

Tutorial para uso do software



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

Brigido Vizeu Camargo e Ana Maria Justo (Laboratório de Psicologia Social da Comunicação e Cognição - UFSC - Brasil)

O IRAMUTEQ é um *software* licenciado por GNU GPL (v2) que permite fazer análises estatísticas sobre *corpus* textuais e sobre tabelas indivíduos/palavras. Ele ancora-se no *software* R (www.r-project.org) e na linguagem *python* (www.python.org).

Para instalar o *software* gratuitamente em seu computador, basta fazer o *download* do *software* R em www.r-project.org e instalá-lo; e em seguida fazer o *download* do *software* IRAMUTEQ em www.iramuteq.org, e instalá-lo também. É necessário que antes de instalar o IRAMUTEQ se instale o R, pois o IRAMUTEQ se utilizará do *software* R para processar suas análises.



Figura 1- Interface inicial do software IRAMUTEQ

Instalação do software para sistema operacional *Windows* usando o "KIT IRAMUTEQ"

O "KIT IRAMUTEQ" está disponibilizado no site: www.laccos.com.br. No *link* "Novidades" há uma notícia sobre este *software* e um *link* "clique AQUI". Baixe todo o kit, que é composto dos softwares, referências e um tutorial.

1- Instale o pacote do *softwares* Open Office

Ele é o equivalente gratuito do pacote Microsoft Office. Dois *softwares* deste pacote Open Office nos interessam: o OpenOffice.org Writer, que cria arquivos de texto tipo ".odt"; e o OpenOffice.org Calc que cria arquivos tipo planilha ".ods". O primeiro é usado para digitar os *corpora* e ler relatórios e resultados, e o segundo para entrar dados sob a forma de matrizes de associação de palavras e também para ler, e exportar resultados.

Jamais abra estes arquivos ou qualquer outro gerado pelo IRAMUTEQ com aplicativos da Microsoft (Word, Excel, WordPad ou Bloco de notas), pois eles produzem bugs com o Unicode (UTF-8), o usado pelo IRAMUTEQ.

2- Instale o *software* "R", pois o IRAMUTEQ ancora-se neste *software* e na linguagem Python.

3- Atualize os pacotes de R.

Abra o *software* R. Clique em "Pacotes" + "Atualizar Pacotes" (Figura 2). Escolher: Brasil (o estado brasileiro mais próximo de seu local). Aguardar alguns segundos (dependendo da velocidade de seu computador e da internet pode demorar um pouco mais). Aparecerá alguns itens selecionados em azul para você atualizar. Clique em "OK" e aguarde a atualização. Assim que a atualização finalizar, feche o *software*.

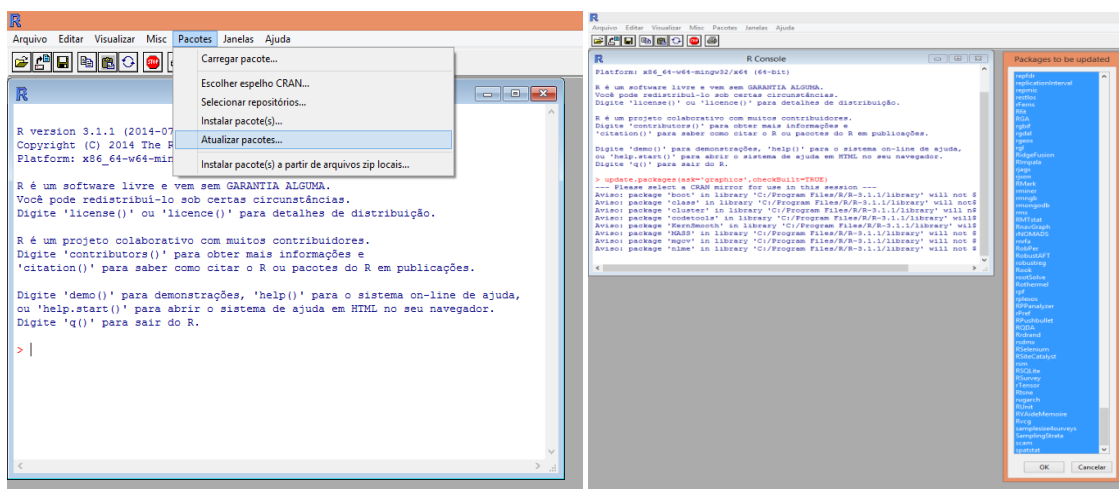


Figura 2- Atualização dos pacotes nas interfaces R

4- Instale o *software* IRAMUTEQ.

5- Atualização das bibliotecas no IRAMUTEQ

Na área de trabalho do seu computador clique no ícone do IRAMUTEQ. **Nesse momento é muito importante estar conectado à internet.** Normalmente aparecerá uma tela apontando que a instalação está incompleta (Figura 3), clique em OK e aguarde a atualização dos arquivos do *software* R.

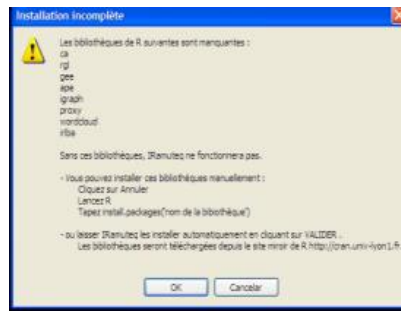


Figura 3- Atualização das bibliotecas na instalação do IRAMUTEQ

Atenção! Caso o *software* não faça essa atualização automaticamente: abra o IRAMUTEQ; clique em "Edição" + "Preferências" + "Atalho de R" e aperte **browse** (Figura 4). **Ache o aplicativo do R** em: Arquivo de programas/R/Bin¹. **Salve e feche** o IRAMUTEQ.

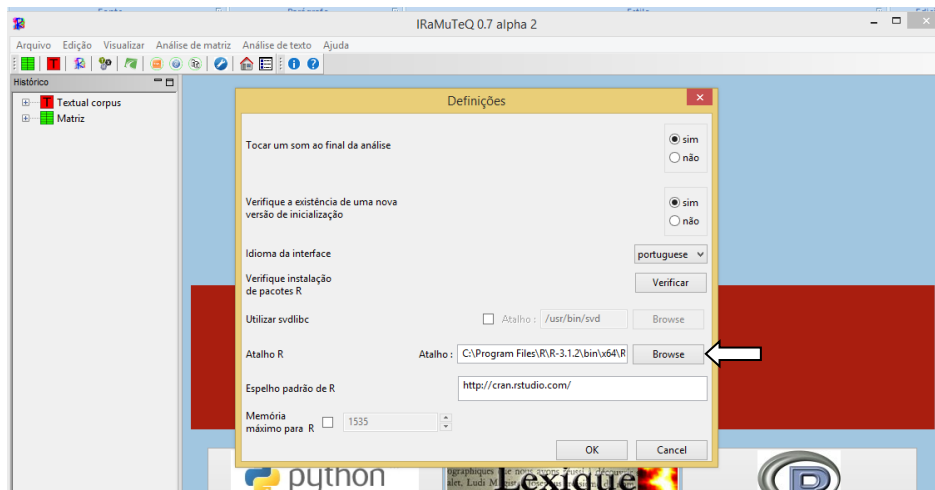


Figura 4- Correção do caminho do R no IRAMUTEQ

Abra novamente o *software* IRAMUTEQ. Clique em "Edição" + "Preferências". Vá em em "Verifique instalação de pacotes R", e clique em "Verificar", espere a instalação (figura 5).

¹ Para usar a interface em português (ainda em preparação) escolha o idioma "português" na interface da figura 4, saia do IRAMUTEQ e o reabra.

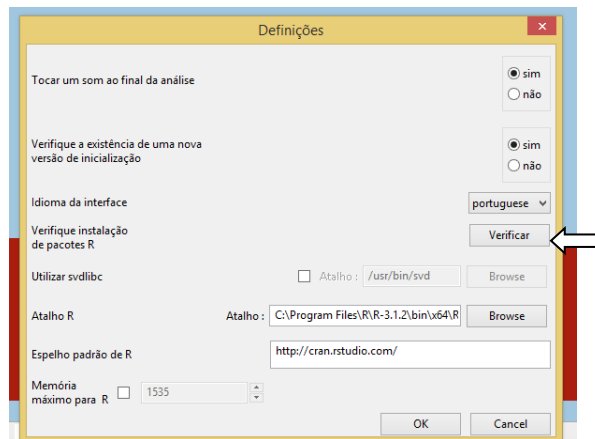


Figura 5- Verificação da instalação das bibliotecas no IRAMUTEQ

Introdução

Trata-se de um *software* que viabiliza diferentes tipos de análise de dados textuais, desde aquelas bem simples, como a lexicografia básica, que abrange sobretudo a lematização e o cálculo de frequência de palavras; até análises multivariadas como classificação hierárquica descendente, análise pós-fatorial de correspondências e análises de similitude. Por meio desse *software*, a distribuição do vocabulário pode ser organizada de forma facilmente compreensível e visualmente clara com representações gráficas pautadas nas análises lexicográficas.

No IRAMUTEQ essas análises podem ser realizadas tanto a partir de um grupo de textos a respeito de uma determinada temática (*corpus* textual) reunidos em um único arquivo de texto; como a partir de matrizes com indivíduos em linha e palavras em coluna, organizadas em planilhas, como é o caso dos bancos de dados construídos a partir de testes de evocações livres.

Parte 1: Análise de corpus textual

A análise textual é um tipo específico de análise de dados, na qual tratamos de material verbal transcrito, ou seja, de textos. Essa análise tem várias finalidades, sendo possível analisar textos, entrevistas, documentos, redações etc. Pode-se a partir da análise textual descrever um material produzido por um produtor, seja individual ou coletivamente, como também podemos utilizar a análise textual com a finalidade comparativa, relacional, comparando produções diferentes em função de variáveis específicas que descrevem quem produziu o texto. Para que se possa

compreender a análise textual, é necessário inicialmente delimitar alguns conceitos importantes.

As noções de: *corpus*, texto e segmento de texto

Corpus

O *corpus* é construído pelo pesquisador. É o conjunto de unidades de contexto inicial que se pretende analisar. Por exemplo, se um pesquisador decide analisar as matérias sobre beleza que saíram numa revista no período de cinco anos; o conjunto destas matérias constituirá um *corpus*. O *corpus* é construído pelo pesquisador.

Texto

A definição destas unidades é feita pelo pesquisador e depende da natureza da pesquisa. No exemplo anterior cada matéria sobre beleza seria um texto. Se a análise for aplicada a um conjunto de entrevistas, cada uma delas será um texto. Caso a análise diga respeito às respostas de "n" participantes a uma questão aberta, cada resposta será um texto e teremos "n" textos. Quando se tratar de pesquisas documentais, atas de reuniões, cartas, etc.; cada exemplar destes documentos será um texto.

Um conjunto de unidades de textos constitui um *corpus* de análise. O *corpus* adequado à análise do tipo Classificação Hierárquica Descendente deve constituir-se de um conjunto textual centrado em um tema. O material textual deve ser monotemático, pois a análise de textos sobre vários itens previamente estruturados ou diversos temas resulta na reprodução da estruturação prévia dos mesmos.

No caso de entrevistas, onde há falas que produzem textos mais extensos, desde que o grupo seja homogêneo, é suficiente entre 20 e 30 textos (Ghiglione e Matalon, 1993). Se o delineamento é comparativo, sugere-se pelo menos 20 textos para cada grupo.

Em se tratando de respostas a questões abertas de um questionário, recomenda-se compor o *corpus* com respostas a uma mesma questão, para garantir que elas se refiram a um mesmo tema. Caso as questões digam respeito a temas ou aspectos diferentes, é necessário realizar uma análise para cada questão. Como mencionado anteriormente, a análise é sensível à estruturação do estímulo que produz o material textual, e isto é uma importante fonte de invalidação das conclusões. Quando as respostas apresentarem uma média em torno de três ou quatro linhas, é necessário um número bem maior de respostas para a constituição de um *corpus* de análise.

