

## Os potenciais do *Iramuteq* para análise de conteúdo de decisões judiciais

### The potentials of *Iramuteq* for content analysis of court decisions

DOI:10.34117/bjdv9n1-336

Recebimento dos originais: 23/12/2022

Aceitação para publicação: 24/01/2023

#### **Ulisses Matheus Braga de Freitas Melo**

Doutorando em Ciência Política pela Universidade Federal de Pernambuco (UFPE)

Instituição: Universidade Federal de Pernambuco (UFPE)

Endereço: Rua Rodrigues Ferreira, 45, 503-A, Recife - PE

E-mail: ulissesgdm@live.com

#### **Laís Oliveira de Souza**

Mestre em Ciência Política pela Universidade Federal de Pernambuco (UFPE)

Instituição: Universidade Federal de Pernambuco (UFPE)

Endereço: R. Diniz Roque de Santana, 845, Terra Firme, Rio das Ostras - RJ

E-mail: lais.oliveirasouza@yahoo.com.br

#### **RESUMO**

O seguinte paper apresenta a técnica de Análise de Conteúdo automatizada aplicada em decisões judiciais. O objetivo é demonstrar os potenciais usos do software *Iramuteq* para o estudo do comportamento jurisdicional dos tribunais, bem como as linhas argumentativas e jurisprudenciais mais utilizadas. O *Iramuteq* possibilita, via uma interface acessível, o desenvolvimento de análises lexicais básicas, análise de especificidades, classificação hierárquica descendente, análise de similitude e nuvem de palavras. Os resultados obtidos a partir de análises iniciais demonstram a capacidade técnica desses métodos, bem como abrem caminho para novas agendas de pesquisa dentro da academia jurídica.

**Palavras-chave:** *Iramuteq*, análise de conteúdo, análise de discurso, Poder Judiciário.

#### **ABSTRACT**

The following paper presents the technique of automated content analysis applied to judicial decisions. The aim is to demonstrate the potential uses of *Iramuteq* software to the study of judicial behavior in courts, as well as argumentative lines and most used jurisprudence. *Iramuteq* enables, via an accessible interface, the development of basic lexical analysis, specificity analysis, descending hierarchical classification, similarity analysis and word cloud. The obtained results from the initial analysis demonstrate the technical potential of these methods and make way to new research agendas in the judicial academy.

**Keywords:** *Iramuteq*, content analysis, discourse analysis, Judicial Power.

## 1 INTRODUÇÃO

A análise de conteúdo é uma metodologia que vem sendo cada vez mais utilizada nas ciências sociais, psicanalíticas e humanas. Trabalhos como o de MOREIRA (2019), CERVI (2018), CHAGAS e CARREIRO (2021), PEREIRA (2021) e RODRIGUES et. al (2020) demonstram como essa linha de pesquisa é capaz de desvelar os arranjos discursivos de uma grande quantidade de texto. Tendo como premissa básica a menor interferência possível do pesquisador no tratamento dos dados estudados, a análise de conteúdo busca, via mecanismos estatísticos, resultados replicáveis capazes de serem testados e reproduzidos por outros pesquisadores, promovendo assim certa profundidade metodológica ao estudo. Este artigo, portanto, visa apresentar para a comunidade acadêmica a possibilidade de utilização do *software Iramuteq* (RATINAUD, 2009) em análises de conteúdo de decisões judiciais.

As decisões judiciais têm uma importância fundamental para a compreensão do *status quo* jurídico. Sentenças colegiadas de tribunais superiores são capazes de criar jurisprudências, uma das fontes de direito, tendo um papel importante para a aplicação e modernização da norma jurídica. O Supremo Tribunal Federal (STF), como órgão de cúpula do poder judiciário tem em sua jurisdição uma ampla gama de poderes institucionais, com sua independência e autonomia garantidas pela Constituição de 1988. Devido a tais prerrogativas o debate de temas como a judicialização da política e o ativismo judicial vêm se desenvolvendo desde o processo de redemocratização. Além disso, a midiaticização de julgamentos, principalmente aqueles de interesse da opinião pública fazem com que cada vez mais os tribunais assumam o centro do debate institucional, devido às consequências de suas decisões para o cenário político e social brasileiro<sup>1</sup>.

A nossa compreensão da atuação jurisdicional dos tribunais nacionais ainda é bastante limitada, muitas perguntas ainda estão em aberto na academia. Diversas teorias tentam responder questões sobre quais são as possíveis motivações políticas e ideológicas por trás de determinado julgado, sobre o impacto da opinião pública nos votos dos ministros em julgamentos colegiados, ou ainda acerca da coesão sobre aplicação de princípios e normas ao longo da jurisprudência. Essas teses, contudo, carecem de testes em larga escala, que comprovem possíveis vieses ou dissonâncias entre os magistrados

---

<sup>1</sup>Alguns exemplos são: julgamento do mensalão, constitucionalidade da união homoafetiva, viabilidade jurídica do impeachment da presidente Dilma Rousseff e suspeição do ex-juiz Sérgio Moro.

Por tais motivos é de suma importância compreender as argumentações utilizadas no processo decisório, não apenas do STF, mas de todos os tribunais do país. Esse texto, portanto, tem como objetivo apresentar à comunidade acadêmica os potenciais do *software Iramuteq* como ferramenta capaz de apontar, com precisão estatística, achados de importante relevância a partir da análise do conteúdo decisório proferido pelo judiciário brasileiro.

Na primeira parte deste artigo descrevemos os conceitos fundamentais da análise textual. No segundo tópico, apresentamos o *software Iramuteq*, discorrendo sobre os diferentes tipos de análises estatísticas e visualizações gráficas proporcionadas pelo programa. Na última parte apresentamos diversas possibilidades de agendas de pesquisa para o futuro.

## 2 ANÁLISE TEXTUAL

### 2.1 ANÁLISE DE CONTEÚDO E APLICAÇÃO NO DIREITO

Segundo Bardin (1977) a análise de conteúdo pode se voltar para o estudo dos ‘significados’, como, por exemplo, na análise temática, como também pode voltar-se à análise dos ‘significantes’, via análise léxica e análise dos procedimentos. A análise textual nada mais é que a ramificação da análise de conteúdo que observa textos escritos para buscar quantificar os dados, utilizando-se da técnica estatística para inferir sobre os significantes que os dados textuais podem demonstrar.

A metodologia de análise textual tem crescente valor para a área de Ciências Humanas e Sociais. Ela nada mais é do que uma forma de análise de dados que estuda o material verbal transcrito. Textos originalmente escritos, documentos, relatos, pronunciamentos e entrevistas são alguns dos objetos de estudo dessa técnica analítica (CAMARGO et al., 2013; NASCIMENTO & MENANDRO, 2006). Por partir da fonte comunicacional, os dados utilizados são compostos essencialmente pela linguagem, possibilitando o estudo sobre opiniões, valores, crenças e pensamentos. A técnica de análise de dados textuais consiste simultaneamente de métodos quantitativos e qualitativos, pois quantifica os elementos textuais e permite o emprego de cálculos estatísticos sobre dados e variáveis substancialmente qualitativas.

Antes de adentrar nos detalhes sobre a análise textual é preciso esclarecer uma diferenciação entre dois tipos de análise que podem se valer desta técnica, e que muitas vezes são confundidas por pesquisadores. De um lado está a análise de discurso que

trabalha mais no domínio da teoria do discurso, focando em interpretações ideológicas, valores do público de destino e filosofias, considerando aspectos específicos do comunicador como crenças, local de fala, posição social, etc. Esse modelo de observação aborda o processo produtivo, a condição de produção de uma determinada forma de comunicação, buscando o sentido do texto sob o contexto histórico e social da produção e voltando a análise qualitativa aprofundada sobre material que já foi produzido. Por outro lado, a análise de conteúdo busca observar o que literalmente foi dito, interpretando o que o texto quer dizer para descobrir um saber por trás da superficialidade do conteúdo. A técnica se funda em metodologia quantitativa e/ou qualitativa, pois sistematiza e descreve o conteúdo literal com complexidade técnica. A integração de ambas as técnicas pode trazer maior robustez e completude para o estudo de determinados temas.

