

IRaMuTeQ

Interface de R pour les Analyses
Multidimensionnelles de Textes et de Questionnaires

Guia de leitura e interpretação da CHD e da AFC no IRaMuTeq

Carlos Alexandre Felício Brito
Susana Nogueira Diniz
Nielce Meneguelo Lobo da Costa

São Paulo
2025

Guia de leitura e interpretação da CHD e da AFC no software IRaMuTeq

Carlos Alexandre Felício Brito

Universidade Municipal de São Caetano do Sul
(USCS)

Susana Nogueira Diniz

Universidade de São Paulo (USP)

Nielce Meneguelo Lobo da Costa

Universidade de São Paulo (USP)

Diagramação: Fernanda Costa Severo

COMO REFERENCIAR:

BRITO, Carlos Alexandre Felício; DINIZ, Susana Nogueira; LOBO DA COSTA, Nielce Meneguelo. **Guia de leitura e interpretação da CHD e da AFC no software IRaMuTeq.** São Paulo, 2025.

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Brito, Carlos Alexandre Felício

Guia de leitura e interpretação da CHD e da AFC no
software Iramuteq [livro eletrônico] /
Carlos Alexandre Felício Brito, Susana Nogueira
Diniz, Nielce Meneguelo Lobo da Costa.
São Caetano do Sul, SP : Ed. dos Autores, 2025.

PDF

ISBN 978-65-01-85842-5

1. Análise fatorial 2. Iramuteq (Software)
3. Pesquisa educacional 4. Software I. Diniz, Susana
Nogueira. II. Costa, Nielce Meneguelo Lobo da.
III. Título.

25-324892.0

CDD-300.72

Índices para catálogo sistemático:

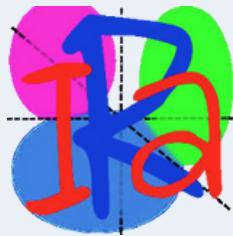
1. Ciências humanas e sociais : Pesquisa 300.72

Camila Aparecida Rodrigues - Bibliotecária CRB -
SP-010133/0

Sumário

Introdução -----	5
Capítulo 1. Classificação Hierárquica Descendente (CHD)	7
O que é a Classificação Hierárquica Descendente (CHD)?.....	8
Antes de gerar a CHD no IRaMuTeQ: passos essenciais	8
O gráfico da CHD no IRaMuTeQ	12
Como ler o dendrograma	13
O que observar em cada classe lexical	16
Passo a passo para a interpretação	18
Dicas e cuidados	19
Checklist	20
Capítulo 2. Análise Fatorial de Correspondência (AFC)	21
O que é a Análise Fatorial de Correspondência (AFC)?.....	22
A representação gráfica da AFC no IRaMuTeQ	22
O que evidencia a AFC	24
Como ler um plano fatorial	26
Como usar em pesquisas	26
Passo a passo para a interpretação	27
Dicas e cuidados	27
Checklist	28
Bibliografia -----	29
Hiperlinks -----	31

Introdução



Introdução

Este guia rápido foi elaborado para dar suporte a estudantes de mestrado e doutorado, bem como a professores e pesquisadores interessados na leitura e na interpretação de duas técnicas de análise textuais amplamente utilizadas nas Ciências Humanas e Sociais: a **Classificação Hierárquica Descendente (CHD)** e a **Análise Fatorial de Correspondência (AFC)**, ambas disponíveis para serem desenvolvidas no software [IRaMuTeQ](#) (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires).

A **CHD** é uma técnica de análise de textos que gera um dendrograma com classes lexicais e a **AFC** é uma técnica estatística que representa graficamente, em um plano, relações entre termos, segmentos de textos e classes lexicais. Ambos serão detalhados mais a frente neste guia.

Enfatizamos que o objetivo deste material é oferecer um **instrumento didático e prático** para facilitar a compreensão dos gráficos e esquemas gerados pelo software e orientar o processo de análise, entretanto, não se propõe a substituir manuais técnicos ou referenciais teóricos.

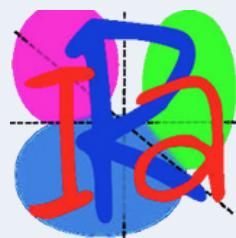
Neste guia, o leitor encontra:

- **Conceitos essenciais** de cada técnica, explicados de forma concisa e objetiva.
- **Passos preparatórios** para a organização do *corpus textual*.
- **Orientações para a leitura dos resultados** – sejam eles dendrogramas, planos fatoriais, resultados do *corpus* etc.
- **Sugestões de interpretação** conectadas à referenciais qualitativos, tais como: Teoria das Representações Sociais, *Survey*, *Grounded Theory*, Etnografia etc.
- **Dicas e cuidados** – sugestões e observações para auxiliar procedimentos.
- **Checklist final** para assegurar que as etapas foram cumpridas adequadamente.

Pretende-se, assim, oferecer um **guia de consulta rápida** para auxiliar tanto na análise de dados de pesquisas em andamento, quanto na formação de professores e de novos pesquisadores. A leitura pode ser feita de forma independente, mas recomenda-se sua utilização em conjunto com atividades de grupo, aulas ou oficinas, nas quais os conceitos possam ser discutidos na prática e à luz de diferentes referenciais e objetos de estudo.

Capítulo 1

Classificação Hierárquica Descendente (CHD)



1

O que é a Classificação Hierárquica Descendente (CHD)?

A **Classificação Hierárquica Descendente (CHD)**, criada por Max Reinert, é uma técnica de análise textual que:

Agrupa segmentos de texto (*STs*) em **classes lexicais** - ou seja em agrupamentos de palavras com vocabulário semelhante.

Permite identificar **fundos tópicos** - ou seja, os agrupamentos de sentido no *corpus textual*.

Mostra **como os discursos se organizam** e quais padrões de linguagem caracterizam diferentes grupos ou perspectivas.

2

Antes de gerar a CHD no IRaMuTeQ: passos essenciais

- 1 **Definir o objeto de estudo** de forma clara e teórica considerando o tema e nele o que se quer investigar, delimitando o foco, espaço e o tempo.

EXEMPLO:

Considerar uma pesquisa sobre “Percepção de profissionais da Educação sobre **inovação tecnológica** no Ensino Superior”.

Nesse estudo temos:

- **Tema** - Inovação tecnológica
- **Foco** - Conhecer a percepção de profissionais da Educação
- **Espaço** - Ensino Superior
- **Tempo** - Era digital

2 **Escolher a estratégia de coleta** mais adequada ao seu objeto:

- a) Entrevistas** – que podem ser tanto estruturadas, como semiestruturadas ou não estruturadas;
- b) Respostas abertas** de questionários;
- c) Documentos formais**, tais como: planos de ensino, diretrizes curriculares, revisões da literatura, artigos científicos, legislação pertinente etc.

3 **Preparar o *corpus* textual:**

- a) Limpar** os caracteres estranhos, tais como, abreviaturas, erros de digitação, gírias, vícios de linguagem (por exemplo: né, tá etc).
- b) Normalizar** o texto, ou seja, padronizar termos utilizados, tais como: termos diferentes para o mesmo significado (exemplo: professor, docente, professora), singular e plural de palavras, corrigir maiúsculas/minúsculas, acentos e duplicações.
- c) Lematizar**, ou seja, reduzir as palavras às suas formas canônicas: verbos no infinitivo, substantivos no singular, adjetivos no masculino singular.
- d) Inserir** marcadores de variáveis.