Os textos são separados por linhas de comando também chamadas de "linhas com asteriscos". No caso de entrevistas, por exemplo, como cada uma delas é um texto, e eles necessariamente devem começar com uma linha de comando, esta linha informa o número de identificação do entrevistado (do produtor do texto que se segue) e algumas características (variáveis) que são importantes para o delineamento da pesquisa (como: sexo, faixa etária, afiliação a determinados grupos, nível social e cultural, etc.). Isto depende de cada pesquisa e o número de modalidades de cada uma destas variáveis depende do delineamento da pesquisa e do número de entrevistas realizadas.

Segmentos de texto

Os segmentos de texto (ST), na maior parte das vezes, tem o tamanho de três linhas, dimensionadas pelo *software* em função do tamanho do *corpus*. Os segmentos de textos são os ambientes das palavras. Podem ser construídos pelo pesquisador, ou automaticamente pelo *software*.

Embora seja o pesquisador que demarca os textos, nem sempre é ele que controla a divisão do corpus em segmentos de texto (ST). Numa análise padrão (standart), após reconhecer as indicações dos textos (pelas linhas com asteriscos) é o *software* que divide o material em ST. Mas o pesquisador pode configurar a divisão dos segmentos, por exemplo: no caso de uma grande quantidade de respostas curtas a uma pergunta aberta de um questionário, aconselha-se cada texto seja definido como um único ST.

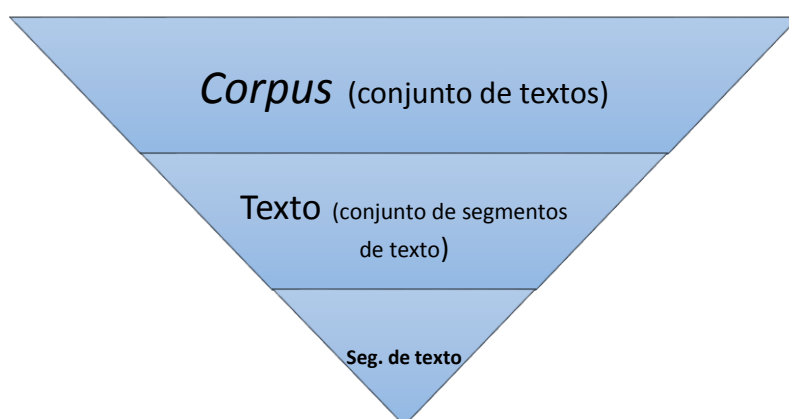


Figura 6- Noções de Corpus, texto e segmento de texto.

Preparação de um *corpus* textual para análise

O primeiro passo para realizar a análise é construir o *corpus* a ser analisado, que deve ser feito de acordo com os seguintes procedimentos:

- 1- Colocar todos os textos (entrevistas, artigos, textos, documentos ou respostas a uma única questão) **em um único arquivo de texto no software OpenOffice.org** (<http://www.openoffice.org/>) **ou LibreOffice** (<http://pt-br.libreoffice.org/>). **Jamais abra estes arquivos ou qualquer outro gerado pelo IRAMUTEQ com aplicativos da Microsoft (Word, Excel, WordPad ou Bloco de notas), pois eles produzem bugs com o Unicode (UTF-8), o usado pelo IRAMUTEQ.**
- 2- Separar os textos com linhas de comando (com asteriscos). Por exemplo, para cada entrevista ser reconhecida pelo *software* como um texto, elas devem começar por uma linha deste tipo. **Observação:** Deixe uma linha em branco antes da primeira linha de comando.

Exemplo de uma linha com asteriscos:

**** *ind_01 *ida_1 *par_2 *fil_2 *temp_2 *caus_1

Digitar quatro asteriscos (sem espaço em branco antes deles), um espaço branco depois, um asterisco e o nome da variável (sem espaço branco entre eles), um traço em baixo da linha e o código da modalidade da variável (também sem espaço branco entre eles), um espaço em branco e depois o asterisco da segunda variável, e assim por diante. Esta linha exemplo foi extraída de uma pesquisa com garotas de *software*, sobre prevenção de doenças sexualmente transmissíveis e gravidez. Ela indica que o material textual que a segue (respostas a uma entrevista) refere-se ao indivíduo nº 1 (utiliza-se dois dígitos, pois a amostra tem mais de 10 indivíduos e menos de 100), sua idade situa-se entre 19 e 26 anos (onde 1= 19 a 26 anos, 2= 27 a 47 anos); ela não tem um parceiro sexual fixo (namorado ou marido) (pois 1= ter parceiro e 2= não ter parceiro); o tempo que ela faz "programas" situa-se entre 13 e 36 meses (onde 1= até 12 meses, 2= de 13 a 36 meses e 3= de 48 a 132 meses) e a causa que ela atribuiu para ser profissional do sexo foi familiar (onde 1= causa familiar, 2= necessidade financeira, 3 = sustento da família, 4= relação amorosa frustrada e 5= não respondeu). Imediatamente após esta linha com asterisco teclar ENTER, e sem tabulação, e linha em branco, digite ou coloque o texto da resposta à entrevista correspondente a este indivíduo n.º 1.

- 3- **Existem duas maneiras de preparar as linhas de um corpus.** A primeira, a original ou monotemática, onde cada linha é seguida por um texto sem separações. Uma segunda maneira, a chamada temática, onde cada linha pode conter duas ou mais temáticas com a inclusão de linhas subordinadas a principal. A análise de *corpus* com divisões temáticas (temas diferentes) nos informa sobre as relações entre o conteúdo de um tema com o outro tema; e pode ser usada como uma análise preliminar de natureza mais exploratória (para se ter uma visão de todo da coleta de material textual, mas deve-se fazer as análises monotemáticas, pois são elas que aprofundam a compreensão do significado do material estudado.

Extrato exemplo de um corpus da maneira original (monotemática)

**** *ind_01 *ida_1 *par_2 *fil_2 *temp_2 *caus_1

Eu utilizo remédios para não engravidar como o anticoncepcional ou injeção, que é mais fácil, porque para quem bebe, ficar tomando remédio é muito difícil, às vezes corta o efeito, então injeção é mais garantido. Utilizo preservativo também, com certeza, porque são bem garantidos. Eu tenho que me cuidar. Acho que devo usar o anticoncepcional, não posso só confiar na camisinha. Posso ter um filho de quem eu mal conheço, ou pegar qualquer outra doença, se isso acontecer, e, eu chegar nessa pessoa e dizer que estou grávida dela. Ele vai me dizer tu és uma prostituta mesmo, está grávida de qualquer um. Eu tenho conhecimento de todas as dst, porque geralmente, quando a gente estuda na escola, tem várias palestras, então ficamos conhecendo todas. Agora, só não se cuida quem não quer mesmo. Para mim, todas são arriscadas, tem que se prevenir de todas, não é fácil andar com um, com outro, porque você sabe que tens que ter cabeça, não é porque você está bêbada, você tem que estar ciente. Às vezes, a gente sai com pessoas aqui dentro que você não suporta, você vê cada coisa, que só a gente mesmo vê para saber como é a realidade. Se você não está ciente, acontece cada coisa que você só vai ver o resultado mais tarde. Às vezes, muitas pessoas não conseguem uma menina na rua ou uma namorada, então vem para uma boate, acham que uma menina da noite é obrigada a aceitar. A menina da noite é considerada vulgar. Eu não me considero uma menina vulgar. Tenho meus motivos, não aceito que chegue aqui e me xingue de nome vulgar, não aceito mesmo, pois não sou, se eu fosse, estaria na rua e topava qualquer coisa. Mas é bem diferente trabalhar em boate, zona e cabaré. Tem muita diferença. Já tive corrimento, que é normal, mas dst não. Corrimento é uma coisa normal, que você pode pegar com o lubrificante da camisinha ou com qualquer outra coisa, com o sabonete, roupas, mas nunca tive dst.

**** *ind_02 *ida_2 *par_1 *fil_2 *temp_3 *caus_2

Eu utilizo camisinha sempre, porque além de prevenir a gravidez, previne a aids e outras doenças, não tem como não usar. Uso camisinha sempre, ela é boa para tudo. Eu conheço várias doenças venéreas, gonorreia, cancro, chato, são tantas, inclusive eu já peguei uma vez, mas não foi na noite, foi de um namorado que eu tinha, a tempos atrás, peguei gonorreia e eu não sabia. Eu não batalhava na época, não imaginava que meu namorado fosse passar para mim. Como foi passando o tempo e comecei a sentir dor, então fui procurar um médico, tive que fazer uma cirurgia, pois já estava bem avançado. Foi assim, que fiquei sabendo que peguei gonorreia. Não tenho medo de falar porque qualquer um corre o risco de pegar, todas essas doenças são ruins, não tem muito o que classificar. Para me proteger das dst faço sexo oral e normal somente com camisinha, e, anal eu não faço de jeito nenhum.

**** *ind_03 *ida_1 *par_2 *fil_1 *temp_1 *caus_2

Eu tomo anticoncepcional e uso camisinha. Eu não tomo propriamente a pílula, tomo é injeção, porque a gente sempre esquece. Por isso eu tomo os dois, porque se romper a (CONTINUA)

Extrato exemplo de um corpus da maneira temática

**** *ind_01 *ida_1 *par_2 *fil_2 *temp_2 *caus_1

-*tema_prevenção

Eu utilizo remédios para não engravidar como o anticoncepcional ou injeção, que é mais fácil, porque para quem bebe, ficar tomando remédio é muito difícil, às vezes corta o efeito, então injeção é mais garantido. Utilizo preservativo também, com certeza, porque são bem garantidos. Eu tenho que me cuidar. Acho que devo usar o anticoncepcional, não posso só confiar na camisinha. Posso ter um filho de quem eu mal conheço, ou pegar qualquer outra doença, se isso acontecer, e, eu chegar nessa pessoa e dizer que estou grávida dela. Ele vai me dizer tu és uma prostituta mesmo, está grávida de qualquer um.

-*tema_dst

Eu tenho conhecimento de todas as dst, porque geralmente, quando a gente estuda na escola, tem várias palestras, então ficamos conhecendo todas. Agora, só não se cuida quem não quer mesmo. Para mim, todas são arriscadas, tem que se prevenir de todas, não é fácil andar com um, com outro, porque você sabe que tens que ter cabeça, não é porque você está bêbada, você tem que estar ciente. Às vezes, a gente sai com pessoas aqui dentro que você não suporta, você vê cada coisa, que só a gente mesmo vê para saber como é a realidade. Se você não está ciente, acontece cada coisa que você só vai ver o resultado mais tarde. Às vezes, muitas pessoas não conseguem uma menina na rua ou uma namorada, então vem para uma boate, acham que uma menina da noite é obrigada a aceitar. A menina da noite é considerada vulgar. Eu não me considero uma menina vulgar. Tenho meus motivos, não aceito que chegue aqui e me xingue de nome vulgar, não aceito mesmo, pois não sou, se eu fosse, estaria na rua e topava qualquer coisa. Mas é bem diferente trabalhar em boate, zona e cabaré. Tem muita diferença. Já tive corrimento, que é normal, mas dst não. Corrimento é uma coisa normal, que você pode pegar com o lubrificante da camisinha ou com qualquer outra coisa, com o sabonete, roupas, mas nunca tive dst.

**** *ind_02 *ida_2 *par_1 *fil_2 *temp_3 *caus_2

-*tema_prevenção

Eu utilizo camisinha sempre, porque além de prevenir a gravidez, previne a aids e outras doenças, não tem como não usar. Uso camisinha sempre, ela é boa para tudo.