A análise textual torna possível a verificação intra-relações e inter-relações a partir dos textos observados por parte do pesquisador. Essas abordagens possibilitam a descrição de um material produzido por um comunicador tanto de forma individual como coletiva (intra-relações), mas também permite a comparação de produções diversas em função de variáveis específicas que descrevem quem produziu o texto (inter-relações) (CAMARGO e JUSTO, 2013).

Apesar do uso ainda incipiente na seara dos conhecimentos sobre poder judiciário, diversos estudos sobre comportamento judicial têm sido desenvolvidos a partir das técnicas de análise de discurso e de conteúdo. CARVALHO & ROESLER (2019), por exemplo, buscando verificar a incidência do argumento de autoridade no STF, analisaram 2.800 votos de ministros em acórdãos, identificando um “forte traço de irracionalidade e subjetivismo no discurso jurídico”. Já OLIVEIRA (2004), utilizou da análise de conteúdo para examinar 8.578 notícias jornalísticas e traçar o perfil de relevância política do tribunal de cúpula para processo de redemocratização brasileiro. Enquanto OLIVEIRA (2012) apresenta o argumento, a partir da análise de conteúdo e uma subsequente abordagem estatística-descritiva de votos em acórdãos.

Usualmente os instrumentos de aplicação da jurisdição que recebem mais atenção são as ações que discutem a constitucionalidade de leis, os votos dos ministros e as decisões judiciais em temas determinados. No entanto, as audiências públicas podem também ser fonte de importantes achados para a atuação dos tribunais, pois são instrumentos de consulta pública, onde sociedade civil, especialistas e terceiros interessados têm a oportunidade de serem ouvidos pela corte sobre determinado tema

polêmico. Estudos como os de Gomes, et al. (2014) e LEAL, et al. (2017), identificam as falhas deste instrumento. GOMES, et al. (2014) verifica a ineficiência de políticas públicas na aplicação dos princípios do SUS (Sistema Único de Saúde) em casos concretos (individuais), considerando 51 palestras de audiências sobre o tema. LEAL, et al. (2017), por sua vez, analisa a última década de audiências públicas e aponta o uso do mecanismo de consulta a especialistas como “inconsistente e incoerente”.

Importante dar destaque para o trabalho de PEREIRA (2021), que selecionou aleatoriamente 10 sentenças condenatórias de improbidade administrativa proferidas pelo TJ-RN, e as analisou a partir da lexicometria - utilizando-se do *software* Iramuteq - tais atos decisórios e sua fundamentação.

Tais estudos demonstram a importância do desenvolvimento do conhecimento no campo da atuação jurisdicional para o cenário político e sociedade. No entanto, apesar de grande parte dos autores recorrerem à utilização da análise de conteúdo e da análise de discurso - a não ser por PEREIRA (2021) - há pouca transparência no que se refere ao modelo adotado durante a observação dos textos e não há indicação de possíveis programas computacionais utilizados para as análises.

Este trabalho busca, portanto, apresentar um *software* acessível e de fácil assimilação por parte dos pesquisadores de conteúdo, capaz de auxiliar o cientista com técnicas replicáveis e metodologicamente acuradas (ver KING, 2015). Intenta-se com isso, incentivar outros pesquisadores a adotar estratégias mais seguras e transparentes no processo de coleta e tratamento dos dados textuais.

## 2.2 DELIMITAÇÕES CONCEITUAIS BÁSICAS

Antes de avançar no detalhamento do *software*, é necessário delimitar alguns conceitos metodológicos que são fundamentais para a compreensão e utilização da análise textual. Num primeiro momento, o pesquisador que se utiliza da análise de conteúdo constrói um *corpus*, que é o conjunto total de textos que se pretende estudar. Esse *corpus* corresponde à base de dados sobre a qual se debruça a análise. Como exemplo, citamos o conjunto de todos os votos dos ministros do STF em decisão colegiada sobre determinado recurso. Outro *corpus* mais amplo, pode ser constituído pelo conjunto de decisões judiciais que estruturam jurisprudência para uma questão jurídica específica, por exemplo, sentenças sobre permissões e proibições da interrupção voluntária da gravidez.

Os textos são as unidades da análise e sua essência depende da natureza da pesquisa. Portanto, se o estudo se funda em entrevistas, por exemplo, cada uma delas deve ser tratada pelo pesquisador como um texto. Outro exemplo está no caso de estudos de jurisprudências, onde cada decisão de tribunais superiores sobre o tema definido é um texto a ser agregado. O *corpus* da análise corresponde então ao conjunto de textos, que será o “N” da pesquisa. Cabe lembrar que quanto maior o N estudado mais válidas serão as conclusões obtidas.

Por fim, é importante destacar que a metodologia da análise de conteúdo propõe diversas formas de subdividir o objeto analisado, com o intuito de garantir uma melhor observação dos detalhes presentes no escrito. O *segmento de texto* é uma unidade de registro bastante utilizada nos estudos que se utilizam desse método. Estes segmentos podem ser dimensionados pelo *software* ou definidos pelo próprio pesquisador. Camargo e Justo (2013), entendem o segmento de texto como sendo o ambiente das palavras, podendo alcançar o tamanho de até três linhas.

A análise textual se expandiu ainda mais com o avanço tecnológico de programas computacionais. Novos *softwares* têm possibilitado a análise de grandes montantes de dados, que antes era tarefa árdua aos pesquisadores, pois era necessária a leitura atenta e quantificação manual de todos os acervos textuais, o que tornava a análise suscetível a diversos erros e vieses. A seguir apresentamos o *software* livre *Iramuteq*, desenvolvido sob a lógica de *open source*, e que conta com diversas vantagens computacionais para análise de conteúdo, pois oferece diversos tipos de técnicas de análise textual, entre elas a análise lexicográfica clássica ou análise estatística, a análise de especificidades, a nuvem de palavras, o método da classificação Hierárquica Descendente (CHD) ou método Reinert, a análise fatorial por correspondência e a análise de similitude.

### 3 O IRAMUTEQ

O *software Iramuteq (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires)* é um programa informático gratuito e de fácil utilização capaz de realizar diversas técnicas de análise de conteúdo via programação em R e na linguagem *python* ([www.python.org](http://www.python.org)). Ele é capaz de organizar a distribuição do vocabulário analisado de forma facilmente compreensível e visualmente clara, já possuindo dicionário em português (CAMARGO e JUSTO, 2013).

O *Iramuteq* permite que pesquisadores analisem uma grande quantidade de texto em pouco tempo. Como ele próprio realiza o processamento dos dados (via R), os usuários não precisam ter profundo conhecimento em programação, sendo necessário apenas acompanhar o passo a passo disponibilizado por manuais como CAMARGO e JUSTO (2018) e SALVIATI (2017), dominando assim a manipulação do *software*. O pesquisador que for observar o processo decisório do judiciário, no entanto, deverá ficar atento a três fatores prévios: a escolha da base de dados analisada, o tratamento dos dados e a atualização do *software*.

Recomenda-se, inicialmente, que sejam escolhidos para a análise do conteúdo decisório, julgamentos que sejam centrados em torno do mesmo tema jurídico. É esperado que decisões que tratam de temáticas jurídicas diferentes tenham, obviamente, argumentações diversas. A utilidade do *Iramuteq* está, justamente, em encontrar diferenças argumentativas em processos – aparentemente – similares, que giram em torno do mesmo tema legal.

Posteriormente, é importante relembrar a necessidade de tratar o *corpus* textual que será analisado pelo *software*. Os guias produzidos por CAMARGO e JUSTO (2018) e SALVIATI (2017) delimitam ponto a ponto o que o pesquisador deve fazer ao tratar sua base de dados para que ela seja devidamente compreendida pelo programa. Será necessário modificar alguns termos, e deletar certos algoritmos que não são lidos pelo *software*, indica-se que seja produzida uma tabela com todas as modificações realizadas pelos pesquisadores, o que irá possibilitar a replicação do estudo.

Por fim, recomenda-se ainda que o pesquisador verifique a última atualização do *software* e a qual versão do programa R ele está adaptado. Geralmente há um intervalo de tempo entre as versões mais recentes do R e as atualizações do *Iramuteq*, o que pode fazer com que o *software* não seja compatível com os últimos *uploads* da ferramenta de programação.