EXEMPLO:

Considerando um *corpus* formado por textos elaborados por profissionais da Educação de várias áreas, tais como: Humanas, Exatas e da Saúde, **os marcadores de variáveis** na sintaxe do software são caracterizadas da seguinte forma:

**** *Área_Humanas; **** *Área_Exatas; **** *Área_Saúde.
Os quatro asteriscos, espaço e mais um asterisco precedendo a Área.

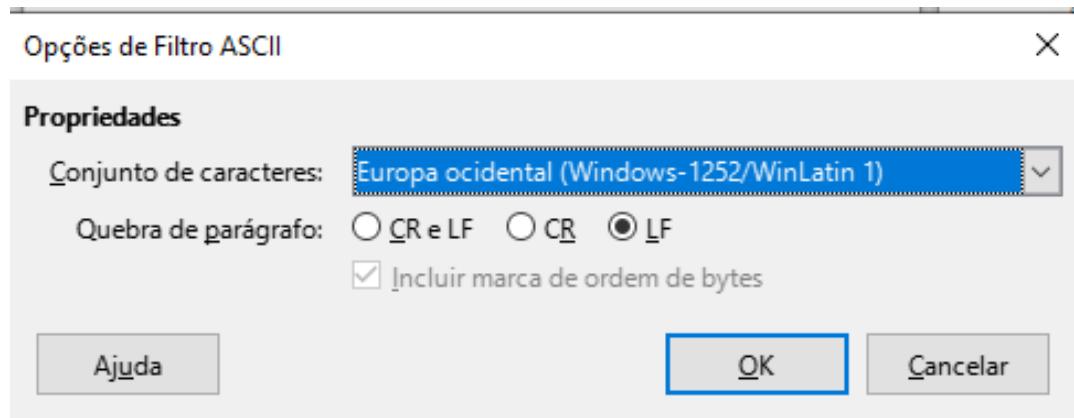
Assim sendo, cada um dos marcadores acima separa os textos dos sujeitos enunciadores da respectiva área, o que possibilita identificar similaridades ou dessemelhanças entre esses subgrupos de participantes.

 Para mais detalhes, consultar [Guia Iramuteq](#).

4 **Formatar o arquivo** em UTF-8, como indicado no tutorial de [Camargo & Justo \(2021\)](#):

Ao formatar o arquivo segundo as regras, tais como: títulos sem espaço, linhas estreladas etc., e salvá-lo no Libre Office, utilizando a opção de “escolha a codificação”. Ao abrir essa opção, aparece como *default* a imagem da figura 1.

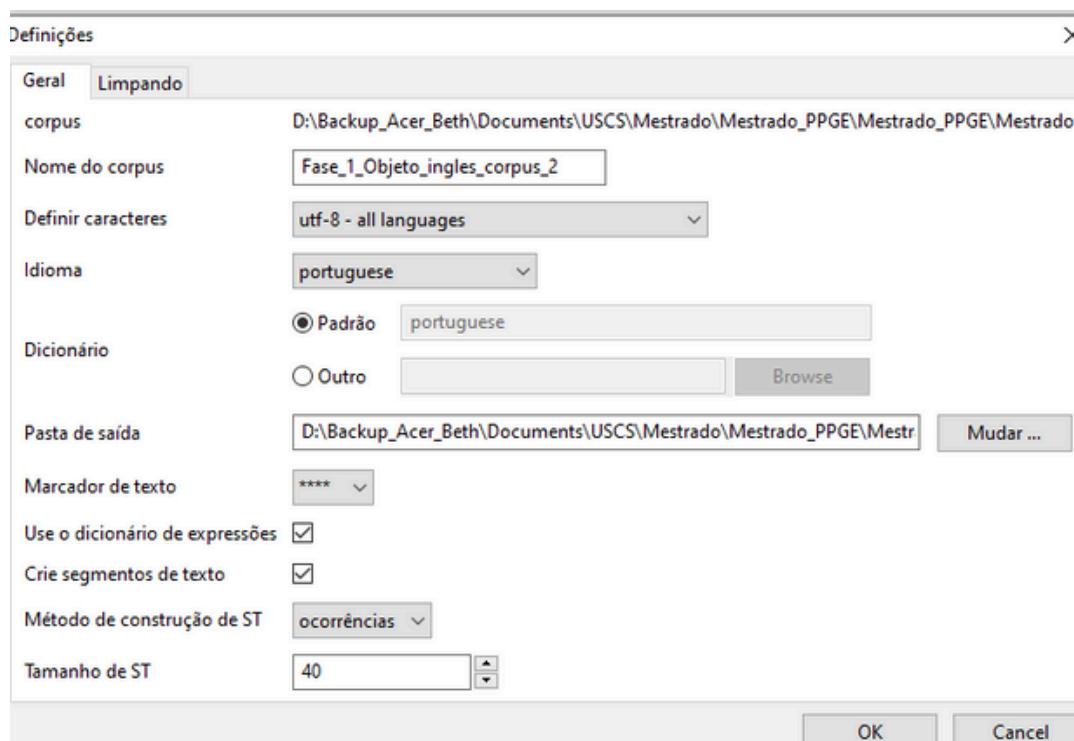
Figura 1 - Opção no IRaMuTeQ para codificação do texto



Fonte: elaboração própria com o software IRaMuTeQ

Observa-se o conjunto de caracteres *default* na figura. O software decodifica o arquivo, nomeando-o com a extensão UTF-8, como se pode observar na figura a seguir.

Figura 2 - Opção no IRaMuTeQ para a extensão UTF-8



Fonte: elaboração própria com o software IRaMuTeQ

EXEMPLO:

A seguir, na figura 3, um exemplo de *corpus* formatado em UTF-8 de uma pesquisa que investigou o aprendizado de inglês.

Figura 3 – Corpus formatado

**** *Aprender_ingles

É ter domínio e prática sobre a língua.

Saber falar em inglês, escrever e conseguir se comunicar e entender.

Aprimorar habilidades de leitura, escrita, fala e escuta em inglês.

Desenvolver a habilidade de se comunicar em outro idioma, entender culturas diferentes, ampliar oportunidades pessoais e profissionais.

Fonte: elaboração própria com o software IRaMuTeQ



Observa-se no exemplo, que a denominação: “**** *Aprender_ingles”, indica os STs advindos dos discursos dos sujeitos enunciadores com a variável (ou modalidade) “aprender_ingles”.

No caso, cada ST foi representado por uma cor, os verbos estão no infinitivo, adjetivos no masculino.

Em síntese, apresenta-se, a seguir, um exemplo de *corpus* preparado para rodar no IRaMuTeQ.

EXEMPLO:

Considerar uma investigação que identificou conhecimentos prévios de pesquisadores iniciantes na Área de Humanas sobre a Ética na pesquisa com seres humanos, a qual utilizou para a coleta de dados um questionário aberto. Após a coleta de dados foi gerado um arquivo limpo, normalizado e lemetizado, com as variáveis inseridas conforme se apresenta na figura 4.

