.3 LC aplicada à ACD

A ACD pode ser entendida como “um movimento acadêmico, uma forma de se fazer análise do discurso a partir de uma perspectiva crítica, em geral focada em conceitos teóricos, tais como poder, ideologia e dominação” (BAKER et al., 2008, p. 273)¹². A área sofre críticas por dois aspectos. Em primeiro lugar, devido à sua suposta fraqueza metodológica, uma vez que, tradicionalmente, se ocupa da análise apenas de fragmentos de textos. Em segundo lugar, devido à subjetividade de seus resultados, que podem decorrer de ideias preconcebidas do analista (cf. CHENG, 2013). A LC, área de estudo que enfoca um conjunto de métodos para o estudo da língua em uso (McENERY; HARDIE, 2012, p. 1), também é criticada. Muitos acreditam que se resume a análises puramente quantitativas, pautadas unicamente por dados estatísticos. De fato, as duas áreas em questão podem apresentar lacunas, que não nos cabe tratar aqui. Acreditamos, contudo, que, ao ajudar na identificação de amostras extraídas semiautomaticamente, a partir de textos autênticos, uma metodologia subjacente à investigação da língua em uso possibilita uma análise mais objetiva do que aquela realizada a partir de

¹¹ “[...] *what matters, from a political point of view, is whether the speaker has ‘credibility’*”

¹² “[...] *an academic movement, a way of doing discourse analysis from a critical perspective, which often focuses on theoretical concepts such as power, ideology and domination.*”

excertos selecionados aleatoriamente, ou, ainda mais preocupante, identificados propositalmente, a fim de se “confirmarem” hipóteses prévias.

Apesar de pesquisas combinando a metodologia da LC aplicada à ACD não serem novidades, essa associação ainda não resulta em grandes números de pesquisas, ao menos se compararmos com a contribuição que a LC desempenha em áreas como lexicografia, descrição gramatical e registro (cf. PARTINGTON, 2004), tradução (BAKER, 1993; ZANETTIN, 2012), terminologia (PEARSON, 1998) e ensino de línguas (CARTER et al., 2007), para citar apenas algumas. Biber et al. (1998, p. 106) também observam que “[...] embora quase todos os estudos em discurso se baseiem na análise de textos reais, não se trata particularmente de investigações baseadas em *corpus* [...]”¹³, pois não usam métodos quantitativos como ponto de partida. Para Sanderson (2008), a escassez de pesquisas que aliam LC e Discurso tem explicação:

A combinação da Linguística de *Corpus* com a metodologia analítica do Discurso é incomum. No passado, os linguistas de *corpus* não se interessaram particularmente pelo Discurso, preferindo concentrar-se na análise lexical e morfosintática. Da mesma forma, os analistas do discurso raramente se utilizavam de *corpora*, preferindo métodos como a introspecção, a elicitación e a coleta não sistemática de evidências anedóticas. (SANDERSON, 2008, p. 59)¹⁴

O pesquisador conclui, portanto, que se trata de via de mão dupla: se, de um lado, o Discurso é preterido na LC, de outro, os estudos em discurso não costumam adotar a LC como metodologia. Tal associação, contudo, tem se mostrado produtiva. Nesse sentido, merecem destaque pesquisas como as de Partington (2004), Partington et al. (2004), Baker (2006, 2012), Baker et al. (2008), entre outras.

Para Baker e McEnery (2005), a análise semiautomática de *corpora* pode desempenhar papel importante, uma vez que leva os pesquisadores a identificar padrões em textos autênticos com maior objetividade e ajuda a enfatizar padrões de associação – colocações – que, em geral, extrapolam a capacidade interpretativa realizada por meio de leitura sequencial de textos. Neste estudo, buscamos o suporte da Linguística de *Corpus* para evidenciar padrões linguísticos nas falas dos deputados e sua representação na construção de discursos da mídia em massa e dos usuários de redes sociais.

¹³ “[...] although nearly all discourse studies are based on analysis of factual texts, they are not typically corpus-based investigations [...]”

¹⁴ “The combination of corpus linguistic with discourse analytic methodology is unusual. In the past, corpus linguists have not been particularly interested in discourse, preferring to concentrate on lexical and morphosyntactic analysis. Similarly, discourse analysts have seldom worked with corpora, preferring methods such as introspection, elicitation and the unsystematic collection of anecdotal evidence.”

3 Metodologia

A fim de analisar as falas dos deputados no processo de *impeachment*, realizamos um levantamento semiautomático, possibilitado por ferramentas computacionais. Nesse levantamento, utilizamos, especialmente, listas de palavras-chave, colocados e linhas de concordância. Em seguida, detalharemos a compilação do *corpus* e as análises quantitativas e qualitativas realizadas.

3.1 O corpus de estudo

A análise de *corpora* orais, produzidos em contextos naturais, é escassa, se comparada à de textos escritos. Uma metodologia baseada em LC pressupõe que o *corpus* esteja em formato eletrônico, para ser, a princípio, analisado quantitativamente, por meio de ferramenta computacional. Para tanto, textos orais devem ser transcritos, e esse processo demanda tempo e, muitas vezes, investimento financeiro (cf. PARTINGTON, 2004; BAKER, 2006). A depender da pesquisa, a transcrição da fala deve considerar as características interacionais, como entonação, mudança de turno, pausa etc., como se observa em Prado (2015), que analisa interações entre pilotos e controladores de tráfego aéreo em comunicações radiofônicas realizadas durante situações anormais. Nesse caso, a pesquisadora utilizou codificações para marcar quebras prosódicas, hesitações e outras características da interação oral, processo extremamente trabalhoso – para transcrever um minuto de áudio foram necessários de trinta a quarenta minutos. Neste estudo, que visa a analisar a escolha lexical dos deputados durante seus votos no processo de *impeachment*, tal marcação não se mostrou necessária. Convém mencionar, contudo, que alguns deputados leram suas falas. Nesse caso, é esperado que o uso de marcadores discursivos, por exemplo, seja menor do que em textos falados espontaneamente (BAKER, 2006).

Desde 2007 a Coordenação de Histórico de Debates do Departamento de Taquigrafia da Câmara dos Deputados gerencia a *webpage Escrevendo a História*, que torna acessíveis “discursos memoráveis” proferidos desde 1946 em sessões no Congresso e na Câmara. Esse material pode ser disponibilizado em formato *.pdf no Diário da Câmara ou mesmo reproduzido em áudio¹⁵. Em geral, a íntegra das sessões é liberada algumas horas depois de seu encerramento, após a revisão e redação final das notas taquigráficas. Vale ressaltar que o método de transcrição realizado pela Câmara pode ser denominado “idealizado”, uma vez que reproduz a língua escrita padrão. Também desconsidera hesitações, pausas preenchidas, repetições, entre outros marcadores próprios da língua oral. Contudo, como

¹⁵ Informações obtidas do *site*: <<http://www2.camara.leg.br/atividade-legislativa/plenario/discursos/escrevendohistoria>>. Acesso em: 10 out. 2017.

já mencionado, o interesse desta pesquisa recai na escolha lexical. Sendo assim, marcadores de oralidade não se mostraram essenciais aqui.