A seguir serão apresentadas as diferentes técnicas disponibilizadas pelo *Iramuteq*, tendo como exemplo um estudo exploratório<sup>2</sup> realizado a partir dos votos dos ministros Edson Fachin e Gilmar Mendes no HC 164.493, que julgou como suspeito o ex-juiz Sérgio Moro<sup>3</sup>. A impetração deste habeas corpus se deu em decorrência de investigações

---

<sup>2</sup> Todo banco de dados utilizado e as suas aplicações no software *Iramuteq* estão disponíveis para replicação: <https://github.com/ulissesgdm/Os-potenciais-do-IRAMUTEQ-para-an-lise-de-conte-do>

<sup>3</sup> Para compreender melhor o contexto do julgamento recomendamos a leitura da matéria a seguir: <https://www.conjur.com.br/2021-jun-23/moro-suspeito-julgar-lula-decide-stf-votos>



contra o ex-presidente Lula, acusado de corrupção no contexto da Operação Lava-Jato. Este caso em específico apresenta um certo potencial para uma observação mais metódica, já que se trata de um julgado onde os votos dos dois ministros marcaram formas bem distintas de análise processual, enquanto Mendes entendeu pela parcialidade do ex-magistrado, Fachin manteve o posicionamento de que Moro agiu com imparcialidade. Todavia, a análise de conteúdo exploratória que será realizada pode demonstrar como a construção de cada voto é feita com base em uma fundamentação jurídica e factual bem diferente.

Além disso, destacamos as vantagens e desvantagens de cada técnica disponibilizada, bem como damos dicas de como aplicar o método em casos ligados à análise jurisdicional.

### 3.1 ANÁLISE ESTATÍSTICA

A primeira, e mais simples, técnica de tratamento de dados oferecida pelo *Iramuteq* é a análise estatística. Ela permite a construção de um dicionário de formas ativas (verbos, substantivos e adjetivos) e suplementares (pronomes, preposições e verbos auxiliares), possibilitando a exploração de características do vocabulário utilizado (SOUSA et al., 2020), além de informar a quantidade de termos (ocorrências) e formas (segmentos textuais) utilizadas no corpus. Essa primeira análise também apresenta a quantidade de hepax (termos não reconhecidos pelo software), a presença de hepaxes é algo comum em toda base de dados, já que o *software* não irá reconhecer palavras que não fazem parte do idioma previamente selecionado, números de processos e alguns outros termos. O tratamento dos dados é algo fundamental para diminuir esse problema, mas dificilmente o pesquisador conseguirá eliminar completamente os hepaxes.

A termo de exemplo, analisamos os votos de Gilmar Mendes e Edson Fachin referentes ao HC 164.493. O *corpus* obtido a partir do voto do primeiro encontrou 29.352 palavras, 3.893 segmentos e 1.820 hepax, que representavam 6,20% de todas as formas do texto. Já o voto de Fachin (que foi unido com a reiteração), apresentou 9.997 palavras, 1.876 frases e 888 hepax, que representavam 8,88% das palavras do texto. Ambas as análises apresentam um percentual acima de 90% sobre a compreensão dos termos usados no *corpus*, o que torna viável a obtenção de conclusões significativas do texto.



Tabela 1 – Voto do ministro Gilmar Mendes

Forma	Freq. ↓	Tipos
juiz	135	nom
processo	111	nom
direito	105	nom
magistrado	97	nom
penal	92	adj
decisão	91	nom
fato	88	nom
advogado	84	nom
caso	69	nom
interceptação	67	nom

Tabela 2 – Voto do ministro Edson Fachin

Forma	Freq. ↓	Tipos
magistrado	61	nom
suspeição	51	nom
habeas_corpus	43	nr
paciente	43	nom
penal	42	adj
hc	41	nr
decisão	38	nom
pr	36	nr
art	35	nr
fato	30	nom

A análise estatística, contudo, possibilita a observação de alguns detalhes a mais. As tabelas 1 e 2 apresentam os resultados obtidos a partir do voto proferido pelo ministro Gilmar Mendes no dia 09 de março de 2021 (à esquerda) e do voto do ministro Edson Fachin e a sua reiteração, proferidas respectivamente nos dias 23 de março de 2018 e 09 de março de 2021 (à direita), elas elencam os 20 termos ativos<sup>4</sup> mais citados em ambos os votos. Podemos observar algumas informações iniciais, como, por exemplo, a importância do termo juiz/magistrado para ambos os votos. Tal informação não surpreende, já que o tema central do Habeas Corpus era a atuação do ex-juiz Sérgio Moro na condução dos processos que analisavam a conduta do ex-presidente Luiz Inácio Lula da Silva.

<sup>4</sup> Os termos ativos são as palavras que carregam mais significado consigo, neste caso são os substantivos, adjetivos e verbos. O Iramuteq também inclui possíveis siglas entre os termos ativos (hc, art, pr...). Os termos suplementares, no entanto, são numerais, advérbios, dentre outras classes gramaticais mais utilizadas como conectores entre termos ativos, ou então são genéricos ao se referir a um determinado objeto, como os pronomes.

Outros dados, contudo, apontam para diferenças nos votos, como, por exemplo, a presença dos termos *HC* e *Habeas\_Corpus*, ambos referentes ao instrumento processual discutido, no voto de Fachin. Os mesmos termos não estão entre os mais citados de Gilmar, que parece dar prioridade aos nomes dos atores que são o centro do debate processual (*Lula* e *Sérgio Moro*) e a interceptação judicial realizada sobre os advogados do ex-presidente (*interceptação, advogado*).

### 3.2 ANÁLISE DE ESPECIFICIDADES OU ANÁLISE DE CONTRASTES

O *Iramuteq*, no entanto, consegue aprofundar o estudo estatístico do *corpus* textual a partir da utilização da análise de especificidades. Esta ferramenta em particular é bastante útil para o estudo comparativo entre dois ou mais textos. Para utilizá-la de forma correta o pesquisador necessitará de um *corpus* subdividido por variáveis, isso permitirá a comparação estatística - via lei hipergeométrica<sup>5</sup> - das palavras e formas gramaticais mais utilizadas em textos diferentes. As subdivisões do conteúdo podem seguir critérios de autoria, data, posicionamento do autor, ou qualquer outra temática que o pesquisador queira comparar.

Para exemplificar, iremos utilizar novamente o exemplo do já citado HC 164.493, observando comparativamente os votos de Gilmar Mendes e Edson Fachin. A tabela 3 mostra os termos estatisticamente mais utilizados por Gilmar, se comparado ao conteúdo do voto de Fachin. A tabela 4 realiza a mesma análise, mas priorizando as palavras mais utilizadas por Emerson Fachin. É importante explicar que a análise de contrastes ressalta os termos que foram muito citados por um autor, mas que foram ignorados por outro. Caso ambos os textos contassem com a presença forte da mesma palavra, a análise de especificidades não daria destaque a esse termo. Esse foi o caso dos termos “magistrado” e “penal”, que foram bastante citados por ambos os ministros em seus votos, e por isso não obtiveram destaque na análise de contrastes.

---

<sup>5</sup> Em teoria das probabilidades e estatística, a distribuição hipergeométrica é uma distribuição de probabilidade discreta que descreve a probabilidade de sucessos em retiradas, sem reposição, de uma população de tamanho que contém exatamente sucessos, sendo cada retirada um sucesso ou um fracasso. Ver mais em: [https://pt.wikipedia.org/wiki/Distribui%C3%A7%C3%A3o\\_hipergeom%C3%A9trica](https://pt.wikipedia.org/wiki/Distribui%C3%A7%C3%A3o_hipergeom%C3%A9trica)

Tabela 3 – voto do ministro Gilmar Mendes

formes	*gru_01 ↓	*gru_02
lula	6.0734	-6.0734
sérgio_moro	5.2212	-5.2212
deltan	4.9747	-4.9747
ex_juiz	4.7336	-4.7336
mpf	4.6739	-4.6739
nao	4.2629	-4.2629
pessoa	4.1206	-4.1206
conversa	4.1206	-4.1206
juiz	4.0127	-4.0127
membro	3.9783	-3.9783

- Elaboração própria

Tabela 4 – voto do ministro Edson Fachin

formes	*gru_01	*gru_02 ↓
elaboração	-13.9013	13.9013
suspeição	-13.0224	13.0224
pr	-8.4948	8.4948
impetração	-7.4555	7.4555
exceção	-7.2602	7.2602
via	-6.964	6.964
hc	-6.6991	6.6991
instância	-6.5661	6.5661
jurisdicional	-6.4779	6.4779
habeas_corpus	-6.2477	6.2477

- Elaboração própria

A partir do exemplo dado, pode-se observar mais uma vez que, enquanto Fachin priorizou termos referentes à processualística do HC (*suspeição*, *impetração*, *pr*, *hc*, *instância*, *habeas\_corpus*, *agravo*, *causa*, *provimento* e *regimental*), Gilmar tendeu a citar os principais atores do processo, seja nominalmente, ou via pronomes (*lula*, *sérgio*, *deltan*, *ex\_juiz*, *MPF*, *juiz*, *roberto* e *teixeira*).

