

# Tutorial (Básico) de Utilização do Iramuteq<sup>1</sup>

## INTRODUÇÃO

O Iramuteq é um software de análise textual baseado em estatísticas (Utilizando o software estatístico R) que revelem ligações e outras características textuais, o posicionamento e a estruturação de palavras no texto, de forma que sejam retornados indicadores e visualizações intuitivas sobre a estrutura e ambientes do texto proposto para análise.

## INSTALAÇÃO

Para instalar o Iramuteq é necessário **primeiro** instalar o software estatístico **R** que se encontra no seguinte link: <https://cran.r-project.org/bin/windows/base/>

### R-3.2.2 for Windows (32/64 bit)

[Download R 3.2.2 for Windows](#) (62 megabytes, 32/64 bit)  
[Installation and other instructions](#)  
[New features in this version](#)

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the [md5sum](#) of the .exe to the [true fingerprint](#). You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

#### Frequently asked questions

- [How do I install R when using Windows Vista?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

#### Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

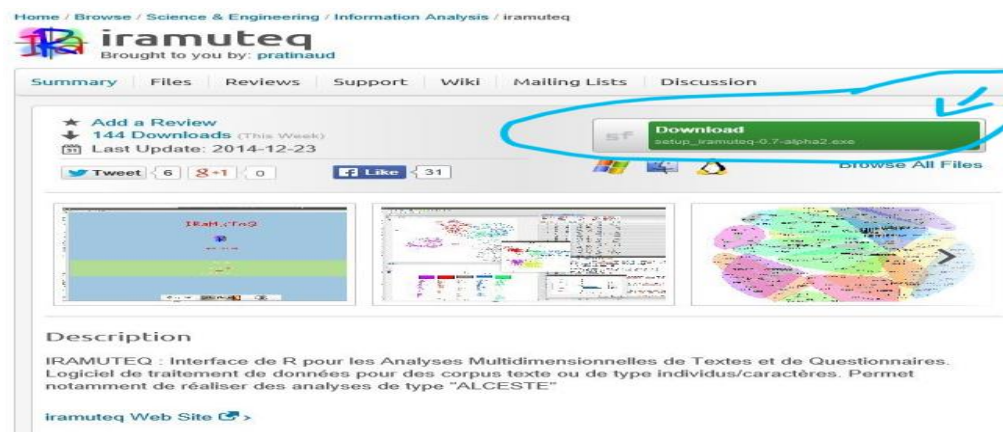
Note to webmasters: A stable link which will redirect to the current Windows binary release is [CRAN.MIRROR>bin/windows/base/release.htm](#).

Last change: 2015-08-14, by Duncan Murdoch

### Download R

A instalação do software R, é bem simples, clique em avançar e escolha as configurações recomendadas.

A pós instalar o R corretamente, chegou a hora de instalar o Iramuteq, ele se encontra para download neste link: <http://sourceforge.net/projects/iramuteq/>



Home / Browse / Science & Engineering / Information Analysis / iramuteq

**iramuteq**  
Brought to you by: pratinaud

Summary | Files | Reviews | Support | Wiki | Mailing Lists | Discussion

★ Add a Review  
↓ 144 Downloads (This Week)  
Last Update: 2014-12-23

Tweet 6 | +1 0 | Like 31

**Download**  
setup\_iramuteq-0.7-alpha2.exe  
Browse All Files

**Description**  
IRAMUTEQ : Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. Logiciel de traitement de données pour des corpus texte ou de type individus/caractères. Permet notamment de réaliser des analyses de type "ALCESTE"

[Iramuteq Web Site](#)

### Download Iramuteq

<sup>1</sup> Autor: Luis Felipe Rosa de Oliveira  
E-mail: luisfelipeprf@gmail.com

A instalação do Iramuteq também é bem simples, só clicar em avançar e esperar a instalação do software.

Ao abrir o Iramuteq pela primeira vez, ele deve automaticamente informar que são necessárias algumas bibliotecas do R, e então vai começar a fazer a sincronização e após isso o software está pronto para uso.

## PASSOS P/ UTILIZAÇÃO

**1º PASSO** – Tratamento do texto: O Iramuteq tem um padrão específico para o processamento de texto. Para inserir diferentes textos de uma vez na análise, é necessário definir as variáveis que nomearão os diferentes *corpus* de texto:

Exemplo:

```
**** *Corpus_1_Mussum
```

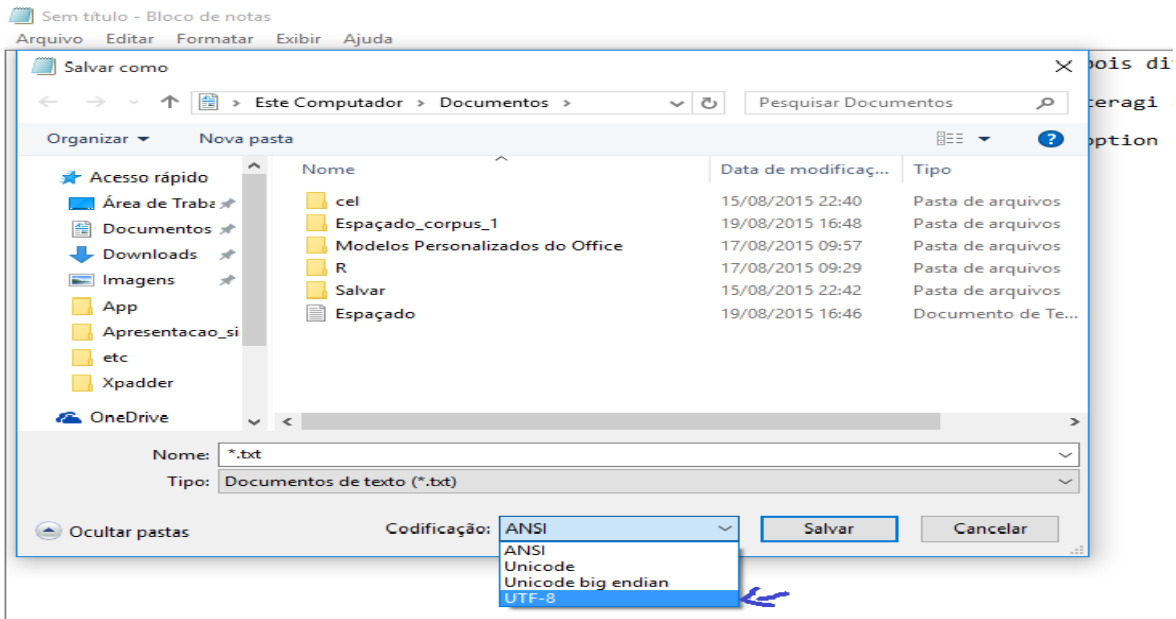
Mussum ipsum cacilds, vidis litro abertis. Consetis adipiscings elitis. Pra lá, depois divoltis porris, paradis. Paisis, filhis, espiritis santis. Mé faiz elementum girarzis, nisi eros vermeio, in elementis mé pra quem é amistosis quis leo. Manduma pindureta quium dia nois paga. Sapien in monti palavrís qui num significa nadis i pareci latim. Interessantiss quisso pudia ce receita de bolis, mais bolis eu num gostis.

