

**Manual do Aplicativo Iramuteq  
(versão 0.7 Alpha 2 e R Versão 3.2.3)**

**Anexo – Exemplo de uma aplicação**

**Maria Elisabeth Salviati**

Planaltina, DF março de 2017

## Introdução

Este exemplo tem por objetivo mostrar todos os procedimentos realizados, desde a preparação do corpus até os resultados. Uma vez que o Iramuteq é um software que se aplica ao estudo de diversas áreas, escolheu-se um corpus com assunto atual e de fácil compreensão aos pesquisadores de diversas áreas: “Impeachment da Presidente da República Dilma Rousseff”. Foram selecionados 15 artigos de jornal (O Globo, Correio Brasiliense e UOL) no período de 11 de abril a 17 de maio de 2016.

## Preparação do Corpus

Os artigos foram copiados eletronicamente da internet diretamente no Word e com esse aplicativo eles foram formatados separadamente: um arquivo para cada artigo. A fonte e tamanho foram padronizados, bem como a eliminação de negritos, itálicos e parágrafos.

Eliminaram-se, também, todos os sinais especiais não recomendados, tais como: hifens; aspas simples e duplas.

As palavras originalmente com hifens tiveram os hifens substituídos por *underscore*. Palavras compostas como Constituição Federal; Congresso Nacional; Advocacia Geral da União foram unidas por *underscore*.

Colocaram-se linha estrela e temática em cada arquivo individual, conforme exemplo abaixo, Figura 1:

```
**** *artigo_01 *jornal_globo *data_19/04/2016
_*senado
```

**Figura 1 – Linha estrela e temática**

As variáveis escolhidas foram:

- artigo: numeração crescente a partir de 01 para cada artigo.
- jornal: nome mnemônico e padronizado do jornal que publicou o artigo.
- data: data de publicação do artigo.
- temática: conforme o assunto predominante do artigo, classificaram-se nas seguintes temáticas: senado; câmara; congresso\_nacional; presidência\_da\_república, STF e trâmite.

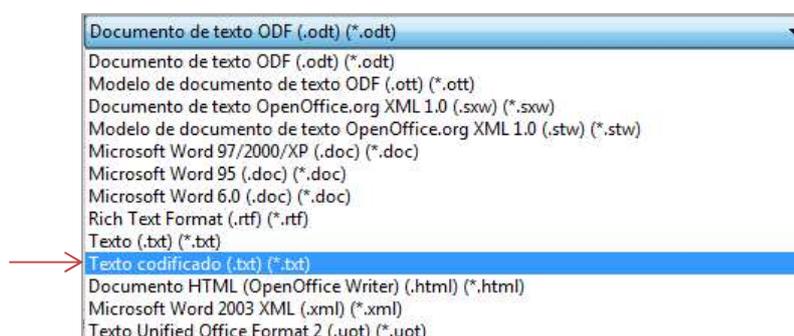
```
------(artigo 14)-----
-----
Temer. No caso de Dilma_Rousseff, as sugestões para a composição do colegiado foram feitas em menos de um dia.
**** *artigo_15 *jornal_correio_brasiliense *data_09/05/2016
_*câmara
Waldir_Maranhão anula tramitação do processo de impeachment de Dilma_Rousseff.
Waldir_Maranhão está no exercício do cargo desde sexta_feira após o STF afastar o
-----
```

**Figura 2 – Junção de arquivos**

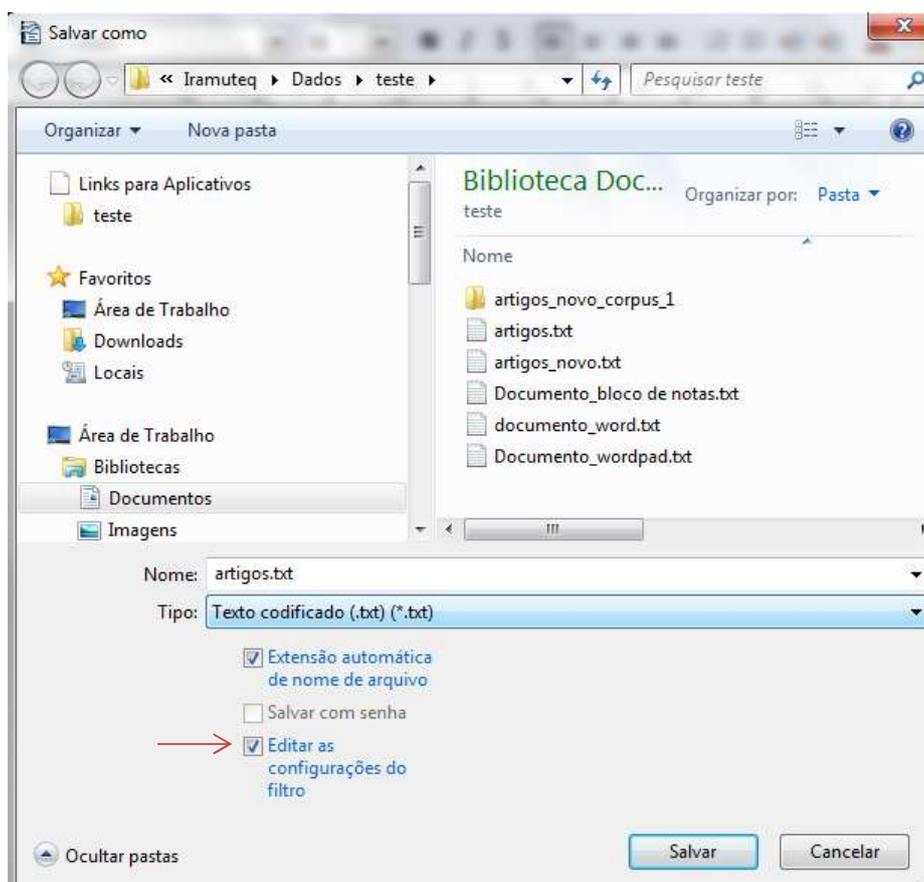
Em seguida, juntaram-se todos os arquivos em um único arquivo, obedecendo-se a numeração em ordem crescente dos artigos. Cada artigo iniciou em uma nova linha, conforme exemplo abaixo, Figura 2.