-*tema_dst

Eu conheço várias doenças venéreas, gonorreia, cancro, chato, são tantas, inclusive eu já peguei uma vez, mas não foi na noite, foi de um namorado que eu tinha, a tempos atrás, peguei gonorreia e eu não sabia. Eu não batalhava na época, não imaginava que meu namorado fosse passar para mim. Como foi passando o tempo e comecei a sentir dor, então fui procurar um médico, tive que fazer uma cirurgia, pois já estava bem avançado. Foi assim, que fiquei sabendo que peguei gonorreia. Não tenho medo de falar porque qualquer um corre o risco de pegar, todas essas doenças são ruins, não tem muito o que

classificar. Para me proteger das dst faço sexo oral e normal somente com camisinha, e, anal eu não faço de jeito nenhum.

**** *ind_03 *ida_1 *par_2 *fil_1 *temp_1 *caus_2

-*tema_prevenção

Eu tomo anticoncepcional e uso camisinha. Eu não tomo propriamente a pílula, tomo é injeção, porque a gente sempre esquece. Por isso eu tomo os dois, porque se romper a (CONTINUA)

Observação: Após preparar o corpus, recomenda-se que se leia o mesmo atentamente, especialmente no que se refere às linhas de comando. Esta verificação precisa ser realizada pelo pesquisador para que o texto possa ser processado.

- 4- **Corrigir e revisar todo o arquivo**, para que os erros de digitação ou outros não sejam tratados como palavras diferentes.
- 5- **A pontuação deve ser observada**, no entanto sugere-se não deixar parágrafos (devido à dificuldade entre nós no uso correto dos mesmos).
- 6- No caso de entrevistas ou questionários, as perguntas e o material verbal produzido pelo pesquisador (intervenções e anotações) devem ser suprimidos para não entrar na análise. Ao suprimir recupere os referentes.
- 7- Não justifique o texto, não use negrito, nem itálico ou outro recurso semelhante.
- 8- É desejável certa **uniformidade em relação às siglas**, ou as usa sempre ou coloque tudo por extenso unido por traço *underline*. Por exemplo: ou oms ou organização_mundial_de_saúde.
- 9- **As palavras compostas hifenizadas** quando digitadas com hífen são entendidas como duas palavras (o hífen vira espaço em branco). Caso necessite-se analisar palavras compostas hifenizadas ou não, **una-as com um traço underline**. Ex: "alto-mar" fica "alto_mar"; "terça-feira" fica "terça_feira"; e "bate-papo" fica "bate_papo".
- 10- Todos os **verbos que utilizem pronomes** devem estar **na forma de próclise**, pois o dicionário não prevê as flexões verbo-pronominais. Ex: No lugar de "torneime", a escrita deve ser: "me tornei".
- 11- Caso possa, **evite uso de diminutivos**, pelas características do dicionário.
- 12- Números devem ser mantidos em sua forma algarísmica. Ex: usar "2013", no lugar de "dois mil e treze"; "70" no lugar de "setenta".
- 13- **Não usar em nenhuma parte do arquivo** dos textos os seguintes caracteres: aspas ("), apóstrofo ('), hífen (-), cifrão (\$), porcentagem (%), reticências (...), e nem asterisco (*). Este último é usado somente nas linhas que antecedem cada texto (linhas de comando).
- 14- **O arquivo com o corpus** preparado no *software* OpenOffice.org ou no LibreOffice deve ser **salvo em uma nova pasta criada no desktop**, somente para a análise, com um nome curto e como **texto codificado** (nome_do_arquivo.txt). No OpenOffice.org esta opção abre uma primeira janela e devemos escolher "manter

formato atual”, e uma segunda janela onde as opções “Conjuntos de caracteres” e “Quebra de parágrafo” devem ser respectivamente “Unicode (UTF- 8)” e “LF”.

- 15- A cada nova análise com o mesmo corpus **não reutilize o arquivo txt (texto codificado)** precedente, crie um novo a partir do arquivo tipo odt (que é o formato que deverá ser arquivado).

Tipos de análise de corpus textual IRAMUTEQ

O IRAMUTEQ oferece a possibilidade de diferentes formas de análise de dados textuais, desde aquelas bem simples, como a lexicografia básica (como cálculo de frequência de palavras), até análises multivariadas (classificação hierárquica descendente).

I) Análises lexicográficas clássicas – Identifica e reformata as unidades de texto, transformando textos em ST, identifica a quantidade de palavras, frequência média e *hapax* (palavras com frequência igual a um), pesquisa o vocabulário e reduz das palavras com base em suas raízes (formas reduzidas), cria do dicionário de formas reduzidas, identifica formas ativas e suplementares.

II) Especificidades e Análise Fatorial de Correspondência – Associa textos com variáveis, ou seja, possibilita a análise da produção textual em função das variáveis de caracterização. Oferece uma análise fatorial de correspondência para variáveis com no mínimo 3 modalidades.

III) Método da Classificação Hierárquica Descendente (CHD) – Os ST são classificados em função dos seus respectivos vocabulários, e o conjunto deles é repartido em função da frequência das formas reduzidas. A partir de matrizes cruzando formas reduzidas e ST (em repetidos testes do tipo χ^2), aplica-se o método de CHD e obtém-se uma classificação definitiva. Esta análise visa obter classes de ST que, ao mesmo tempo, apresentam vocabulário semelhante entre si, e vocabulário diferente dos segmentos das outras classes. A partir dessas análises o *software* organiza a análise dos dados em **um dendograma** que ilustra as relações entre as classes.

O *software* executa cálculos e fornece resultados que nos permite a descrição de cada uma das classes, principalmente, pelo seu vocabulário característico (léxico) e pelas suas palavras com asterisco (variáveis). Além disto, o *software* fornece uma outra forma de apresentação dos resultados, através de uma análise fatorial de correspondência feita a partir da CHD. Com base nas classes escolhidas, o *software*

calcula e fornece os ST mais característicos de cada classe permitindo a contextualização do vocabulário típico de cada classe.

Estas classes de palavras e segmentos de texto, em nível do *software* são compostas de vários segmentos em função de uma classificação segundo a distribuição do seu vocabulário. Em nível interpretativo depende do marco teórico de cada pesquisa. Reinert (1990), ao estudar a literatura francesa considerou cada classe como uma noção de "mundo", enquanto um quadro perceptivo-cognitivo com certa estabilidade temporal associado a um ambiente complexo. Em pesquisas no campo da linguística estas classes foram interpretadas como campos lexicais (Cros, 1993) ou contextos semânticos. Em pesquisas no campo da psicologia social, particularmente aquelas interessadas em estudar o conhecimento do senso comum, tendo em vista o estatuto que elas conferem às manifestações linguísticas, estas classes podem indicar representações sociais ou campos de imagens sobre um dado objeto, ou somente aspectos de uma mesma representação social (Veloz, Nascimento-Schulze e Camargo, 1999).

Na maior parte das vezes não há coincidência entre o número de classes e o número de representações sociais envolvidas, como ocorreu no estudo relativo que se acabou de citar. O que vai definir se elas indicam representações sociais ou apenas uma representação social é o seu conteúdo, e sua relação com fatores ligados ao plano geral de cada pesquisa, geralmente expresso na seleção diferenciada dos participantes segundo sua afiliação grupal, suas práticas sociais anteriores, etc.

IV) Análise de similitude – Esse tipo de análise baseia-se na teoria dos grafos e é utilizada frequentemente por pesquisadores das representações sociais. Possibilita identificar as coocorrências entre as palavras e seu resultado traz indicações da conexão entre as palavras, auxiliando na identificação da estrutura do conteúdo de um corpus textual. Permite também identificar as partes comuns e as especificidades em função das variáveis descritivas identificadas na análise (Marchand & Ratinaud, 2012).

V) Nuvem de palavras – Agrupa as palavras e as organiza graficamente em função da sua frequência. É uma análise lexical mais simples, porém graficamente interessante.

Processando a análise no software

Inicialmente, abra o *software* para trabalhar em sua interface, e importe o *corpus*. Na barra de ferramentas superior clique em *Arquivo* e *Abrir um corpus textual*, conforme indica a Figura 7. Localize e selecione o *corpus* que deseja analisar e clique em *Abrir*.



Figura 7- Importação do corpus de análise.

No momento em que o *software* importar o *corpus*, uma nova janela será aberta (Figura 8).

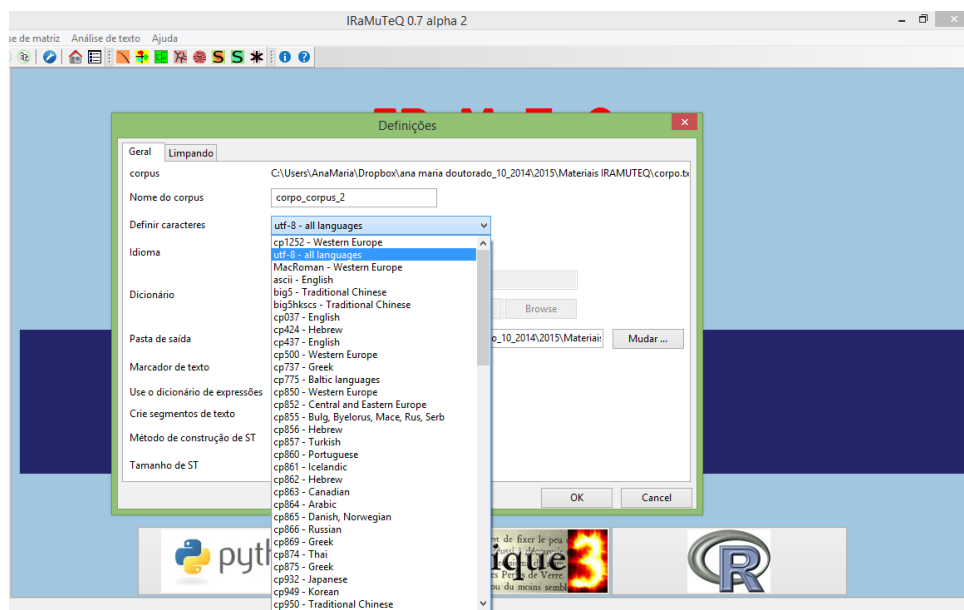


Figura 8- Configurações de análise – codificação do corpus.

Nessa janela (Figura 8) podem ser observadas algumas configurações do software para analisar os dados textuais. A maior parte das configurações, na aba Geral, pode ser mantidas conforme o padrão, com exceção de duas que precisam ser modificadas. A primeira refere-se à codificação (*Definir caracteres*) do texto, que deve ser a segunda opção de cima para baixo: “*utf-8 – all languages*”.

A outra configuração é a da língua (*Idioma*). Conforme a Figura 9, selecione a língua: português no caso de o texto estar nesta língua, ou escolha a língua correspondente ao caso.

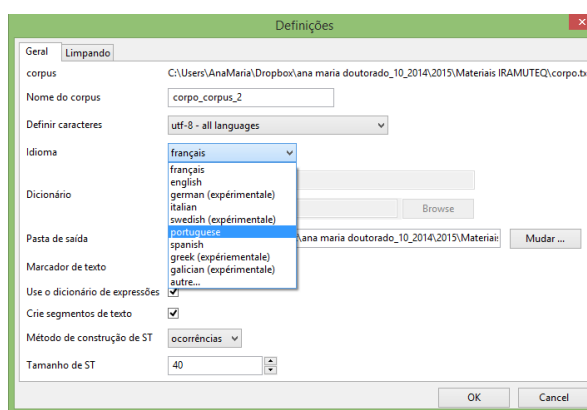


Figura 9- Configurações de análise – língua.

Clique em OK e aguarde alguns segundos para que se processe importação dos dados. Em seguida, na grande janela da direita aparecerá uma breve descrição do corpus, como indicado na figura 10, onde se pode verificar, o número de Textos e de Segmentos de texto, Formas identificadas, Ocorrências, e Frequência de Hapax.