Figura 4 – Trecho do corpus formatado em UTF-8 no software

```
**** "Ética
São pesquisas pautadas em um cuidado onde, se apontar os riscos e benefícios com os seres humanos.
A ética em pesquisa permeia considerar a segurança, o anonimato e os direitos humanos. É um regulamento que protege a pessoa em seus aspectos humanos em busca do desenvolvimento científico, preservando a saúde e a vida.
A ética em pesquisa consiste em proteger os direitos dos pesquisadores (público-alvo) durante e após a pesquisa. Existem comitês de ética que são formados por pessoas de variadas ordens que julgam se a pesquisa respeita os princípios de Heisenberg. Através desta comissão o pesquisador tem respaldo da promoção. Todo trabalho realizado com seres humanos necessita da aprovação do comitê de ética antes da sua realização.
Entendo por ética em pesquisa envolvendo seres humanos. Respeito, dignidade ao mesmo. Se tratando de pesquisa deve haver um segmento à seguido, através do comitê de ética e autorização da parte envolvida.
É respeitar, as condições propostas para o estudo.
Ética em pesquisa é a veracidade, idoneidade das informações, dados.
Inicialmente é respeitar o conhecimento a ser pesquisado sem fazer juízo de valor sobre qualquer opinião fundamentada sobre o indivíduo pesquisado, levando como base fundamental o conhecimento científico e experiências vinculadas de cada ser humano.
Ética é preservar os indivíduos envolvidos na pesquisa tanto psicologicamente como fisicamente.

**** "Avaliação ética
Sim. Os aspectos éticos devem ser sempre considerados, para assim os riscos serem diminuídos e às vezes serem excluídas a pesquisa caso o risco seja grande para o ser humano.
Toda pesquisa deve considerar o aspecto ético, pois, ao lidar com a vida estamos problematizando os direitos da mesma.
Dependerá se o trabalho a ser realizado será embasado em pesquisa em grupos de seres humanos.
Esta resposta eu não teria no momento. Penso que não necessita, por não se tratar de um ser humano.
Faz-se necessário considerar os aspectos éticos, pois em uma pesquisa se propõe explorar informações que podem gerar dilemas que cause impactos significativos em um determinado grupo de sociedade.
Sim. Para respaldo de cunho técnico científico, do contrato, a pesquisa ramifica- se pelo campo empírico e perde credibilidade.
Sim, na área de ensino e relevante considerar a avaliação ética, que não traga algum desconforto aos participantes.
Sim, a ética se faz necessária em todos os segmentos tendo como principal papel o elemento norteador de princípio e conduta

Ln 41, Col 143 9.534 caracteres
```

100% Unix (LF) UTF-8

Fonte: Corpus da pesquisa de Lobo da Costa, Diniz e Brito (2021)

3

O gráfico da CHD no IRaMuTeQ

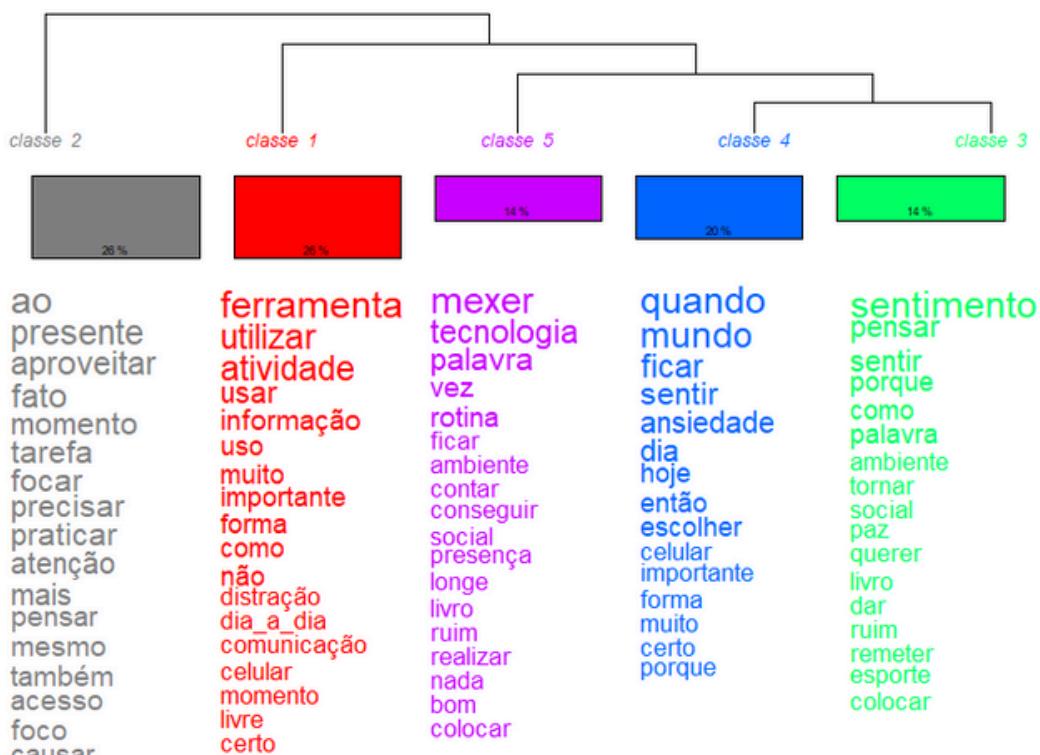
Ao rodar a CHD no software IRaMuTeQ o gráfico gerado é um **dendrograma**, ou seja, é um diagrama de árvore desenvolvido a partir de similaridades, formando clusters muito usados para hierarquização.

EXEMPLO:

Considerar uma pesquisa cujo objetivo foi “compreender a **Nomophobia** (fobia causada por estar sem tecnologias, tais como celular etc.) enquanto um fenômeno socialmente construído, analisando suas representações sociais no contexto educacional e os impactos da regulação tecnológica na prática pedagógica”.

O gráfico gerado pela CHD no IRaMuTeQ está na figura 5.

Figura 5 – Dendrograma resultante da CHD do *corpus* textual da pesquisa



Fonte: elaboração própria com o software IRaMuTeQ



Observa-se que se trata de um **diagrama de árvore, com cinco clusters**. A interpretação detalhada do dendrograma está apresentada na seção a seguir.

4

Como ler o dendrograma

O dendrograma pode ser lido em duas direções: horizontal e vertical. Cada direção pode ser lida em dois sentidos. A seguir, apresentam-se as orientações de leitura.

Na direção horizontal



⚠️ **Atenção:** Ler da direita para a esquerda ⚠️

As classes mais à Direita representam o que há de mais geral no *corpus*, identificando classes ainda amplas, com sentidos mais genéricos.

As classes mais à Esquerda representam o que há de mais característico dos discursos dos sujeitos enunciadores no *corpus*, identificando Classes mais específicas e homogêneas. Elas indicam discursos particulares ou visões menos hegemônicas.