**Nota:** Acrescentou-se, no início do arquivo, uma linha em branco, pois se notou que o Iramuteq estava desprezando o primeiro artigo. O acréscimo dessa linha em branco resolveu o problema.

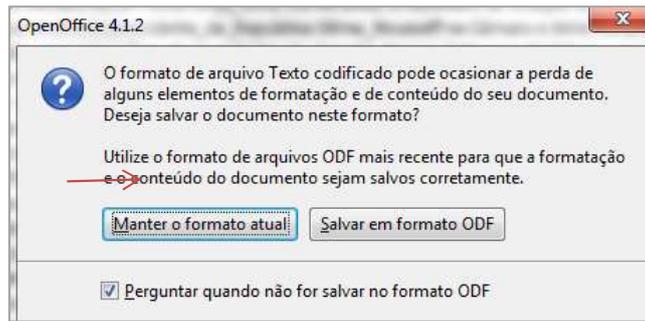
O arquivo salvo no Word com a extensão docx foi, então, aberto no OpenOffice Writer e salvo com a opção de “Salvar como” “Texto codificado (.txt)”, “Editar configurações do filtro” e “Manter formato atual: Figuras 3, 4 e 5.



**Figura 3 – Texto codificado**

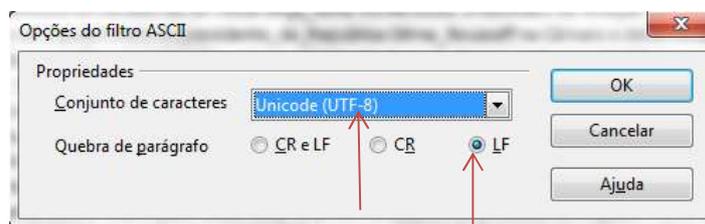


**Figura 4 – Editar configurações do filtro**



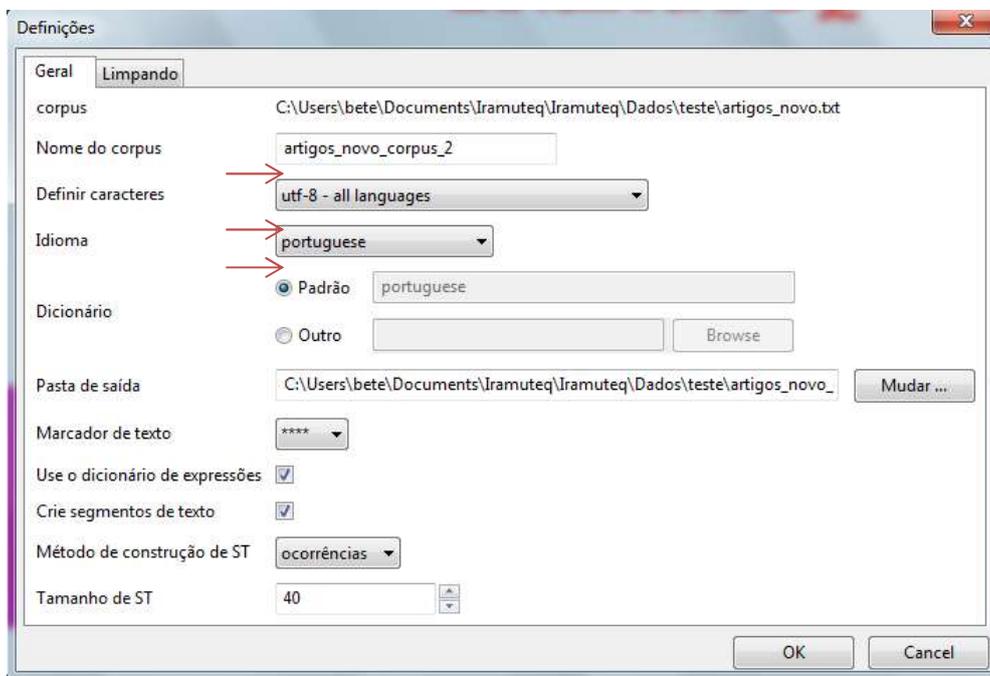
**Figura 5 – Manter o formato atual**

Em seguida, escolheu-se o formato Unicode (UTF-8) e LF (Line Feed), ou seja, alimentação de linha, Figura 6.



**Figura 6 – Formato UTF-8 e LF**

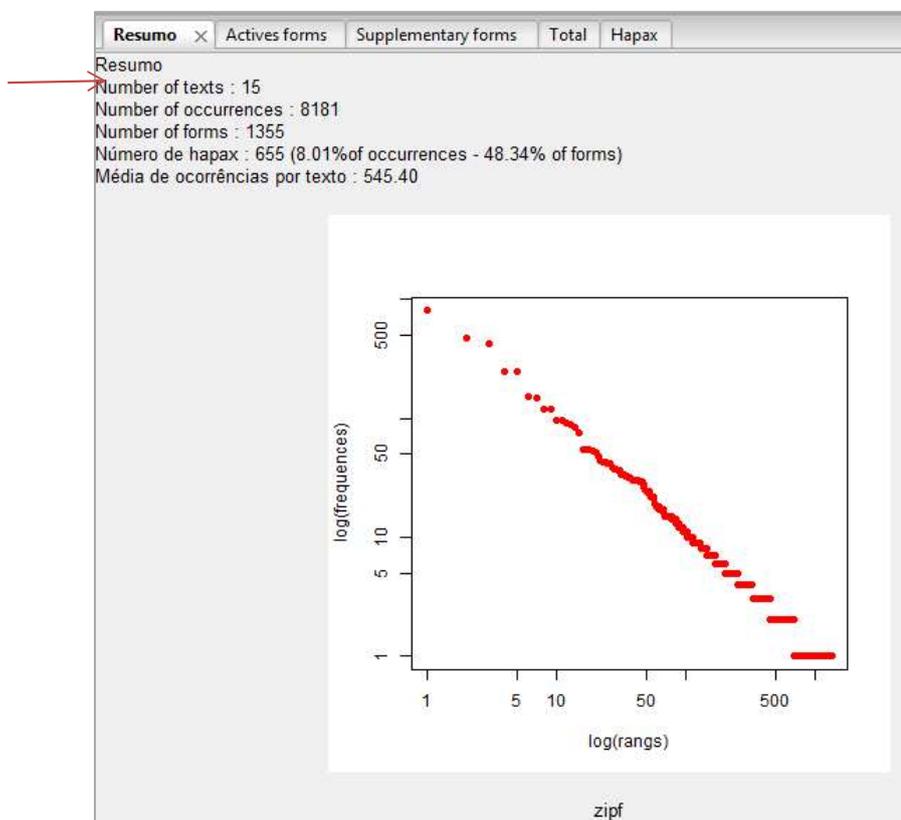
O arquivo novo gerado com extensão txt foi aberto no Iramuteq e escolhidas as opções de abertura conforme figura 7.