Suco de cevadiss, é um leite divinis, qui tem lupuliz, matis, aguis e fermentis. Interagi no mé, cursus quis, vehicula ac nisi. Aenean vel dui dui. Nullam leo erat, aliquet quis tempus a, posuere ut mi. Ut scelerisque neque et turpis posuere pulvinar pellentesque nibh ullamcorper. Pharetra in mattis molestie, volutpat elementum justo. Aenean ut ante turpis. Pellentesque laoreet mé vel lectus scelerisque interdum cursus velit auctor. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam ac mauris lectus, non scelerisque augue. Aenean justo massa.

```
**** *Corpus_2_Lorem
```

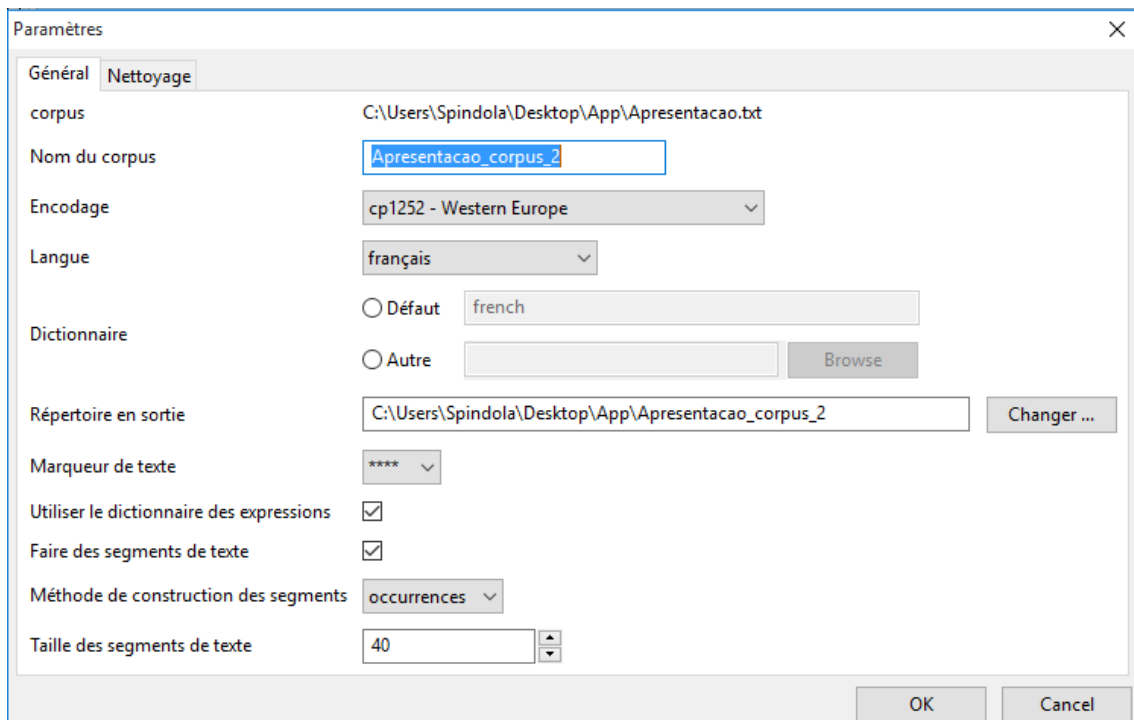
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras quam ante, vulputate vel eros id, laoreet finibus lorem. Nulla ac pretium magna. Suspendisse dictum ultrices enim quis varius. Mauris et lobortis eros, ut ornare quam. Donec vulputate congue maximus. Interdum et malesuada fames ac ante ipsum primis in faucibus. Etiam sagittis quam nibh, nec suscipit lectus venenatis id. Phasellus ut felis felis

Ao estruturar o(s) texto(s) dessa forma, é indicado que seja salvo em bloco de notas e em formato txt. Atente-se à classificação do texto em formato UTF-8:



\*OBS: Talvez seja necessário eliminar os parágrafos e concentrar o corpus de um tema/variável em um único texto corrido.

**2° PASSO** – Importação do Texto pelo Iramuteq: Para importar o texto já estruturado em formato txt, inicie o iramuteq, e vá em “Fichier”(Arquivo) → “Ouvrir un corpus texte”(Abrir um corpus de texto) → Escolha o arquivo e vá em abrir. Vai aparecer a seguinte janela de parâmetros de importação:



- Na aba “Général”, Geral:

- Em “Corpus” temos o endereço do arquivo importado.
- Em “Nom du Corpus”, é possível definir o nome dado ao arquivo principal de análise.
- Em “Encodage”, têm-se as configurações de codificação do texto. **IMPORTANTE: Se você salvou o arquivo TXT em formato UTF-8, é necessário selecionar neste ponto a codificação UTF-8 (última opção) ao invés de cp1252.**
- Em “Langue” é necessário selecionar a linguagem do texto a ser analisado. O Iramuteq tem dicionários de análise completo das seguintes línguas: francês, inglês, italiano, português e espanhol; As línguas: alemão, sueco, grego e gálico ainda estão em experimentação.
- Em “Dictionnaire”, deixe por padrão o dicionário da língua escolhida para análise.
- Em “Répertoire en sortie” você pode escolher em qual diretório deseja salvar a pasta com as análises feitas.
- Em “Marquer du texte” deixe por padrão os quatro asteriscos, que delimitam os corpus de textos.
- Em “Utiliser le dictionnaire des expressions”, deixe selecionado para utilizar o dicionário de expressões.
- Em “Faire des segments de texte”, deixe selecionado para que o software separe o texto em segmentos de texto (processo em que o software realiza uma separação do corpus textual em segmentos de texto para a análise fracionada e identificação de ambientes léxicos).
- Em “Méthode de construction des segments”, selecione “ocurrences” para segmentar os corpus textuais pela frequência das palavras, “caractères” para segmentar a partir dos caracteres e “paragraphes” para segmentar por parágrafos (por padrão deixe por frequência).
- Em “Tailles des segments de texte” selecione o número que definirá o tamanho dos segmentos de texto. (Por padrão deixe 40, se quiser que os ambientes de análise sejam menores ou maiores, aumente ou diminua o número respectivamente).