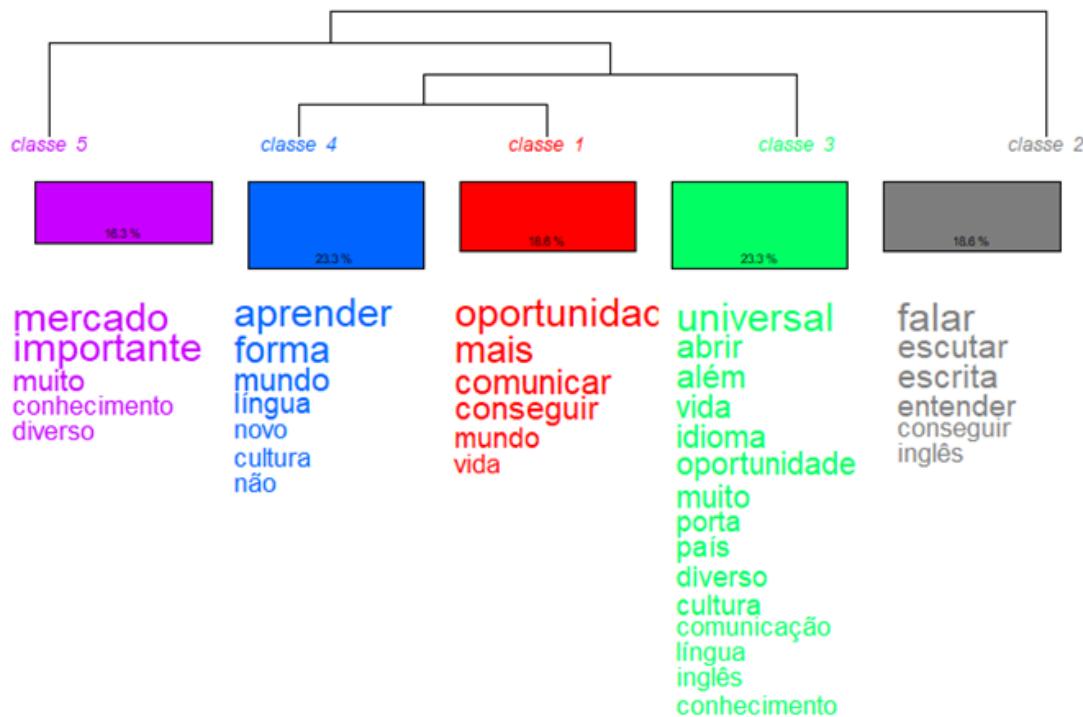
Esse movimento, de estar mais à direita ou mais à esquerda, não indica “mais importante” ou “menos importante”, mas sim que se tem uma representação que aponta **do geral ao específico**.

EXEMPLO:

Considerar uma pesquisa cujo objetivo foi “Analisar como estudantes do Ensino Médio Técnico representam e aprendem conteúdos gramaticais de **Língua inglesa** por meio da produção de vídeos, a partir de uma perspectiva semiótica e de multiletramentos, articulando evidências quantitativas (perfil e autoavaliação) e qualitativas (representações discursivas) em ciclos de *Design-Based Research*,”, ao rodar o *corpus* formatado no IRaMuTeQ.

O gráfico gerado pela CHD no IRaMuTeQ está na figura 6.

Figura 6 – Dendrograma resultante da CHD do *corpus* textual da pesquisa



Fonte: MOTA, J. S. et al. (2025), p. 12.



No gráfico nota-se, pela leitura da direita para a esquerda, que:

Na classe mais geral (à direita) estão palavras como “falar”, “escrita” e “escutar”, as quais indicam habilidades linguísticas (ler, escrever, falar e ouvir).

Na classe intermediária estão palavras como “universal”, “oportunidade” e “aprender”, as quais indicam comunicação, cultura e oportunidades, em que aprender inglês é ampliar possibilidades de interação social e acesso a informação.

Na classe mais específica (à esquerda): palavras como “mercado”, “importante” e “conhecimento” indicam mercado de trabalho e usos práticos, relacionando o idioma a empregabilidade e atividades de lazer digital.

Na direção vertical



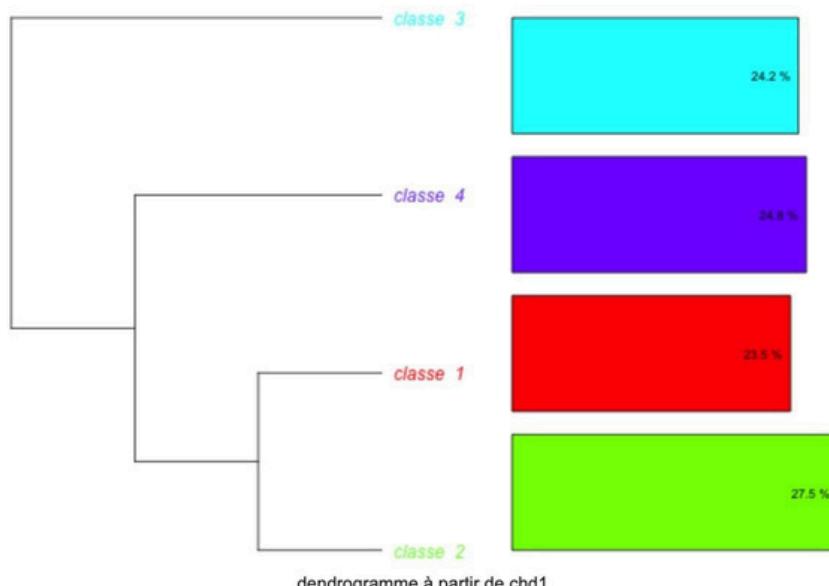
→ **Atenção:** Ler de cima para baixo ←

- Trata-se de uma organização gráfica.
- A posição superior ou inferior não representa relevância.

EXEMPLO:

O dendrograma gerado pelo software IRaMuTeQ pode se apresentar na vertical, como descrito logo abaixo na figura 7

Figura 7 – Dendrograma resultante da CHD do *corpus* textual da pesquisa



Fonte: elaboração própria com o software IRaMuTeQ



Observa-se que, no caso, o dendrograma deste *corpus*, está representado no sentido vertical e deve ser lido de cima para baixo.

Neste exemplo os percentuais de cada uma das classes estão visíveis. **A classe 3 (azul)** com 24,2% **a classe 4 (azul cobalto)** com 24,8%, **a classe 1 (vermelha)** com 23,5% e **a classe 2 (verde)** com 27,5%. Nota-se, portanto, que a posição superior não representa a mais relevante em termos de STs.

As classes 1 e 2 formam um cluster e com a classe 4 formam um segundo cluster, todos ligados à classe 3 formando o terceiro cluster.

Em conclusão: as quatro classes estão conectadas, porém com sentidos similares, muito embora com características próprias que as definem de alguma maneira as suas propriedades qualitativas, em que [Max Reinert](#) comprehende como um **mundo lexical** para os sujeitos-enunciadores em função do objeto representado.

5

O que observar em cada classe lexical

Para interpretar uma classe, considerar:

- 1 **Palavras com maior valor de qui-quadrado (χ^2)** → essas são as mais representativas dessa classe lexical.

O IRaMuTeQ usa o χ^2 porque ele é a ferramenta estatística adequada para medir **associação entre categorias** (neste caso, palavras e classes).

A definição de Vieira (2003) – ver hiperlink [**qui-quadrado \(\$\chi^2\$ \)**](#) – reforça a base conceitual: o χ^2 serve justamente para verificar independência ou associação entre variáveis — o mesmo raciocínio aplicado às classes lexicais no método Reinert.

- 2 **Segmentos de texto exemplares** → esses são os que evidenciam como os termos aparecem no discurso.

- 3 **Variáveis associadas (opcional)** → essas identificam se há predominância de determinado perfil (área de atuação profissional, grau de instrução, idade, gênero, região geográfica etc.).

- 4 **Coerência interna** → verificar se os termos e trechos têm relação semântica estabelecida.