**Figura 7 – Opções de abertura do corpus**

Não foi reportado nenhum problema e o arquivo foi importado corretamente. Em seguida, solicitou-se a execução da Análise Estatística (Análise de texto/Estatística).

O sistema selecionou as palavras ativas, suplementares e hápax (frequência 1) apresentou o relatório da Figura 8.



**Figura 8 – Análise estatística textual**

É importante, nesse ponto, que o usuário verifique se o total de textos considerados é o mesmo que o corpus preparado.

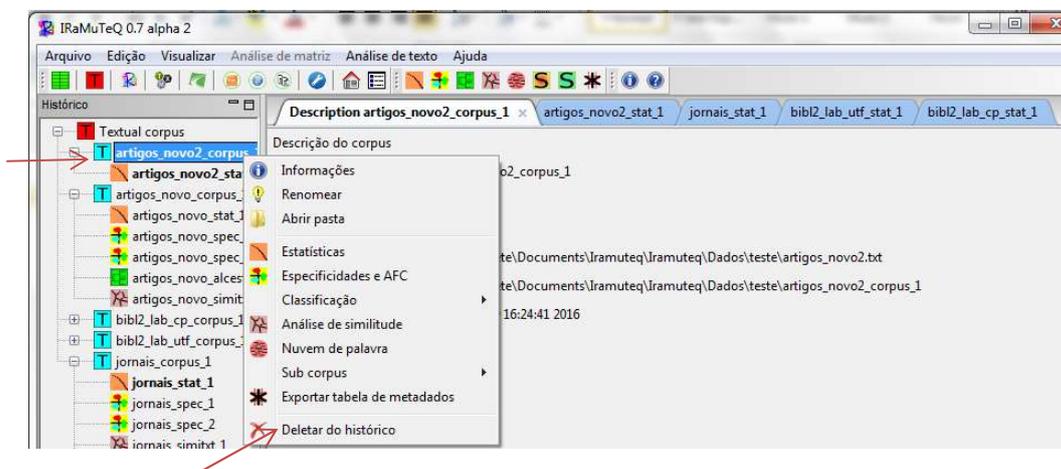
Em seguida, abriu-se a Lista de Palavras Ativas e verificou-se a lista criada. Nesse momento, notaram-se problemas na contabilização das frequências devido à existência de muitos sinônimos para um mesmo conceito. A Tabela 1 traz algumas das correções que foram necessárias. O corpus foi todo corrigido e, então, carregado novamente no Iramuteq.

Palavra	Substituição
Casa	Senado ou Câmara conforme o caso.
Petista	Presidente_da_República Dilma_Rousseff
Dilma	Presidente_da_República Dilma_Rousseff
Dilma Rousseff	Presidente_da_República Dilma_Rousseff
Presidente Dilma	Presidente_da_República Dilma_Rousseff
Renan Calheiros	presidente_do_Senado Renan_Calheiros
Renan	presidente_do_Senado Renan_Calheiros
Presidente do Senado	presidente_do_Senado Renan_Calheiros
Cunha	presidente_da_Câmara Eduardo_Cunha
Eduardo Cunha	presidente_da_Câmara Eduardo_Cunha
Presidente da Câmara	presidente_da_Câmara Eduardo_Cunha
Levandowski	presidente_do_STF Levandowski
Presidente do STF	presidente_do_STF Levandowski
Presidente do Supremo	presidente_do_STF Levandowski
comissão	comissão_especial
comissão especial	comissão_especial

**Tabela 1 – Padronização de palavras no corpus**

Para não ficarem duas bases no Iramuteq, excluiu-se primeiro a base de dados do Iramuteq, Figura 9.

Selecionou-se do lado esquerdo o arquivo (da hierarquia maior) e solicitou-se “Deletar do histórico”. Além disso, utilizando-se o Windows Explorer excluiu-se a pasta criada para os resultados do corpus. Esse procedimento evita confusões por manter vários dados do mesmo corpus ativos no Iramuteq.



**Figura 9 – Deletar a base de dados do Iramuteq**

A Figura 10 mostra a lista de formas ativas. Nela, verificou-se a necessidade de colocar os Advérbios (palavras selecionadas na lista) como formas suplementares para não prejudicar a análise das formas mais importantes encontradas.

Resumo	Actives forms	Supplementary forms	Total
Forma	Freq.	Tipos	
não	89	adv	
impeachment	79	nom	
dilma_rousseff	61	nr	
processo	61	nom	
senado	45	nom	
câmara	43	nom	
ao	42	adv	
deputado	42	nom	
presidente_da_repúbli...	38	nr	
stf	35	nr	
eduardo_cunha	34	nr	
comissão	33	nom	
senador	31	nom	
votação	31	nom	
voto	31	nom	
temer	29	ver	
decisão	27	nom	
plenário	27	adj	
como	25	adv	
até	24	adv	
partido	23	nom	
governo	22	nom	
sessão	22	nom	
votar	22	ver	
mais	20	adv	
presidente_da_câmara	20	nr	
pedido	19	nom	
já	18	adv	