- Na aba “Nettoyage” (Limpeza):

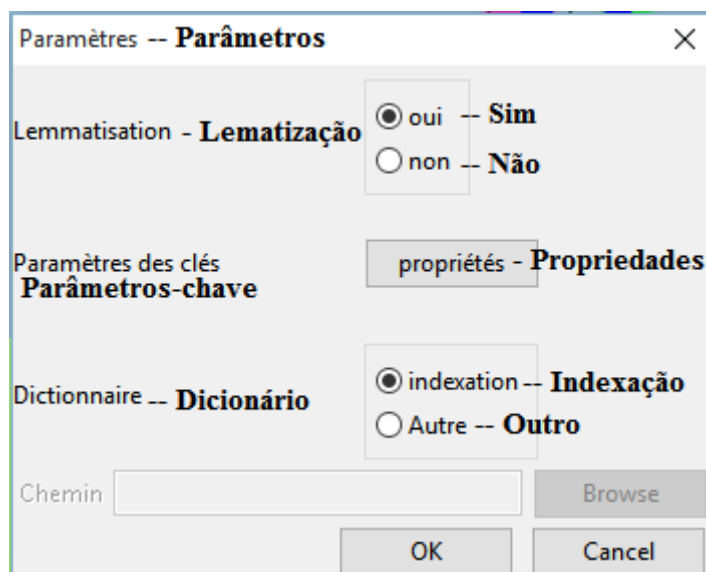
- Em “Mettre le texte en minuscule”, deixe marcado para padronizar o texto em letras minúsculas.
- Em “Remplacer les apostrophes par des espaces”, deixe selecionado para substituir as aspas simples por espaços (em caso de análise por tabela).
- Em “Replacer les tirets par des espaces”, deixe selecionado para substituir traços por espaços.
- Em “Conservar la ponctuation”, selecione somente se desejar conservar a pontuação do texto.

- Em “Pas d’espace entre deux formes”, selecione somente se desejar que os espaços entre duas formas sejam removidos.

**3º PASSO – ANÁLISES DO TEXTO:** Existem 5 tipos de análises textuais que o Iramuteq processa: Estatísticas textuais, Especificidades e AFC, CHD, Análise de Similitude, Nuvem de Palavras, cada uma tem características e reflexões específicas.

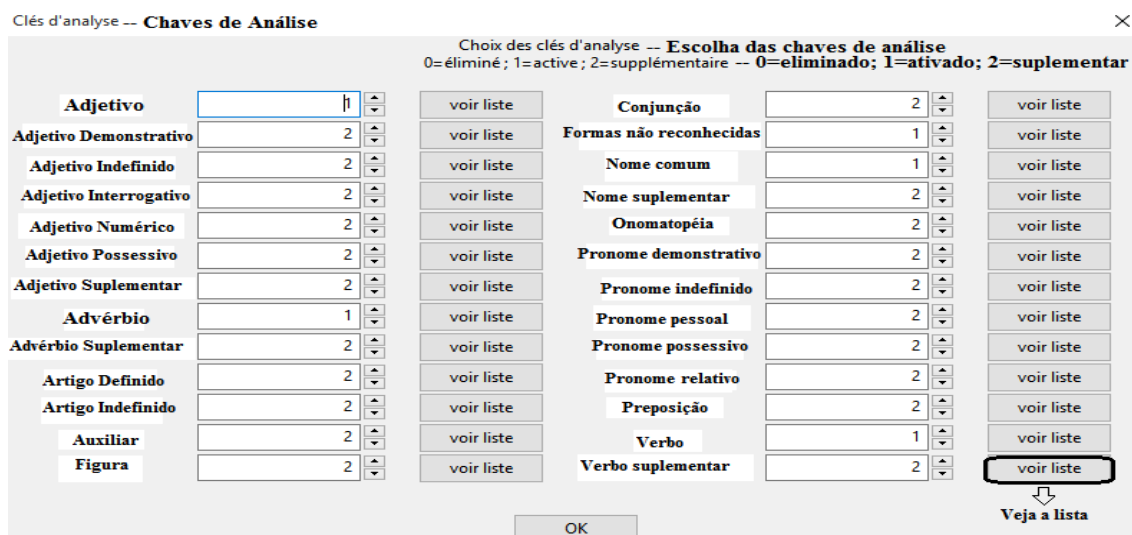
Para iniciar as análises clique em “Analyse de texte”, esta é a parte em que tratamos as análises do texto em si.

\*OBS: Antes de cada análise aparece uma janela para a definição dos parâmetros:



Escolha **Sim** em **Lematização** se quiser que o Iramuteq faça uma classificação de formas reduzidas de algumas palavras, por exemplo: criação, criativo, criacionismo serão classificados como criar, porém, mesmo assim cada uma será classificada de acordo com seu tipo gramatical (Recomenda-se que deixe marcado sim, para melhor aproveitamento das análises).

Ao selecionar **Propriedades** em **Parâmetros-chave**, é aberta uma janela com algumas opções, em que você pode modificar como o Iramuteq analisa as palavras\*:

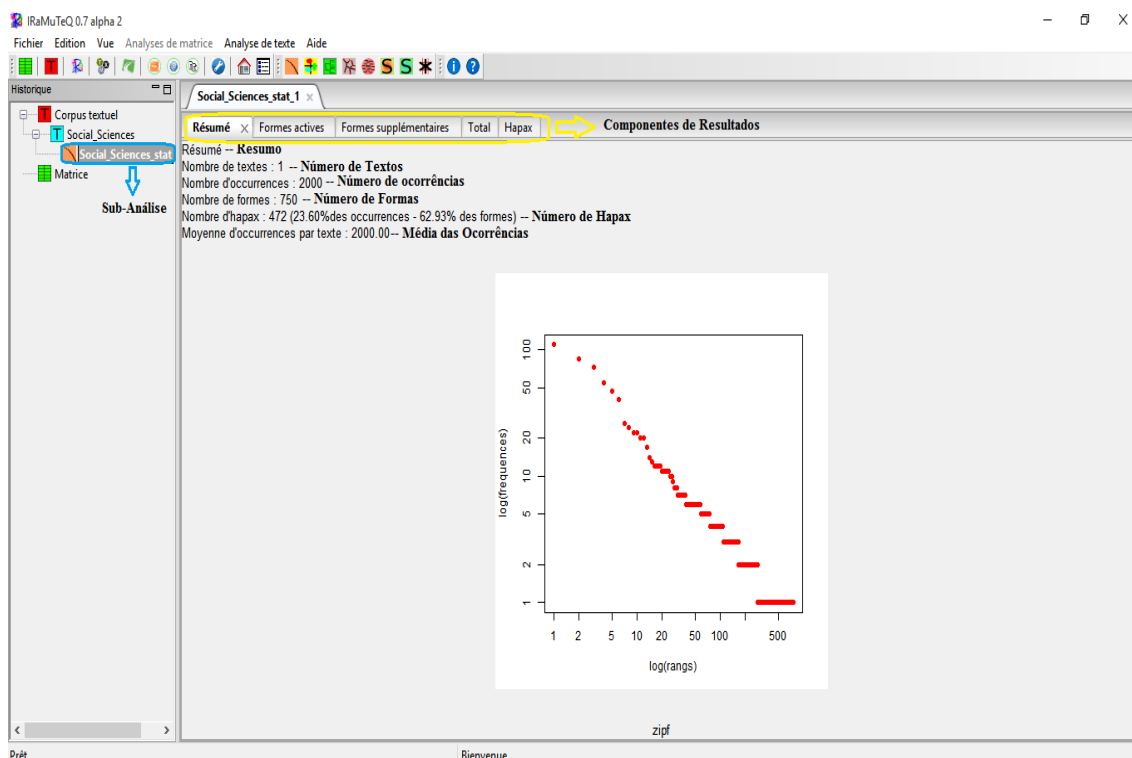


\*0 para eliminar a o tipo, 1 para deixar o tipo como ativo e 2 para classifica-lo como complementar – mais a frente você terá a opção de escolher analisar somente os ativos, suplementares ou os dois.

Deixe selecionado **Indexação** em **Dicionário** para que o Iramuteq utilize o dicionário padrão da língua, possibilitando que seja feito o cruzamento entre as formas do dicionário e a do corpus analisado.

## TIPOS DE ANÁLISES:

1º - **“STATISTIQUES” (ESTATÍSTICAS)** – Esta opção gera uma das sub-análises do corpus com 5 componentes (abas) de resultados:



O **gráfico** apresentado na aba **“Résumé”** mostra a relação entre a frequência e a quantidade de formas do corpus em análise. Por exemplo, no gráfico apresentado acima, quanto maior é a frequência de uma forma/palavra, menor é a sua quantidade no texto. Sendo assim, neste caso as palavras com frequência 1 existem em grande quantidade no texto, e palavras com frequência 100 aparecem em menor quantidade (Padrão normal de corpus textuais menores).

Nas **demais abas**: **“Formes actives”** (Formas ativas), **“Formes Supplémentaires”** (Formas Suplementares) e **“Total”**, é possível identificar as frequências das palavras de cada categoria, e em **“Hapax”** é possível visualizar as palavras com frequência igual a 1.

2º - “**SPÉCIFICITÉS ET AFC**” (ESPECIFICIDADES E AFC), nesta análise são retornados as frequências e os valores de correlação Qui<sup>2</sup> de cada palavra do corpus a partir da frequência pré-definida. (Recomenda-se mais de uma variável para melhor proveito desta análise).

Ao realizar esta análise aparece uma tela de **parametrização**:

Nesta janela você pode escolher quais formas quer utilizar, se prefere selecionar por variáveis ou o todo o corpus, escolher o tipo de índice e a frequência mínima das palavras a serem analisadas.

**OBS.** Ambos os índices são usados para mostrar a probabilidade de existência da correlação entre as formas/palavras e as variáveis trabalhadas ou o corpus.

**OBS<sup>2</sup>.** Para fazer análise entre variáveis, escolha a seleção por modalidades e selecione as variáveis que deseja comparar.

Os resultados retornam 7 componentes de análise, cada um com seus diferentes resultados:

Formes	*Business	*Social_Sciences
financial	3.2858	-3.2858
information	2.1309	-2.1309
such	1.6312	-1.6312
word	1.6312	-1.6312
the	1.4751	-1.4751
we	1.1173	-1.1173
datum	-1.0272	-1.0272
on	0.7857	-0.7857
s	0.7497	-0.7497
paper	0.7497	-0.7497
in	0.7014	-0.7014
context	0.4869	-0.4869
which	0.4801	-0.4801
approach	0.4757	-0.4757
this	0.4069	-0.4069
to	0.3997	-0.3997
for	0.3817	-0.3817
is	-0.3745	-0.3745
can	-0.3626	-0.3626
two	0.3535	-0.3535
social	0.3535	-0.3535
as	0.3491	-0.3491
that	0.3491	-0.3491
are	0.3027	-0.3027
result	0.2556	-0.2556
using	0.2437	-0.2437
with	-0.2382	0.2382
their	-0.2542	0.2542
experience	-0.2542	0.2542

As abas “Formes” (Formas), “Formes banales” (Formas Banais), “Fréquences des formes” (Frequência das formas), mostram a frequência das palavras no corpus ou variáveis selecionadas e os índices (hipergeométrico/qui<sup>2</sup>). As abas “Types” (Tipos) e “Fréquences des types” (Frequências dos Tipos), mostra o valor do índice e a frequência dos tipos das formas/palavras respectivamente: sendo **sw** = suplementar/artigos\*, **nom** = sujeito/nome, **ver** = verbo, **nr** = não reconhecido, **num** = numeral, **adj** = adjetivo. E as abas sobre “Fréquences relatives” (Frequências relativas), mostram as frequências relativas das formas e dos tipos (em sequência).