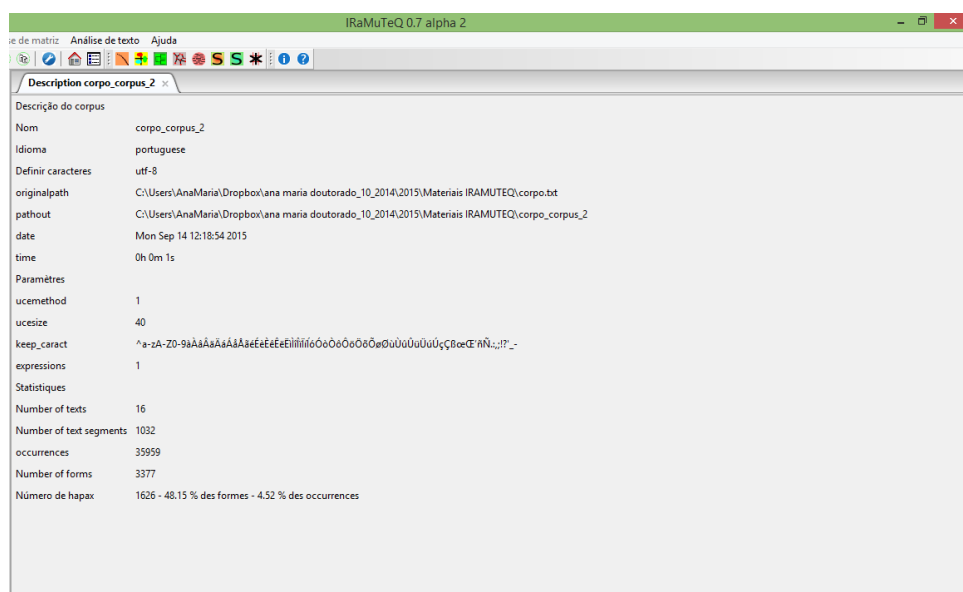


Figura 10- Resultados preliminares, descrição do corpus.

Tendo sido realizada a importação do corpus, as análises já podem ser iniciadas. Para realizá-las, na barra de ferramentas superior, selecione *Análise de texto*, e aparecerão as possibilidades de análise (Figura 11).

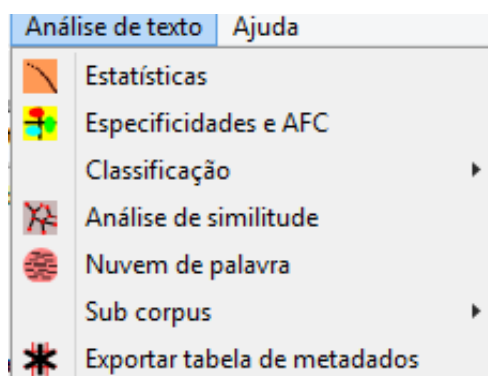


Figura 11- Escolha da análise textual

Toda a vez que for escolhida uma análise, surgirá uma nova janela perguntando se você deseja manter a *Lematização*. Deixe selecionado SIM, pois assim o *software* utilizará o dicionário de formas reduzidas para processar a análise. Nessa janela você também poderá editar as formas ativas e suplementares, se assim desejar, clicando em *Propriedades Chave*. É indicado que o pesquisador selecione quais as classes gramaticais ele deseja considerar ativas na análise (**0= palavras são eliminadas; 1= palavras são ativas; 2= palavras são suplementares**). Uma vez feita essa alteração nas preferências da lematização, ela se manterá nas análises subsequentes para um mesmo *corpus*. O pesquisador pode alterá-las novamente no momento que desejar. Após escolher as classes gramaticais clique em *Ok*, e novamente em *Ok* que a análise será realizada.

Sugestão de parametragem das Propriedades Chave:

Sugere-se aos estudos de psicologia que se utilize a parametragem conforme a ilustração da figura 12. Esta parametragem traz uma boa limpeza para pesquisa onde o conteúdo do texto é o mais importante. A lógica é trabalhar com os elementos de linguagem "plenos" como ativos: adjetivos, formas não reconhecidas, nomes (substantivos), verbos; e com nomes (substantivos) e verbos auxiliares como complementares (suplementares); eliminando as "palavras instrumento". Além disto selecionem as palavras na análise de similitude e na nuvem, e desconsiderem também as palavras com maior frequência associadas as perguntas.

Clés d'analyse

Choix des clés d'analyse
0=éliminé; 1=active; 2=supplémentaire

Adjectif	1	voir liste	Conjonction	0	voir liste
Adjectif démonstratif	0	voir liste	Formes non reconnues	1	voir liste
Adjectif indéfini	0	voir liste	Nom commun	1	voir liste
Adjectif interrogatif	0	voir liste	Nom supplémentaire	2	voir liste
Adjectif numérique	0	voir liste	Onomatopée	0	voir liste
Adjectif possessif	0	voir liste	Pronom démonstratif	0	voir liste
Adjectif supplémentaire	0	voir liste	Pronom indéfini	0	voir liste
Adverbe	0	voir liste	Pronom personnel	0	voir liste
Adverbe supplémentaire	0	voir liste	Pronom possessif	0	voir liste
Article défini	0	voir liste	Pronom relatif	0	voir liste
Article indéfini	0	voir liste	Préposition	1	voir liste
Auxiliaire	0	voir liste	Verbe	1	voir liste
Chiffre	0	voir liste	Verbe supplémentaire	2	voir liste

OK

Figura 12- Parametragem de palavras ativas, suplementares e eliminadas da análise.

Análise: Estatísticas textuais

Na primeira opção de análise, "Estatísticas textuais", o *software* fornece o número de textos e segmentos de textos, ocorrências, frequência média das palavras, bem como a frequência total de cada forma; e sua classificação gramatical, de acordo com o dicionário de formas reduzidas. Na interface dos resultados você poderá visualizar o diagrama de Zipf (Figura 13), que apresenta o comportamento das frequências das palavras no *corpus*, num gráfico que ilustra a distribuição de frequência X rang.

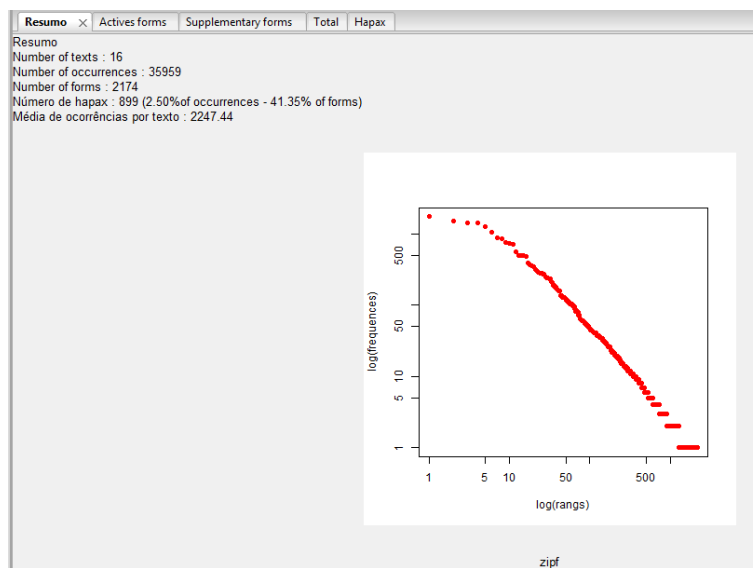


Figura 13- Diagrama de Zipf

Na coluna que se apresenta à esquerda, na interface do *software*, você identifica essa análise como: **NOME DO CORPUS_stat_1**. Colocando o cursor sobre esse nome, você pode clicar com o botão direito do mouse sobre o mesmo e selecionar algumas opções, dentre elas, exportar o dicionário de formas reduzidas, o qual será salvo na pasta em que foi salvo o corpus inicial, dentro de uma sub-pasta denominada: **NOME DO CORPUS_stat_1**.

Ao gerar o dicionário, o *software* classifica as palavras em formas gramaticais, com a seguinte codificação, a qual será utilizada ao longo de todas as análises:

adj = adjetivo
 adj_num = adjetivo numeral
 adj_sup = adjetivo colocado em forma suplementar
 adv = advérbio
 adv_sup = advérbio colocado em forma suplementar
 art_def = artigo definido
 conj = conjunção
 nom = nome
 nom_sup = nome colocado em forma suplementar
 nr = não reconhecida
 ono = onomatopéia
 pro_ind = pronome indefinido
 pre = preposição
 ver = verbo
 verbe_sup = verbo colocado em forma suplementar

Análise: Especificidades e AFC

Ao selecionar o modo "Especificidades e AFC", você deverá escolher a variável categorial em função da qual deseja realizar a análise. Selecione-a na janela que aparece na interface e clique em *Ok*. Aguarde alguns instantes e os resultados aparecerão na janela principal, conforme a figura 14.

Formas comuns	Tipos	Formas frequências	Tipos de frequências	Frequência relativa das formas	Tipos de frequências relativas
formas	*ctb_1	*ctb_2			
gordo	12.2941	-12.2941			
ela	9.3787	-9.3787			
pro	9.014	-9.014			
mídia	7.5278	-7.5278			
padrão	7.2615	-7.2615			
bonito	6.8863	-6.8863			
magro	6.3627	-6.3627			
feliz	5.5467	-5.5467			
pessoa	5.2417	-5.2417			
barbie	4.685	-4.685			
querer	4.4499	-4.4499			
tv	3.6808	-3.6808			
baixo	3.6808	-3.6808			
autoestima	3.6654	-3.6654			
beleza	3.6493	-3.6493			
me	3.4625	-3.4625			
botar	3.3461	-3.3461			
muito	3.2157	-3.2157			
criança	3.0957	-3.0957			
assim	2.8884	-2.8884			
perfeito	2.8459	-2.8459			
aceitar	2.8413	-2.8413			
comprar	2.8413	-2.8413			
que	2.8388	-2.8388			
expressão	2.7234	-2.7234			
expressar	2.7132	-2.7132			
tal	2.6566	-2.6566			
ser	2.5995	-2.5995			
ninguém	2.5417	-2.5417			

Figura 14- Resultados, especificidades e AFC.

Ao clicar com o botão direito do mouse sobre qualquer uma das palavras apresentadas na tabela (Fig. 14) e em seguida em *Concordância*, uma nova janela se abrirá e nela é possível identificar os segmentos de textos que contém a palavra, recuperando o seu contexto.

Análise: Classificação hierárquica descendente (CHD)

Ao escolher a CHD (*Análises, Classificação, Método de Reinert*), você pode optar por três possibilidades de análise na janela que aparecerá na interface do IRAMUTEQ.

- DUPLA SOBRE RST – não utilizada, pois usualmente tem baixo aproveitamento do *corpus*.
- SIMPLES SOBRE ST– que equivale a uma análise sobre os segmentos de texto, delimitados pelo *software* (Análise Standart), recomendada para respostas longas.
- SIMPLES SOBRE TEXTOS – que realiza a análise considerando a os textos, sem dividi-los em segmentos de texto. Recomendada para respostas curtas².

Escolha uma das modalidades de classificação. Nas demais configurações (parametragens) não é necessária nenhuma modificação. Clique em *OK* e aguarde alguns segundos até que a análise seja finalizada. Na interface de resultados aparecerão alguns dados importantes à CHD (Fig. 15), seguidos do dendograma (Fig. 16).

