

# Review sobre as análises do Iramuteq

L3P – Laboratório de Políticas Públicas Participativas

Luis Felipe Rosa de Oliveira

Data:07/10/2016



# ● Introdução

---

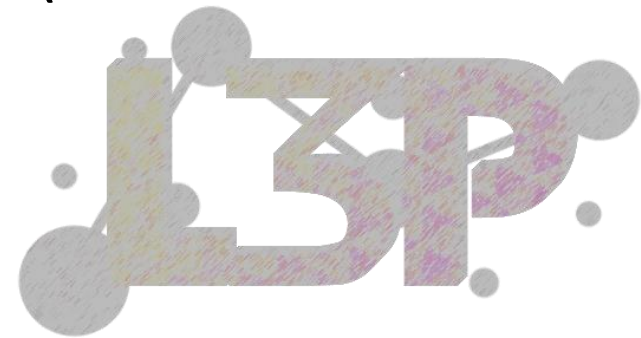
1. Análise Textual
2. Especificidades e AFC
3. CHD - Classificação Hierárquica Descendente
4. Análise de Similitude
5. Nuvem de Palavras



# 1. Estatísticas

---

- Esta análise apresenta algumas estatísticas descritivas do corpus textual.
- O gráfico de resumo apresenta a relação entre a frequência das formas e a quantidade de formas com tal frequência.
- As demais abas apresentam a frequência das formas ativas, complementares e o total, além de mostrar as formas hápax (formas com frequência 1).



# Interface - Estatísticas

IRaMuTeQ 0.7 alpha 2

Fichier Edition Vue Analyses de matrice Analyse de texte Aide

Historique

- Corpus textuel
- Social\_Sciences
- Social\_Sciences\_stat
- Matrice

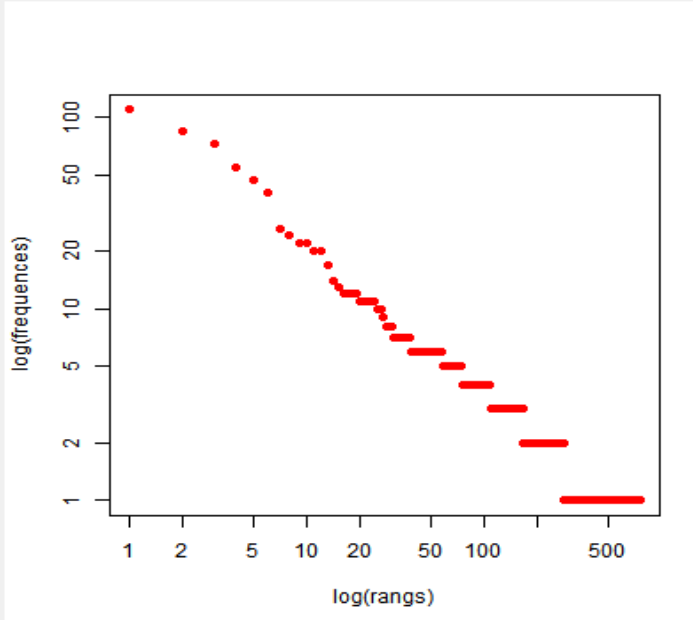
Sub-Análise

Social\_Sciences\_stat\_1 x

Résumé x Formes actives Formes supplémentaires Total Hapax

**Résumé -- Resumo**  
Nombre de textes : 1 -- **Número de Textos**  
Nombre d'occurrences : 2000 -- **Número de ocorrências**  
Nombre de formes : 750 -- **Número de Formas**  
Nombre d'hapax : 472 (23.60% des occurrences - 62.93% des formes) -- **Número de Hapax**  
Moyenne d'occurrences par texte : 2000.00 -- **Média das Ocorrências**