Nesta pesquisa, utilizamos a transcrição das falas dos deputados durante a sessão que votou o processo de *impeachment* de Dilma Rousseff, conforme disponibilizada pela Câmara em formato de planilha do Google¹⁶, formada por seis colunas: (i) nome do(a) deputado(a), (ii) partido, (iii) Estado, (iv) voto, (v) gênero e (vi) transcrição da fala de cada um. Consideramos para a análise os textos da sexta coluna, ao passo que o conteúdo das outras cinco foi mantido no cabeçalho, conforme ilustrado a seguir (Quadro 1). Os textos, salvos em formato *.txt, a fim de serem inicialmente processados por meio da ferramenta Word Smith 6 (WS) (SCOTT, 2012), foram subdivididos em três *subcorpora*, de acordo com a modalidade do voto – sim, não e abstenção. Vale ressaltar que as informações guardadas no cabeçalho (*header*) possibilitam outras pesquisas, por exemplo, sobre as palavras significativamente mais recorrentes nas falas de representantes de cada partido, se houve diferença entre as falas de homens e mulheres etc. O Quadro 1 exemplifica o modelo de cabeçalho utilizado:

Quadro 1 – Modelo de cabeçalho

```
<header>
<evento> Sessão de votação para aprovação do processo de impedimento da câmara dos de-
putados </evento>
<data> 17 de abril de 2016 </data>
<deputado> Abel Mesquita Jr </deputado>
<partido> DEM </partido>
<estado> RR </estado>
<voto> SIM </voto>
<gênero> M </gênero>
</header>

<transcrição> Roraima, verás que o filho teu não foge à luta! O povo brasileiro merece respeito!
Por um Brasil com justiça, igualdade social e sem corrupção, por uma Roraima desacorrentada,
para que possamos exercer o direito constitucional de ir e vir e por todas as famílias roraimen-
ses, eu voto sim, Sr. Presidente.
</transcrição>
```

Fonte: Elaborado pela autora

¹⁶ A íntegra das falas dos deputados utilizadas nesta pesquisa pode ser obtida em formato de Planilha Google em: <<https://docs.google.com/spreadsheets/d/1rTFC9kbPqeHWrr2JalcvehgRXdAu03BwyBTgTL6oFC8/edit#gid=177016461>>. Acesso em: 28 nov. 2017.

A Tabela 1 resume o conteúdo do *corpus* de estudo:

Tabela 1 – *Corpus* de estudo

	Nº textos	Nº palavras (<i>tokens</i>)
SIM	367	19.249
NÃO	137	7.836
ABSTENÇÃO	7	299
Total	511	27.384

Fonte: Elaborado pela autora

O número de votos, e, conseqüentemente, o número de palavras é bastante discrepante entre as modalidades de voto, o que pode ficar ainda mais claro por meio do Gráfico 1:

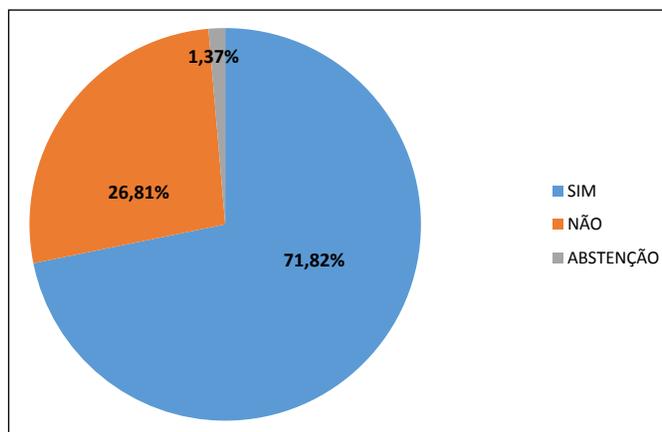


Gráfico 1 – Proporção entre os votos nas três modalidades

Fonte: Elaborado pela autora

Portanto, uma análise quantitativa baseada simplesmente em recorrência de itens lexicais, como aquelas publicadas na mídia e citadas anteriormente, não se mostraria adequada para o estudo das falas dos deputados. Assim, procedemos com o levantamento das palavras-chave, ou seja, palavras significativamente mais recorrentes no *corpus* de estudo, quando comparadas a um *corpus* de referência. Neste estudo, utilizamos como *corpus* de referência textos da língua geral¹⁷. Por meio da função *KW Database*, do WS, foi possível identificar as palavras-chave

¹⁷ O *corpus* de referência utilizado é uma parte (aproximadamente 2 milhões de palavras) do *Lácio-Ref corpus* que faz parte do projeto *Lácio-Web* (disponível em: <<http://www.nilc.icmc.usp.br/nilc/projects/lacio-web.htm>>).

recorrentes nas três modalidades de voto (coluna *Texts*) – neste caso, 100% –, a que denominamos palavras-chave-chave (coluna *KW*). A coluna *Overall Freq.* apresenta o número total de ocorrências da palavra no *corpus* de estudo, conforme apresentado na Tabela 2 em ordem decrescente de frequência:

Tabela 2 – Palavras-chave-chave nos três *subcorpora*

N	KW	Texts	%	Overall Freq.	No. Ass.
1	voto	3	100	543	129
2	Presidente	3	100	494	129
3	eu	3	100	441	129
4	meu	3	100	435	129
5	Sr	3	100	384	129
6	Brasil	3	100	320	129
7	minha	3	100	312	129
8	contra	3	100	157	129
9	país	3	100	151	129
10	<i>impeachment</i>	3	100	131	129
11	respeito	3	100	109	129
12	todos	3	100	91	129
13	me	3	100	73	129
14	corrupção	3	100	66	129
15	partido	3	100	62	129
16	favor	3	100	38	129
17	porque	3	100	36	129
18	votar	3	100	36	129
19	vou	3	100	27	129
20	Bahia	3	100	24	129

Fonte: Elaborado pela autora

É preciso enfatizar que o *corpus* de estudo não recebeu etiquetagem morfosintática. Por se tratar de *corpus* de pequenas proporções, palavras grafadas de forma idêntica, porém pertencentes a diferentes categorias gramaticais, por exemplo, puderam ser identificadas manualmente. É o caso de “voto”, palavra-chave-chave com maior número de ocorrências. A visualização das linhas de concordância, possibilitada pela função Concord, do WS, evidenciou que a palavra é utilizada tanto como substantivo – como em “meu voto é” – quanto como conjugação do verbo “votar” na primeira pessoa do presente do indicativo – “[eu] voto sim/não”, conforme ilustrado na Figura 1:

N	Concordance
1	e pelo meu País. Que Deus nos abençoe! Voto sim ao impeachment! Janeiro, da minha
2	foi denunciada, o que será confirmado adiante, voto não, pelo não prosseguimento da
3	— essas velhas raposas que estão aí. Voto sim ao impeachment. da política do
4	que estamos movendo, pelo meu Amazonas, voto sim. M
5	panfleto. Em segundo lugar, em respeito ao voto popular, em respeito à democracia, eu
6	o povo, para as instituições. Em respeito ao voto popular, em respeito à segurança das
7	em defesa da democracia e do respeito ao voto do cidadão brasileiro, eu voto com toda
8	Brasil, pela democraciae pelo respeito ao voto soberano do povo brasileiro, que elegeu
9	Brasil, pela democraciae pelo respeito ao voto soberano do povo brasileiro, que elegeu
10	Padre João PT MG NÃO M Pelo respeito ao voto popular, pela Presidenta Dilma, que não
11	em defesa da democracia e do respeito ao voto do cidadão brasileiro, eu voto com toda
12	pares, voto com o Relator Jovair Arantes. Voto pela reconstrução do Brasil. Voto sim!
13	e à justiça. Eu voto pelo povo baiano, voto pela minha mulher Maria Luísa, pelos
14	Estado e por uma esperança para o Brasil, voto sim. SIM M
15	região, ao Estado de São Paulo e ao Brasil. Voto sim, Sr. Presidente! eu quero, em

Figura 1 – Visualização parcial das linhas de concordância da palavra de busca “voto”
 Fonte: Elaborado pela autora

Também recorrente nas três modalidades de voto é o vocativo “Sr. Presidente”, comumente utilizado pelos parlamentares ao se dirigirem ao representante da Câmara dos Deputados¹⁸. Repetem-se, também, entre as palavras-chave “Brasil”, “Bahia”, “país”, “partido” e “todos”, resultando em frases como “em/a favor do Brasil”, “pelo meu Brasil”, “pela minha Bahia”, “pelo meu País”, “por todos os brasileiros” e “em respeito ao meu partido”, utilizados para justificar as escolhas. Também não faltaram justificativas em nome de familiares, como em “meu pai, que me ensinou”, “em nome da minha família”. Votos “contra/em combate à/pelo fim da corrupção” também foram frequentes. A conjunção explicativa “porque” foi bastante utilizada para explicar o voto: “porque é constitucional/incompetente/necessário”.

Devido ao tamanho reduzido do *subcorpus* “ABSTENÇÃO”, o levantamento das palavras-chave-chave, ou seja, aquelas recorrentes nos três *subcorpora*, apresentou poucos resultados. Além disso, consideramos importante também analisar o entorno dessas palavras, ou seja, seus colocados. E um levantamento dos colocados das palavras estatisticamente recorrentes no *subcorpus* “ABSTENÇÃO” retornou um número insignificante de palavras lexicais. Além de palavras gramaticais – preposições, artigos e pronomes –, a única palavra de conteúdo que aparece

¹⁸ Na época da votação do processo de *impeachment* de Dilma Rousseff, o presidente da Câmara era o então deputado Eduardo Cunha.

como colocado (da palavra-chave-chave “me”) é “abstenho”, como na sentença “eu me abstenho [de votar]”. Portanto, decidimos dar continuidade na análise considerando somente os *subcorpora* “SIM” e “NÃO”.

3.2 Palavras-chave

A fim de revelar as diferenças e semelhanças nos *subcorpora* “SIM” e “NÃO”, partimos para o levantamento das palavras-chave exclusivas de cada um deles – apresentados nas duas primeiras colunas da Tabela 3, a seguir –, assim como das palavras-chave recorrentes nos dois *subcorpora* – apresentadas na terceira coluna. Entre parênteses, encontra-se o número de ocorrências de cada palavra. Na coluna “SIM” e “NÃO”, os números em parênteses se referem, respectivamente, a cada modalidade de voto:

Tabela 3 – Palavras-chave nos *subcorpora* “SIM” e “NÃO”

SIM	NÃO	SIM e NÃO
abeneçoe (11), abraço (7), acima (10), agora (13), agricultores (7), Amazonas (9), amigos (14), amor (13), Arantes (7), Campos (9), Catarina (12), certeza (8), colegas (7), crescimento (10), dar (11), desempregados (10), dias (13), do (524), economia (9), eleitores (26), em (289), emprego (12), especial (15), especialmente (7), Espírito (9), esposa (18), fazer (20), fez (9), filha(s) (19), fim (20), fui (9), futuro (44), gerações (8), Goiás (12), gostaria (8), Grande (33), Grosso (10), honra (11), Janeiro (22), João (9), Jovair (7), jovens (8), mãe (13), Mato (10), me (59), melhor(es) (35), memória (10), mil (9), mim (20), mineiros (10), momento (40), Moro (8), mudança (12), mudar (10), muita (9), nacional (12), natal (7), neto(s) (23), Norte (9), nova(o)s (23),	aceito (3), agrária (7), aí (5), aquilo (4), ausência (3), (3), Brizola (3), cadeira (6), Cãmara (3), campo (5), canalhas (4), cidadão (3), cinco (5), coerência (4), companheiros (8), conseguiu (4), contas (6), convocação (3), coragem (4), cumprir (7), da (144), daqueles (6), defender (6), deixar (3), democraticamente (3), democrático (5), deram (3), deveria (5), direito(s) (22), ditadura (7), elegeu (3), eleição(ões) (7), eleita (3), em (170), esse (27), estiveram (5), Exa. (6), fácil (5), falar (9), farsa (10), ficar (3), fizeram (3), fraude (3), golpistas (10), Governador (4), hipocrisia (8), história (13), homens (7), honesta (9), honrada (8), ilegítimo (4), injustiça (3), instituições (3), isso (29), jurei (7), lava-jato (5), legalidade (3), legitimamente (3), legitimidade (4), liberdade (6), luta (18), lutando (3), lutar (4), lutaram (7), mãos (6), mesa (3), Michel (13), movimentos (5), muitos (7), não (270), nas (24),	à (100) (53), ao(s) (207) (92), aqueles (13) (6), aqui (85) (36), Bahia (13) (9), Brasil (272) (46), brasileiro(a)(s) (162) (58), Casa (39) (26), cidade (44) (9), cometeu (13) (9), consciência (15) (4), Constituição (33) (38), contra (65) (89), corrupção (47) (17), crime(s) (27) (28), Cunha (8) (29), defesa (15) (49), democracia (14) (98), deputado(a)(s) (86) (41), desta(e) (30) (11), Deus (49) (7), dia (13) (8), dignidade (11) (6), digo (16) (5), Dilma (58) (45), direito (12), dizer (31) (16), e (632) (251), é (238) (139), Eduardo (12) (23), em (289) (170), esperança (52) (4), esta(e) (80) (35), está (34) (29), Estado (137) (22), estamos (15) (5), estão (29) (20), estou (14) (4), ética (8) (3), eu (303) (129), família(s) (125) (11), favor (31) (5), federal(ais) (24) (12), filho(s) (60) (7), fora (26) (6), Gerais (31) (3), golpe (8) (87), Governo(s) (46) (4), hoje (48) (13), homenagem (32) (28), <i>impeachment</i> (92) (36),