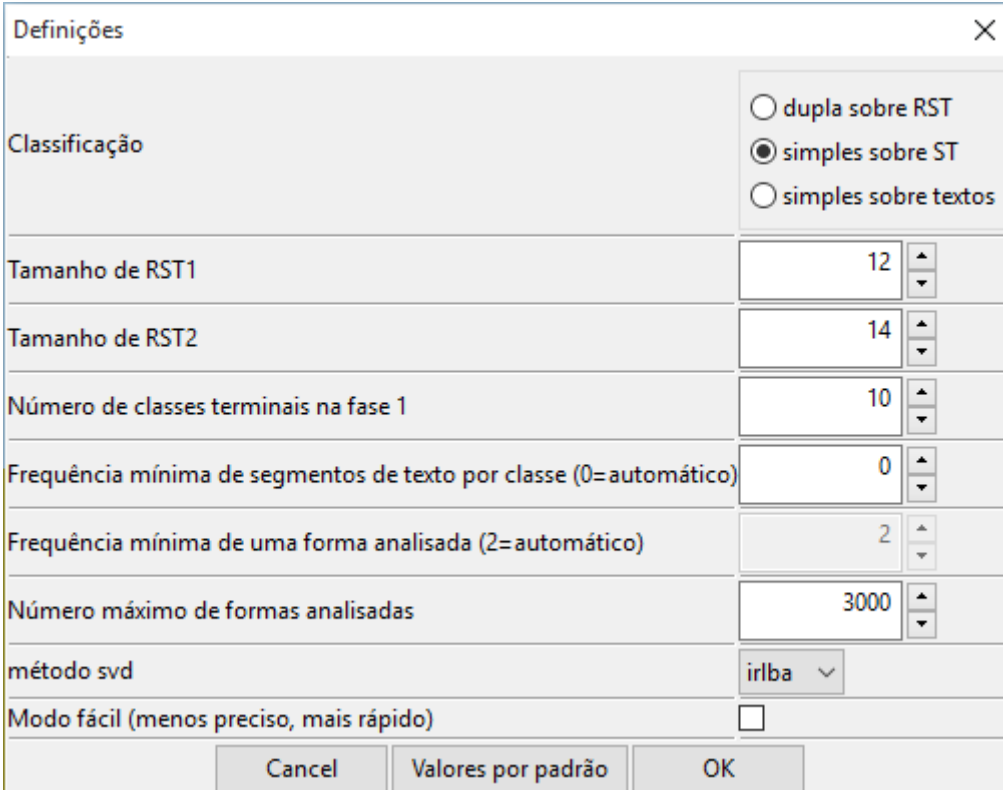
\*O significado de **sw** pode não estar correto.

### 3° - “CLASSIFICATION → MÉTHOD REINERT” (CLASSIFICAÇÃO → MÉTODO REINERT) OU CHD (CLASSIFICAÇÃO HIERÁRQUICA DESCENDENTE):

Esta é uma das análises mais importantes do Iramuteq, nela o software, ao utilizar da lógica de correlação, utiliza as segmentações do corpus textual, juntamente com a lista de formas reduzidas e o dicionário embutido para apresentar um esquema hierárquico de classes.

Ou seja, o Iramuteq, processa o texto de modo que possam ser identificadas classes de vocabulário, sendo assim, é possível inferir quais ideias o corpus textual deseja transmitir. (Vale lembrar que essa análise é feita a partir de uma lógica estatística processada por computador e aplicada de forma lexical. Resultados como este não podem ser comparados à uma análise de conteúdo ou coisa parecida.)

Ao selecionar esta opção, será aberta a seguinte janela de **preferências**:



A janela de Definições do Iramuteq apresenta as seguintes configurações:

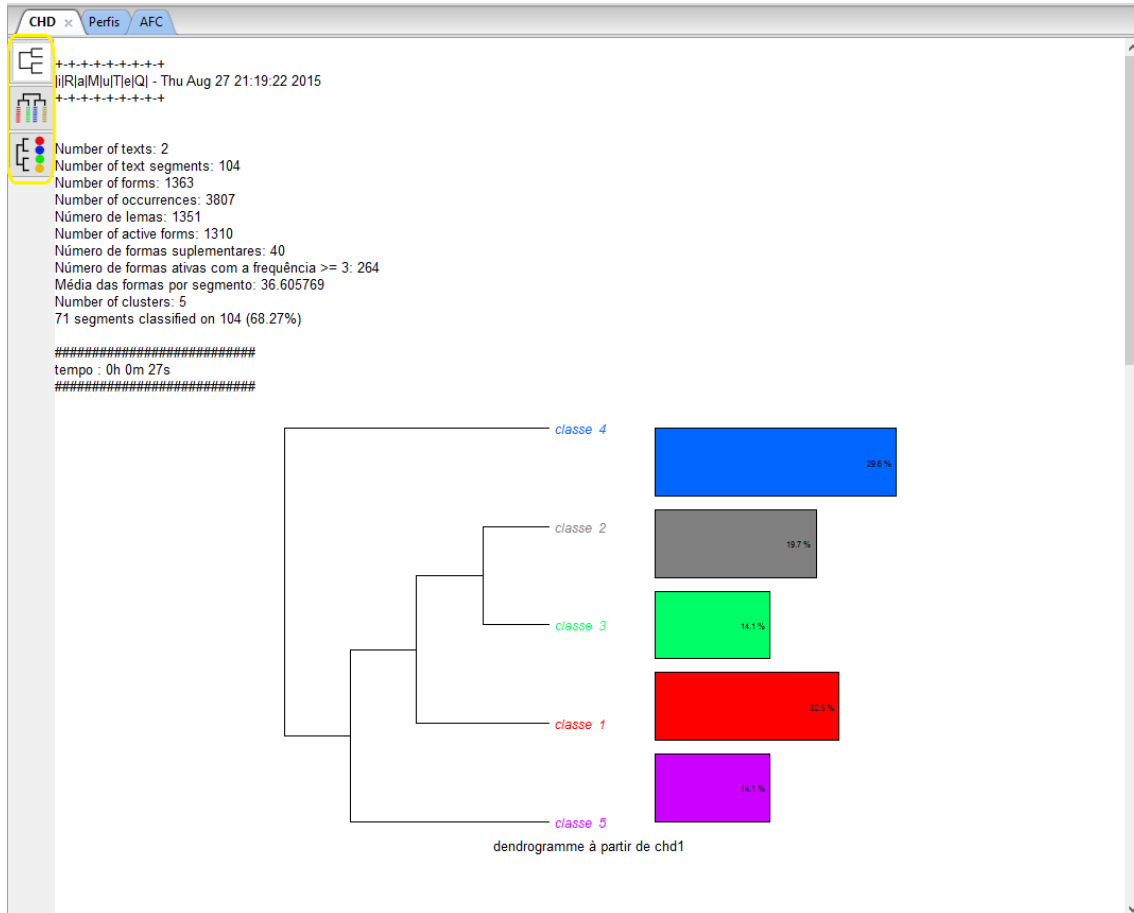
Parâmetro	Valor
Classificação	<input type="radio"/> dupla sobre RST <input checked="" type="radio"/> simples sobre ST <input type="radio"/> simples sobre textos
Tamanho de RST1	12
Tamanho de RST2	14
Número de classes terminais na fase 1	10
Frequência mínima de segmentos de texto por classe (0=automático)	0
Frequência mínima de uma forma analisada (2=automático)	2
Número máximo de formas analisadas	3000
método svd	irlba
Modo fácil (menos preciso, mais rápido)	<input type="checkbox"/>

Botões de ação: Cancel, Valores por padrão, OK



- **Em classificação**, é recomendado que deixe marcado “simples sobre ST” (análise sobre segmentos de texto), pois o programa processa por padrão os segmentos de texto, a primeira opção (“dupla sobre RST”) tem baixo aproveitamento do corpus e a terceira (“simples sobre textos”), é mais indicada para respostas curtas.
- Deixe o **tamanho de RST** como padrão.
- No **número de classes** deixe como padrão 10, porém se ocorrer erro, tente abaixar este número até que funcione.
- **O resto** deixe como está por padrão.