Componentes de Resultados



zipf

Prêt

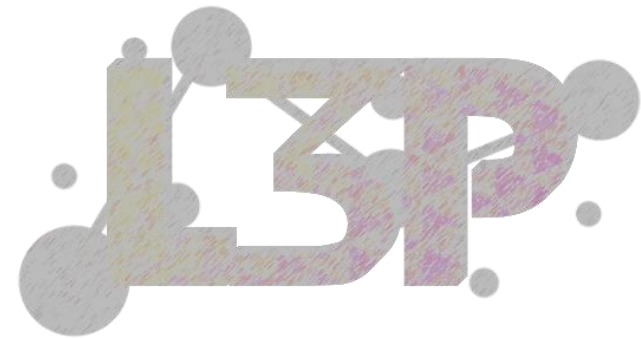
Bienvenue



## 2. Especificidades e AFC

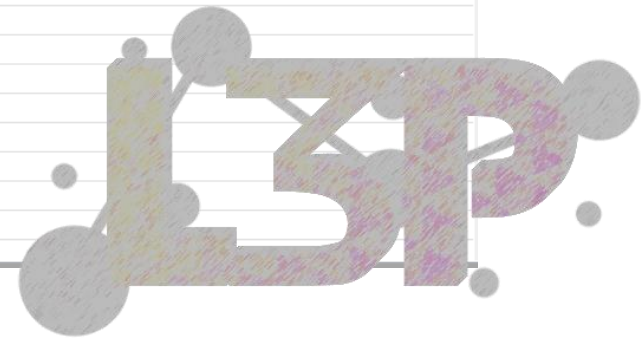
---

- Nessa etapa de análise temos indicadores de correspondências das palavras com o corpus textual (Hipergeométrico/ $Qui^2$ ), frequências e frequências relativas.
- Temos também indicadores para os tipos de formas:
  - **sw** – complementar/artigo\*
  - **nom** – nome/substantivo
  - **ver** – verbo
  - **nr** – não reconhecido
  - **num** - numeral
  - **adj** - adjetivo



# Interface – Especificidades e AFC

Formas	Formas comuns	Tipos	Forms frequências	Tipos de frequências	Frequência relativa das formas	Tipos de frequências relativas
formas		*Business	↓	*Social_Sciences		
finacial		4		-4		
information		3		-3		
the		2		-2		
we		2		-2		
such		2		-2		
word		2		-2		
for		1		-1		
two		1		-1		
as		1		-1		
using		1		-1		
s		1		-1		
in		1		-1		
context		1		-1		
approach		1		-1		
result		1		-1		
is		1		-1		
which		1		-1		
are		1		-1		
that		1		-1		
paper		1		-1		
can		1		-1		
this		1		-1		
on		1		-1		
social		1		-1		
datum		1		-1		
to		1		-1		
an		-1		1		
with		-1		1		
of		-1		1		
by		-1		1		
be		-1		1		
has		-1		1		
and		-1		1		
their		-1		1		
research		-1		1		
its		-1		1		