Resumo	Actives forms	Supplementary forms	Total
Forma	Freq.	Tipos	
dia	15	nom	
vice_presidente	15	nr	
ainda	14	adv	
cargo	14	nom	
especial	14	adj	
parlamentar	14	nom	
presidente	14	nom	
bloco	13	nom	
crime	13	nom	
decidir	13	ver	
denúncia	13	nom	
lewandowski	13	nr	
líder	13	nom	
responsabilidade	13	nom	
assumir	12	ver	
dar	12	ver	
pmdb_rj	12	nr	
querer	12	ver	
renan_calheiros	12	nr	
rito	12	nom	
analisar	11	ver	
antes	11	adv	
hora	11	nom	
nota	11	nom	
parecer	11	nom	
prazo	11	nom	
também	11	adv	
terça_feira	11	nr	

Figura 10 – Lista de Palavras Ativas

As formas suplementares também são importantes e podem ser utilizadas em várias análises. Proporcionam maior robustez à análise textual. Porém, ao analisar a lista criada, Figura 11, notou-se a necessidade de eliminar alguns tipos gramaticais, considerados como suplementares, por não acrescentarem nada ao texto e ao mesmo tempo prejudicarem a visualização dos resultados. Eles estão selecionados nas listas da Figura 11.

Resumo	Actives forms	Supplementary forms	Total
Forma	Freq.	Tipos	
ser	148	ver_sup	
não	97	adv	
ir	88	ver_sup	
ter	53	ver_sup	
ao	48	adv	
poder	43	ver_sup	
dizer	36	ver_sup	
haver	31	ver_sup	
como	30	adv	
fazer	25	ver_sup	
até	24	adv	
mais	23	adv	
estar	22	ver_sup	
já	21	adv	
dois	18	adj_num	
segundo	17	adj_num	
_	15	nr	
ainda	15	adv	
dever	15	ver_sup	
essa	14	pro_dem	
isso	14	pro_ind	
todo	14	pro_ind	
antes	12	adv	
também	12	adv	
porque	11	adv	
cada	10	pro_ind	
maioria	10	nom_sup	

Resumo	Actives forms	Supplementary forms	Total
Forma	Freq.	Tipos	
muito	9	adv	
quando	9	adv	
primeiro	8	adj_num	
apenas	7	adv	
depois	6	adv	
outro	6	pro_ind	
qualquer	6	pro_ind	
quanto	6	pro_ind	
além	5	adv	
assim	5	adv	
entanto	5	adv	
então	5	adv	
inicialmente	5	adv	
tempo	5	nom_sup	
três	5	adj_num	
último	5	adj_sup	
agora	4	adv	
algum	4	pro_ind	
menos	4	adv	
qual	4	pro_ind	
quatro	4	adj_num	
sim	4	adv	
absolutamente	3	adv	
apesar	3	ono	
bem	3	adv	
dez	3	adj_num	
esse	3	pro_dem	

Figura 11 – Lista de Palavras Suplementares

Foram, portanto, eliminados os seguintes tipos gramaticais, Figura 12:

- artigos indefinidos
- artigos definidos
- pronomes relativos
- pronomes pessoais
- pronomes possessivos
- números
- adjetivos numéricos
- preposições

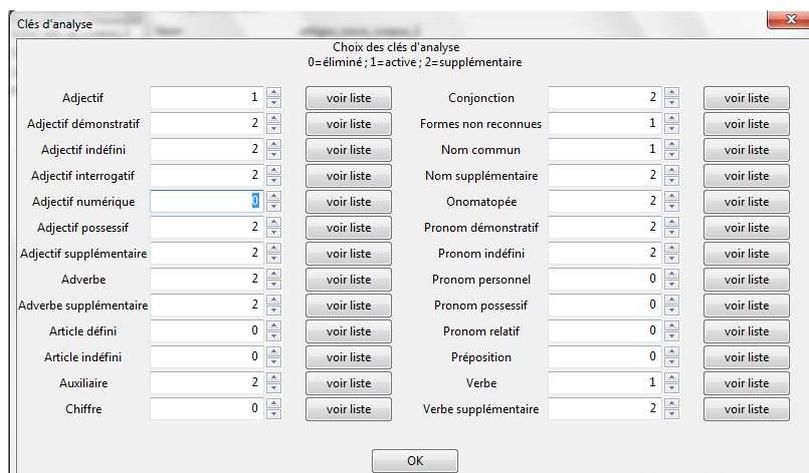


Figura 12 – Tipos Gramaticais Eliminados

## Resultados das Análises

Apresentam-se, a seguir, algumas análises realizadas sobre o corpus. O pesquisador deve observar, no entanto, que nem todos grafos gerados são úteis para elucidar o conteúdo textual, por isso, é importante analisa-lo e, então, decidir sobre sua utilidade ou não.

### 1) Análise de Especificidades e AFC

Foram utilizadas as seguintes configurações: formas ativas e suplementares; variável: jornal, lei hipergeométrica. O grafo gerado foi reeditado, conforme configuração da Figura 13.