Após as definições será obtido um **resultado** como este:



As classes divididas em cores, mostram quantos vocabulários estão presentes no corpus do texto e a porcentagem indica em qual a abrangência deste vocabulário.

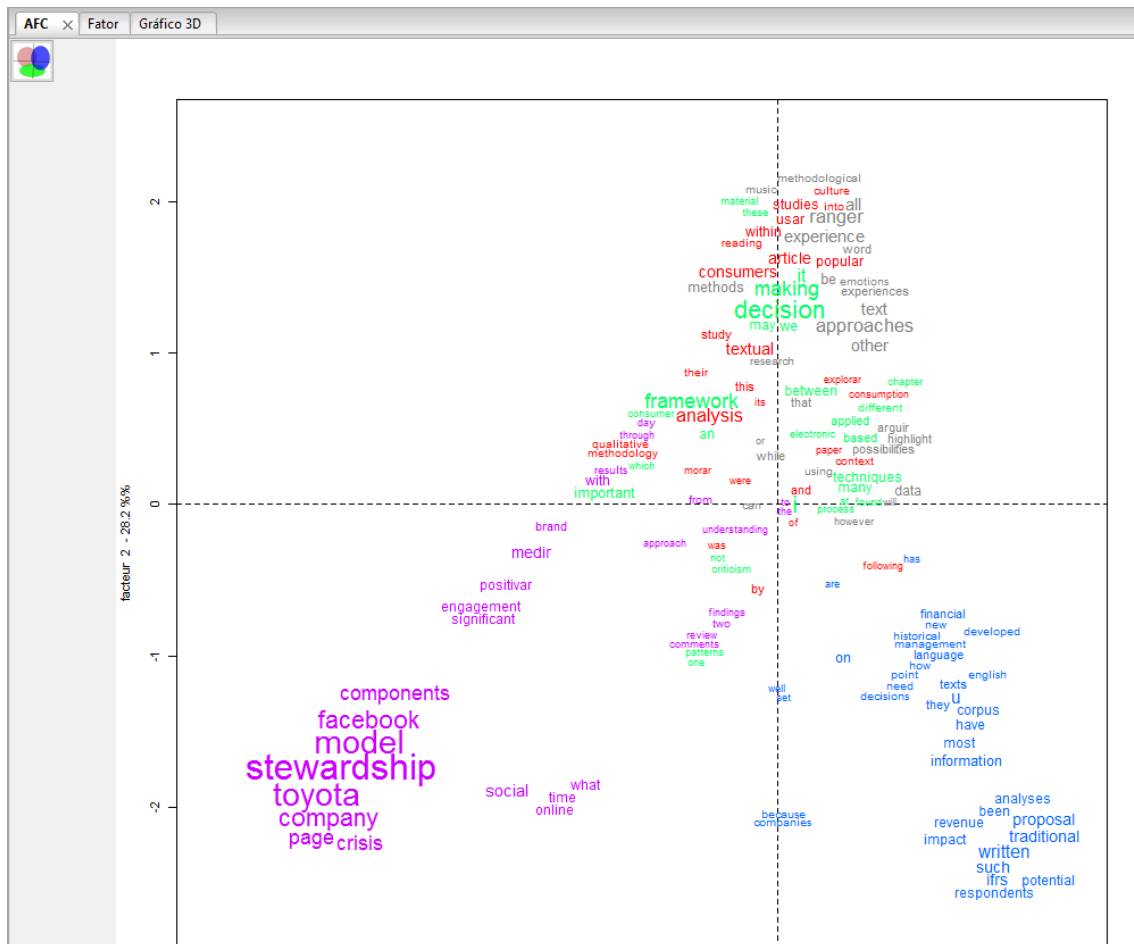
As **opções demarcadas no canto esquerdo superior** mostram outras opções de visualização, sendo que **a primeira** indica um visualização da esquerda para a direita das classes com preenchimento em cores e em barras horizontais (visualização acima); **a segunda** apresenta uma visualização das classes de cima para baixo, com preenchimento em barras na vertical e quais as palavras que compõe cada classe; e **a terceira** mostra uma visualização da esquerda para a direita das classes, com uma pequena nuvem de palavras de cada classe.

## Na aba **Perfis**:

CHD		Perfis		AFC					
1 Classe 1 16/71 22.54%		2 Classe 2 14/71 19.72%		3 Classe 3 10/71 14.08%		4 Classe 4 21/71 29.58%		5 Classe 5 10/71 14.08%	
n...	eff. s.t.	eff. total	pourcentage	chi2	Type	forme	p		
0	11	20	55.0	16.81	nr	analysis	< 0.0001		
1	10	19	52.63	13.46	adj	textual	0.00024		
2	3	3	100.0	10.77	nr	article	0.00103		
3	3	3	100.0	10.77	nr	consumers	0.00103		
4	4	6	66.67	7.31	adj	popular	0.00685		
5	3	4	75.0	6.68	nr	studies	0.00973		
6	3	4	75.0	6.68	nr	within	0.00973		
7	5	9	55.56	6.44	ver	usar	0.01117		
8	16	56	28.57	5.53	nr	and	0.01866		
9	4	7	57.14	5.33	nr	study	0.02098		
10	9	24	37.5	4.65	nr	this	0.03103		
11	6	14	42.86	4.13	nr	by	0.04224		
12	2	3	66.67	3.49	nr	reading	NS (0.06156)		
13	2	3	66.67	3.49	nr	into	NS (0.06156)		
14	2	3	66.67	3.49	nr	qualitative	NS (0.06156)		
15	2	3	66.67	3.49	nr	methodology	NS (0.06156)		
16	2	3	66.67	3.49	nr	culture	NS (0.06156)		
17	3	6	50.0	2.83	nr	their	NS (0.09240)		
18	3	6	50.0	2.83	nr	context	NS (0.09240)		
19	12	41	29.27	2.52	pre	in	NS (0.11241)		
20	12	31	38.71	8.25		*Social_Sciences	0.00408		

Temos os indicadores das palavras que compõe cada classe, referentes à: quantidade, quantidade, porcentagem, qui<sup>2</sup>, tipo, forma, e probabilidade.