o (548), Oeste (7), oportunidade (11), orgulho (11), pai(s) (29), Paraíba (9), Paraná (19), Paulo (37), pedir (9), possa (8), PRB (7), precisamos (11), princípios (7), PT (23), pública (12), região (19), representando (10), retomada (8), Rio (40), Rousseff (14), Santa (16), Santo (9), saúde (10), Sul (18), suplente (7), tão (12), temos (14), tenha (8), terra (10), toda(o) (38), viva (14), vontade (9)	nenhum (8), Neves (5), nunca (5), oposição (3), ouvi (7), pares (3), passaram (4), PCdoB (5), PDT (5), plenário (4), pobres (8), popular (12), porque (12), posição (8), posse (3), presidência (3), presidindo (4), primeiro (10), processo (24), quem (8), querem (13), quilombolas (4), razão (5), reconhecimento (3), reforma (10), respeitando (4), respeitar (3), resultado (6), réu(s) (8), sei (4), sem-terra (3), senhor(es) (8), sentado (3), séria (3), sertão (3), sessão (6), soberania (5), sociais (9), solução (6), Suíça (3), Supremo (6), Tancredo (3), tem (19), Temer (23), ter (10), tirar (7), tomei (3), trabalhador(a)(s) (es) (45), traidor (7), Tribunal (9), universidades (3), urnas (5), V.[Vossa] (6), venho (3), vi (4), vice-presidente (3), vocês (8), vota (3)	impedimento (9) (4), justiça (8) (4), juventude (8) (6), liberdade (10) (6), Lula (9) (7), maioria (17) (6), mandato (11) (6), meu(s) (461) (73), milhares (10) (5), milhões (28) (11), Minas (36) (4), minha(s) (275) (48), mulher(es) (9) (18), nação (31) (3), nesta(e) (62) (22), ninguém (14) (3), nome (105) (30), nós (61) (21), nossa(o)(s) (115) (29), país (115) (34), Pará (11) (4), parlamentares (13) (5), Partido (48) (9), pela(o)(s) (734) (151), Pernambuco (9) (8), política (24) (12), população (20) (11), por (242) (87), povo (192) (55), Presidenta (8) (22), Presidente (371) (119), que (529) (259), querida(o) (67) (8), quero (45) (26), República (20) (7), respeito (68) (39), responsabilidade (35) (13), rua(s) (27) (24), sim (393) (8), sou (19) (22), Sr(a)(s) (377) (102), tenho (19) (3), todos (72) (17), vai (21) (15), vamos (14) (5), vida (25) (14), votamos (7) (14), votando (9) (3), votar (22) (11), voto (384) (155), votos (9) (13), vou (16) (9)
---	--	--

Fonte: Elaborado pela autora

Fazer a análise de cada palavra-chave dos dois *subcorpora* fugiria ao escopo – e ao limite! – deste capítulo. Decidimos, portanto, focar as palavras que geraram maior polêmica na mídia e nas redes sociais, quais sejam, aquelas relacionadas a “Deus”, “família” e “nação”. Julgamos relevante, antes de tudo, verificar até que ponto os resultados levantados em nosso *corpus* de estudo coincidem com as conclusões de Chilton (2004). Para tanto, iniciamos a análise quantitativa a partir das palavras-chave presentes em “SIM” ou “NÃO”.

3.2.1 “SIM” ou “NÃO”

Seguindo a dicotomia das estratégias de legitimação/deslegitimação identificadas por Chilton (2004), observamos nos dois *subcorpora* diferentes formas de autorrepresentação positiva e representação negativa do opositor. “Amor”,

“crescimento”, “honra” e “orgulho” são algumas das escolhas lexicais dos votantes pró-*impeachment* para se referir a si mesmos, como podemos observar na fala transcrita a seguir (grifo nosso):

- (1) “Sr. Presidente, quanta **honra** o destino me reservou de poder da minha voz sair o grito de esperança de milhões de brasileiros.”

Já os eleitores contrários ao *impeachment* utilizaram, entre outras, palavras como “convicção”, “coragem”, “honesta” e “honrada” para se referir a si mesmos e à então presidente:

- (2) “Defender a Constituição em momentos contra majoritários é para quem tem **coragem.**”

Em relação à deslegitimação, é possível observar, do lado daqueles que apoiavam o processo, um vocabulário de crítica e acusação ao governo da época e de expectativa de progresso, caso o *impeachment* se concretizasse:

- (3) “[...] pensando também nos 10 milhões de brasileiros que estão **desempregados** [...].”
- (4) “[...] pelo meu querido povo mineiro e pela **retomada do crescimento** do Brasil [...].”

“Fraude”, “farsa”, “hipocrisia” e “golpistas” são exemplos de insultos dos governistas representando a situação, em relação a seus opositores:

- (5) “E durmam com essa, **canalhas!**”

Identificamos, portanto, que muitas das estratégias identificadas por Chilton (2004), em sua análise de discursos e entrevistas envolvendo políticos ingleses, norte-americanos, entre outros, também foram reconhecidas de forma sistemática nas falas dos deputados brasileiros, tanto contrárias quanto favoráveis ao processo em votação.

Passemos, então, à análise de palavras estatisticamente recorrentes nas duas modalidades de votos.

3.2.2 “SIM” e “NÃO”

Entre as palavras-chave que se repetem na coluna “SIM e NÃO”, encontram-se palavras funcionais, tais como preposições, artigos, conjunções etc. Em geral, essas palavras são preteridas pelas lexicais, ou de conteúdo, na análise do

corpus. Contudo, duas categorias merecem nossa atenção. São elas (i) o pronome em primeira pessoa do plural “nós” e seu respectivo pronome adjetivo possessivo (“nosso(a)(s)”) e (ii) a combinação da preposição “per” e o artigo definido “o/a”, resultando em “pelo(a)(s)”, com significado de “em respeito a”, “em nome de”. O primeiro caso coincide com a observação de Chilton (2004), que afirma que, no discurso político, é comum a estratégia de aproximação com o interlocutor por meio desse recurso linguístico. Vejamos as declarações abaixo, que representam, respectivamente, votos “SIM” e “NÃO”:

- (6) “[...] em favor de melhorar a economia do **nosso** País [...].”
- (7) “E golpe, **nós** não podemos votar por ele.”