É importante ressaltar que tanto a operação *spoofing*<sup>6</sup>, como os grampos telefônicos realizados pelo então juiz Sérgio Moro no escritório de advocacia que defendia o então investigado, o ex-presidente Lula, são citados diretamente no voto de Mendes (um dos motivos da proeminência do termo *deltan*), enquanto o voto de Fachin gira em torno dos instrumentos processuais passíveis de debate, nesse caso, a viabilidade do tratamento da matéria julgada via HC. Isso pode ser decorrente das estratégias

<sup>6</sup>Investigação que se debruçou sobre o hackeamento de uma grande quantidade de celulares e computadores. Dentre os aparelhos invadidos alguns pertenciam a procuradores do Ministério Público Federal. Ver mais em: [https://pt.wikipedia.org/wiki/Opera%C3%A7%C3%A3o\\_Spoofing](https://pt.wikipedia.org/wiki/Opera%C3%A7%C3%A3o_Spoofing)

argumentativas de cada um dos magistrados, enquanto Mendes analisou as provas levantadas pela defesa, Fachin questionou a validade do Habeas Corpus, já que o mesmo defendia a perda de objeto do instrumento processual<sup>7</sup>.

A Análise de especificidades aprofunda então o estudo estatístico do conteúdo textual, sendo recomendada para pesquisadores que vão realizar análise comparativa entre autores diferentes, ou textos escritos pelo mesmo autor em momentos distintos, podendo ainda distinguir o *corpus* a partir de variáveis sociodemográficas, como a raça, o gênero ou a idade do autor (CAMARGO e JUSTO, 2013).

### 3.3 NUVEM DE PALAVRAS

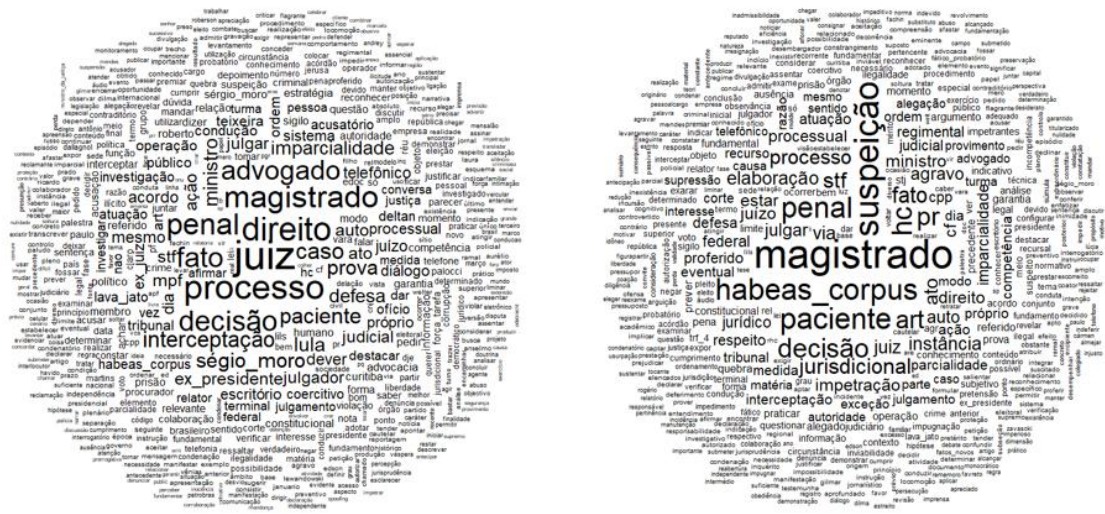
A técnica da nuvem de palavras permite a visualização gráfica das ocorrências no *corpus* (SOUSA et al., 2020). Ela pode ser utilizada de forma complementar a análise estatística, já que tem por base o mesmo método de observação. Recomendamos a utilização da nuvem de palavras como parte do estudo das decisões judiciais, mas não como único método de observação devido à sua simplicidade e limitação analítica. Vejamos a seguir as nuvens de palavras obtidas a partir dos votos de Gilmar Mendes e Edson Fachin, respectivamente.

Podemos observar novamente a predominância dos termos *magistrado/juiz* em ambos os *corpus* textuais, da mesma forma que as palavras *habeas\_corpus* e *hc* têm maior destaque no voto de Fachin, enquanto os termos *morar*, *interceptação* e *advogado* tem maior destaque no voto de Mendes.

---

<sup>7</sup> Ver mais sobre o ponto de vista de Fachin:  
<http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=461870&ori=1>

Figura 1 – Nuvens de palavras decorrentes dos votos de Gilmar Mendes e Edson Fachin, respectivamente.



### 3.4 MÉTODO REINERT

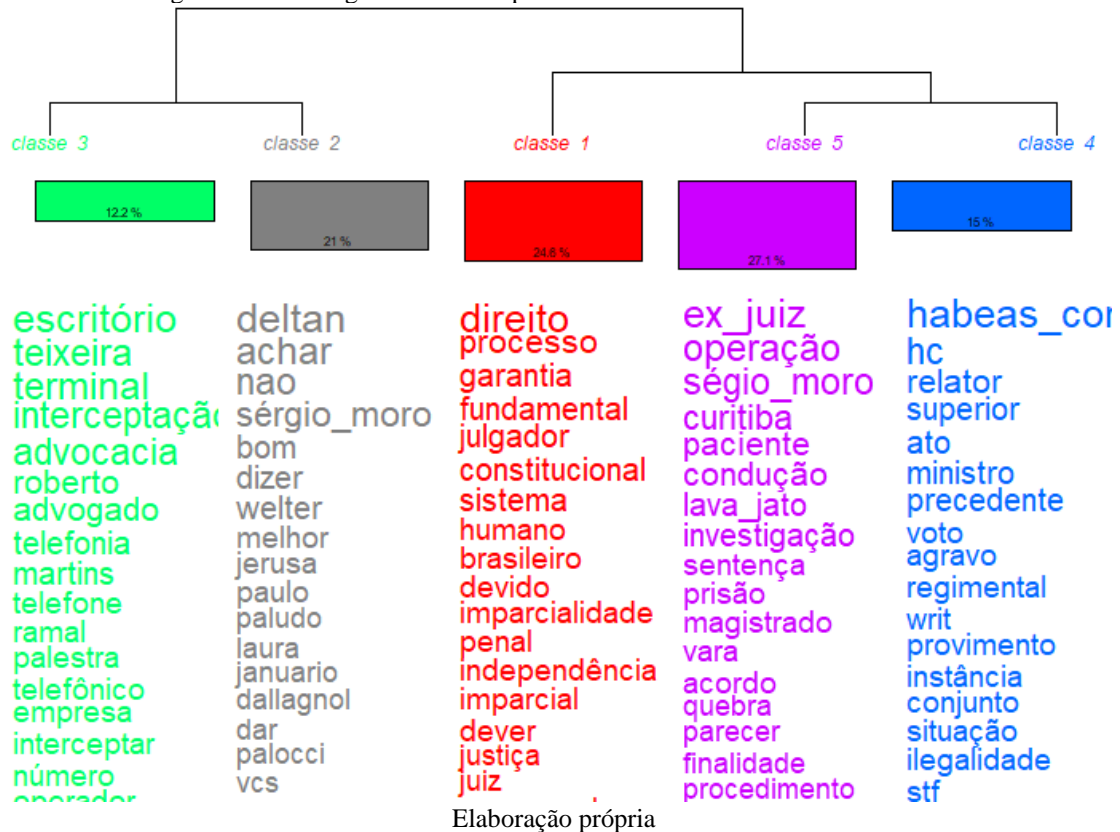
O Método Reinert (REINERT, 1990), também conhecido como Classificação Hierárquica Descendente (CHD), foi criado por Max Reinert e tem como intuito observar a existência de correlação entre termos dentro de um mesmo segmento de *corpus* textual (CERVI, 2018). Inicialmente essa técnica era aplicada via algoritmo no *software Alceste*<sup>8</sup>, sendo posteriormente disponibilizada no *Iramuteq*. Trata-se de um método que auxilia na análise de conteúdo de uma grande quantidade de texto, tendo a capacidade de indicar classes de palavras recorrentes no *corpus* estudado.