EXEMPLO:

Considerar, na figura 8, a **Classe 6** na tabela que apresenta os perfis de todas as classes lexicais identificadas em uma pesquisa sobre “Formação em Enfermagem” que investigou, na área da Saúde a variável “Cuidado Integral” do indivíduo.

Figura 8 – Perfil das classes identificadas pelo software.

Classificação - Sentidos_da_Avaliação_corpus_2		Classificação - Fase_1_Objeto_ingles_corpus_1		Classificação - Justificativa_corpus_1	
CHD	Perfis	AFC			
1 Classe 1 6/43 13.95%	2 Classe 2 5/43 11.63%	3 Classe 3 5/43 11.63%	4 Classe 4 8/43 18.6%	5 Classe 5 7/43 16.28%	6 Classe 6 5/43 11.63% × 7 Classe 7 7/43 16.28%
n...	eff. s.t.	eff. total	pourcentage	chi2	Type
0	4	5	80.0	25.74	nom exemplo
1	4	5	80.0	25.74	nom apoio
2	3	3	100.0	24.51	nom alimentação
3	2	3	66.67	9.51	adv também
4	2	3	66.67	9.51	nom enfermeiro
5	2	3	66.67	9.51	ver dar
6	2	3	66.67	9.51	adv além
7	2	3	66.67	9.51	nom diabetes
8	3	7	42.86	7.94	nom saúde
9	3	11	27.27	3.52	adv como
10	2	6	33.33	3.2	adv quando
11	3	7	42.86	7.94	pre sobre
12	4	13	30.77	6.64	pre com
13	2	4	50.0	6.32	ver_sup ter
14	3	9	33.33	5.22	pro_pos sua
15	2	6	33.33	3.2	nom_s... pessoa
16	5	16	31.25	9.55	*Cuidado_integral

Fonte: elaboração própria com o software IRaMuTeQ



Nessa tabela, observa-se o número de cada classe (de 1 a 7), o número de STs de cada classe – no caso a **Classe 6** (– 5/43 – 11,63%) tem cinco ST de um total de 43 no *corpus*, o que representa 11,63% dos ST – e, por último, apresenta os valores respectivos de qui-quadrado (χ^2) de cada palavra representativa da classe lexical (na sintaxe do software **chi2**). Por exemplo, na Classe 6 a palavra “alimentação” tem qui-quadrado com valor 24,51%.

Na próxima figura está um exemplo dos segmentos representativos da **Classe 6** (*typical texts segments* – Classe 6) gerado pelo software, no qual é possível observar em destaque (cor vermelha) as palavras com maiores valores de χ^2 , seu **score**, ou seja, a força lexical, bem como a variável associada (**** *Cuidado_integral) ao perfil do *corpus* textual.

Figura 9 – STs da Classe 6 organizado pela sua força lexical (Score).

Typical text segments - Classe 6

**** *Cuidado_integral
score : 119.18

18_cuidado_integral é **como** por **exemplo** quando um paciente com **diabetes** além de receber a medicação **também** recebe orientação sobre **alimentação** exercícios e **apoio** emocional para lidar com a condição visando melhorar sua **saúde** de forma global não só controlando o açúcar no sangue

**** *Cuidado_integral
score : 95.01

um **exemplo** seria o atendimento a um idoso com **diabetes** que não envolve apenas o controle da glicemia mas **também** orientações sobre **alimentação** atividade física **apoio** psicológico e envolvimento da família promovendo o bem-estar global

**** *Cuidado_integral
score : 72.79

o **enfermeiro** conversa com ela sobre sua rotina vê se ela tem **apoio** da família **como** está se sentindo emocionalmente e **dá** dicas de **alimentação** saudável e exercícios

**** *Cuidado_integral
score : 68.93

8_cuidado_integral é **quando** o **enfermeiro** cuida da pessoa **como** um todo não só do problema de **saúde** um **exemplo** simples seria uma pessoa idosa com pressão alta **além** de **dar** o remédio certo para controlar a pressão

**** *Cuidado_integral
score : 59.42

por **exemplo** em um tratamento de câncer a pessoa se encontra muito fragilizada emocionalmente e ter um acompanhamento e **apoio** emocional por parte dos seus profissionais de **saúde** é essencial para sua plena recuperação

Fonte: elaboração própria com o software IRaMuTeQ



Observa-se que o ST representativo com maior score (119,18) é relativo à variável “**Cuidado integral**” e são palavras representativas do discurso as seguintes: “**alimentação**”, “**apoio**”, “**saúde**”. No exemplo, percebe-se a frequência de repetição dos léxicos representativos.

A coerência interna é o quarto ponto a observar em cada classe lexical, que depende da perspectiva filosófica e teórica do pesquisador.

6

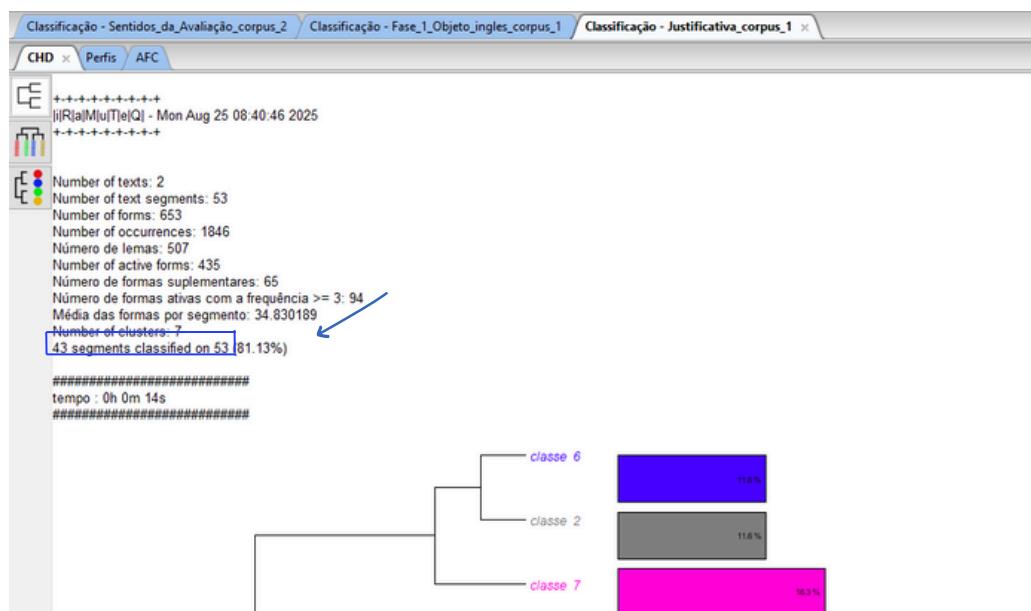
Passo a passo para a interpretação

- 1 Verificar a **taxa de aproveitamento** dos segmentos de textos (STs) que deve ser $\geq 70\%$. Essa taxa é a indicada pelo método Reinert.

EXEMPLO:

Considerar uma pesquisa que gerou um *corpus* textual denominado “Justificativa_corpus_1”. O software IRaMuTeQ gera uma síntese do *corpus*, apresentando os números totais relativos aos STs, às ocorrências, aos lemas, às formas ativas, às formas suplementares, bem como a taxa de aproveitamento das classes, como ilustra a figura 10.