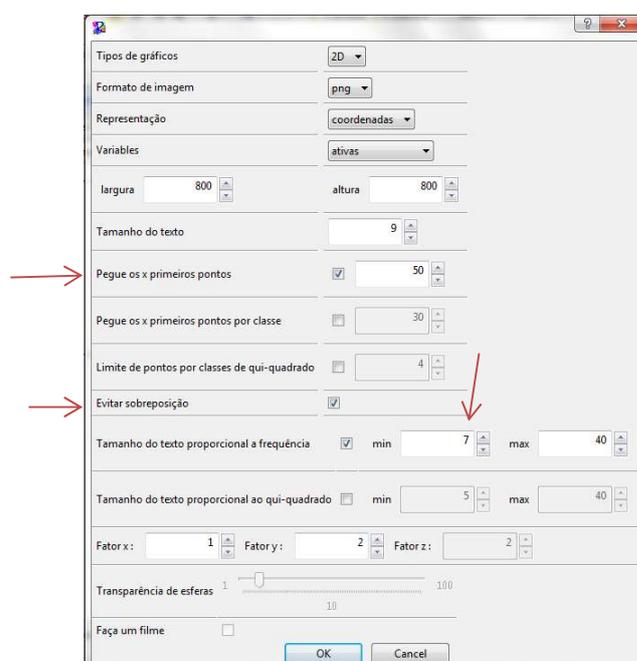
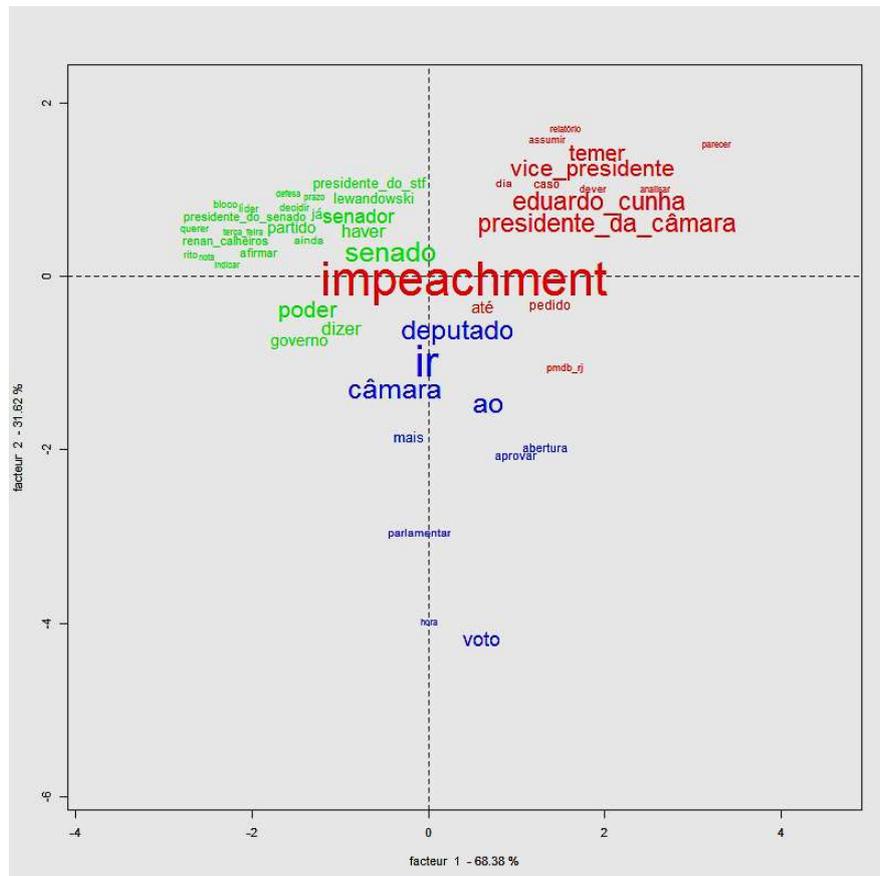


Figura 13 – Opções de configuração do grafo de AFC

As setas vermelhas mostram os parâmetros que foram modificados. O grafo resultante é mostrado na Figura 14.



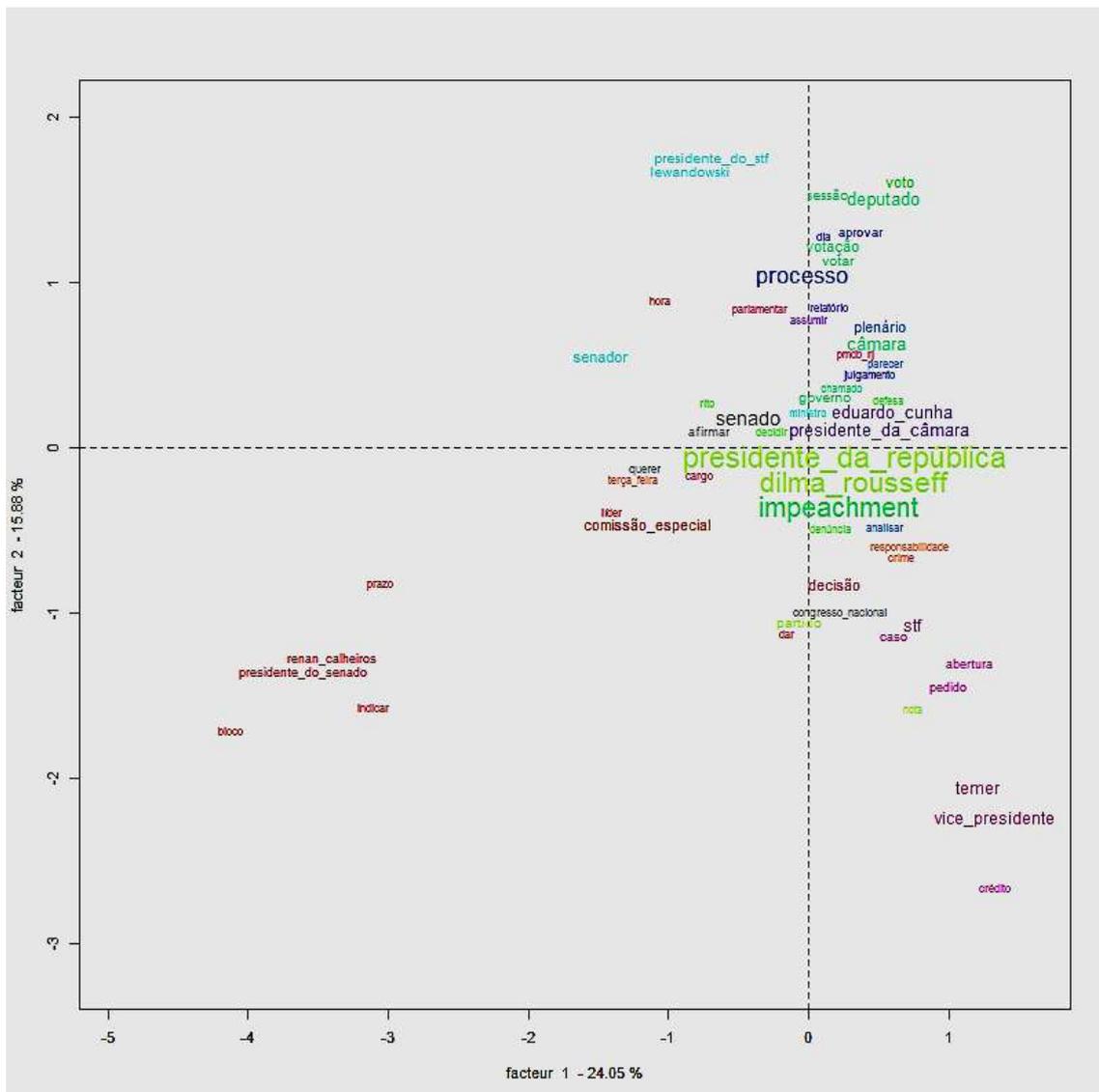
**Figura 14 – Grafo AFC por jornal**

Observe que o grafo de AFC representa o conteúdo de cada jornal: Correio Brasiliense em vermelho; O Globo em verde e em azul o UOL (apenas um artigo). No centro aparece a palavra Impeachment, tema principal do corpus.