Legitimar o voto associando-o a outrem também pôde ser observado nas duas modalidades de voto:

- (8) “**Pelo** Brasil, **pelo** meu Estado e **pela** honra da minha família, eu voto sim.”
- (9) “**Por** aquela trabalhadora que conseguiu ter uma carteira assinada; **por** aquele trabalhador que conseguiu colocar seu filho em Harvard ou no MIT; **pelo** trabalhador rural que recebeu energia elétrica na sua casa; **pelo** fim da hipocrisia, meu voto é não, Sr. Presidente.”

Para Chilton (2004), o controle político está relacionado ao controle da informação. Esse controle pode ser qualitativo ou quantitativo. No segundo caso, o falante fornece dados estatísticos que (supostamente) comprovam a informação apresentada. Nas falas dos deputados que votaram a favor e contra o processo, observamos recorrência de palavras utilizadas para se referir a grandes quantidades. Vejamos os excertos (10), de um voto a favor do processo, e (11), contra:

- (10) “Quero fazer homenagem aqui aos brasileiros de bem, àqueles **milhões** que foram às ruas para reivindicar mudanças [...].”
- (11) “[...] em respeito aos **milhares** e **milhares** de brasileiros e brasileiras que votaram em Dilma [...]”.

3.2.3 Tríade “Deus”, “família” e “nação”

Voltemo-nos, agora, às palavras que deram vazão a diversas matérias jornalísticas e *posts* em redes sociais, em especial de crítica às falas pró-*impeachment*, a saber “Deus”, “família” e “nação”. Por meio do levantamento das palavras-chave significativamente recorrentes nos *subcorpora* “SIM” e “NÃO”, é possível observar que (i) palavras relacionadas a familiares – “filhos” e “netos” –, (ii) palavras que se

referem ao território nacional – nação, país, Brasil –, e (iii) “Deus” recorrem de forma estatisticamente significativa nos dois *subcorpora*.

Contudo, é preciso ter cautela para não incorreremos nas mesmas generalizações feitas pela mídia e pelos internautas, que simplesmente calcularam o número de vezes que determinada palavra foi pronunciada e a associaram a uma das modalidades de voto, em geral àquela favorável ao *impeachment*. Afinal, frequência, apenas, não constitui análise confiável em um *corpus* como o deste estudo, em que não há balanceamento do número de palavras nos *subcorpora* que o compõem. Portanto, a fim de analisar as palavras que formam a “triade” tão discutida, recorreremos, quando possível, à análise dos colocados dessas palavras, a fim de se constatar ou não diferenças nessas modalidades de voto.

3.3 Colocação

Em conformidade com McEnery e Hardie (2012, p. 123), neste estudo colocação (*collocation*) é entendida como “padrões de coocorrência observados em dados do *corpus*”¹⁹. Várias são as ferramentas computacionais que medem o grau de proximidade entre dada palavra de busca e aquelas em seu entorno. Alguns exemplos são WS e AntConc (ANTHONY, 2016)²⁰. Neste estudo, optamos pela ferramenta GraphColl (BREZINA et al., 2015)²¹, que permite a análise de redes de colocados. Segundo Brezina et al. (2015, p. 141), “Colocados de palavras não ocorrem de forma isolada, mas fazem parte de uma rede complexa de relações semânticas que acaba revelando seu significado, assim como a estrutura semântica de um texto ou *corpus*”²².

Assim como outros *softwares* de análise textual, o GraphColl oferece diferentes testes estatísticos para medir a atração entre palavras. Contudo, como em Baker (2016), optamos pelo MI (*mutual information*), uma vez que esse teste privilegia as relações entre palavras lexicais, preterindo as combinações com palavras gramaticais, que ocorrem com maior frequência. O MI considera a razão Observado/Esperado para medir a força de associação entre a palavra de busca e o colocado (HUNSTON, 2002), ou seja, calcula a frequência de coocorrência das palavras no *corpus* dentro de dada janela (*span*) (razão Observado), e a compara à frequência esperada de coocorrência, considerando o tamanho do *corpus* e a

¹⁹ “[...] *co-occurrence patterns observed in corpus data.*”

²⁰ Para uma análise do levantamento de colocados por essas ferramentas, sugerimos a leitura de Baker (2016).

²¹ Para detalhes sobre o potencial de uso da ferramenta, ver Brezina et al. (2015) e Baker (2016). Para informações detalhadas sobre os testes estatísticos disponíveis no GraphColl para medir a associação entre as palavras, recomendamos a leitura de Brezina et al. (2015).

²² “*Collocates of words do not occur in isolation, but are part of a complex network of semantic relationships which ultimately reveals their meaning and the semantic structure of a text or corpus.*”

frequência relativa de cada uma das palavras, caso ocorressem aleatoriamente no *corpus* (razão Esperado) (cf. CLEAR, 1993). Portanto, utilizando o teste estatístico MI (*Mutual Information*) (Stat: 03 – MI), ajustamos a ferramenta para identificar como colocados palavras que ocorressem ao menos cinco vezes em uma janela de cinco à direita e cinco à esquerda da palavra de busca (Span: 5 < > 5).

A título de ilustração, apresentamos, na Figura 2, o levantamento de colocados da palavra de busca “*Impeachment*” no *subcorpus* “SIM”:

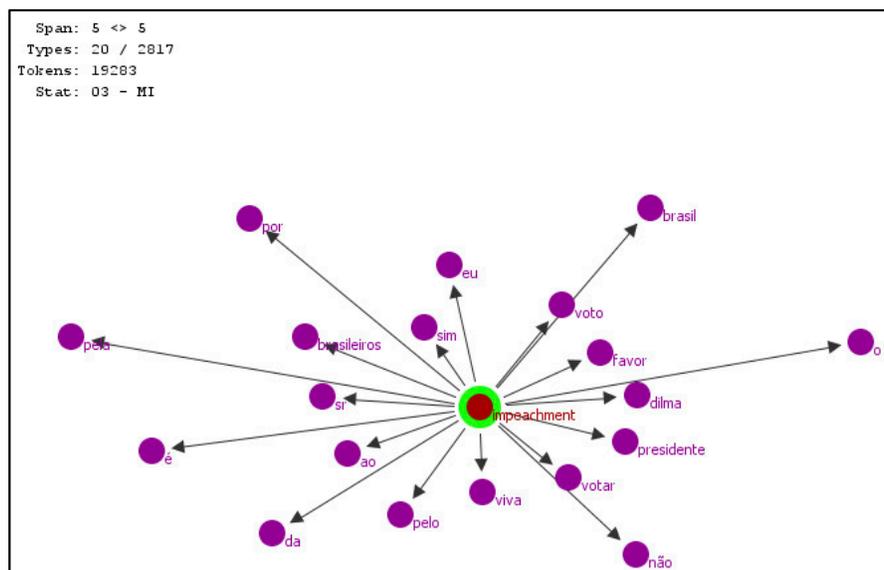


Figura 2 – Colocados de “*Impeachment*” no *subcorpus* “SIM”

Fonte: Elaborado pela autora

A figura apresenta a palavra de busca (*node word*) em destaque, ligada a seus colocados por setas. O tamanho das linhas que separam a palavra de busca de seus colocados indica a proximidade entre eles: quanto mais curta a linha, mais forte a ligação entre eles.