Figura 10 – Síntese do *corpus* textual gerada pelo software IRaMuTeQ



Fonte: elaboração própria com o software IRaMuTeQ



Neste caso, a taxa de aproveitamento foi de 81,13% dos STs (43/53), portanto o resultado é adequado para aplicar a CHD pelo método Reinert.

-
- 2 **Identificar os termos-chave** → são os que apresentam o maior qui-quadrado χ^2 ;
 - 3 **Ler os trechos originais** correspondentes aos segmentos de texto selecionados pelo software a partir do *corpus* textual para entender o contexto de uso;
 - 4 **Agrupar ideias centrais**, identificando o que cada classe revela sobre o fenômeno;
 - 5 **Comparar classes lexicais**, verificando se: há oposições, complementaridades ou tensões;
 - 6 **Relacionar com a teoria e objetivos da pesquisa** - a análise deve vincular as evidências encontradas à teoria de base da pesquisa de modo a atingir os objetivos propostos.

7

Dicas e cuidados

- ! Não confundir posição no dendrograma com relevância teórica. A posição representa os conceitos mais gerais – à direita – para os mais específicos – à esquerda.
- ! Voltar sempre ao *corpus* textual original para interpretar — não se basear apenas nas palavras-chave identificadas.
- ! Integrar a CHD a outros métodos (ex.: análise prototípica, AFC, análise temática).
- ! Contextualizar sempre a interpretação no escopo da pesquisa.

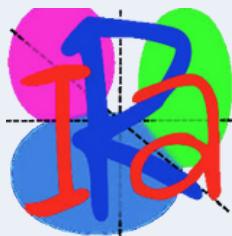
8

Checklist

- ✓ Objeto de estudo bem definido.
- ✓ *Corpus* preparado e formatado corretamente.
- ✓ CHD rodada com taxa de aproveitamento satisfatória.
- ✓ Leitura horizontal feita da direita para a esquerda.
- ✓ Interpretação baseada em termos, trechos e variáveis.
- ✓ Conexão com a teoria e os objetivos da pesquisa.

Capítulo 2

Análise Fatorial de Correspondência (AFC)



1

O que é a Análise Fatorial de Correspondência (AFC)?

A **Análise Fatorial de Correspondência (AFC)** é uma técnica de interpretação de textos por meios estatísticos, baseada na representação gráfica dos dados no **plano fatorial**. Essa representação permite **visualizar as relações entre classes lexicais, termos e segmentos de texto**.

A AFC tem como **função central**:

Identificar **proximidades** e **oposições** nos discursos e auxiliar a interpretá-los

A AFC tem como **papel didático**:

Evidenciar tensões, consensos e variações entre discursos de diferentes grupos de respondentes ou entre diferentes dimensões do corpus e auxiliar a interpretá-los.

Enfatizando: a AFC Permite visualizar sob a forma de um plano fatorial, as posições resultantes da CHD, como ensinam Marchand e Ratinaud (2012).

No IRaMuTeQ, a AFC é frequentemente usada após a CHD de modo a projetar as classes em um **plano fatorial**.

2

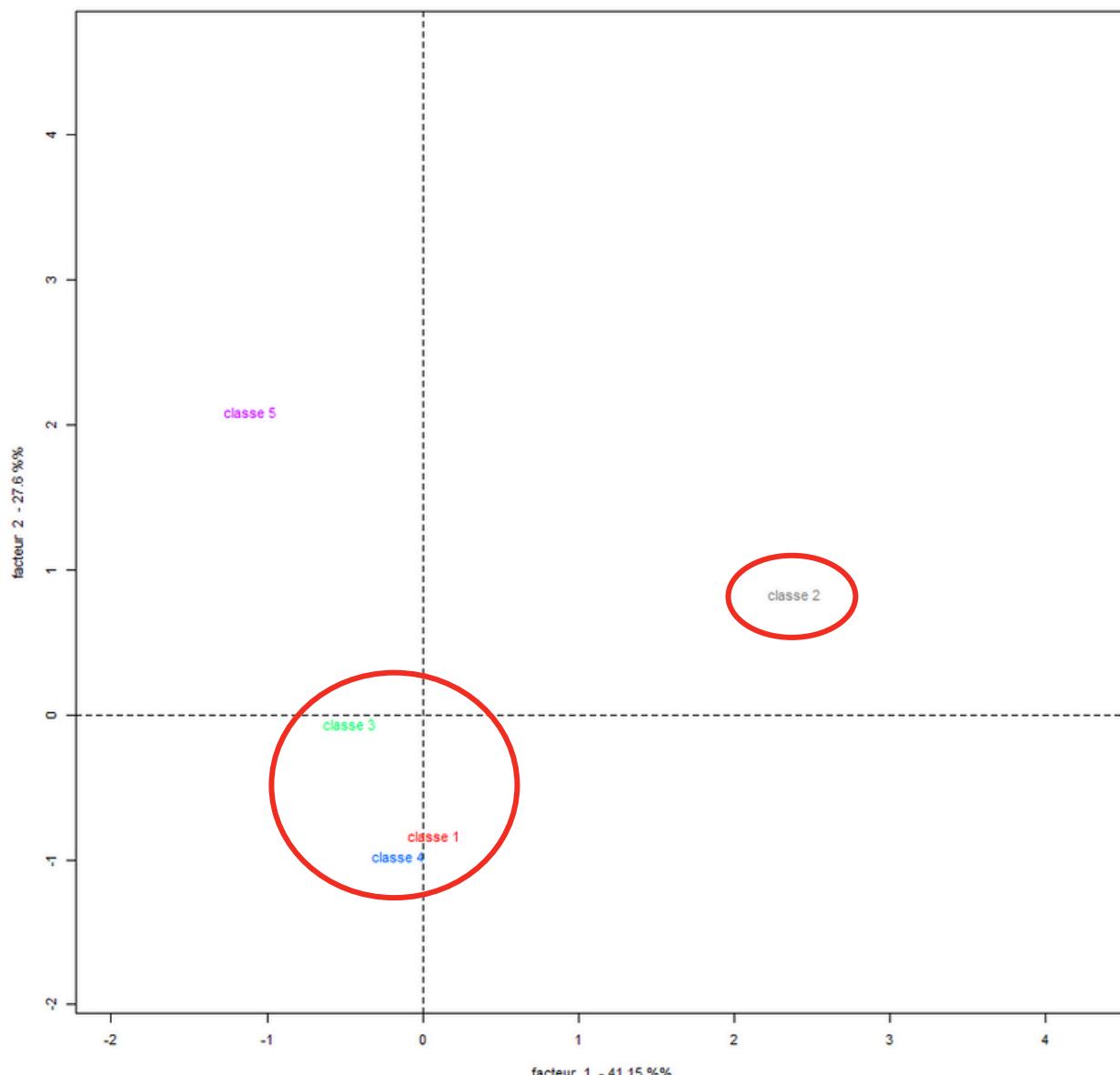
A representação gráfica da AFC no IRaMuTeQ

Ao rodar a AFC no software IRaMuTeQ o gráfico é gerado em um **plano fatorial**, ou seja, em um espaço bidimensional no qual são considerados dois eixos: um horizontal e outro vertical, os quais dividem o plano em quatro quadrantes. Neles são representadas variáveis/modalidades distintas, criando uma representação das correspondências entre tais variáveis/modalidades. Assim sendo, pontos próximos significam alto grau de associação entre tais variáveis/modalidades.