Nas palavras de Emerson Cervi (2018), o método Reinert é uma das técnicas mais importantes para a análise léxica automatizada de conteúdos de textos e documentos. Ela permite um avanço nas descrições, sendo muito mais do que uma simples observação da presença e quantidade de léxicos. Ela consegue verificar a associação entre as palavras através do contexto e da presença de termos, possibilitando diferenciar linhas argumentativas em função das posições das palavras nos segmentos textuais.

A CHD é uma das ferramentas presentes no *Iramuteq* com maior potencial para o estudo de decisões judiciais, principalmente quando a intenção do pesquisador é acompanhar os principais argumentos dos magistrados durante o processo decisório. É possível distinguir os polos discursivos de cada *corpus* textual, bem como comparar o posicionamento - e a proximidade - de juízes e desembargadores a partir do conteúdo dos seus votos e sentenças.

<sup>8</sup> Programa de análise quantitativa de dados textuais utilizado em representações sociais. Ver mais em: [https://pt.wikipedia.org/wiki/Alceste\\_\(software\)](https://pt.wikipedia.org/wiki/Alceste_(software))

Figura 2 – Dendrograma obtido a partir do voto do ministro Gilmar Mendes



O *Iramuteq* oferece diversas formas de visualização da CHD. A primeira é através de um dendrograma que apresenta as classes de palavras encontradas pela técnica, quais palavras são predominantes em cada classe e o quanto cada classe representa do conteúdo do *corpus*. A figura 3 apresenta o dendrograma obtido a partir do voto do ministro Gilmar Mendes no HC 164.493.

É possível observar a partir da imagem que a CHD encontrou cinco classes de palavras utilizadas no *corpus*. A primeira grande divisão entre as classes, contudo, se dá entre as classes 1, 4 e 5 e as classes 2 e 3. Isso indica que esses grupos de palavras estão isolados dentro do texto. A classe 3 apresenta os termos *teixeira*, *escritório*, *interceptação*, *terminal* e *advocacia*, como principais, representando 12,2% do *corpus* textual. Já a classe 2 tem como termos mais representativos as palavras *Deltan*, *achar*, *Sérgio\_Moro*, *Welter*, *Jerusa*..., nomes presentes nas conversas inicialmente vazadas pela Vaza-Jato<sup>9</sup>. Essas duas classes indicam os momentos onde Gilmar Mendes cita as principais bases probatórias para o seu voto a favor da suspeição do ex-juiz Sérgio Moro:

<sup>9</sup> Ver mais sobre a Vaza-Jato em: <https://theintercept.com/series/mensagens-lava-jato/>



o grampo telefônico realizado no escritório de advocacia que fazia a defesa do réu<sup>10</sup>, o ex-presidente Lula, e as conversas vazadas pela operação *Spoofing*. Podemos tirar essas conclusões também a partir das citações dos termos *telefone*, *telefonía* e *deltan* presentes nessa classe<sup>11</sup>.

A segunda divisão observada ocorre entre as classes 4 e 5 e a classe 1. As palavras mais significativas desse último grupo são *direito*, *processo*, *fundamental*, *garantia* e *juiz*. Ela representa 24,8% do *corpus* analisado. É nela que vemos parte do desenvolvimento jurídico do voto, podemos tirar essa conclusão também a partir da observação de termos que remetem aos princípios e garantias fundamentais estabelecidas na Constituição de 1988.

Por fim, as classes 4 e 5 se apresentam como as mais próximas e conectadas entre si, isso significa que, geralmente, essas linhas argumentativas se entrecruzam no texto. A classe 5 parece focar mais na narrativa sobre o contexto político em que o processo foi conduzido e os atores institucionais que pautaram o debate na época. Há citações aos termos: *ex-juiz*, *Sérgio\_Moro*, *operação*, *curitiba* e *lava\_jato*, é de conhecimento público que o ministro Gilmar Mendes teceu diversas críticas direcionadas ao ex-juiz Sérgio Moro e à operação Lava-Jato durante os últimos anos, no plenário e fora dele<sup>12</sup>. Os dados colhidos indicam como os eventos - e escândalos - políticos que envolveram a Lava-Jato foram levados em conta neste voto. Essa classe representa 27,1% do *corpus* textual.

Já a classe 4, que engloba 15% dos termos analisados, é mais voltada para o rito do Habeas Corpus. Há citações aos termos *habeas\_corpus*, *hc*, *relator*, *precedente*, *voto* e *agravo*, o que demonstra que o ministro tratou em seu voto do debate sobre a validade do instrumento jurídico e dos argumentos levantados pela defesa do réu.

Por outro lado, ao realizarmos a CHD no voto (e sua reiteração) do ministro Edson Fachin verificamos a presença de 4 classes diferentes de palavras (Figura 4). A primeira grande distinção está entre as classes 1, 2 e 3 e a classe 4. Da mesma forma que no voto de Gilmar, essa linha argumentativa aparenta girar em torno do grampo telefônico

<sup>10</sup> Ver mais sobre o grampo realizado no escritório de advocacia que defendia o ex-presidente em: <http://g1.globo.com/politica/operacao-lava-jato/noticia/2016/03/moro-fez-grampo-ilegal-de-escritorio-de-advocacia-diz-defesa-de-lula.html>

<sup>11</sup> Ver mais sobre o voto de Gilmar Mendes no *habeas corpus* em: <http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=461996&ori=1>

<sup>12</sup> O ministro Gilmar Mendes vem se colocando como crítico da operação Lava-Jato e do ex-juiz Sérgio Moro nos últimos anos, principalmente após o vazamento das mensagens da Vaza-Jato e da participação do ex-juiz no governo Bolsonaro. Ver mais em: [https://www.em.com.br/app/noticia/politica/2021/03/23/interna\\_politica,1249733/gilmar-mendes-voces-comprariam-um-carro-de-moro-e-do-dallagnol.shtml](https://www.em.com.br/app/noticia/politica/2021/03/23/interna_politica,1249733/gilmar-mendes-voces-comprariam-um-carro-de-moro-e-do-dallagnol.shtml) .



realizado no escritório de advocacia que fazia a defesa de Lula, tanto que as palavras *telefônico*, *interceptação*, *advogado* e *ex\_presidente* tem centralidade. Esta classe representa 13,1% do *corpus* textual.



Elaboração própria

A segunda subdivisão observada pela CHD é entre as classes 1 e 2 e a classe 3. Essa última reúne termos mais voltados ao andamento processual do caso, palavras como *termo*, *processo*, *normativo*, *auto*, *resposta* e *ofício* indicam essa tendência. Esse grupo de palavras corresponde a 28,4% do *corpus*.

Por fim, as classes 1 e 2 se apresentam como as mais próximas entre si. A última, que abrange 29,7% do voto do ministro, aponta uma argumentação processualística em torno do caso. Os termos *agravo*, *julgamento*, *instância*, *supressão*, *habeas\_corpus* e *tribunal* indicam também que o autor está abordando temas referentes ao cabimento processual do HC para a matéria julgada, bem como a competência das instâncias judiciais pelas quais os autos do processo passaram. Esse trecho pode estar relacionado à

linha argumentativa em que o ministro defende que o Habeas Corpus não é o instrumento processual devido para a apresentação de novas provas - sem exame prévio - aos autos<sup>13</sup>.

Já a classe 1, que abrange 29,7% do *corpus* textual, aparenta discorrer sobre os eventos que ocorreram em torno do processo, os termos *magistrado*, *recorrente*, *interesse*, *lava\_jato*, *operação*, *contraditório* e *imparcialidade* demonstram que o ministro abordou o debate sobre os questionamentos feitos pela defesa do ex-presidente Lula, tratando dos temas ligados à imparcialidade do ex-juiz Sérgio Moro e a legalidade dos procedimentos realizados pela operação Lava-Jato.