## Na aba **AFC**:



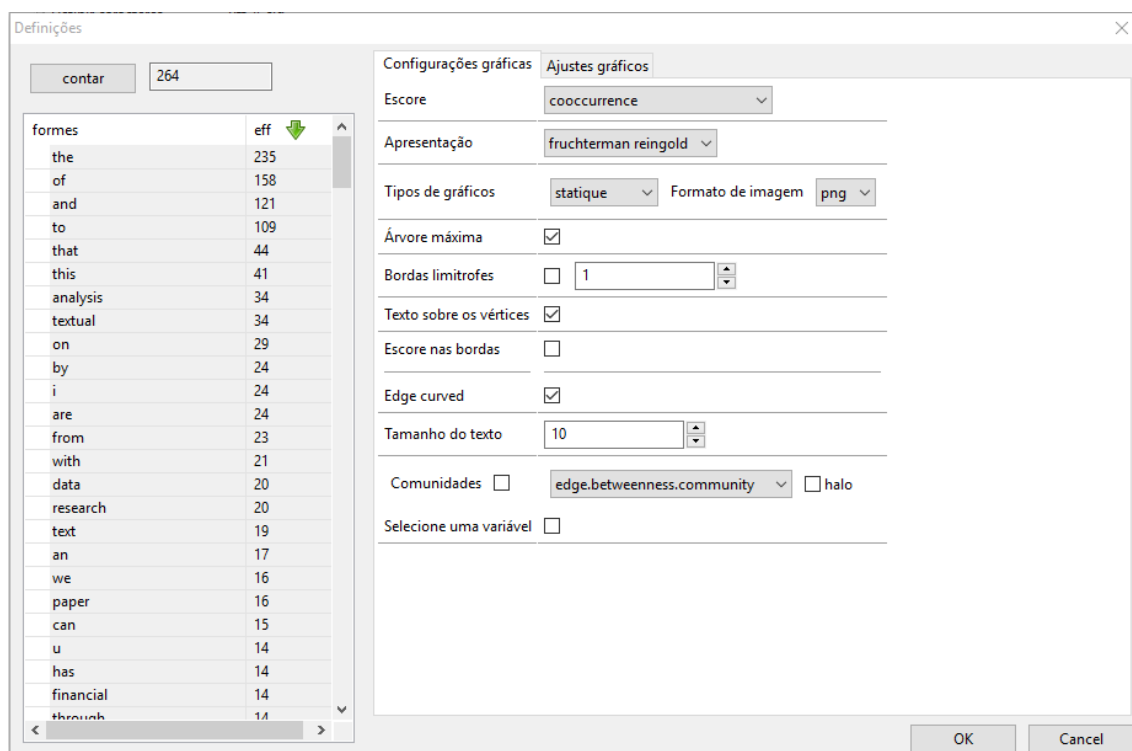
Temos algumas representações gráficas na primeira sub-aba, elas representam o posicionamento das classes de vocábulos no corpus textual, podemos ver quais classes se complementam e concentram o corpus, e quais se distanciam do centro e mostram certa especificidade.

Nas sub-abas, “Fator” e “Gráfico 3D”, temos respectivamente dados sobre os fatores que compõem os eixos dos gráficos e a porcentagem de representação dos mesmos, e, a possibilidade de produzir um gráfico 3D.

#### 4º - ANÁLISE DE SIMILITUDE:

A análise de similitude mostra uma grafo que representa a ligação entre as palavras do corpus textual. A partir desta análise é possível inferir a estrutura de construção do texto e temas de relativa importância.

Ao selecionar esta análise é aberta a seguinte **janela de configuração**:



A esquerda temos em formato de tabela duas colunas mostrando as palavras utilizadas e um indicador de quantidade (peso) delas.

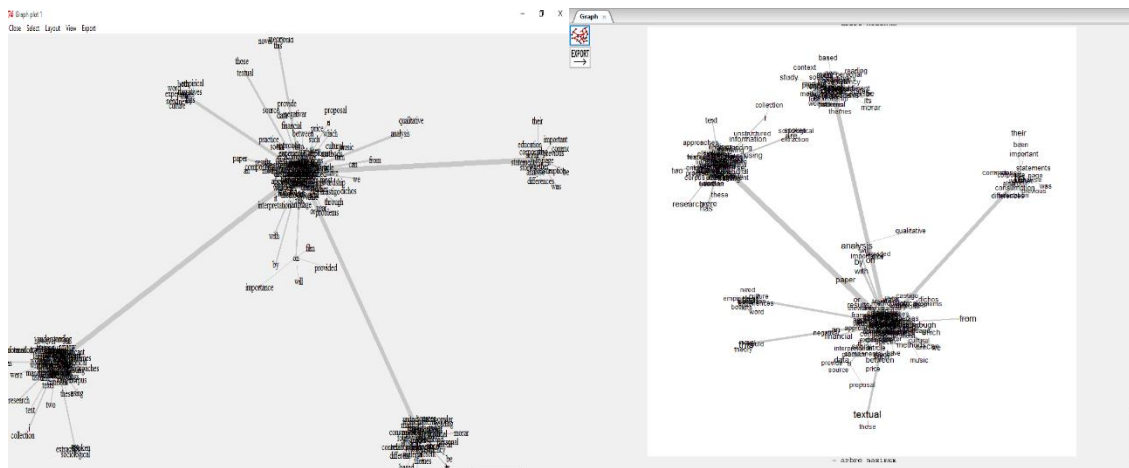
Na aba “Configurações gráficas” temos:

- “Escore”: possibilita escolher com qual indicador estatístico será processada a análise.
- “Apresentação”: permite escolher o algoritmo de visualização do grafo.
- “Tipos de Gráficos”: Permite escolher entre grafo estático, dinâmico (permite interação), ou 3D, além do formato de imagem de exportação do grafo.
- “Árvore máxima”: define que o grafo não terá estrutura de árvore e ramificações.
- “Bordas limítrofes”: define se existira limite nas arestas e a quantidade deste limite.
- “Texto sobre os vértices”: permite escolher colocar as palavras como indicação dos vértices.