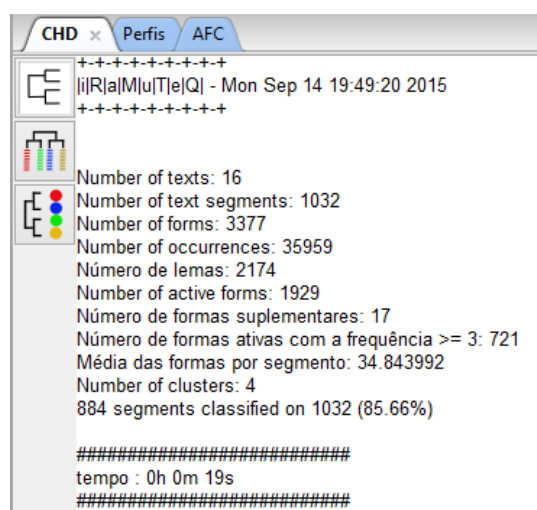


Figura 15- Principais pontos da CHD a serem considerados

²Nesse caso, é necessária uma parametragem anterior. Logo ao importar o corpus, além de indicar a codificação e a língua, selecione “*parágrafos*” **como método de construção dos ST**.

Nessa parte da descrição dos resultados, as principais características da análise a serem consideradas são as seguintes:

- Número de textos (*Number of texts*) = 16 (o *software* reconhece a separação do corpus em 16 unidades de texto iniciais).
- Número de segmentos de textos (*number of text segments*) = 1.032 (o *software* reparte em 1.032 segmentos de texto)
- Número de formas distintas (*number of forms*) = 3.377
- Número de ocorrências (*number of occurrences*) = 35.959
- Número de formas ativas: 1.929
- Número de classes (*number of clusters*) = 4
- Retenção de segmentos de texto: **884 segments classified on 1.032 (85,66%)**

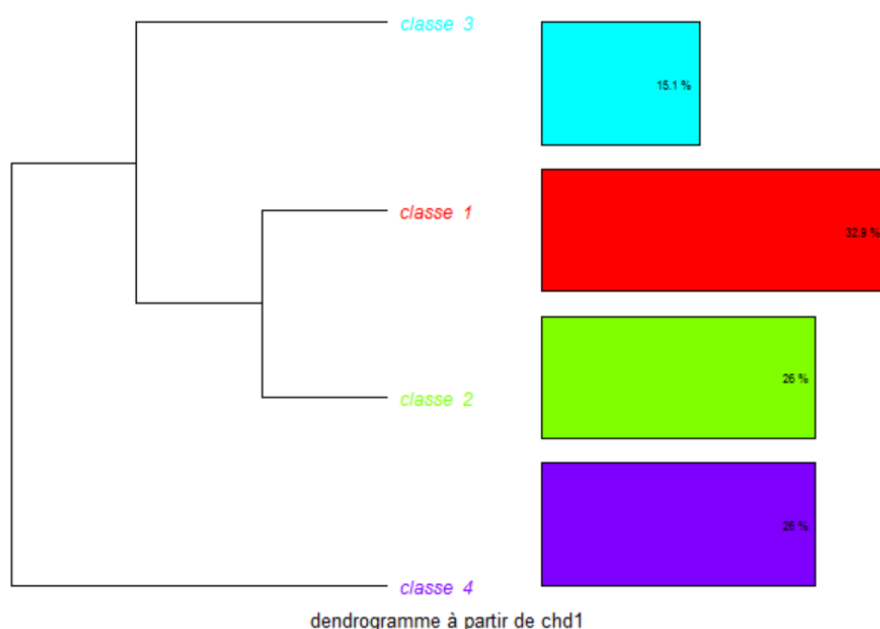


Figura 16- Dendrograma da CHD.

Na aba CHD dos resultados, é possível ter acesso ao dendrograma, que apresenta as partições que foram feitas no *corpus* até que se chegasse às classes finais. Lê-se o dendrograma da esquerda para a direita. No exemplo da figura 16 num primeiro momento, o *corpus* "Corpo", utilizado aqui como exemplo, foi dividido (1ª partição ou iteração) em dois *sub-corpus*, separando a classe 4 do restante do material. Num segundo momento o *sub-corpus* maior foi dividido, originando a classe 3 (2ª partição ou iteração). Num terceiro momento há uma partição originando as classes 1 e 2. A CHD parou aqui, pois as 4 classes mostraram-se estáveis, ou seja, compostas de unidades de segmentos de texto com vocabulário semelhante.

Uma segunda forma de apresentação do mesmo dendograma é apresentada na figura 17.

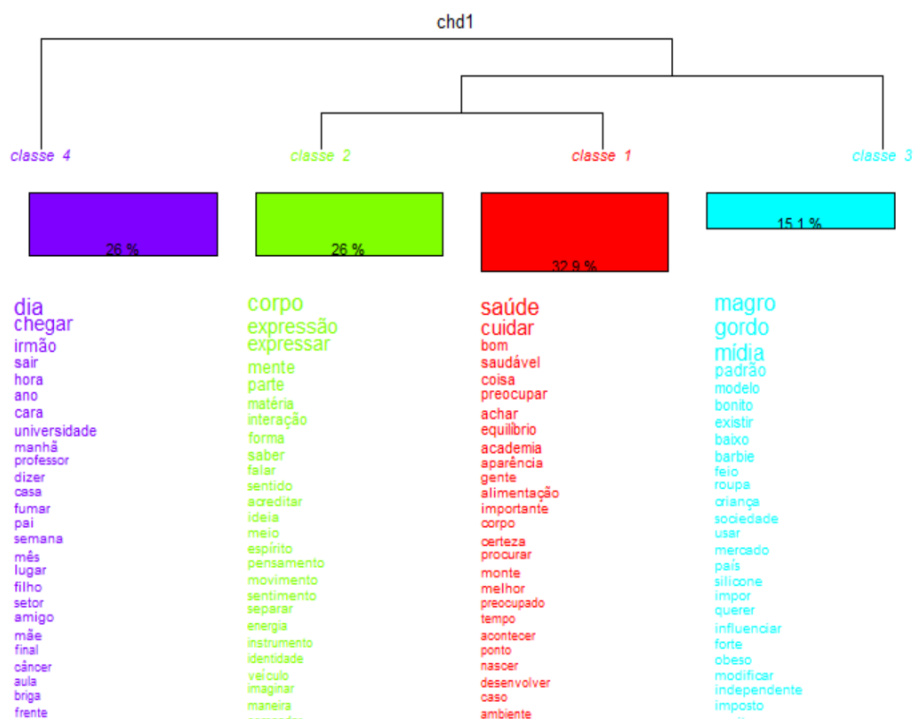


Figura 17- Dendograma da CHD.

Explorando os dados da CHD

Além do dendograma, essa interface de resultados também possibilita que se identifique o conteúdo lexical de cada uma das classes (para acessá-lo, basta clicar na aba *Perfis*) e uma representação fatorial da CHD (para acessá-la, basta clicar na aba AFC).

Na aba *Perfis*, para cada classe encontram-se dados referentes ao seu conteúdo: *n.* (número que ordena as palavras na tabela); *eff. st* (número de segmentos de texto que contêm a palavra na classe); *eff. total* (número de segmentos de texto no *corpus* que contêm, ao menos uma vez, a palavra citada); *pourcentage* (percentagem de ocorrência da palavra nos segmentos de texto nessa classe, em relação a sua ocorrência no *corpus*); *chi2* (X^2 de associação da palavra com a classe); *Type* (classe gramatical em que a palavra foi identificada no dicionário de formas); *Forme* (identifica a palavra) e *P* (identifica o nível de significância da associação da palavra com a classe). A figura 18 ilustra a aba "perfis".

1 Classe 1		2 Classe 2		3 Classe 3		4 Classe 4	
291/884		230/884		133/884		230/884	
32.92%		26.02%		15.05%		26.02%	
n...	eff. ...	eff. total	pourcentage	chi2	Type	forme	p
100	169	465	36.34	5.21	ver_sup	ter	0.02241
106	162	438	36.99	6.51		*sex_2	0.01075
13	149	372	40.05	14.81	nom	corpo	0.00011
99	120	310	38.71	7.25	ver_sup	estar	0.00708
102	109	293	37.2	3.64	ver_sup	ir	NS (0.05639)
10	105	238	44.12	18.5	nom	gente	< 0.0001
4	96	208	46.15	21.58	nom	coisa	< 0.0001
6	91	196	46.43	20.82	ver	achar	< 0.0001
1	57	83	68.67	53.03	ver	cuidar	< 0.0001
103	49	125	39.2	2.6	ver_sup	ver	NS (0.10677)
93	48	123	39.02	2.41	ver	ficar	NS (0.12039)
0	47	59	79.66	62.55	nom	saúde	< 0.0001
51	44	100	44.0	6.27	ver	pensar	0.01227
95	39	98	39.8	2.36	adj	só	NS (0.12441)
38	39	84	46.43	7.67	nom	vida	0.00560
81	36	87	41.38	3.13	adj	mesmo	NS (0.07695)
62	36	82	43.9	4.94	ver	dar	0.02627
108	31	67	46.27	5.85		*gru_13	0.01556
2	28	42	66.67	22.74	adj	bom	< 0.0001
107	25	51	49.02	6.35		*gru_04	0.01171
110	24	58	41.38	2.01		*gru_03	NS (0.15603)
109	23	54	42.59	2.44		*gru_01	NS (0.11846)
104	23	39	58.97	12.54		*gru_11	0.00039
12	23	35	65.71	17.75	adj	importante	< 0.0001
3	19	25	76.0	21.62	adj	saudável	< 0.0001
89	18	40	45.0	7.77	ver	parar	NS (0.06601)

Figura 18- Exibição das formas associadas à classe 1.

Para a análise descritiva do vocabulário de cada classe, uma sugere-se utilizar dois critérios simultâneos: 1) reter a atenção nas palavras não instrumentais com frequência maior do que a frequência média do conjunto de palavras da totalidade do corpus (neste exemplo 35.959 ocorrências divididas por 3.377 formas distintas, o que resulta no valor de 10,65) e 2) considerar aquelas palavras com χ^2 de associação à classe $\geq 3,84$ (pois $p < 0,05$).

Na coluna da esquerda na interface, clicando com o botão direito do mouse sobre a análise denominada **NOME DO CORPUS_alceste_1**, você pode ter acesso a mais alguns resultados da análise. Dentre eles, os mais importantes são:

- **Corpus Colorido** - o qual abrirá uma interface do seu navegador da internet que permitirá que você visualize os segmentos de texto característicos de cada classe, identificando-a pelas cores das classes, conforme as apresentadas no dendograma (ver Figura 19).
- **Relatório**– que criará um documento em .txt, denominado **Rapport**, dentro da pasta que contém o corpus, em uma sub-pasta denominada **NOME DO CORPUS_alceste_1**. Esse documento, que poderá ser visualizado em qualquer editor de texto, contém a descrição lexical de cada uma das classes formadas pela CHD, numa espécie de relatório simplificado da análise.

**** *gru_01 *ctx_1 *ida_1 *sex_2

a palavra que me veio agora na cabeça foi cartão de visitas pelas imagens do vídeo é a ideia de que o corpo é a primeira impressão é o cartão de visitas é o que vai apresentar

então tem que estar com os cabelos bem cuidados bem vestido com o corpo em forma para mim veio também essa questão de identidade mesmo porque é assim que a gente se mostra

não só a forma de se vestir mas a forma como você se expressa o jeito como anda tudo é uma questão de como as pessoas te percebem e como acaba virando meio que um consumismo

as pessoas querem parecer bem e por isso elas compram bastante investem corpo como um objeto essa questão da identidade é uma coisa também imposta socialmente tu tens que ser bonita tu tens que ser magra tu tens que estar sempre bem vestida

uma coisa que me veio foi esta tentativa de padronização padrão de beleza uma coisa que marcou no vídeo é que 95 das imagens eram de pessoas bonitas magras homens malhados e ai tinha três gordinhos só três gordinhos

foi bem para esse lado mesmo isso também me marcou e também se eu estou realmente satisfeita com o meu corpo ou se eu tento o que as pessoas estão satisfeitas porque todo mundo é assim

será que eu realmente gosto de ser assim ou eu sou assim porque todo mundo e é o padrão não tem que ser magro eu acho que estou sendo influenciada eu quero estar com a minha barriga retinha

Figura 19- Corpus colorido.

