

Processamento de Linguagem Natural sobre obra Clássica de Psicologia - The Interpretation of Dreams de Sigmund Freud

Charles Belchior Gabriel da Silva^{1*}; Juliano Schimiguel²

¹ Universidade Cruzeiro do Sul. Bacharel em Psicologia. Av. Dr. Ussiel Cirilo, 93 - São Miguel Paulista, 08060-070, São Paulo/SP, Brasil

² Universidade Cruzeiro do Sul. Doutorado e Mestrado em Ciência da Computação pelo Instituto de Computação da UNICAMP. Rua Galvão Bueno, 868, Liberdade, 01506-000, São Paulo/SP, Brasil

*autor correspondente: charlesbelchiorgs@gmail.com

Processamento de Linguagem Natural Sobre Obra Clássica de Psicologia - The Interpretation of Dreams by Sigmund Freud

Resumo

O avanço da chamada indústria 4.0 tornou a tecnologia ainda mais relevante e presente no cotidiano humano, possibilitando que processos e soluções que antes eram limitadas e muitas vezes artesanais fossem automatizados e aprimorados. Um dos fatores que contribuíram para que isso acontecesse foi sem dúvida o aumento da capacidade de processamento de dados, possibilitando novas formas de lidar com grandes volumes de informações, e nesse sentido, uma das faces da tecnologia que sem dúvidas tem seu grau de notoriedade é o Processamento de Linguagem Natural (PLN). O objetivo maior deste trabalho foi realizar a análise da obra "A Interpretação dos Sonhos" de Sigmund Freud utilizando técnicas de PLN a fim de extrair informações relevantes. Para isso, foram aplicadas algumas das técnicas de PLN por meio da linguagem R. Por ser uma obra bem conhecida, o processo de análise foi facilitado, mas ainda assim muito interessante. Tais resultados foram obtidos por meio de análise textual e morfológica, e abrem espaço para estudos mais avançados e abrangentes no que se trata da intersecção entre psicologia e ciência de dados.

Palavras-chave: Psicanálise; Iramuteq; Palavras; Dados; Análise.

Natural Language Processing About Classical Work of Psychology - The Interpretation of Dreams by Sigmund Freud

Abstract

The advancement of the so-called Industry 4.0 has made technology even more relevant and prevalent in human daily life, enabling the automation and enhancement of processes and solutions that were previously limited and often artisanal. One of the factors that unquestionably contributed to this development was the increase in data processing capacity, allowing new ways to handle vast amounts of information. In this regard, one of the facets of technology that undeniably holds a degree of notoriety is Natural Language Processing (NLP). The primary goal of this work was to analyze Sigmund Freud's work "The Interpretation of Dreams" using NLP techniques to extract relevant information. To achieve this, some NLP techniques were applied through the R language. Due to its well-known nature, the analysis process was facilitated but remained highly interesting. These results were obtained through textual and morphological analysis, opening up avenues for more advanced and comprehensive studies in the intersection of psychology and data science.

Keywords: Psychoanalysis; Iramuteq; Words; Data; Analysis.

Introdução

Sabe-se que as revoluções industriais foram alguns dos marcos históricos mais importantes da humanidade, e que a primeira revolução industrial aconteceu em meados do século XVIII. Todavia, observa-se o alvorecer de uma nova revolução, e com isso um momento de transição que vai além da simples modernização de processos. Essa mudança, chamada de quarta revolução industrial, e também denominada como Indústria 4.0 (Cardoso, 2016), é uma revolução em que a tecnologia não só se torna facilitadora no

processo, mas também a base para todo um novo modo de viver e consumir, criando processos dos quais muitos apenas se tornam possíveis pelo uso de inteligência artificial, robótica, internet das coisas e computação em nuvem (Santos e Lima, 2018), e uma aplicação desses conceitos em áreas mais restritas ao trato humano se torna minimamente relevante, sendo assim, mais do que suficiente para esse estudo.

Existe uma sociedade em constante mudança de paradigma, onde a tecnologia se torna de grande relevância quando se considera a dinâmica social, sua estrutura, e formas com que o indivíduo se relaciona com o meio e com seus semelhantes, e da mesma forma, como altera as dinâmicas existentes dentro das organizações que fazem uso dessas novas tecnologias (Silva, 2014).