A ferramenta também apresenta os colocados em formato de tabela, conforme apresentado a seguir:

Tabela 4 – Colocados de “*impeachment*” no *subcorpus* “SIM”

impeachment				
Freq: 87, Edges: 0 in, 18 out				
Dir	Type	Stat	Freq (within)	Freq (all)
out	Sim	6,654595	5	11
out	Voto	6,289598	6	17
out	ao	5,792098	39	156
out	votar	5,654595	5	22
out	favor	5,645257	7	31
out	Dilma	5,634557	13	58
out	sim	5,191949	63	382
out	Presidente	5,143360	59	370
out	Sr	5,108283	47	302
out	Por	4,866099	5	38
out	pelo	4,758378	37	303
out	eu	4,553312	25	236
out	voto	4,520390	38	367
out	brasileiros	4,470170	5	50
out	da	3,929151	20	291
out	é	3,653334	13	229
out	não	3,582645	6	111
out	o	3,198574	21	507

Fonte: Elaborado pela autora

Agora vejamos o resultado da busca com a mesma palavra, no *subcorpus* “NÃO” (Figura 3):

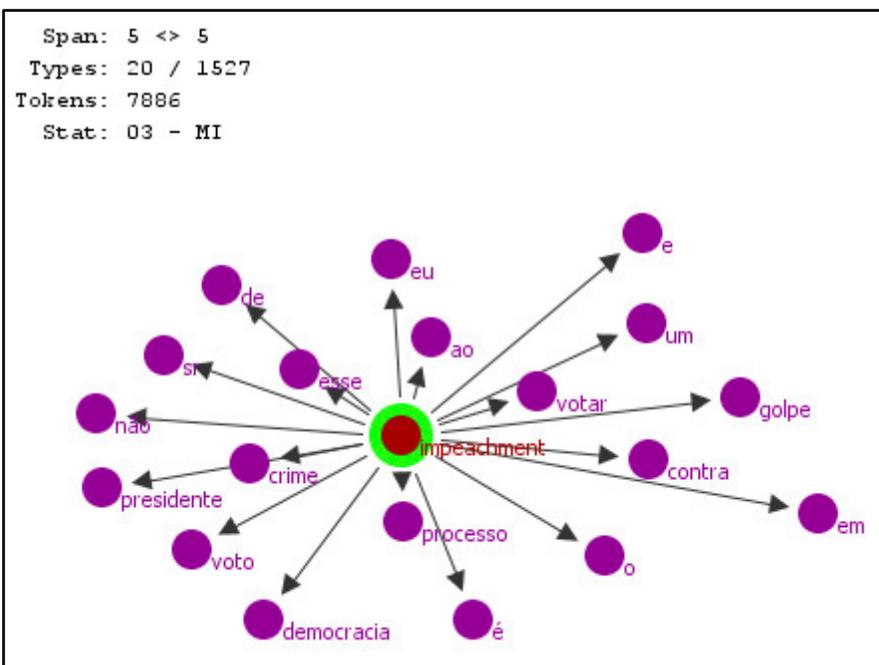


Figura 3 – Colocados de “*impeachment*” no *subcorpus* “NÃO”
Fonte: Elaborado pela autora

Não surpreende que no *subcorpus* dos votos favoráveis ao impedimento o colocado mais próximo de “*impeachment*” seja “*sim*”, seguido de “*presidente*”, “*sr*”, “*voto*” etc. Expandindo a rede de ligações, observamos relações tais como “*sr presidente voto sim ao impeachment*”. Semelhante levantamento no *subcorpus* “NÃO” mostrou que os principais colocados de “*impeachment*” são “*nãõ*”, “*o*”, “*de*”, “*voto*” etc., e que “*nãõ*” se liga fortemente a “*voto*”, “*votar*”, “*é*” etc. No *subcorpus* “NÃO”, coocorrem com “*golpe*” as palavras “*contra*”, “*nãõ*”, “*o*”, “*voto*” etc., formando relações tais como “*contra o golpe eu voto nãõ*”. Já no *subcorpus* “SIM”, os colocados mais significativos com a mesma palavra de busca são “*há*” e “*nãõ*”, formando, entre outras, a relação “*nãõ há golpe*”.

Por meio dos colocados, é possível constatar, portanto, a posição de cada uma das modalidades de voto, representadas, aqui, pelos *subcorpora* “SIM” e “NÃO”. Contudo, até aqui não há novidade em relação aos resultados obtidos com as outras ferramentas. Observamos, inclusive, colocados idênticos nos dois levantamentos. Nesses casos, vale a pena fazer uma busca em segundo nível, a fim de identificar as redes de colocações. Vejamos, a seguir, a Figura 4:

- (12) “**Pela família e pela** inocência das crianças em sala de aula, que o PT nunca teve...”.
- (13) “**Em respeito à** minha **família** e à Constituição e **por** uma democracia plena no nosso País, eu voto não.”

A análise dos colocados da palavra “família” mostrou que, entre aqueles lexicais, encontram-se, no *subcorpus* “SIM”, “amigos”, “esposa”, brasileiro(a), “filhos”, “pai”, “respeito”, “Deus”, “Estado”, “sim”, “Brasil”, “povo” e “voto”. Já no *subcorpus* “NÃO”, os colocados de “família” são apenas “minha” e “Bolsa”, este compondo “Bolsa Família”, programa social que visa a suprir as necessidades básicas de famílias de baixa renda. Também recorrente nos dois *subcorpora*, “filho(s)” forma colocação com “pelo” e “meu” no *subcorpus* “SIM”. Já no *subcorpus* “NÃO”, o número de ocorrências da palavra de busca não é suficiente para revelar colocados a partir dos ajustes estabelecidos.

3.3.2 “nação”

Em conformidade com Chilton (2004), observou-se, nos dois *subcorpora*, recorrência de palavras utilizadas como forma de se apelar ao patriotismo – “nação”, “país” e “Brasil” – além de “pátria”, que, com apenas oito ocorrências não aparece nas listas de palavras-chave. Os excertos (14) e (15) ilustram, respectivamente, os votos “SIM” e “NÃO” (grifo nosso):

- (14) “É nessa direção, com respeito ao povo de São Paulo e por amor à **Nação** brasileira, que eu voto sim.
- (15) “Em defesa da minha **Nação**, do Nordeste, do Piauí, da minha cidade de Oeiras, mas, principalmente, pelo combate à corrupção representada por Eduardo Cunha e Michel Temer, eu digo não a esta corrupção ridícula que envergonha o meu País”.