Estes dois exemplos demonstram a capacidade interpretativa que o método Reinert possui. Importante ressaltar que foram utilizados textos que possuíam entre 10.000 e 30.000 palavras, que poderiam ser lidos em um dia, no entanto, o grande trunfo desta técnica está em viabilizar uma análise de conteúdo em larga escala, transformá-la em representação gráfica de fácil compreensão para o leitor, subdividir um texto em categorias, além, é claro, de diminuir as possibilidades de viés, já que a classificação é feita a partir de análise estatística.

### 3.5 ANÁLISE FATORIAL POR CORRESPONDÊNCIA

A Análise Fatorial por Correspondência (AFC) permite a visualização das relações entre os termos e classes presentes no *corpus*, deixando mais clara a “distância” entre as linhas argumentativas presentes nos textos, representando os termos encontrados no *corpus* em um plano fatorial. Ela pode ser utilizada tanto em conjunto com a análise de especificidade como de forma complementar à CHD. Vejamos a seguir as análises fatoriais obtidas dos votos dos ministros Gilmar Mendes e Edson Fachin a partir do método Reinert.

Pode-se observar no anexo A os detalhes da argumentação de cada voto proferido. Tanto Gilmar como Fachin reservaram parte do seu voto para discorrer sobre a interceptação telefônica realizada contra os advogados do réu. É possível notar também como algumas classes se sobrepõem em ambos os textos, mesmo tendo sido classificadas como diferentes.

---

<sup>13</sup> “[...] ainda que se admitisse a possibilidade, em tese, de uso em favor do acusado de prova ilícitamente obtida, essa providência inserir-se-ia no campo da validade. Nada obstante, a confiabilidade desses elementos, a meu ver, dependeria de prévio exame e reconhecimento, em sede jurisdicional própria, de sua autenticidade e integridade, aspectos que refogem à espacialidade deste habeas corpus.” FACHIN, Edson. *Reiteração de Voto*. Habeas Corpus nº164.493/PR, STF, Brasília, 09 de março de 2021.

Outra possibilidade fornecida pela AFC é verificar quais são as classes centrais ao texto, ou seja, aquele grupo de palavras que interconecta os argumentos que os magistrados utilizam em seus votos. No caso do voto de Gilmar Mendes é possível verificar que a classe 5 - focada no contexto político que circundou o processo - é central para a argumentação do ministro. Já para Fachin é a classe 1 - focada na abordagem temática do HC - que se interliga, principalmente, com as classes 2 e 3. Fica claro, a partir dessas leituras, que, diferentemente de Mendes, Fachin evitou abordar o contexto político no qual o processo se encontrava, abordando apenas a temática questionada pela defesa. Isso pode ser consequência da predominância do voto de 2018 e da atualização tímida que o ministro fez em 2021. Cabe dizer ainda que Fachin se preocupou mais em discutir se o habeas corpus impetrado era o instrumento jurídico correto para a discussão da contenda.

### 3.6 ANÁLISE DE SIMILITUDE

A análise de similitude se baseia na teoria dos grafos, possibilitando a identificação de ocorrências entre as palavras (CAMARGO e JUSTO, 2013). Ela promove a visualização das correlações entre os termos mais citados de um *corpus*, tendo potencial de representar como os autores utilizaram os conceitos centrais de cada texto. Vejamos os exemplos das análises de similitude das quarenta<sup>14</sup> palavras mais citadas de cada voto da nossa base de dados experimental.

Os grafos apresentados no anexo B representam a análise de similitude dos votos de Gilmar Mendes e Edson Fachin. A partir delas podemos observar, por exemplo, como a figura do juiz (magistrado) é central em ambos os votos, no entanto, enquanto Fachin interliga essa figura central ao *habeas corpus*, Mendes desenvolve várias linhas de raciocínio, aproximando a figura do julgador de temas ligados ao debate político da época.

A análise de similitude tem um menor potencial comparativo que outras técnicas aqui apresentadas -como a estatística e a de especificidades -, no entanto, seu grande trunfo está no potencial de estudo de uma grande base de dados. Essa técnica é capaz de sistematizar de forma didática as linhas argumentativas de um grande *corpus* textual, algo que seria inviável a partir da observação manual de uma série de decisões judiciais.

---

<sup>14</sup>A escolha da quantidade de termos incluídos na análise de similitude é de escolha do pesquisador. Recomenda-se um número que possibilite uma visualização limpa e, ao mesmo tempo, representativa do corpus textual.

#### 4 CONSIDERAÇÕES FINAIS

O estudo sobre o poder judiciário brasileiro tem se tornado nos últimos anos essencial para compreendermos em completude o contexto político, social e jurídico do país. Apesar de abordagens teóricas e doutrinárias serem mais tradicionalmente adotadas, defendemos que métodos mais automatizados e rigorosos de análise tem potencial de trazer mais robustez ao debate, possibilitando o avanço do conhecimento jurídico ao suplantar conflitos puramente interpretativos.

Voltar a atenção da comunidade jurídica à discussão dos fatos e efeitos reais da atuação judicial na sociedade é uma tendência imparável. Mais que entender a normatividade do direito é fundamental aprofundarmos sobre a aplicação empírica da norma, os discursos políticos e sociais por trás das decisões judiciais que a todos vinculam e o impacto da performance jurídica sobre o entorno do destinatário final, o cidadão.

Seguindo esta linha, a análise de conteúdo e a análise de discurso são técnicas que se apresentam como boas aliadas aos trabalhos que buscam compreender os mais diversos instrumentos judiciais, como decisões, acórdãos, audiências públicas, entre outros.

A análise de conteúdo como método misto de pesquisa - quantitativo e qualitativo - tem capacidade de instrumentalizar âmbitos antes inexplorados estatisticamente, em conhecimento científico replicável. *Softwares* de iniciativa *open source* como *Iramuteq* são ferramentas imprescindíveis ao desenvolvimento e democratização de pesquisas que analisam dados textuais.

Diversos são os potenciais de pesquisas que utilizam a análise de conteúdo, através do *Iramuteq*, no campo de estudo sobre o Poder Judiciário. A partir de uma perspectiva comparativa é possível analisar o conteúdo textual de decisões colegiadas de tribunais superiores em ações específicas relevantes à jurisprudência, como as que conferem repercussão geral, como as ADINs, ADOs, ADPFs, etc., identificando os argumentos favoráveis e contrários. Outras possíveis agendas de pesquisa concernem a análises intertemporais de jurisprudências, identificando a evolução do pensamento jurídico ao longo do tempo. Além disso, a análise do conteúdo decisório de determinada temática pode ser abordada a partir da comparação entre as diferentes instâncias – na cadeia hierárquica do judiciário – e entre diferentes foros, considerando o fator regional.

Pode-se também verificar a correlação entre o conteúdo decisório e variáveis sociodemográficas aplicadas as partes, ou até mesmo aos magistrados. Verificar se a raça, religião, idade, formação e gênero do réu ou do julgador interferem de alguma forma no

conteúdo decisório dos tribunais. As potencialidades são incontáveis, vez que o judiciário se depara diariamente com ampla abrangência de assuntos, desde discussões sociais, políticas e econômicas a conflitos relativos ao Legislativo, Executivo, e ao próprio Judiciário.

A academia jurídica ainda carece desse tipo de análise no campo dos estudos sobre o poder judiciário, mesmo tendo grande potencial de explicar o modo como a norma jurídica é aplicada, auxiliando na compreensão sobre o tema do comportamento judicial, dinâmica interna, usos do tribunal e desenho institucional. Nota-se então que esta técnica é plenamente adaptável ao tradicional estudo jurisprudencial e estatístico-descritivo. Este artigo, portanto, buscou apresentar alguns mecanismos capazes de oferecer um método mais aprofundado, capaz de expor as nuances do nosso judiciário e como a instituição fundamenta suas decisões a partir do próprio conteúdo textual por ela produzido.