É importante salientar que as análises do tipo CHD, para serem úteis à classificação de qualquer material textual, requerem uma retenção mínima de 75% dos segmentos de texto (alguns autores, falam da possibilidade de se considerar o aproveitamento de 70% dos segmentos de textos). Caso a CHD ofereça uma classificação com retenção inferior a esta, a mesma não pode ser considerada uma análise adequada aos dados coletados, pois fornece apenas uma classificação parcial. Nesses casos (quando a retenção de segmentos de texto for inferior a 75%), sugere-se que o método de CDH seja abandonado e a análise do material seja feita com outros recursos, como por exemplo, a análise de especificidades.

Ainda na aba *Perfis*, o conteúdo de cada uma das classes pode ser explorado pelo pesquisador a partir de mais recursos disponibilizados pelo *software*, conforme ilustra a figura 20. Para ter acesso a esses recursos basta clicar com o botão direito no mouse sobre qualquer palavra pertencente à classe que você deseja explorar. A parte superior da janela aprofunda os dados referentes à palavra selecionada, enquanto a parte inferior fornece informações referentes à respectiva classe.

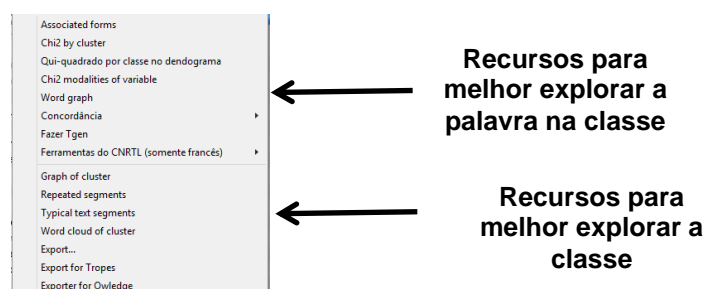


Figura 20- Recursos para interpretar cada uma das classes.

Os recursos disponibilizados pela janela representada na figura 20 além de possibilitarem acesso às palavras associadas à forma (a partir do dicionário de formas reduzidas), permitem a visualização de gráficos que ilustram frequência, associação e coocorrência de uma palavra específica, bem como dos segmentos de texto em que a palavra aparece na classe. Já no que se refere à classe como um todo, é possível visualizar uma árvore de similitude da classe (*Graph of cluster*), as expressões ou combinações de palavras que mais se repetem (*Repeated Segments*), os segmentos de textos característicos da classe (*Typical text segments*) conforme ilustrados pela figura 21, bem como exportar os segmentos associados à classe (*Export*).

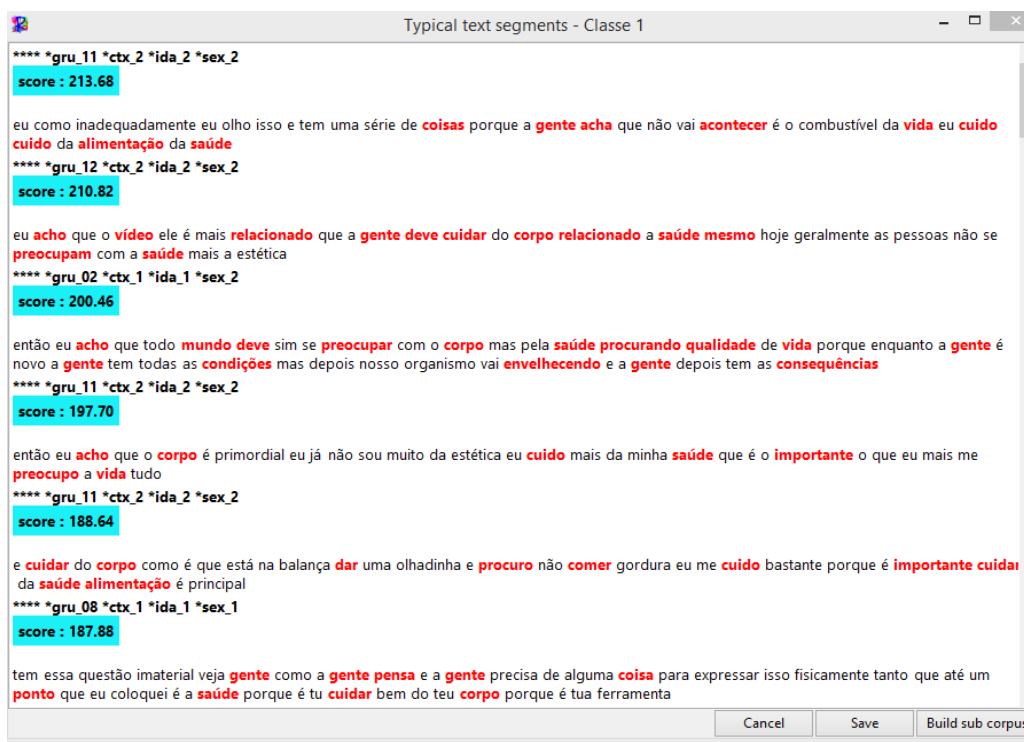


Figura 21- Segmentos de texto típicos da classe 1.

CHD a partir de respostas curtas (questionário)

Quando temos uma grande quantidade de respostas curtas a uma questão aberta de questionário, temos que parametrar a CHD de forma específica (veja a figura 22). Ao importar um corpus deste tipo, além de identificar a codificação e a língua, selecione "parágrafos" como método de construção de segmentos de texto (ST). E depois escolha *Classificação* "simples sobre textos", para que o software não segmente o texto de cada resposta. Assim o segmento de texto será considerado o próprio texto ou resposta curta a pergunta de um questionário.

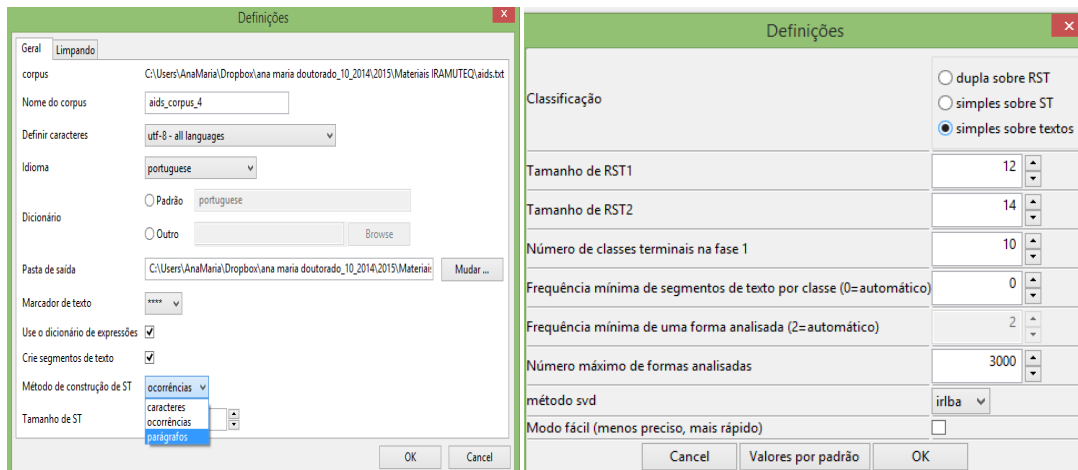


Figura 22- Configuração do método de construção de ST.

Análise: Similitude

Ao escolher a análise de similitude, uma nova janela se abrirá (Fig. 23), possibilitando que sejam escolhidos alguns parâmetros para a construção da árvore de coocorrências. Em *Configurações Gráficas*, você pode editar a análise, trocar o índice de coocorrências por algum outro, escolher se será uma árvore máxima ou não, bem como selecionar uma variável descritiva que seja apresentada em destaque na árvore. Clicando em *Comunidades + Halo*, você pode solicitar que as palavras mais associadas fiquem agrupadas, envoltas por nuvens coloridas. E em *Escores nas Bordas* é possível visualizar no gráfico os valores relativos às coocorrências entre as palavras. Na aba *Ajustes Gráficos*, por sua vez, é possível fazer edições gráficas (tamanho do texto, tamanho das arestas, cores, etc.).

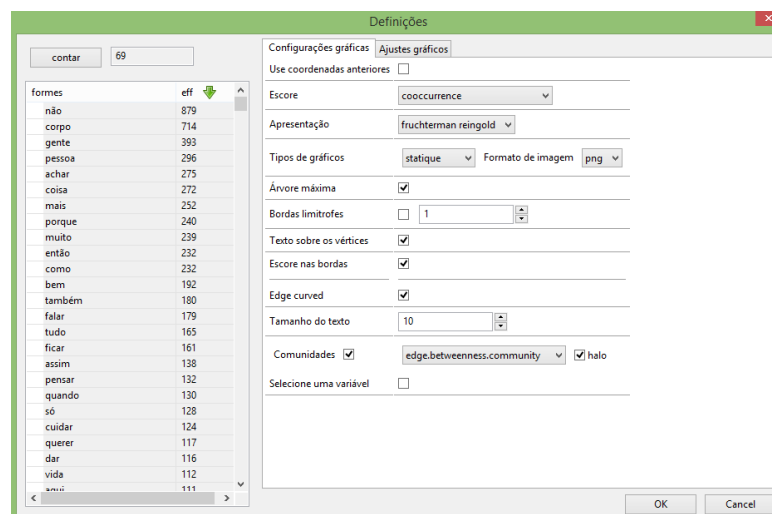


Figura 23- Janela de edição dos parâmetros para Análise de similitude

Na coluna da esquerda é possível selecionar as palavras que estarão na análise. E ao marcar “*Selecione uma variável*” é possível escolher uma variável categorial a participar da análise de similitude, podendo identificar diferenças entre grupos.

Tendo escolhido os parâmetros clique em *OK* e aguarde enquanto a análise se finaliza.

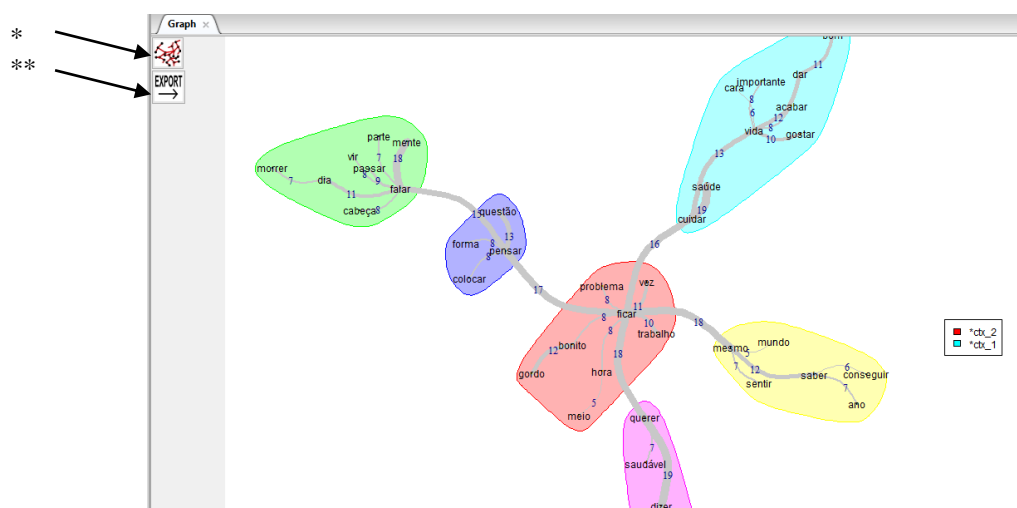


Figura24- Resultados da Análise de similitude

Conforme se observa na Figura 24, a árvore é apresentada na interface dos resultados. No canto superior esquerdo dessa janela, aparecem dois botões. O primeiro deles (*) com traços vermelhos e pontos pretos permite que se modifique a parametragem da análise, abrindo novamente a janela para edição dos parâmetros. O segundo botão (**), no qual está escrito *EXPORT*, exportará a imagem para a pasta das análises, dentro de uma sub-pasta denominada **NOME DO CORPUS_ simitxt_1**.