No segundo grafo de Especificidades e AFC, Figura 15, solicitou-se a análise de cada artigo. Utilizou-se a mesma configuração anterior, mas alterando-se a variável para artigo.

Nessa figura, o centro indica o tema: impeachment, Dilma Rousseff, Presidente da República.

Ao redor estão os principais atores do processo: Eduardo Cunha; Renan Calheiros, Levandowski e Temer vice-presidente, bem como as principais palavras-chaves do texto: crime de responsabilidade, denúncia, decisão, congresso nacional, processo, comissão especial.



**Figura 15 – Grafo AFC por artigo**

## 2) Classificação: Método Reinert

As Figuras 16 e 17 mostram o Dendrograma em dois formatos diferentes e na Figura 18, a seguir, a AFC.

Foram encontradas cinco classes de discurso, bastante equilibradas entre si. Porém apenas duas classes formam a hierarquia superior. A Classe 5 em lilás e todas as outras quatro classes. A classe 5 trata das ações do Senado Federal. A classe 4 trata do vice-presidente Michel Temer e o pedido para seu impeachment. Em conjunto está a classe 3, as ações da Câmara com Eduardo Cunha e Waldir Maranhão para o impeachment da Presidente Dilma Rousseff. As classes 1 e 2 estão unidas hierarquicamente. A primeira trata do parecer favorável para a instauração do processo de afastamento. A classe dois trata da sessão de votação realizada no domingo 17/04/2016. Essas duas se juntam também as outras duas classes (4 e 3).

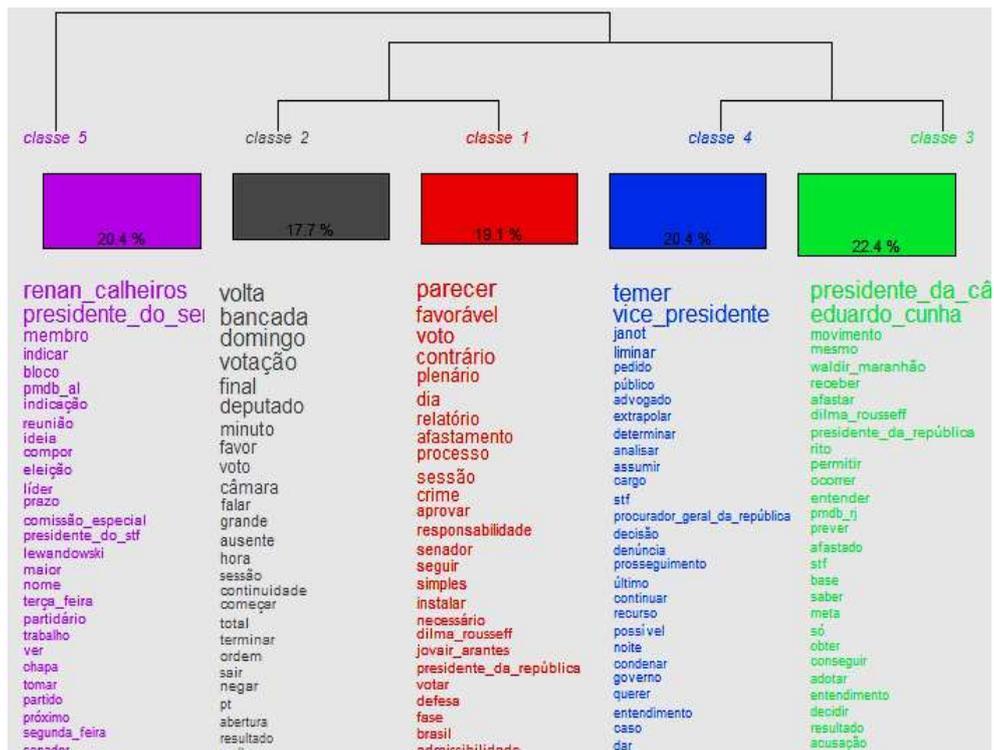


Figura 16 – Dendrograma (formato 1)

A Figura 8 traz as mesmas informações em um formato diferente.