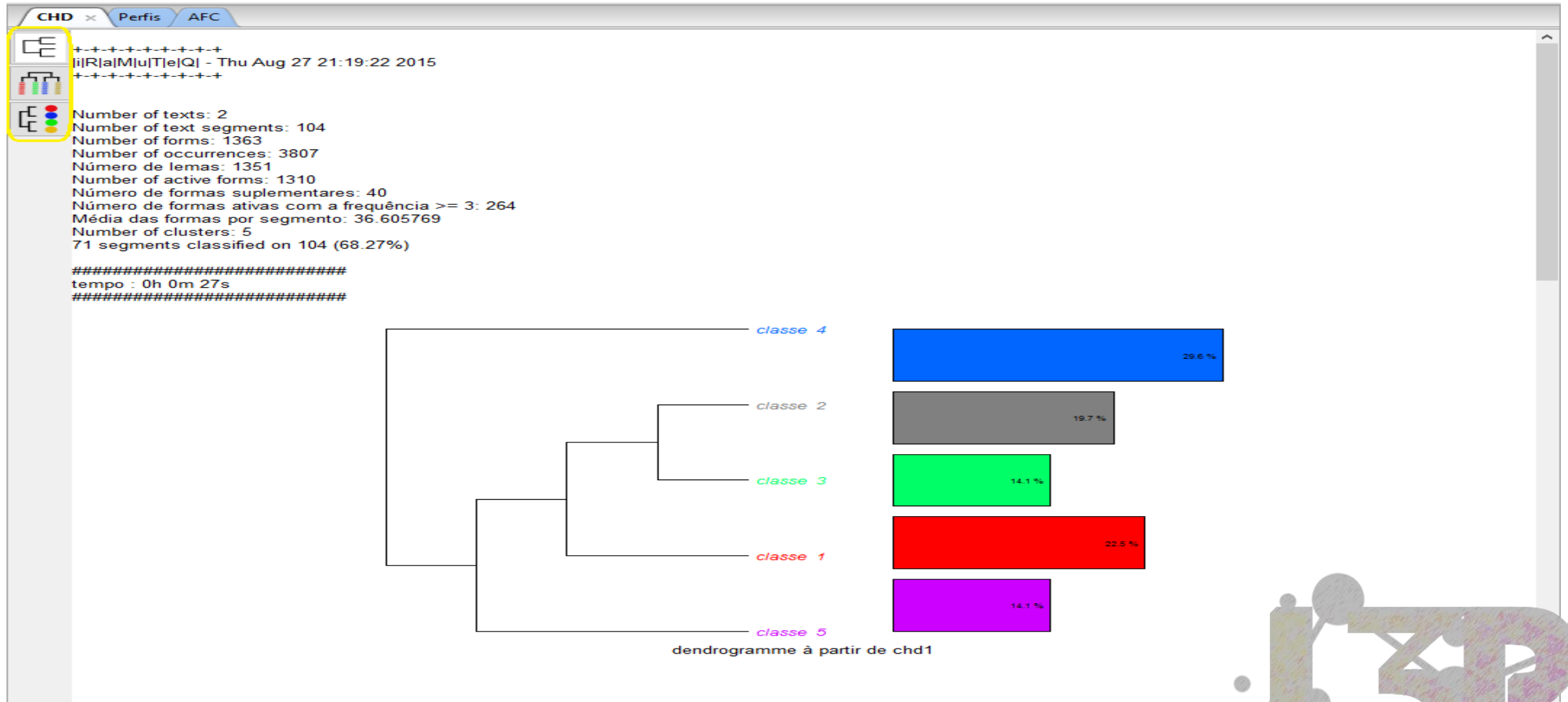
# 3. CHD – Classificação Hierárquica Descendente

---

- A CHD é uma das mais importantes análises do Iramuteq, ela define classes através da análise de correspondência entre as listas de formas reduzidas, o total de palavras e do dicionário da língua, e as cruza tanto com as segmentações de texto (ambientes) quanto com o corpus completo.
- Assim, é possível identificar classes de vocabulário diferentes no texto, e qual a posição desta classe no corpus.



# Interface - CHD





# 4. Análise de Similitude

---

- A análise de similitude é processada através de indicadores estatísticos que apresentam as relações entre as palavras, ou seja, forma uma árvore de palavras com ramificações a partir da relação que uma tem com a outra.
- Como resultado é apresentado um grafo com várias possibilidades de visualização da ligação entre as palavras.





# 5. Nuvem de Palavras

---

- A nuvem de palavras é uma das análises mais antigas quando se fala de análise de texto por processamento computacional.
- São utilizados indicadores estatísticos para inferir relevâncias a certas palavras.
- O resultado é apresentado em um “emaranhado” de palavras em que as mais relevantes de acordo com os indicadores são destacadas pelo tamanho ou até cor



# Interface – Nuvem de Palavras