A tecnologia está cada vez mais presente no nosso cotidiano, e uma das tecnologias que causaram maior impacto e que continuam a causar mudanças são os smartphones e toda a tecnologia mobile, contribuindo de forma significativa de muitas formas para essas mudanças, não só transformando o ambiente, mas também transformando nossa forma de lidar com ele e com todos à nossa volta. O que de início era desenvolvida, transformada e recriada pela ação do ser humano, hoje é transformadora e geradora de novos comportamentos, estruturas de relacionamento, culturas, formas de ver o mundo, e em certa instância, também é criadora de novas tecnologias (Coutinho, 2014).

Entre as muitas tecnologias presentes nesse processo de evolução da indústria, umas que vêm recebendo notoriedade pela usabilidade e versatilidade são os chamados bots, que de forma generalista, são robôs que automatizam processos virtuais. Dentro desse nicho, existem também os chamados Chatbots que utilizam Inteligência Artificial (IA) que possibilita a automatização de respostas em chats via texto e em alguns casos, também em áudio (Maciel, 2009).

Algumas tentativas de reproduzir o papel do psicoterapeuta por meio de chatbots não são recentes, a maioria porém utilizam de outras abordagens que não seja a Psicanálise para isso, como é o caso do chatbot Beck, baseado na terapia cognitivo-comportamental com o objetivo de apoiar adolescentes com depressão (Almeida, 2017).

Outro chatbot que também faz uso da terapia cognitivo-comportamental é o Rumi, que foi desenvolvido com um chatbot psicoeducativo para elucidar e diminuir sintomas de ansiedade, ruminação e depressão (Oliveira, 2021).

Nesse contexto, há poucas produções que tentam aproximar de algum modo a psicanálise e o Processamento de Linguagem Natural (PLN), sendo o comum a aplicação das técnicas sobre o discurso do paciente, e não sobre a produção dos teóricos (Kaszubowski, 2013).

Todas essas são tentativas de tornar o processo terapêutico de certa forma automatizado, porém, sem se preocupar de fato em aplicar técnicas de processamento de dados sobre as teorias psicológicas, ou do efeito que isso pode ter sobre a construção de conhecimento.

Em meio a todo esse contexto tecnológico, os chatbots fazem uso de algumas técnicas de processamento de dados, sendo que muitas existem desde muito antes do salto tecnológico causado pelos chatbots. Essas técnicas de processamento de dados são agrupadas no que se chama de Processamento de Linguagem Natural (PLN), uma área que já é conhecida dos cientistas da computação, cientistas de dados, linguistas e estatísticos (Indurkha, 2010).

O PLN engloba muitas técnicas de processamento de dados que foram desenvolvidas e até ajustadas para o manejo da informação em formato de texto, ou seja, em palavras. O que essencialmente se difere muito de informações contidas em números e de como manipulá-los. Uma frase possui estrutura diferente de uma equação à primeira vista, porém, apenas à primeira vista, já que a PLN e todo o conjunto de esforço computacional atrelado a ela tem como principal objetivo entregar análises que possam transformar dados em informação (Bird et al., 2009).

Ao se aplicar as técnicas de PLN, é possível correlacionar as palavras entre si, criando então uma grande equação probabilística e estatística, fazendo assim com que o computador possa trabalhar com as palavras quase tão bem quanto trabalha com números.

Neste trabalho será explorado mais dessa área, apresentando de forma geral, e na medida do possível em seus pormenores, o que são, como são, e para que são aplicadas algumas técnicas do processamento de linguagem natural, entre elas, a Contagem de Palavras, Nuvens de Palavras e Análise de Similitude. Para isso, tais técnicas foram aplicadas na análise de um dos - se não o - mais famoso texto da psicanálise: A Interpretação dos Sonhos, escrito por Sigmund Freud em 1899, e uma das obras mais influentes da modernidade. A obra original foi escrita em alemão e foi traduzida para inúmeros idiomas, entre os quais o inglês, sendo este um dos idiomas mais populares do mundo e que servirá como padrão das análises deste trabalho (Gazeau, 2018).