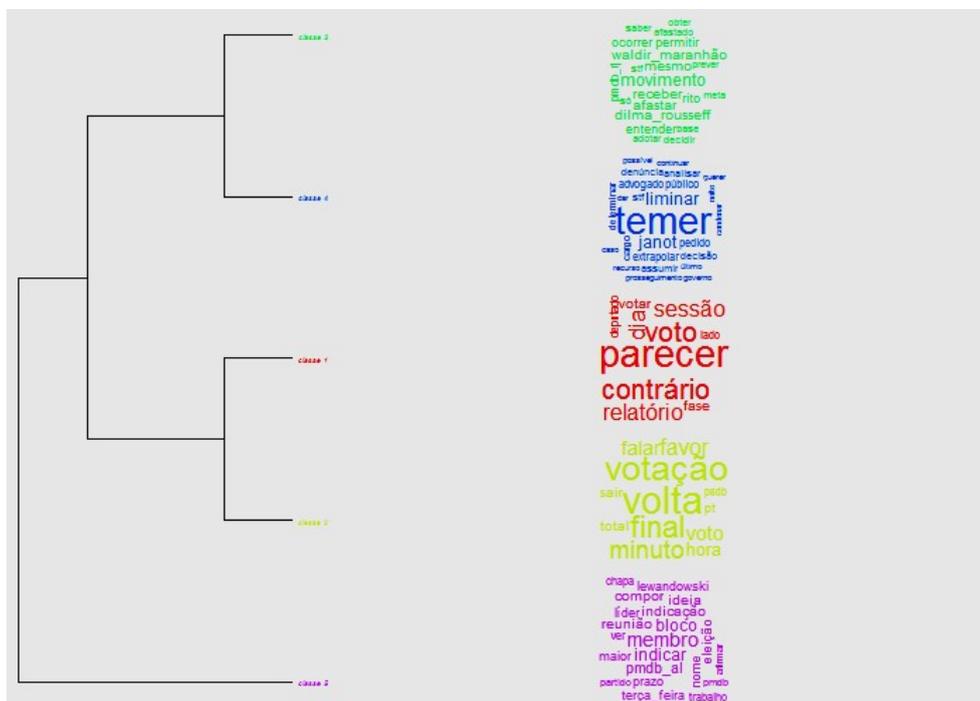


Figura 17 – Dendrograma (formato 2)

A Figura 18 mostra a Análise Fatorial por Correspondência pelas cinco classes do Dendrograma. Notam-se as cores correspondentes a cada classe.

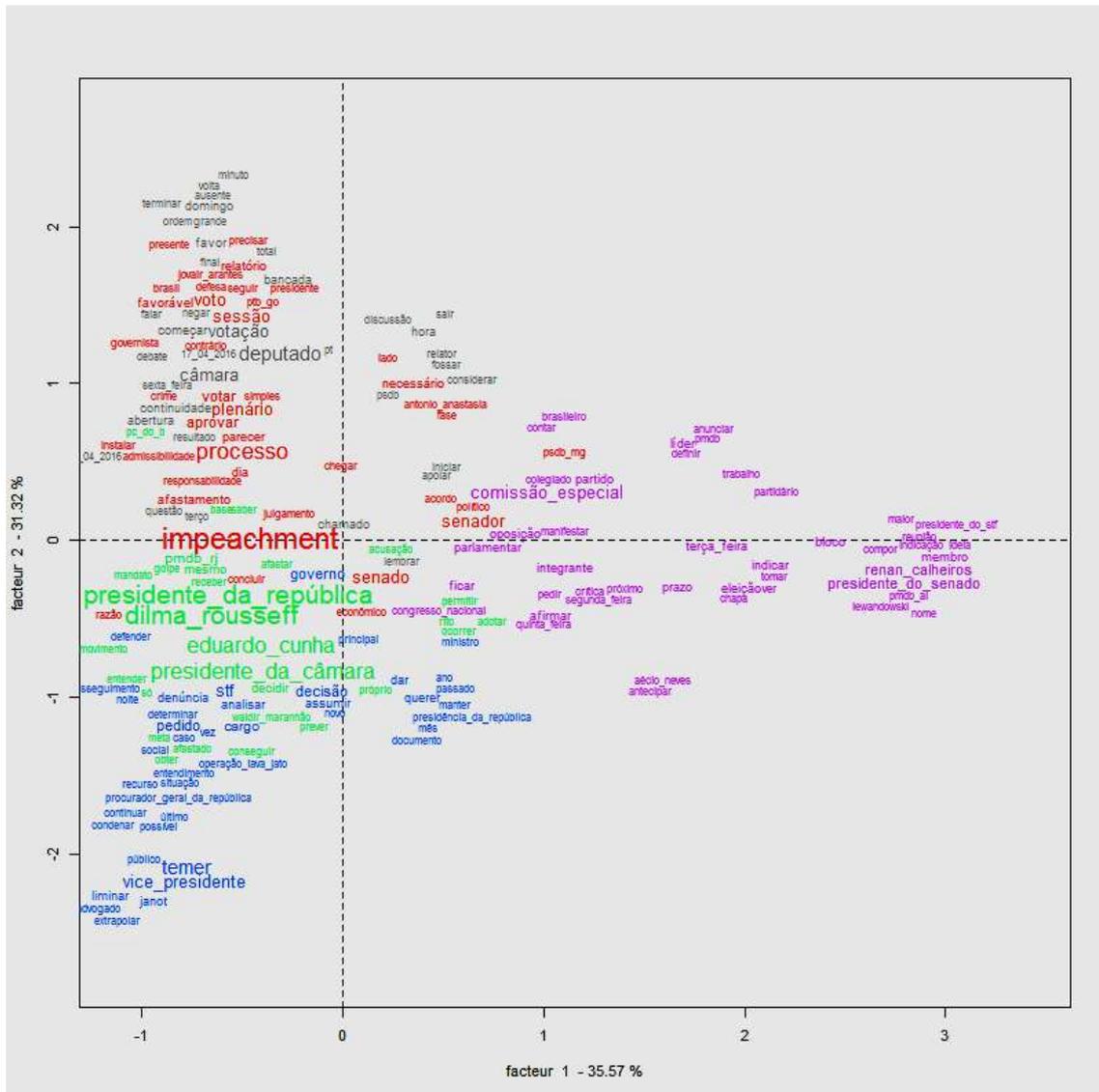


Figura 18 – AFC pelas cinco classes

### 3) Análise de Similitude

A Figura 19 mostra as configurações adotadas para o grafo de similitude: mostra a escolha das palavras desejadas na janela do lado esquerdo; e, no lado direito, a configuração do grafo. O resultado está apresentado na Figura 20.