EXEMPLO:

Considerar a pesquisa que investigou o “Aprendizado de inglês” (CHD ilustrada na figura 5). A AFC, apresenta-se no plano factorial na figura a seguir.

Figura 11 – Representação no plano factorial das categorias de variáveis



Fonte: MOTA, J. S. et al. (2025) p. 16.

Observa-se nesse gráfico no plano factorial, chamado de **mapa de correspondências** ou **biplot**, que as Classes 1, 3 e 4 que estão associadas em áreas de contraste, o que está indicando a especificidade, explicada como 41,15% da variância (fator 1), sendo que a Classe 2 – que é mais geral – corresponde a 27,6% da variância (fator 2).

3

O que evidencia a AFC

A associação de termos ou classes identificados pela proximidade espacial:

- termos ou classes que aparecem próximos no plano fatorial têm **associação mais expressiva**.

Os discursos ou perspectivas em tensão identificados pela localização em

- eixos opostos: elementos de textos situados em lados opostos de um eixo, no plano fatorial, representam significados em tensão ou oposição.

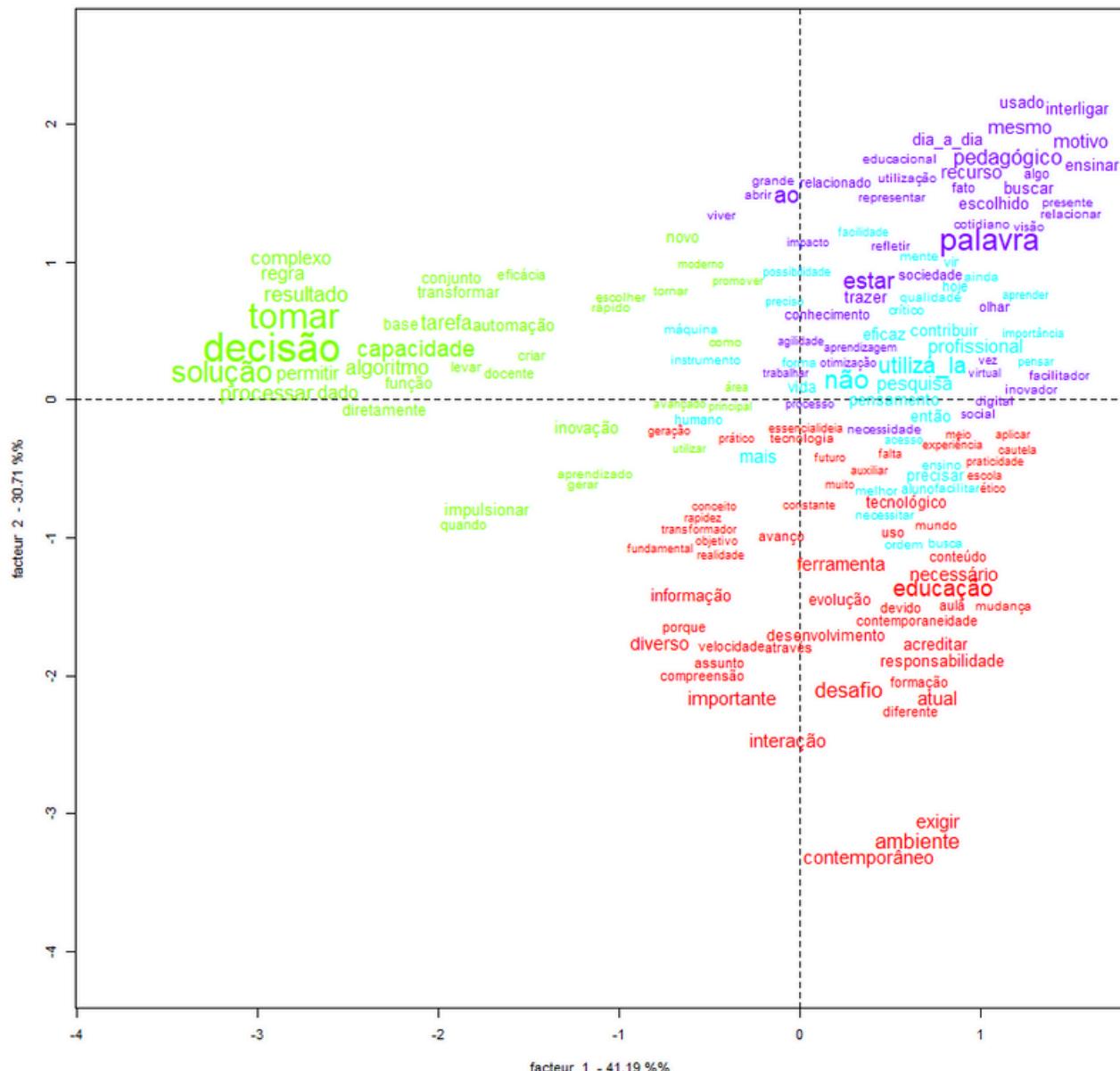
O significado do valor percentual da Variância: Eixos apresentando maior

- valor percentual representam dimensões mais fortes de **diferenciação discursiva**.

EXEMPLO:

Considerar a pesquisa que investigou “Inteligência Artificial na Educação” (Brito & Sá, 2025). Na figura 11 apresenta-se outro tipo de representação da AFC gerada pelo IRaMuTeQ, na qual as palavras representativas de cada classe estão expostas com diferentes cores.

Figura 12 – Representação da AFC por léxicos nas diversas classes



Fonte: Apresentação de Brito e Sá (2025)



Observa-se a distribuição das quatro classes lexicais, sendo que a **Classe 1 (vermelha)** – mais geral – em contraste com a **Classe 2 (verde)**. A **Classes 4 (roxa)** e **3 (azul)** estão intermediando o pensamento do mais geral para o mais específico. Mais detalhes da leitura, estão na próxima seção.

4

Como ler um plano fatorial

O plano fatorial pode ser lido a partir dos dois eixos e dos quadrantes. Eixo 1 horizontal e Eixo 2: vertical. Cada direção nos eixos pode ser lida em dois sentidos: positivo ou negativo. A seguir, apresentam-se as orientações de leitura:

Eixo 1 (horizontal)



Representa geralmente a dimensão mais forte de **oposição** discursiva.

Eixo 2 (vertical)



Segunda dimensão relevante, **menos intensa** que a primeira.

Quadrantes



Revelam agrupamentos de sentidos, permitindo interpretar contrastes (por exemplo.: tradicional × inovador, individual × coletivo etc.).

EXEMPLO:

Considerar a pesquisa sobre IA na Educação (ver figura 11) considerando a TRS observa-se que na Classe 1 os sujeitos enunciadores falam enfaticamente da IA no **ambiente** da **Educação contemporânea**, portanto apresentando uma visão ampla do objeto, tensionando seus discursos até a especificidade para tomar **decisão** e **solução** de problemas, a partir da compreensão dos **algoritmos** impulsionando a **inovação**.

5

Como usar em pesquisas

Na análise de **entrevistas**: verificar se determinados perfis de respondentes ocupam regiões distintas no plano.

Em **questionários abertos**: observar como diferentes termos se alinham a grupos de respondentes.

Na Teoria das Representações Sociais (TRS): identificar como as representações podem se organizar em torno de **pólos simbólicos** em disputa (por exemplo: avaliação como “controle” × avaliação como “transformação”).