Sendo assim, esse trabalho tem por objetivo apresentar algumas técnicas de PLN aplicados na obra "A Interpretação dos Sonhos" de Sigmund Freud e seus resultados.

Material e Métodos

Todo o projeto apresentado segue como uma pesquisa exploratória (Gil, 2002), tentando apresentar a implementação das técnicas de PLN, porém sem se apegar a uma

biblioteca ou ferramenta, mas sim tendo foco maior em sua aplicabilidade e utilidade no que se refere a extrair informações dos dados analisados.

Para tal propósito, foram utilizados textos livremente distribuídos, materiais bibliográficos encontrados na literatura de referência psicanalítica que, de acordo com a lei brasileira sobre direitos autorais, expressa na Lei nº9. 610 de 1998, já estão submetidas à lei de domínio público. A priori, limitando-se a obras de referência da psicanálise freudiana, que se enquadram nesta lei desde 2010.

Das obras referidas, foram utilizados textos que correspondem à contribuição feita por Sigmund Freud entre 1900 e 1940, que fazem parte de uma coleção composta por 23 volumes. Das obras dessa coleção, foi escolhido um livro em específico, sendo sua obra intitulada “A Interpretação dos Sonhos” (Freud, 1899), aqui trabalhada em sua versão em inglês, intitulada: “The Interpretation Of Dream”. Essa obra, tão importante na modernidade e no estudo e entendimento da psicologia, serviu de fonte para a aplicação das técnicas de PLN (Processamento de Linguagem Natural) e interpretação que serão apresentadas a seguir.

Antes de iniciar o processo de explanação da análise, é preciso compreender alguns termos e conceitos que estão intimamente ligados à execução das análises textuais, e sem os quais não se poderia avançar. Começando por definir de fato o que é Processamento de Linguagem Natural (PLN), que segundo Finatto (2015) é uma subárea da Inteligência Artificial que se define como a habilidade de processar a linguagem humana utilizada no cotidiano pelo computador, sendo seu principal objetivo criar soluções referidas a reprodução e reconhecimento da linguagem humana.

Outro conceito importante é o de *corpus*, que é o conjunto de textos que se pretende analisar, sendo esses conjuntos de textos podem ser compostos pelos mais variados textos, nos mais variados assuntos e vindo de inúmeras fontes, sendo tal corpus construído pelo analista ou cientista de dados (Camargo, 2013).

Tokenização é o processo de dividir o corpus em *tokens*, sendo que um token pode ser definido como a menor unidade do texto faça sentido na análise, sendo comumente palavras, mas que também podem ser símbolos, frases inteiras, ou qualquer outro aglomerado que possuam significado (Jurafsky, 2000).

Stopwords são palavras em forma de tokens que, apesar de terem sentido não possuem relevância de significado no texto, são palavras que podem ser desconsideradas na análise devido a alta frequência comumente encontrada em todos os textos. Em suma, palavras comuns, como: “que”, “para”, “por”, “um”, “uma”, entre tantas outras (Silge, 2017).

Lemmatização é o processo de agrupar todas as flexões de uma palavra em uma única representação morfológica dessa palavra (Manning, 1999).

Sendo assim, ao se trabalhar com PLN é possível seguir um cronograma de processos, sendo que alguns processos podem ser ignorados, e outros não. O que não necessariamente torna o passo-a-passo uma regra, mas sim recomendações que sendo seguidas elevam significativamente a qualidade das análises posteriores, demonstrada por meio da representação de um pequeno fluxograma desses processos na Figura 1.