## REFERÊNCIAS

- BARDIN, Laurence. *Análise de Conteúdo*. Ed. 70. Presses Universitaires de France, 1977.
- CAMARGO, Brigido Vizeu; JUSTO, Ana Maria. *IRAMUTEQ: Um software gratuito para análise de dados textuais*. Temas em Psicologia, v.21, n.2, pp. 513-518, 2013.
- CAMARGO, Brigido Vizeu; JUSTO, Ana Maria. *Tutorial para o uso do software Iramuteq*. Disponível em: <http://www.iramuteq.org> . Acesso em 19 de abril de 2018.
- CARVALHO, A. G. P. & ROESLER, C. R. *O argumento de autoridade no Supremo Tribunal Federal: uma análise retórica em perspectiva histórica*. Revista Direito, Estado e Sociedade, n. 55, 2019. DOI: 10.17808/des.55.907. Disponível em: <https://revistades.jur.puc-rio.br/index.php/revistades/article/view/907>. Acesso em 05 de julho de 2021.
- CERVI, Emerson U. *Análise de conteúdo automatizada para conversações em redes sociais online: uma proposta metodológica*. 42º encontro anual da Anpocs, GT17 (Mídias, política e eleições), Caxambu, 22-26 de outubro de 2018.
- CHAGAS, Viktor; CARREIRO, Rodrigo. *Macarthismo no Zap: Como se comporta a rede anticomunista de apoiadores de Jair Bolsonaro*. In: CERVI, Emerson U.; WEBER, Maria Helena. Impactos Político-comunicacionais nas eleições brasileiras de 2018. Curitiba: CPOP, 2021.
- GOMES, D. F.; SOUZA, C. R.; SILVA, F. L.; PÔRTO, J. A.; MORAIS, I. A.; RAMOS, M. C.; SILVA, E. N. *Judicialização da saúde e a audiência pública convocada pelo Supremo Tribunal Federal em 2009: o que mudou de lá para cá?* Revista Saúde em Debate, 2014. DOI: 10.5935/0103-104.20140008.
- KING, Gary. *Replicação, Replicação*. Revista Eletrônica de Ciência Política, vol.6, n.2, 2015.
- LEAL, F.; HERDY, R. & MASSADAS, J. *Uma década de audiências públicas no Supremo Tribunal Federal (2007-2017)*. Revista de Investigações Constitucionais, 2017. Disponível em: <https://www.scielo.br/j/rinc/a/mLpfbct74p4rHPbPLY4J9mc/?lang=pt> Acesso em 05 de julho de 2021.
- OLIVEIRA, E. S. *O Sistema político brasileiro hoje: o governo do Supremo Tribunal Federal e a legitimidade democrática*. Revista Sociologias, 2012. Disponível em <https://www.scielo.br/j/soc/a/dryZFtM6NHcGJsc6mJP45CH/abstract/?lang=pt> Acesso em 05 de julho de 2021.
- OLIVEIRA, F. L. *O Supremo Tribunal Federal no processo de transição democrática: uma análise de conteúdo dos jornais Folha de S. Paulo e o Estado de S. Paulo*. Revista de Sociologia e Política, 2004. Disponível em <https://www.scielo.br/j/rsocp/a/dH6zwtjSNSPk8WvCM8zTr4b/abstract/?lang=pt> Acesso em 05 de julho de 2021.
- PEREIRA, Rebeka Souto Brandão. *A normatividade em Improbidade Administrativa e a sua aplicabilidade nas decisões judiciais*. 2021. 156f. Dissertação (Mestrado em Direito)

- Centro de Ciências Sociais Aplicadas, Universidade Federal do Rio Grande do Norte, Natal, 2021.

RODRIGUES, Ana Paula Lopes da Silva; MOURÃO, Victor Luiz Alves; LOPES, Ivonete da Silva. “Eficácia, segurança e qualidade”: *Parâmetros discursivos nas audiências públicas da ANVISA sobre regulamentação e pesquisas com cannabis para fins medicinais*. Teoria e Cultura, v.12, n.2, 2020.

RATINAUD, Pierre. *IRAMUTEQ: Interface de R pour les analyses Multidimensionnelles de textes et de questionnaires* [Computer Software]. Retirado de <http://www.iramuteq.org>. 2009.

REINERT, Max. *Alceste, une méthodologie d'analyse des données textuelles et une application: Aurelia de Gerard de Nerval*. Bulletin de Methodologie Sociologique, v.26, p.24-54, 1990.

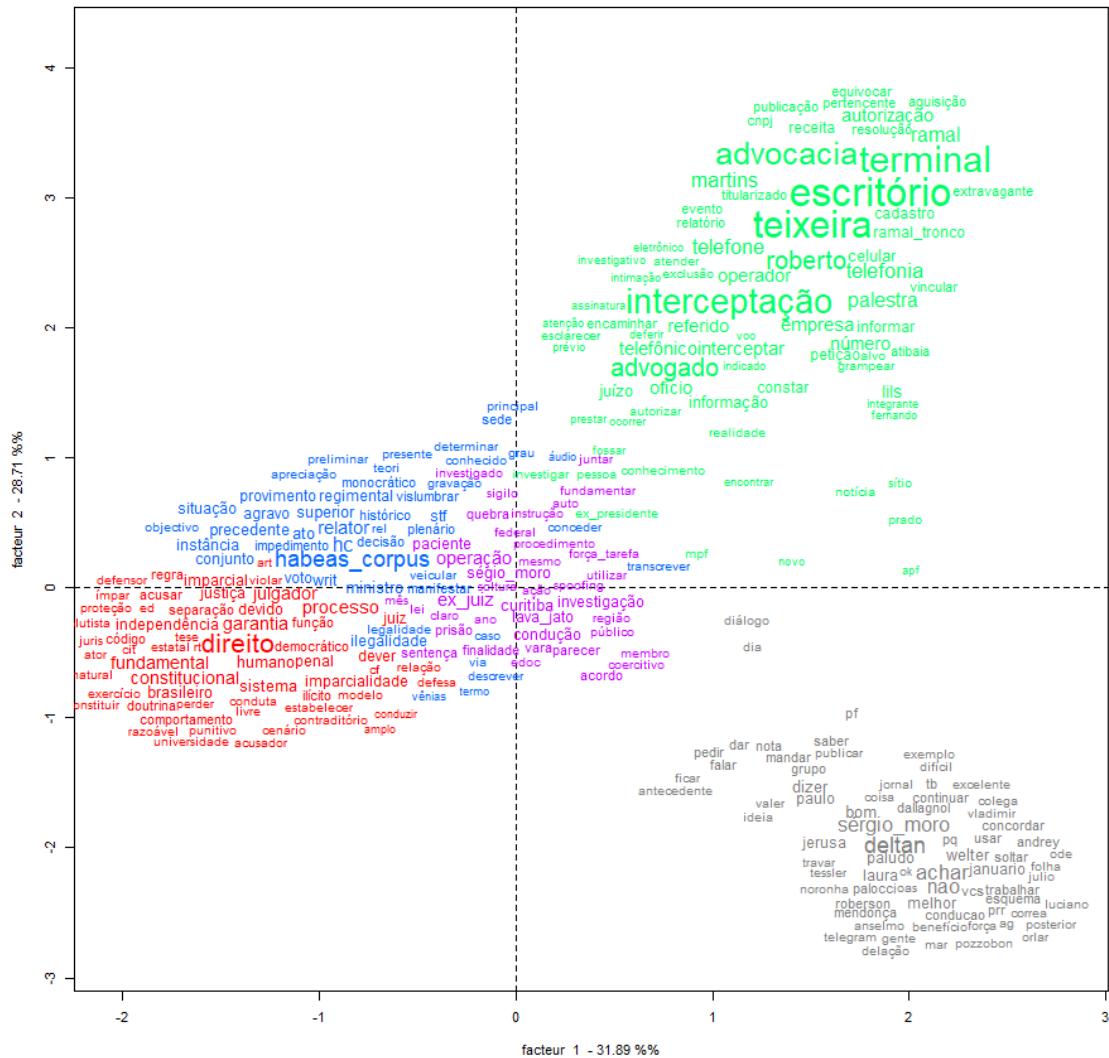
SALVIATI, Maria Elisabeth. *Manual do aplicativo IRAMUTEQ (versão 0.7, Alpha 2 e R versão 3.2.3)*. Planaltina, 2017.

SOUZA, Yuri Sá Oliveira et. al. *O uso do software Iramuteq para análise de dados de entrevistas*. Pesquisa e Práticas Psicossociais, v.15, n.2, 2020.