Quanto aos colocados lexicais de “nação”, observamos, no *subcorpus* “SIM”, “mudar”, “Deus”, “Brasil” e “voto”. Já no *subcorpus* “NÃO”, não houve retorno de colocados com essa palavra de busca. Em relação à palavra “Brasil”, com maior número de ocorrências entre as palavras que se referem a “nação”, observamos “desistir” como principal colocado no *subcorpus* “SIM”, formando uma rede de colocados com “não”, “vamos” e “do”. Já no *subcorpus* “NÃO”, os principais colocados são “voto”, “não”, “democracia” e “contra”, sendo que o principal é “quero”, que forma a rede de colocados “dizer”, “eu”, “Presidente”, “que” e “não”.

3.3.3 “Deus”

A linguagem do discurso político costuma se entrelaçar com crenças religiosas (cf. CHILTON, 2004). Corroborando essa conclusão, “Deus” aparece como palavra-chave nos dois *subcorpora*. Vejamos dois excertos, (16) e (17), que representam, respectivamente, as modalidades “SIM” e “NÃO” (grifo nosso):

(16) “Que **Deus** abençoe o nosso país.”

(17) “[...] eu rogo a **Deus** que ilumine os caminhos da Paraíba e os caminhos do Brasil.”

A análise dos colocados, contudo, revela diferenças entre o uso da palavra “Deus” entre aqueles favoráveis e contrários ao processo. Observamos, no *subcorpus* “SIM”, “abençoe”, “senhor”, “feliz”, “nação”, “país”, “família”, “sim”, “voto”, “Presidente”, “povo” e “Brasil”, só para mencionar os colocados lexicais de “Deus”. Já em relação ao *subcorpus* “NÃO”, o número baixo de ocorrências impossibilitou a busca por colocados, ao menos a partir dos ajustes feitos na ferramenta. Portanto, observemos as linhas de concordância (Quadro 2):

Quadro 2 – Linhas de concordância com a palavra “Deus” no *subcorpus* “NÃO”

N	Concordance
1	e dando razão a V.Exa. quando pediu a Deus que tenha misericórdia deste País, e tem que ter
2	e Srs. Deputados, primeiro, eu rogo a Deus que ilumine os caminhos da Paraíba e os caminhos
3	quarto e ao sétimo mandamentos da lei de Deus . Quero dizer também, colegas Deputadas e
4	tão curto, eu ouvi tantas vezes o nome de Deus ser usado em vão, como se fosse um panfleto.
5	coisas mais diversas, inclusive o nome de Deus . Não aludem ao crime de responsabilidade, que
6	Meu Deus! Quanta hipocrisia! Não é Dilma que tem que sair do
7	entes, em primeiro lugar, eu oro para que Deus abençoe a nossa querida Nação, o Brasil. Em segundo

Fonte: Elaborado pela autora

Analisando-se os contextos de uso, verificamos que, das sete ocorrências da palavra “Deus” no *subcorpus* “NÃO”, a palavra é utilizada (i) para criticar as falas daqueles que apoiaram o *impeachment* (linhas 1, 3, 4 e 5), (ii) como interjeição (linha 6) e (iii) para invocar ajuda divina (linhas 2 e 7).

4 Discussão

Conforme exposto na introdução deste capítulo, o que motivou a pesquisa foi o interesse em comparar as informações publicadas na mídia e nas redes sociais sobre as escolhas lexicais dos deputados durante votação do processo de

impeachment de Dilma Rousseff, em geral com críticas à recorrência de palavras relacionadas a Deus, nação e família na justificativa dos votos favoráveis ao prosseguimento do processo para o Senado.

Segundo os ajustes feitos na ferramenta para o levantamento das palavras que ocorrem com frequência significativamente mais alta no *corpus* de estudo do que no *corpus* de referência, observamos que palavras relacionadas à chamada “tríade” “família”, “Deus” e “nação”, apenas para mencionar aquelas que mais chamaram a atenção da mídia e dos internautas, foram recorrentes nas falas dos deputados que votaram a favor e contra o processo de *impeachment*, durante a votação realizada em 17 de abril de 2016, ainda que não ocorressem estatisticamente na mesma proporção: “Deus” ocorre em proporção de 0,25% no *subcorpus* “SIM” e 0,09%, no “NÃO”. Além disso, a análise dos contextos de uso da palavra mostrou que não foi usada com o mesmo conceito nas duas modalidades de voto. Em todas as ocorrências da palavra no *subcorpus* “SIM”, observamos o apelo para que Deus beneficie aqueles a quem o político alega defender, corroborando as conclusões de Chilton (2004) em sua análise de discursos políticos proferidos por líderes de diferentes nações. Já no *subcorpus* “NÃO”, a palavra raramente foi utilizada com a mesma intenção.

Já as palavras relacionadas a território – “Brasil”, “nação” e “país” – totalizam 2,17% das palavras do *subcorpus* “SIM” e 1,05% do “NÃO”, ainda em conformidade com Chilton (2004), que identifica no discurso político o uso estratégico de referências patriotas. Justificativas associadas a familiares também recorrem nos dois *subcorpora*. Portanto, excluindo-se as palavras vinculadas aos polos contrários, quais sejam (i) o de autorrepresentação por meio de palavras positivas e (ii) o de desqualificar o opositor com palavras negativas, em geral observamos escolhas lexicais semelhantes nos dois *subcorpora*, evidenciando “[...] divergências entre o que o falante professa e aquilo em que realmente parece acreditar”²³ (PARTINGTON, 2004).

Para Van Dijk (2001 apud BAKER, 2006, p. 73-74), duas questões essenciais permeiam a ACD: a primeira diz respeito a como os grupos mais poderosos controlam o discurso público; a segunda, a como esse discurso influencia a forma de pensar e agir dos grupos menos poderosos. A mídia tem papel preponderante na produção e reprodução de discursos (BAKER, 2006). Assim, a representação do discurso na mídia deve ser entendida como um importante processo ideológico (FAIRCLOUGH, 1995) e, como tal, deve ser revelado. Ainda que se declare neutra, cada publicação tende a privilegiar algum lado. Em relação à política, em geral as publicações se posicionam a partir de ideologias relacionadas à direita ou à esquerda. Sendo a então Presidente da República, Dilma Rousseff, representante de partido de esquerda – PT (Partido dos Trabalhadores) –, não surpreende que

²³ “[...] *divergencies between what a speaker professes and what they really seem to think.*”

jornais e revistas, impressos ou *on-line*, tenham se posicionado de forma contrária ou favorável ao impedimento. Vale ressaltar, contudo, que este estudo não pretende abarcar questões ideológicas das dicotomias políticas, nem mesmo das publicações utilizadas para justificar este estudo, mas simplesmente analisar o discurso da mídia e do público após o processo, no que tange especificamente às menções às escolhas lexicais utilizadas durante as falas dos deputados.