Análise: Nuvem de palavras

Ao escolher a nuvem de palavras, uma nova janela se abrirá, semelhante àquela da análise similitude, também possibilitando que sejam escolhidos alguns parâmetros para a análise, os quais não necessariamente precisam ser editados. Esta é uma análise mais simples, que trabalha com a representação gráfica em função da frequência das palavras. Tendo escolhido os parâmetros, clique em *OK* nas duas janelas que aparecerão e aguarde alguns instantes.



Figura 25- Resultados da Nuvem de palavras

Na interface dos resultados (Figura 25) você poderá visualizar a nuvem de palavras, a qual também pode ser visualizada, dentro da pasta de análises, na sub-pasta **NOME DO CORPUS_wordcloud_1**, em arquivo de imagem denominado "nuage_1".

Todos os resultados das análises, incluindo as figuras e os gráficos estarão localizadas também dentro da pasta na qual foi salvo inicialmente o corpus de análise. Cada análise (estatísticas, especificidades, CHD, similitude e nuvem de palavras) terá uma sub-pasta com os documentos relativos à mesma.

Parte 2: Análise de matrizes

O IRAMUTEQ permite que se trabalhe com matrizes que envolvam variáveis categoriais e listas de palavras, tais quais aquelas utilizadas para analisar tarefas de evocações livres. Nesse caso, o software viabiliza contagem de frequência, análise prototípica e análise de similitude. Para isso, trabalha-se em um banco de dados montado a partir de um arquivo do Open Office Calc, conforme ilustra a figura 26.

participante	sexo	esc	evoc	rang	evoc	rang	evoc	rang	evoc	rang	evoc	rang	evoc	rang
1	1	1	2 imunidade	1 preservativo	2 macaco	3 vírus	4 preconceito	5						
2	2	2	2 cuidados	1 mais_atençã	2 descuido_co	3 preconceito	4 pessoa_nor	5						
3	3	2	2 medo de_p	1 pena	2 vírus	3 sofrimento	4 hiv	5						
4	4	2	2 despreveni	1 tratamento	2 exclusão	3 preconceito	4 falta_de_ate	5						
5	5	2	2 respeito	1 preconceito	2 igualdade	3 cuidado	4 proteção	5						
6	6	1	2 tratamento	1 descuidado	2 falta_de_ate	3 despreveni	4 preconceito	5						
7	7	2	2 preconceito	1 sofrimento	2 discriminaçã	3 amor	4 problemas	5						
8	8	1	2 preconceito	1 vírus	2 homossexual	3	4 camisinha	5						
9	9	1	2 preservativo	1 doente	2 irresponsáv	3 desinforma	4 preconceito	5						
10	10	1	2 preconceito	1 camisinha	2 sexo	3 discriminaçã	4 doente	5						
11	11	2	2 contaminaçã	1 doença	2 irresponsabi	3 pena	4 cuidado	5						
12	12	1	2 preconceito	1 irresponsabi	2 perigo	3 preocupação	4 igual	5						
13	13	1	2 preconceito	1 tratamento	2 preservativo	3 vírus	4 doente	5						
14	14	2	2 discriminaçã	1 medo	2 sofrimento	3 preconceito	4 dor	5						
15	15	1	2 doença	1 macaco	2 sem proteçã	3 gays	4 problemas	5						
16	16	1	2 preconceito	1 sexo	2 perigo	3 exclusão_so	4 discriminaçã	5						
17	17	2	2 sexo	1 prevenção	2 preservativo	3 sangue	4 risco	5						
18	18	1	2 sexo	1 camisinha	2 preconceito	3 vírus	4 remédio	5						
19	19	1	2 doente	1 sexo	2 camisinha	3 viagra	4 prostitutas	5						
20	20	1	2 preconceito	1 falta_de_info	2 responsabili	3 exclusão_so	4 medo	5						
21	21	1	2 sofrimento	1 fim_da_vida	2 solidão	3 sem_chão	4 tristeza	5						
22	22	2	2 baixa	1 inconsciênci	2 boemia	3 falta_de_info	4 pobreza	5						
23	23	2	2 mulher	1 preconceito	2 coquetel	3 melhoria	4 doença	5						
24	24	1	2 fragilidade	1 insegurança	2 morte	3 hiv	4 vírus	5						
25	25	2	2 vírus	1 transmissão	2 contágio	3 sexo	4 falta_de_info	5						
26	26	2	2 preconceito	1 DST	2 falta_de_pro	3 sexo	4 cuidado	5						
27	27	2	2 preconceito	1 DST	2 sexo	3 falta_de_pro	4 respeito	5						
28	28	1	2 discriminaçã	1 preconceito	2 sofrimento	3 perigo	4 afastamento	5						

Figura 26- Modelo banco de dados para análise de matrizes.

Aconselha-se que o banco de dados siga as seguintes indicações:

- A formatação do arquivo de entrada seja: ods; csv; xls (não usar xlsx – Excel atual, pois ele é incompatível com o IRAMUTEQ). Já a codificação deve ser a mesma usada para as análises de texto: **UTF 8 all languages**.
- O banco de dados não pode conter os caracteres: : ; ‘ “.
- Não conter espaços nas células (use *underline* para ligas mais de uma palavra).
- Não conter acentos ou caracteres especiais no nome do arquivo.
- As variáveis numéricas podem ser apresentadas no arquivo, mas elas não poderão ser utilizadas nas análises (salvo para os “rangs” nas análises prototípicas).

- Caso tenha-se a informação da ordem de aparecimento ou de importância das palavras, esta deve ser acrescentada em uma coluna logo após a palavra.
- É necessária uma ampla revisão do corpus, uma vez que esse tipo de análise não realiza a lematização.

Após ter o banco de dados salvo em uma pasta exclusiva para a análise, ao abrir o IRAMUTEQ, selecionar o ícone *Arquivo*, e em seguida *Abrir uma matriz*. Localize o arquivo que contém seu banco de dados e clique em *Abrir*. Para a importação dos dados, uma outra janela se abrirá (conforme figura 27) e nela você poderá indicar alguns parâmetros do seu banco de dados. São eles: A primeira linha da planilha contém os nomes das colunas (indicado); a primeira coluna é um identificador (indicado); separador de coluna (será, no caso de formato CSV); separador de texto ("); codificação dos caracteres (aconselha-se utf-8- all languages).

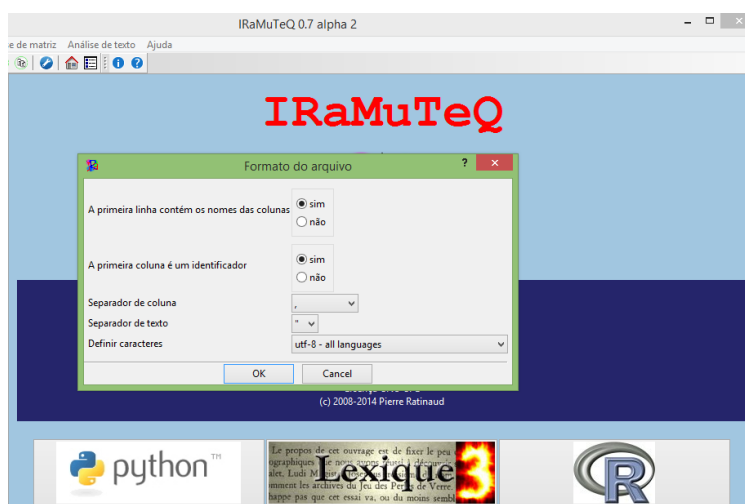


Figura 27- Importação do banco de dados de matriz

Após selecionar os parâmetros e clicar em OK e a janela seguinte mostrará a matriz importada pelo IRAMUTEQ, como na figura 28. As análises possíveis de serem realizadas com os bancos de dados de matrizes envolvem cálculos de frequências, classificação hierárquica descendente (aconselhada apenas nos casos em que o número de participantes é bastante alto), análise de similitude e análise prototípica.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	par	sexo	esc	pessoa	conheci	atitude	evoc	rang	evoc	rang	evoc	rang	evoc	rang
2	1	1	2	2	1	3	imunidade	1	preservativos	2	macaco	3	virus	4
3	2	2	2	1	2	1	cuidados	1	mais_atençao	2	descuido_com	3	preconceito	4
4	3	2	2	2	2	3	medo_de_peg	1	pena	2	virus	3	sofrimento	4
5	4	2	2	2	2	1	desprevenida	1	tratamento	2	exclusão	3	preconceito	4
6	5	2	2	2	2	1	respeito	1	preconceito	2	igualdade	3	cuidado	4
7	6	1	2	2	2	3	tratamento	1	descuidoado	2	falta_de_atenç	3	desprevenido	4
8	7	2	2	2	2	1	preconceito	1	sofrimento	2	discriminação	3	amor	4
9	8	1	2	2	1	1	preconceito	1	virus	2	homossexual	3		4
10	9	1	2	2	1	1	preservativo	1	doente	2	irresponsavel	3	desinformado	4
11	10	1	2	2	2	3	preconceito	1	camisinha	2	sexo	3	discriminação	4
12	11	2	2	2	2	1	contaminação	1	doença	2	irresponsabilid	3	pena	4
13	12	1	2	2	2	1	preconceito	1	irresponsabilid	2	perigo	3	preocupação	4
14	13	1	2	2	1	1	preconceito	1	tratamento	2	preservativo	3	virus	4
15	14	2	2	2	2	1	discriminação	1	medo	2	sofrimento	3	preconceito	4
16	15	1	2	2	1	1	doença	1	macaco	2	sem_proteção	3	gays	4
17	16	1	2	2	1	1	preconceito	1	sexo	2	perigo	3	exclusão_socia	4
18	17	2	2	2	1	1	sexo	1	prevenção	2	preservativo	3	sangue	4
19	18	1	2	2	1	1	sexo	1	camisinha	2	preconceito	3	virus	4
20	19	1	2	2	2	1	doente	1	sexo	2	camisinha	3	viagra	4
21	20	1	2	2	2	1	preconceito	1	falta_de_infor	2	responsabilida	3	exclusão_socia	4
22	21	1	2	2	2	1	sofrimento	1	fim_da_vida	2	solidão	3	sem_chão	4
23	22	2	2	1	1	3	baixa	1	inconsciência	2	boemia	3	falta_de_infor	4
24	23	2	2	1	1	3	mulher	1	preconceito	2	coquetel	3	melhoria	4
25	24	1	2	1	1	1	fragilidade	1	insegurança	2	morte	3	HIV	4
26	25	2	2	2	1	1	virus	1	transmissão	2	contágio	3	sexo	4
27	26	2	2	2	1	1	preconceito	1	DST	2	falta_de_prote	3	sexo	4
28	27	2	2	1	2	1	preconceito	1	DST	2	sexo	3	falta_de_prote	4
29	28	1	2	2	1	3	discriminação	1	preconceito	2	sofrimento	3	perigo	4
30	29	1	2	1	2	1	preconceito	1	morte	2	tempo	3		4

Figura 28- Matriz importada pelo software

Para processar as análises, basta clicar no ícone Análise de matriz e em seguida selecionar a análise desejada (figura 29).