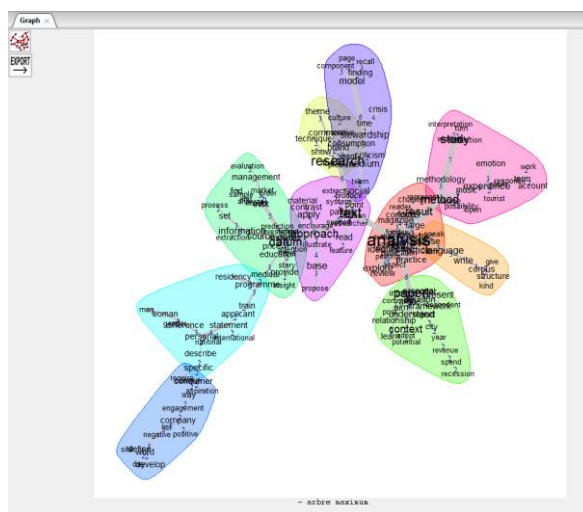
- “Escore nas bordas”: define se o tamanho das arestas será definido por indicadores de peso relacionado à ligação entre as palavras.
- “Edge curved”: permite escolher arestas curvas.
- “Tamanho do texto”: escolhe o tamanho o texto sobre os vértices.
- “Comunidades”: permite escolher formatos de representação em cores, o que realça os grupos de palavras mais relacionados entre si. (Selecionando “Halo”, são criados círculos coloridos de agrupamento).
- “Selecione uma variável”: possibilita selecionar restringir o grafo à variáveis escolhidas posteriormente.

\*Na aba “Ajustes gráficos”, é possível, escolher o tamanho da área do grafo, o tamanho do texto, vértices, e arestas com base em quantidade ou  $qui^2$ , além de cores e transparência.

Ao fim dos ajustes a análise retornara alguns grafos como estes:



Grafo dinâmico (direita)/Grafo estático (esquerda).



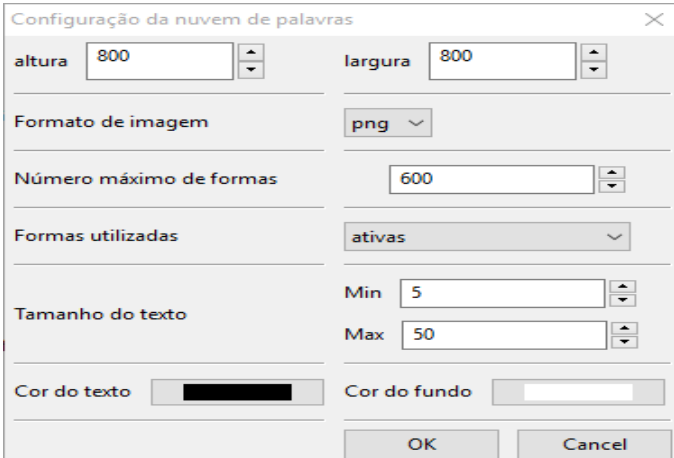
Grafo utilizando visualização por comunidades.

\*Existe a possibilidade de exportação para imagem ou para programas de visualização e análise de redes (Gephi, por exemplo).

## 5º - NUVEM DE PALAVRAS:

A análise por meio de nuvem de palavras é a mais típica quando se trata de análise de texto, ela mostra um apanhado de palavras estruturadas em forma de nuvem, as palavras são apresentadas de tamanho diferentes, ou seja, as palavras maiores são aquelas que detém certa importância no corpus textual (a partir de um simples indicador de frequência ou outro processo estatístico).

Ao selecionar esta análise abrirá a seguinte **janela de configuração**:



Configuração da nuvem de palavras

altura: 800      largura: 800

Formato de imagem: png

Número máximo de formas: 600

Formas utilizadas: ativas

Tamanho do texto: Min: 5      Max: 50

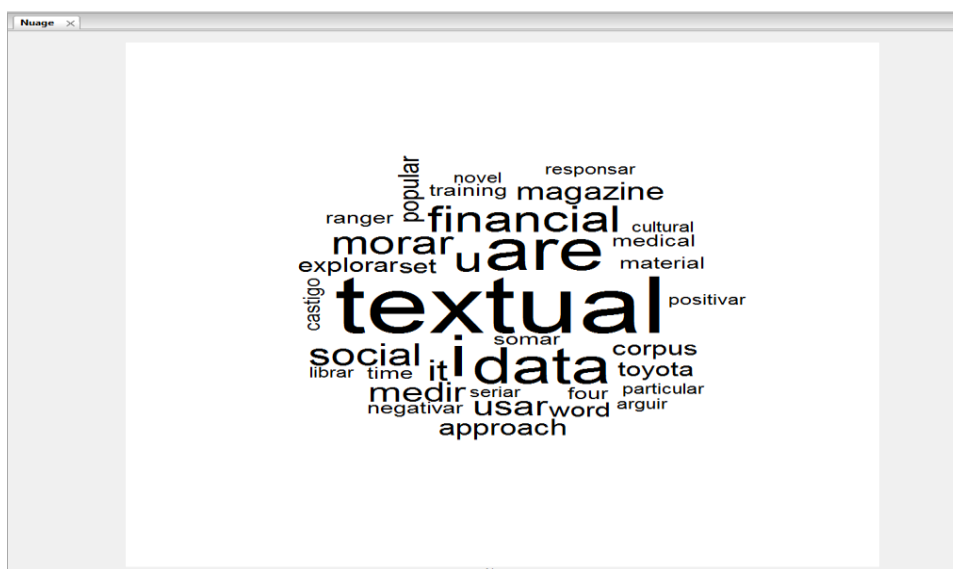
Cor do texto: [preto]      Cor do fundo: [branco]

OK      Cancel

Nesta janela é possível configurar tamanho, formato, número máximo de palavras, qual o tipo das palavras (ativas/complementares/ou as duas), qual o tamanho do texto e as cores.

\*Logo após a configuração aparecerá uma tabela de duas colunas mostrando as palavras e seu indicador de frequência (peso).

Ao terminar as configurações será retornado uma visualização mais ou menos como esta:



## OBSERVAÇÕES

- **Linguagem** – por padrão o Iramuteq vem em francês, porém existe a possibilidade de escolher outra linguagem de interface em: Edition → Préférences → Langue de l’interface.
- **Definir Parâmetros** – Lembre – se, definir os parâmetros (3º Passo) antes de cada análise pode especificar o resultado das análises, ou seja, é importante observar o intuito da análise e modificar os parâmetros de acordo com este intuito e sua necessidade.

## REFERÊNCIAS