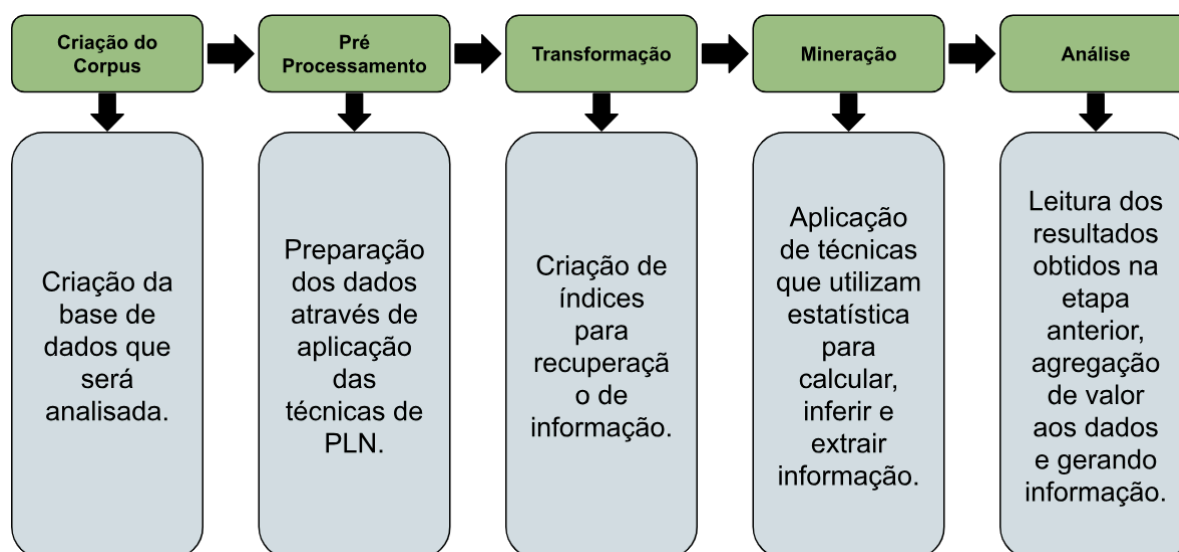


Figura 1. Fluxo do processo de análise de corpus textuais

Fonte: Autoria própria

A obra analisada está presente de forma gratuita no Projeto Gutenberg¹, sendo facilmente acessada por um navegador de internet e/ou por linhas de códigos em diversas linguagens. Na linguagem R, pode ser acessada por meio de seu `gutenberg_id 66048`. Neste trabalho, foram usados arquivos baixados tanto via linhas de códigos quanto por navegadores, dada a natureza da análise.

Para a execução prática foram utilizadas algumas ferramentas, a primeira sendo a linguagem R por meio do software RStudio, e a segunda sendo o software Iramuteq².

Também foi utilizado o software gratuito e open source chamado IRAMUTEQ, que é uma interface gráfica baseada em R e Python, desenvolvido em 2009 originalmente para o tratamento de dados em francês, mas que é multilinguagem no presente momento de publicação deste trabalho. Tal software tem foco na análise estatística de dados presentes

¹ <https://www.gutenberg.org/>

² <http://www.iramuteq.org/>

em corpus textuais feita por meio de cálculos estatísticos que possibilitam que dados qualitativos sejam mensurados e tratados como dados quantitativos (Camargo, 2013).

A escolha deste software ocorreu devido a sua facilidade no manejo, rápida curva de aprendizado e resultados plenamente satisfatórios, além de permitir um grau de comparação entre o manejo dos dados e técnicas. Ressalta-se que a escolha desse software não desmerece a utilização de qualquer outro para os mesmos fins, e esperasse que os resultados e métodos utilizados neste trabalho possam servir para demonstrar quão vasta é a área de processamento de linguagem natural, tanto em técnicas quanto em ferramentas disponíveis.

Antes de se utilizar o software nas análises textuais, foi necessário a preparação do corpus, que ocorre de forma diferente do usualmente conhecido na linguagem R trabalhada no RStudio. Foram utilizadas apenas algumas dessas configurações, a fim de apresentar de forma sucinta como ocorre a preparação do corpus para a análise no software.

Primeiramente deve-se ter o corpus textual em arquivo .txt na codificação Unicode (UTF - 8) o que é possível obtê-lo nessas configurações diretamente no site do projeto Gutenberg. Também deve-se obedecer uma série de pré requisitos de formatação e revisão do corpus, como: substituição do hífen em palavras compostas por traço underline (_), remoção de uma série de caracteres especiais como asteriscos, aspas, cifrão, entre tantas outras que são listados e podem ser consultadas diretamente no manual de utilização do software (Camargo, 2013). Todo esse processo foi feito manualmente, editando diretamente o arquivo em .txt através de um editor de texto com a função “Localizar e Substituir”.