## ANEXOS

### Anexo A – Análise Fatorial por correspondência

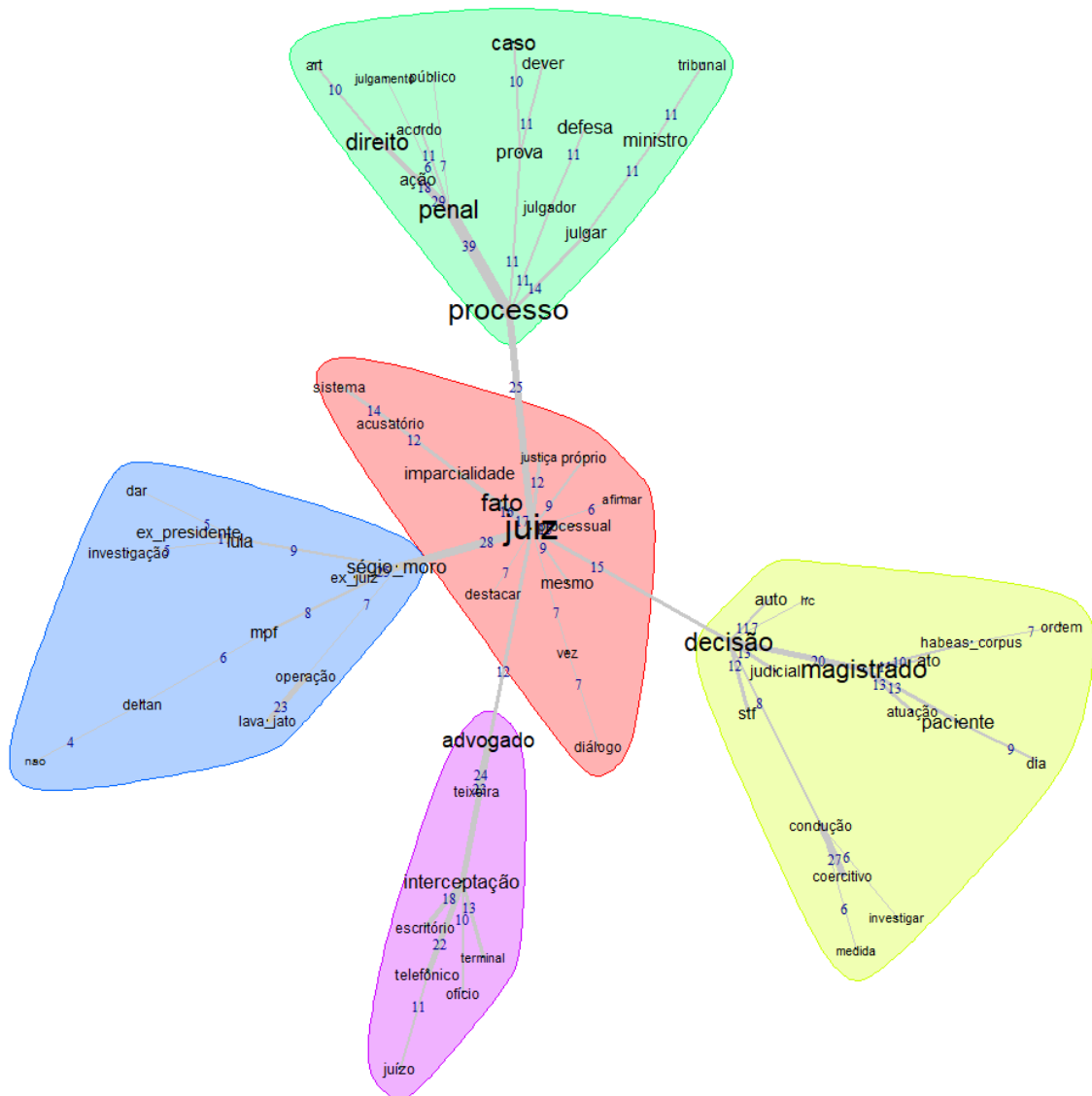


Análise fatorial obtida do voto do ministro Gilmar Mendes a partir do método Reinert – elaboração própria

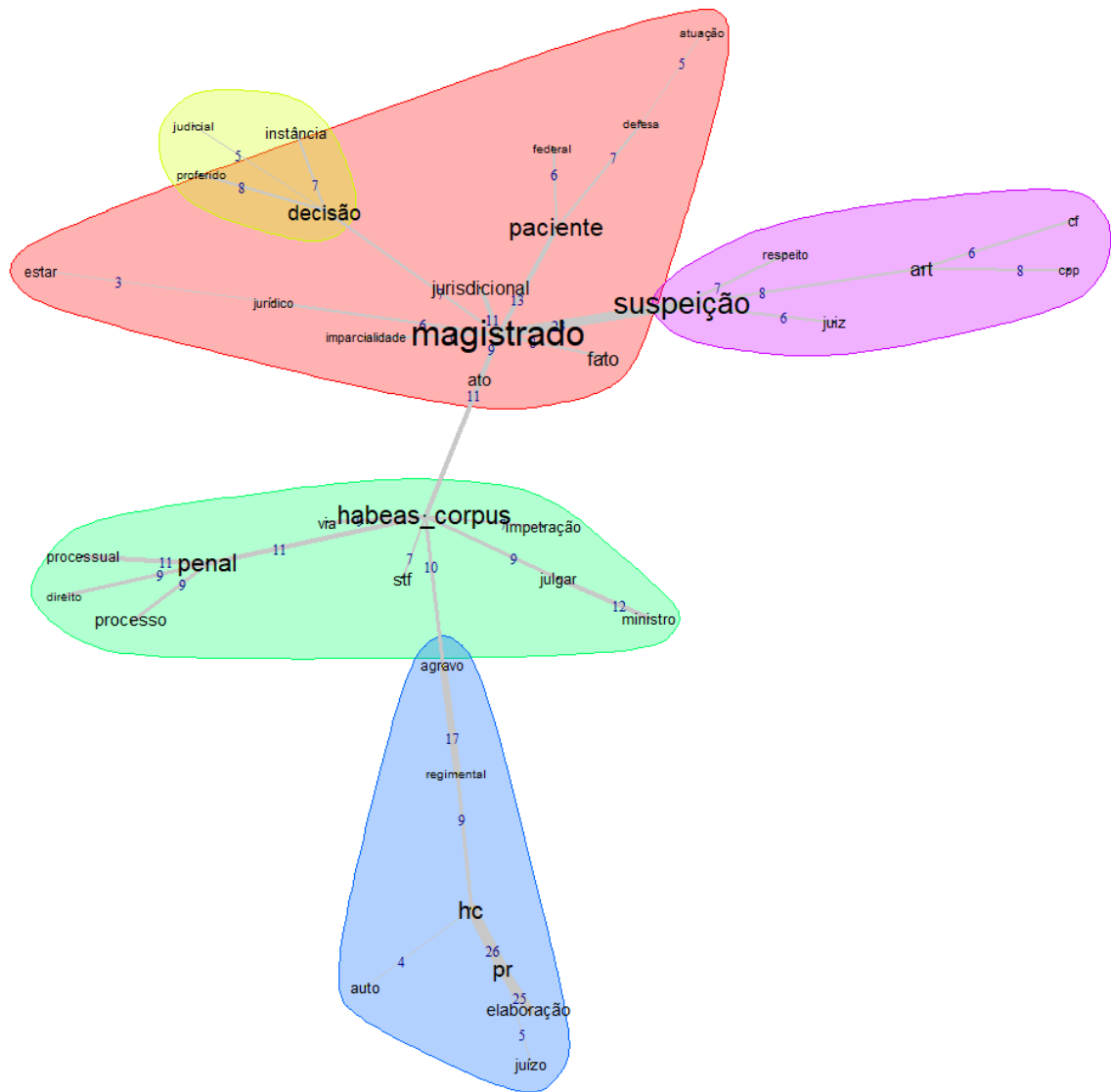
- Legenda:
- Classe 1 - vermelho
  - Classe 2 - verde
  - Classe 3 - azul
  - Classe 4 - roxo



Anexo B – Análise de Similitude



Análise de similitude do voto de Gilmar Mendes – elaboração própria



Análise de similitude do voto de Edson Fachin - elaboração própria