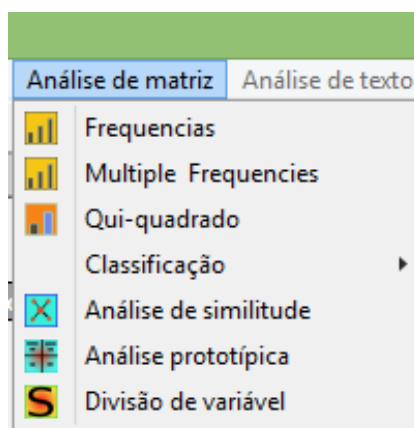


Figura 29- Análises possíveis para as matrizes

A análise mais simples trata-se da análise de frequências. Indica-se a análise de *Frequências* para acessar as frequências das variáveis categoriais da matriz e a análise de *Frequências Múltiplas (Multiple Frequencies)* para obter um relatório de frequência absoluta e relativa das palavras presentes na matriz. Ao selecionar a análise desejada é necessário escolher sobre quais variáveis serão processados os cálculos. Nesse caso, não há interesse no Rang (ordem de evocação) mas apenas nas palavras e eventualmente em variáveis descritivas inseridas na matriz.

A figura 30 ilustra um relatório das frequências múltiplas relativas às palavras evocadas em um teste de associação livre.

mod	freq	percent of total	row number	percent of rows
preconceito	114	7.61	114	38.0
tristeza	66	4.41	66	22.0
sofrimento	54	3.6	54	18.0
morte	52	3.47	52	17.33
sexo	50	3.34	50	16.67
doença	49	3.27	49	16.33
camisinha	39	2.6	39	13.0
medo	33	2.2	33	11.0
doente	31	2.07	31	10.33
tratamento	26	1.74	26	8.67
discriminação	25	1.67	25	8.33
vírus	21	1.4	21	7.0
irresponsabilidade	21	1.4	21	7.0
remédios	20	1.34	20	6.67
cuidados	17	1.13	17	5.67
cuidado	17	1.13	17	5.67
descuido	17	1.13	17	5.67
prevenção	16	1.07	16	5.33
irresponsável	15	1.0	15	5.0
dor	15	1.0	15	5.0
pena	14	0.93	14	4.67
resposta	14	0.93	14	4.67

Figura 30- Análise de frequências

Conforme se observa na figura 30, a análise fornece uma tabela com as palavras ordenadas por sua frequência, assim como a frequência bruta na segunda coluna, seguida da sua proporção em relação ao total de evocações, o número de linhas que contém esta palavra, bem como sua proporção em relação ao número total de linhas. Lembrado, cada linha representa um participante respondente.

A análise prototípica trata-se de uma técnica simples e eficaz desenvolvida especificamente pelo campo de estudo de representações sociais que visa identificar a estrutura representacional a partir dos critérios de frequência e ordem de evocação das palavras provenientes de um teste de evocações livres (Wachelke & Wolter, 2011). A mesma pode ser realizada com o software IRAMUTEQ a partir dos comandos *Análise de matriz* e em seguida *Análise prototípica*.

Figura 31- Definições da análise prototípica

Ao abrir a janela de definições deve-se selecionar (com um clique simples) na parte esquerda as variáveis correspondentes às evocações e na parte direita as variáveis correspondentes ao RANG (seja ele a ordem de evocação ou de

importância atribuída, à escolha segundo os critérios do pesquisador). Os demais parâmetros referem-se aos critérios de cálculo da análise prototípica e podem ser mantidos os padrões automáticos, conforme ilustrado na figura 31.

Definidos os padrões, clique em OK e em alguns segundos será apresentado o produto da análise prototípica (figura 32). Este diagrama de quatro quadrantes representa quatro dimensões da estrutura da representação social. No exemplo em questão, trata-se de uma tarefa de evocação livre com termo indutor Aids. O primeiro quadrante (superior esquerdo) indica as palavras que têm alta frequência (uma frequência maior que a média) e baixa ordem de evocação (aquelas que foram mais prontamente evocadas). Essas seriam as prováveis indicadoras do núcleo central de uma representação.

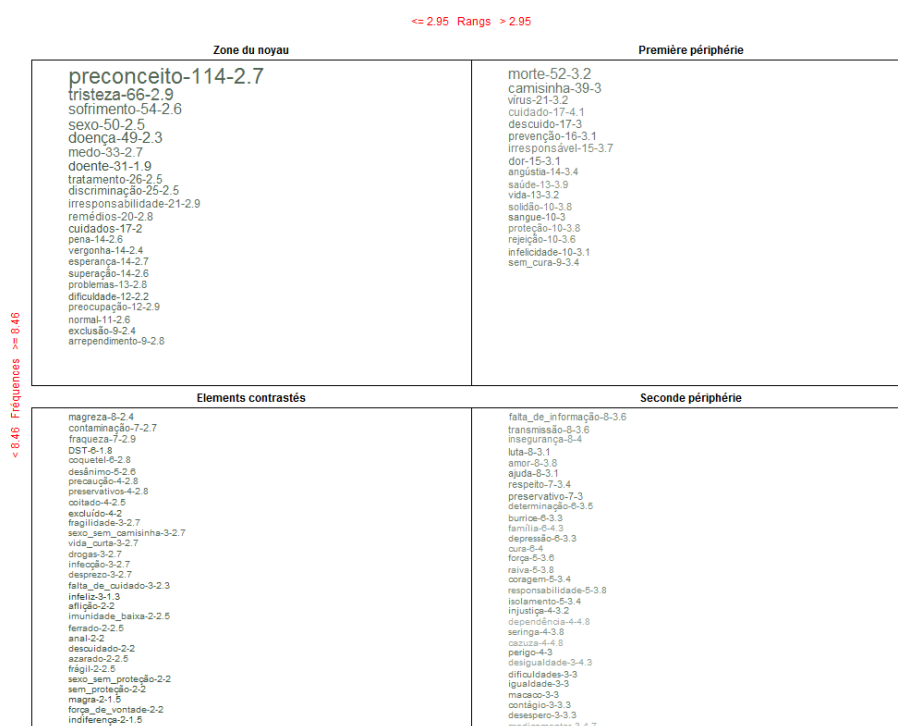


Figura 32- Diagrama dos quatro quadrantes – Análise prototípica

No segundo quadrante (superior direito), temos a primeira periferia, com as palavras que têm alta frequência, mas que tiveram ordem média maior, ou seja, não foram tão prontamente evocadas. No terceiro quadrante (inferior esquerdo), a zona de contraste contém elementos que foram prontamente evocados, porém com frequência abaixo da média. Por fim, a segunda periferia no quarto quadrante (inferior direito) indica os elementos com menor frequência e maior ordem de evocação.

Por fim a análise de similitude, também indicadora da estrutura de uma representação social, pode ser realizada a partir dos comandos: *Análise de Matrizes* e *Análise de Similitude*. O processamento da análise se dá de modo análogo ao

realizado com o material textual, por meio de uma janela de definições conforme ilustra a figura 33.

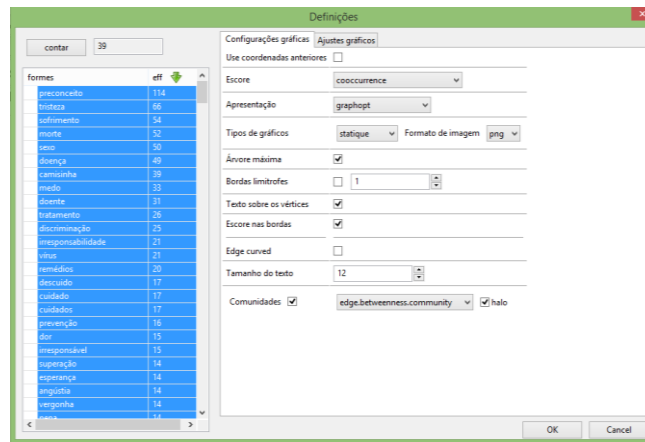


Figura 33- Definições para análise de similitude

A imagem resultante da análise de similitude encontra-se na figura 34, onde o tamanho dos vértices coloridos é proporcional à frequência das palavras e as arestas indicam a força da coocorrência entre as palavras.

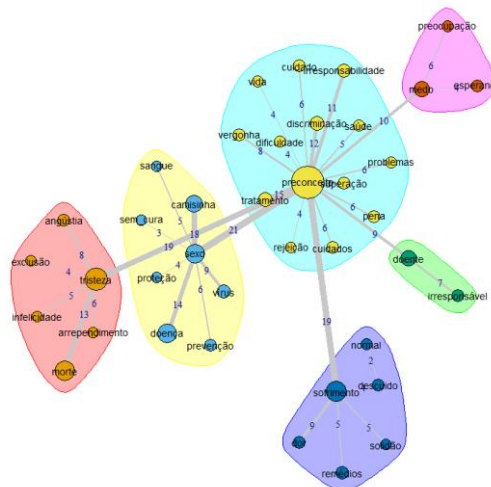


Figura 34- Análise de Similitude

Referências

- Antunes, L. (2013). O papel dos estereótipos nas representações sociais compartilhadas por adolescentes sobre as pessoas que vivem com HIV/aids. *Dissertação de Mestrado* (não publicada). Programa de Pós-Graduação em Psicologia. Universidade Federal de Santa Catarina. Florianópolis, SC.
- Camargo, B. V., Justo, A. M. (2013). *IRAMUTEQ: Um Software Gratuito para Análise de Dados Textuais*. *Temas em Psicologia*, 21 (2), 513-518.

- Cibois, P. (1990). *L'analyse des données en sociologie*. Paris: P.U.F.
- Cros, M. (1993). Les apports de la linguistique: langage des jeunes et sida. In ANRS (Agence Nationale de Recherche sur le Sida). *Les jeunes face au Sida: de la recherche à l'action* (pp. 50-61). Paris: ANRS.
- Ghiglione, R.; Matalon, B. (1993). *O inquérito: Teoria e prática*. Oeiras: Celta.
- Justo, A. M. (2011). Representações sociais sobre o corpo e implicações do contexto de inserção desse objeto. *Dissertação de Mestrado* (não publicada). Programa de Pós-Graduação em Psicologia. Universidade Federal de Santa Catarina. Florianópolis, SC.
- Justo, A. M.; Camargo, B. V. (2014). Estudos qualitativos e uso de softwares para análises lexicais. Em: C. Novikoff; S. R. M. Santos; O. B. Mithidieri (Orgs.). *Cadernos de artigos: X SIAT e II SERPRO Lageres/UNIGRANRIO* (pp. 37-54). Duque de Caxias: UNIGRANRIO.
- Lahlou, S. (2012). Text Mining Methods: An answer to Chartier and Meunier. *Papers on Social Representations*, 20 (38), 1-7.
- Lebart, L. & Salem, A. (1988). *Analyse statistique des données textuelles*. Paris: Dunod.
- Marchand, P.; P. Ratinaud. (2012). L'analyse de similitude appliqué aux corpus textuelles: les primaires socialistes pour l'élection présidentielle française. Em: *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles. JADT 2012*. (687–699). Presented at the 11eme Journées internationales d'Analyse Statistique des Données Textuelles. JADT 2012. Liège, Belgique
- Ratinaud, P. (2009). IRAMUTEQ: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires [Computer software]. Recuperado em 5 março, 2013, de <http://www.iramuteq.org>
- Ratinaud, P., & Marchand, P. (2012). Application de la méthode ALCESTE à de “gros” corpus et stabilité des “mondes lexicaux” : analyse du “CableGate” avec IraMuTeQ. Em: *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles* (835–844). Presented at the 11eme Journées internationales d'Analyse statistique des Données Textuelles. JADT 2012, Liège.
- Reinert, M. (1990). ALCESTE, une méthodologie d'analyse des données textuelles et une application: Aurélia de G. de Nerval. *Bulletin de méthodologie sociologique*, (28) 24- 54.
- Veloz, M. C. T.; Nascimento-Schulze, C. M.; Camargo, B. V. (1999). Representações sociais do envelhecimento. *Psicologia: Reflexão e Crítica*, 12 (2), 479-501.
- Wachelke, J. F. R. & Wolter, R. (2011). Critérios de construção e relato da análise prototípica para representações sociais. *Psicologia Teoria e Pesquisa*, 27 (4), 521-526.