6

Passo a passo para a interpretação

- 1 Observar a **variância explicada pelos eixos** (quanto maior, mais robusta a oposição).
- 2 Identificar **quais classes ou termos** estão concentrados em cada **lado do eixo**.
- 3 Relacionar esses agrupamentos ao **objeto de estudo**.
- 4 Verificar se há **coerência com a CHD** (a AFC não substitui, mas complementa a CHD).

7

Dicas e cuidados

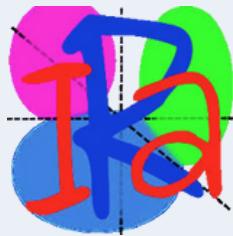
- ! Não interpretar eixos **somente** pelos números da variância — é fundamental articular ao contexto do estudo.
- ! Evitar leituras isoladas de termos: considerar sempre a **relação espacial no plano fatorial**.
- ! A AFC **não substitui** a CHD, mas a complementa; usá-las em conjunto fortalece a interpretação.
- ! Retomar à teoria e aos objetivos da pesquisa para analisar **oposições ou proximidades no gráfico**.
- ! Sempre conferir a **coerência** entre agrupamentos e o *corpus* textual original.

8

Checklist

- ✓ Valor da **variância** (explicada pelos eixos) verificada.
- ✓ **Termos ou classes** concentradas em cada lado do eixo identificados.
- ✓ **Agrupamentos** relacionados ao objeto de estudo identificados.
- ✓ Coerência com os achados da **CHD** constatada.
- ✓ Interpretação articulada à **teoria e aos objetivos** da pesquisa, conferida.

Bibliografia



Bibliografia

BAUER, M. W; GASKELL, G. Pesquisa qualitativa com texto: imagem e som - um manual prático. 13. ed. São Paulo: Vozes, 2015. ISBN 9788532627278.

BRITO, C. A. F.; SÁ, I.R. (2025). 17e édition, la Conférence Internationale sur les Représentations Sociales (CIRS). Diversity of Social Representations of Artificial Intelligence: A Comparative Analysis Across Educators from Different Educational Levels. 2025. (Congresso). Disponível em: <https://cirs2025.sciencesconf.org/?lang=fr>

CAMARGO, B. V.; JUSTO, A. M. Tutorial para uso do software IRaMuTeQ Florianópolis: Universidade Federal de Santa Catarina, 1-18, 2013. 967, 2013. Disponível em: http://www.iramuteq.org/documentation/fichiers/Tutorial%20IRaMuTeQ%20em%20portugues_22.11.2021.pdf/view. Acesso em 23 de set. 2025

LOBO DA COSTA, N. M.; DINIZ, S. N.; BRITO, C. A. F. Conhecimentos Prévios sobre a Ética nas Pesquisas com Seres Humanos. Veredas Revista Interdisciplinar de Humanidades, v. 4, p. 134-155-155, 2021. Disponível em: <https://periodicos.unisa.br/index.php/veredas/article/view/207>. Acesso em 1 out. 2025.

MARCHAND, P.; RATINAUD, P. L'analyse de similitude appliquée aux corpus textuels: les primaires socialistes pour l'élection présidentielle. In: JOURNÉES INTERNATIONALES D'ANALYSE STATISTIQUE DES DONNÉES TEXTUELLES – JADT, 11., 2012, Liège. Actes... Liège: Presses universitaires de Louvain, 2012. p. 687-699.

MOTA, J. S.; SANTOS, J. S.; GOUVEIA, F. A.; GARCIA, P. S.; LIMA, L. B.; BRITO, C. A. F. Aprender inglês em contexto: representações discursivas e especificidades lexicais de contingência no ensino técnico. Cuadernos de Educación y Desarrollo - QUALIS A4, [S. l.], v. 17, n. 11, p. e9943, 2025.

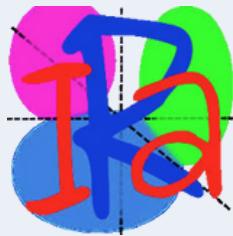
DOI: 10.55905/cuadv17n11-029.

Disponível em:

<https://ojs.cuadernoseducacion.com/ojs/index.php/ced/article/view/9943> Acesso em: 18 nov. 2025.

VIEIRA, S. Bioestatística: tópicos avançados. 5a reimpressão. Ed. Elsevier, 2003. Rio de Janeiro. ISBN: 85-352-0967-0.

Hiperlinks



Hiperlinks

 O **IRaMuTeQ** (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires) é um software livre e de código aberto desenvolvido por Pierre Ratinaud, em 2009, para efetuar análises textuais.

Para mais informações:

[https://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-389X2013000200016#:~:text=Esta%20nota%20visa%20apresentar%20o,por%20Pierr e%20Ratinaud%20\(2009\).](https://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-389X2013000200016#:~:text=Esta%20nota%20visa%20apresentar%20o,por%20Pierr e%20Ratinaud%20(2009).) Acesso em 17 set. 2025

 **Max Reinert** (1944 - 2024) foi um matemático e engenheiro francês especialista na análise de dados textuais e desenvolvedor, nos anos 1970, do software ALCESTE e do método de análise textual – hoje denominado **método Reinert** – que consiste no estabelecimento de classes lexicais em corpora textuais por meio da identificação de polaridades na utilização das palavras pelo sujeito enunciador. O software ALCESTE é o precursor do IRaMuTeQ. Para mais informações acessar:

<https://biblat.unam.mx/hevila/Revistadeeducaopublica/2008/no33/6.pdf>

 O **corpus textual** é um construto primordial na pesquisa qualitativa e deve estar alinhado ao objetivo e a abordagem teórica. Conforme Bauer e Gaskell (2008), construir um corpus significa criar um conjunto estruturado e significativo de dados para compreender as representações, práticas e discursos de determinados grupos sociais.

O corpus textual pode ser criado a partir de diversos tipos de dados, tais como: (a) textos escritos- documentos oficiais, cartas, relatos, discursos, e-mails etc. – (b) registros escritos a partir de interações – entrevistas, grupos focais, observações de campo – (c) textos advindos de materiais digitais- postagens em redes sociais, memes, vídeos, entre outros.

 **Guia IRaMuTeQ** em português disponível em:

http://www.iramuteq.org/documentation/fichiers/Tutorial%20IRaMuTeQ%20em%20portugues_22.11.2021.pdf Acesso em 24 de set. 2025

 A **distribuição χ^2 ou qui-quadrado**, segundo definição de Vieira (2003, p. 59): “O teste de χ^2 (lê-se qui-quadrado, pois χ é uma letra grega, de nome qui) serve para testar a hipótese de que duas variáveis categorizadas são independentes. É um teste clássico e muitas vezes referido, na literatura, como χ^2 de Pearson, uma homenagem ao autor.

Este trabalho está licenciado sob a licença
Creative Commons Atribuição–Não Comercial–Compartilha Igual 4.0
Internacional (CC BY-NC-SA 4.0).
Para ver uma cópia desta licença, visite:
<https://creativecommons.org/licenses/by-nc-sa/4.0/>

ISBN: 978-65-01-85842-5



9 786501 858425

A standard linear barcode representing the ISBN 978-65-01-85842-5. The barcode is enclosed in a rectangular border with rounded corners. Below the barcode, the numbers "9 786501 858425" are printed in a large, bold font.