O software livre Iramuteq permite que as análises sejam feitas por meio da interação com sua interface, sendo a navegação feita por meio de menus e submenus, que expandem a possibilidade de análise.

Nesse trabalho foi utilizado o RStudio versão: 2023.03.0+386 "Cherry Blossom" e Iramuteq versão 0.7 alpha 2. Além da versão do R 4.1.2 utilizada tanto pelo RStudio quanto Iramuteq.

Resultados e Discussão

Para o início das análises pela linguagem R no RStudio, é necessário antes a instalação e carregamento do pacote Gutenberg, e em seguida ser feito download do arquivo de texto referente a obra de Freud. Todo esse processo pode ser feito por meio das seguintes linhas de código presentes na Figura 2.

```
1 #Instalação do pacote Gutenbergr
2 install.packages("gutenbergr")
3
4 #Carregando o pacote:
5 library("gutenbergr")
6
7 #Baixando o livro "THE INTERPRETATION OF DREAMS" em inglês
8 dreams <- gutenberg_download(c(66048))
```

Figura 2. Script R para download dos pré requisitos para análise

Fonte: Autoria própria

Seguindo com a análise, foi realizado o procedimento de remoção de números, stopwords e tokenização.

Em seguida, foi feita uma contagem da frequência de palavras e gerado um gráfico das 10 mais frequentes, utilizando a biblioteca ggplot2, que podem ser vistos na Figura 3.

```
23 #Carrega a biblioteca ggplot2
24 library("ggplot2")
25
26 #Conte a frequência das palavras