Ora, uma vez que o resultado da votação contentou aqueles favoráveis ao processo, é compreensível que as falas dos deputados que permitiram a continuidade do processo não tenham recebido grande atenção nas discussões pós-sessão. Já àqueles contrariados com o resultado restou a chance de criticar essas falas. No entanto, o posicionamento da mídia tem consequências. Os jornalistas exercem influência nos leitores ao produzir seus próprios discursos ou reformular outros já existentes (BAKER, 2006). No entanto, conforme enfatiza Fairclough (1989), os efeitos produzidos pela mídia são cumulativos: só atingem o objetivo desejado por meio da repetição. Além disso, os interlocutores não são passivos, mas interagem com o conteúdo a partir de sua posição ideológica.

5 Considerações finais

A análise semiautomática das falas dos deputados mostrou que, diferentemente do que a mídia publicou – e as redes sociais endossaram –, a tríade “nação”, “Deus” e “família” não ocorre significativamente com maior frequência nos discursos pró-*impeachment*, mas sim em maior quantidade nesses discursos, uma vez que os votos favoráveis ao processo foram 2,68 vezes maiores do que aqueles contrários a ele. Além disso, é importante que se analise também o entorno das palavras, ou seja, os colocados, estes que podem evidenciar as reais diferenças entre os discursos.

A análise quantitativa, baseada em dados estatísticos, aliada à manual, possibilitada pela ACD, ajuda a revelar dados que poderiam ficar restritos apenas ao quantitativo ou à interpretação (tendenciosa) do analista. A metodologia da LC aplicada à ACD possibilita uma análise mais objetiva, a partir de dados revelados por meio de ferramentas computacionais. Contudo, não podemos afirmar que essa análise esteja totalmente livre de subjetividades. O analista sempre faz escolhas. Neste estudo, por exemplo, selecionamos as manchetes dos jornais que justificam nossa pesquisa, priorizamos a análise de algumas palavras-chave e colocados, em detrimento de outros, e escolhemos os excertos do *corpus* para ilustrar os resultados quantitativos obtidos.

A partir dos resultados evidenciados por meio da metodologia explicitada, não temos a pretensão de tecer generalizações sobre o discurso político, tão somente analisar as falas dos deputados durante o processo de *impeachment* de Dilma

Rousseff. Aventamos, para o futuro, a comparação com as falas dos deputados durante o processo de *impeachment* de Fernando Collor, a fim de verificar se houve diferença significativa em relação às falas analisadas neste estudo, uma vez que naquele caso o então presidente não representava ideologias de esquerda. Outra possibilidade, também, seria comparar as falas dos deputados às dos senadores, que deram continuidade ao processo. Enfim, várias são as possibilidades de análise do discurso político, que, naturalmente, não encerra com o término deste capítulo.

Referências

- ANTHONY, L. *AntConc* (3.4.4) [Computer Software]. Tokyo: Waseda University, 2016. Disponível em: <<http://www.laurenceanthony.net/software.html>>. Acesso em: 23 out. 2017.
- BAKER, M. *Corpus Linguistics and Translation Studies: implications and applications*. In: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. (Org.). *Text and technology: in Honour of John Sinclair*. Amsterdam: John Benjamins, 1993, p. 233-250.
- BAKER, P. *Using corpora in Discourse Analysis*. London: Continuum, 2006.
- _____. Acceptable bias? Using *Corpus Linguistics* methods with Critical Discourse Analysis. *Critical Discourse Studies*, v. 9, n. 3, p. 247-256, ago. 2012.
- _____. The shapes of collocation. *International Journal of Corpus Linguistics*, v. 21, n. 2, p. 139-164, 2016.
- BAKER, P.; McENERY, T. A *corpus*-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics*, v. 4, n. 2, p. 197-226, 2005.
- BAKER, P.; GABRIELATOS, C.; KHOSRAVINIK, M.; KRZYŻANOWSKI, M.; McENERY, T.; WODAK, R. A useful methodological synergy? Combining Critical Discourse Analysis and *Corpus Linguistics* to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, v. 19, n. 3, p. 273-306, 2008.
- BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus Linguistics: investigating language structure and use*. Cambridge: Cambridge University Press, 1998.
- BREZINA, V.; McENERY, T.; WATTAM, S. Collocations in context: a new perspective on collocation networks. *International Journal of Corpus Linguistics*, v. 20, n. 2, p. 139-173, 2015.
- CARTER, R.; McCARTHY, M.; O'KEEFFE, A. *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press, 2007.
- CHENG, W. *Corpus-based linguistic approaches to Critical Discourse Analysis*. In: CHAPELLE, C. (Org.). *The Encyclopedia of Applied Linguistics*. London: Blackwell, 2013, p. 1-8.
- CLEAR, J. From Firth principles: computational tools for the study of collocation. In: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. (Org.). *Text and technology: in honour of John Sinclair*. Amsterdam: John Benjamins, 1993, p. 271-292.
- CHILTON, P. *Analyzing Political Discourse: theory and practice*. London: Routledge, 2004.
- FAIRCLOUGH, N. *Language and power*. London: Longman, 1989.
- _____. *Critical Discourse Analysis: the critical study of language*. London: Longman, 1995.

- FETZER, A. Put bluntly, you have something of a credibility problem: sincerity and credibility in political interviews. In: CHILTON, P.; SCHÄFFNER, C. (Org.). *Politics as Text and Talk*. Amsterdam: John Benjamins, 2002, p. 173-201.
- HUNSTON, S. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press, 2002.
- McENERY, T.; HARDIE, A. *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press, 2012.
- PARTINGTON, A. *Corpora and discourse, a most congruous beast*. In: PARTINGTON, A.; MORLEY, J.; HAARMAN, L. (Org.). *Corpora and discourse*. Bern: Peter Lang, 2004, p. 9-18.
- PARTINGTON, A.; MORLEY, J.; HAARMAN, L. (Org.). *Corpora and discourse*. Bern: Peter Lang, 2004.
- PEARSON, J. *Terms in context*. Amsterdam: John Benjamins, 1998.
- PRADO, M. *Levantamento dos padrões léxico-gramaticais do inglês para aviação: um estudo vetorado pela Linguística de Corpus*. 2015. 133 f. Dissertação (mestrado em Estudos Linguísticos e Literários em Inglês). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2015.
- SANDERSON, T. *Corpus, culture, discourse*. Tübingen: Gunter Narr Verlag, 2008.
- SCOTT, M. *Word Smith Tools version 6.0*. Oxford: Oxford University Press, 2012.
- VAN DIJK, T. Critical Discourse Analysis. In: SCHIFFRIN, D. T.; HAMILTON, H. E. (Org.). *The Handbook of Discourse Analysis*. London: Blackwell, 2001, p. 352-71.
- ZANETTIN, F. *Translation-driven corpora: corpora resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome, 2012.