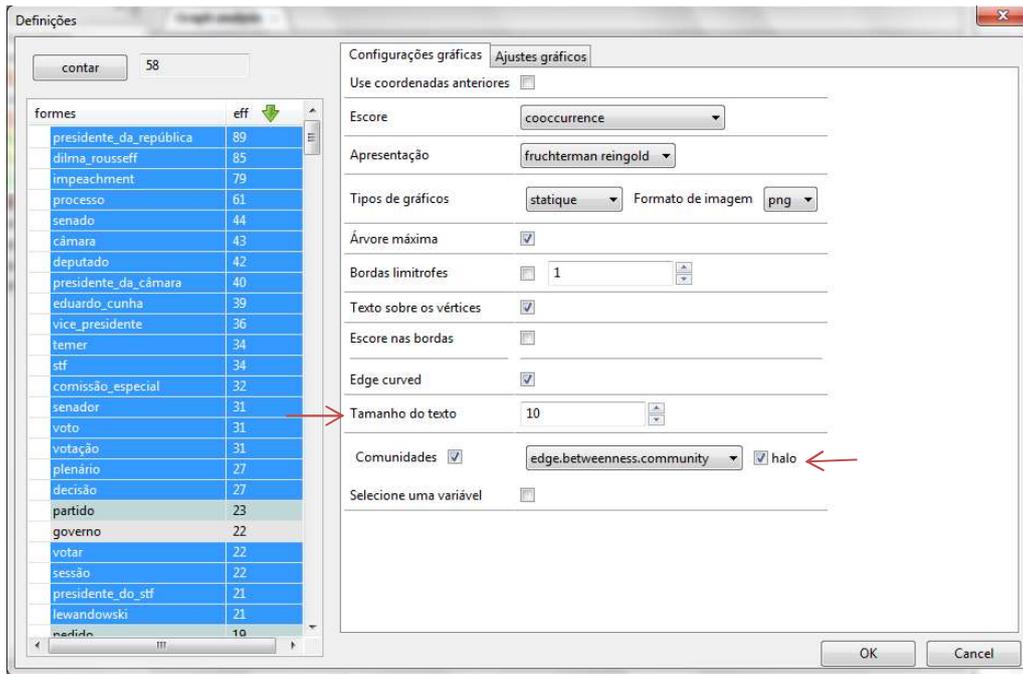


Figura 19 – Configurações do Grafo de Similitude

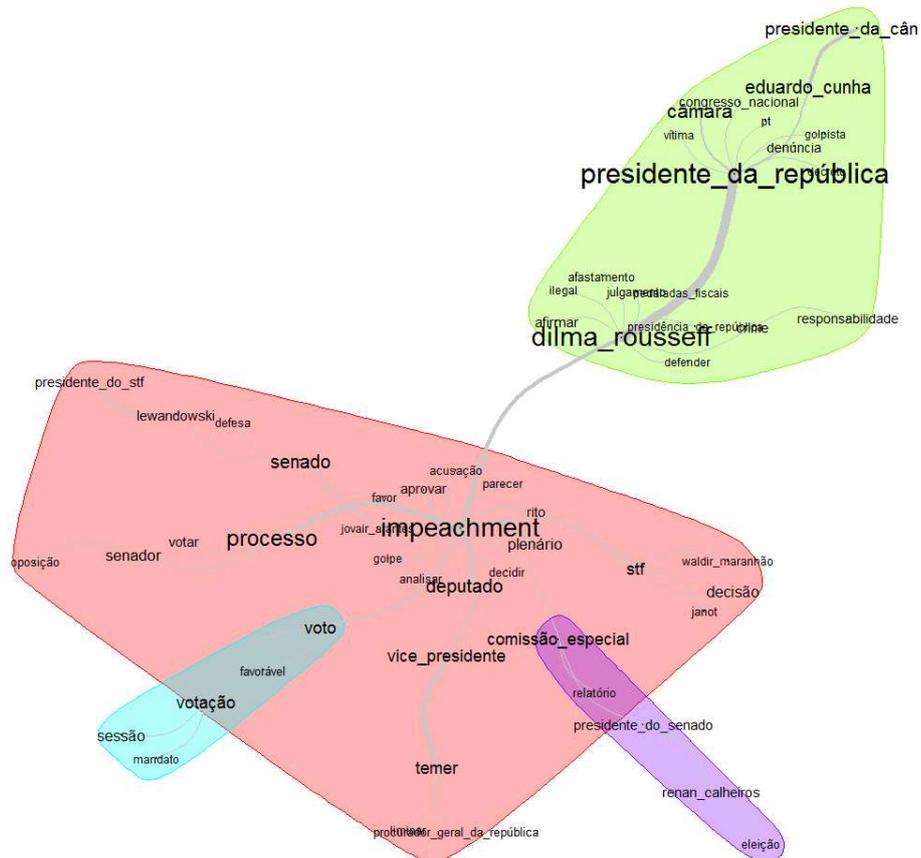


Figura 20 – Grafo de Similitude

#### 4) Nuvem de palavras

Para execução da Nuvem de palavras, utilizaram-se as configurações mostradas nas Figuras 21 e 22. As configurações alteradas, Figura 21, estão marcadas com a seta vermelha.



Figura 21 – Configuração da Nuvem de palavras

A Figura 22 mostra as palavras que foram selecionadas para compor a Nuvem.

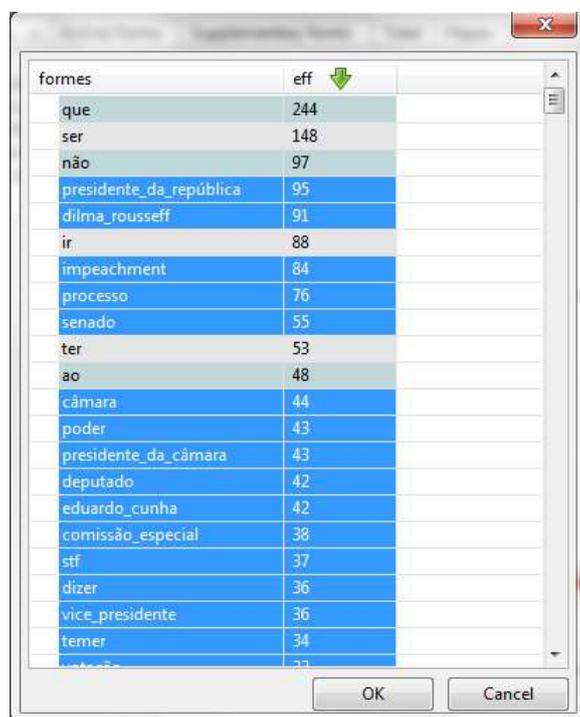


Figura 22 – Escolha das Formas para Gerar a Nuvem