27 freq_palavras <- dreams %>%
28   count(word, sort = TRUE) %>%
29   head(10) #Pegue as 10 palavras mais frequentes
30
31 #Crie o gráfico de barras
32 ggplot(freq_palavras, aes(x = reorder(word, -n), y = n)) +
33   geom_bar(stat = "identity", fill = "dodgerblue") +
34   labs(title = "As 10 Palavras Mais Frequentes",
35        x = "Palavra",
36        y = "Frequência") +
37   theme_minimal() +
38   theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Figura 1. Figura 3. Scripts R para criação de gráfico de frequência

Fonte: Autoria própria

O gráfico gerado por essas linhas de códigos pode ser visto na Figura 4. Interessante notar que o gráfico apresentou uma palavra aquém do texto, sendo ela “footnote”, em tradução livre: “notas de rodapé”, que são notas feitas pelo próprio autor ou comentarista durante revisões e traduções da obra. A identificação dessa palavra pelo analista ressalta a importância de se conhecer os dados analisados, ainda que não em sua totalidade, pois muitas vezes pode ser humanamente inviável, é importante que na medida do possível, tenha-se um certo grau de feeling no tocante ao assunto analisado.

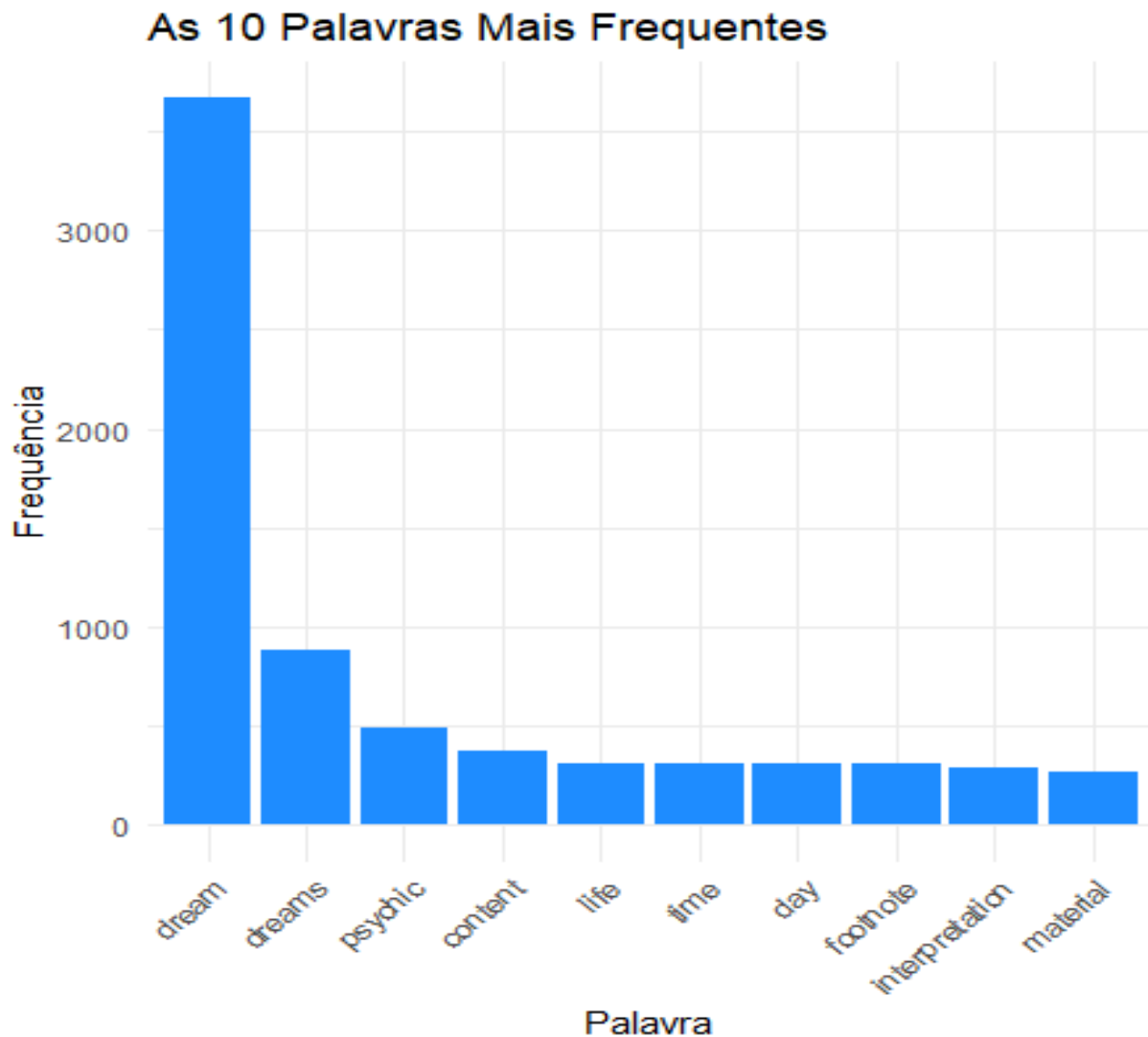


Figura 4. Gráfico de frequência de palavras

Fonte: Autoria própria

Como tal palavra não agrega valor real a análise, pois não diz respeito ao assunto da obra em si, optou-se por retirá-la por meio do seguinte código presente na Figura 5:

```
44 #Remoção da palavra "footnote"
45 remocao <- dreams %>% filter(str_detect(word, "footnote")) %>% select(word)
46
47 dreams <- dreams %>% anti_join(remocao, by = "word")
48
49 #Remoção de stop words e Tokenização
50 dreams <- dreams %>% unnest_tokens(word, word) %>% anti_join(stop_words)
```

Figura 5. Scripts R para remoção da palavra “footnote” do corpus

Fonte: Autoria própria

Após isso, foi gerado um novo gráfico utilizando a mesma estrutura de código que gerou o gráfico anteriormente. O resultado foi um gráfico com uma nova palavra relacionada ao texto em si, a saber: “Walking”, que pode ser visto na Figura 6.

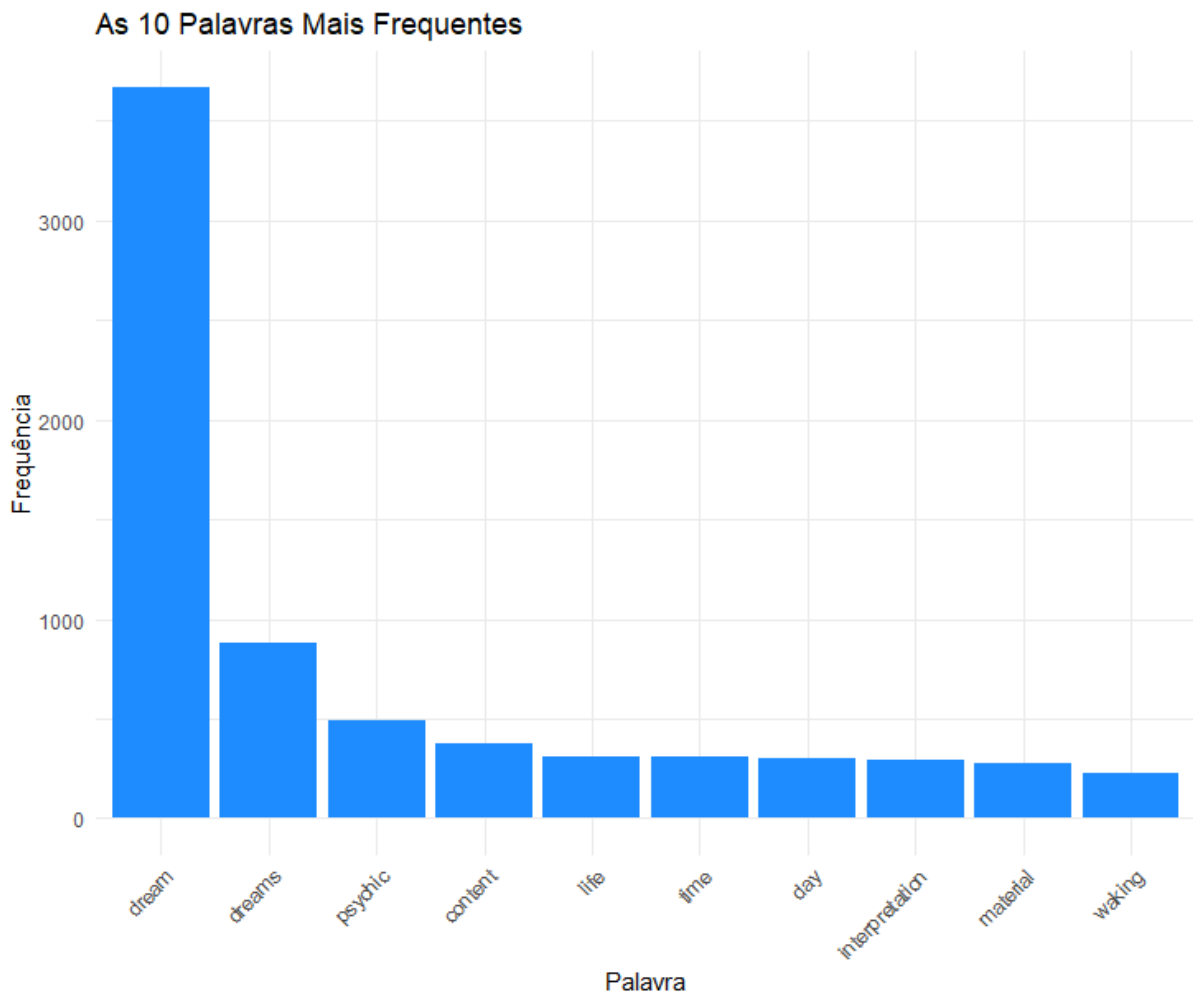


Figura 6. Gráfico de frequência de palavras após remoção da palavra “footnote”

Fonte: Autoria própria

Em seguida, foi aplicado a técnica de criação de Nuvens de Palavras, solicitando as 50 palavras mais frequentes no texto, por meio das linhas de código visíveis na Figura 7.

```

73 #Nuvem de Palavras
74 dreams_wordcloud <- dreams %>% count(word, sort = TRUE)
75
76 #Define a paleta de cores
77 pal = brewer.pal(8, "Dark2")
78
79 #Word Cloud
80 dreams_wordcloud %>% with(wordcloud(word, n, random.order = FALSE, max.words = 50, colors = pal))

```

Figura 7. Scripts R para criação de Nuvem de Palavras

Fonte: Autoria própria

Ao qual resultou em um gráfico um pouco mais rico em informações que pode ser interpretado visualmente de forma rápida. A exemplo, pode-se aferir que o texto analisado fala sobre dreams (sonhos), dream (sonho), psychic (psíquico), interpretation (interpretação), memory (memória), theory (teoria), entre outras. Visível na Figura 8.

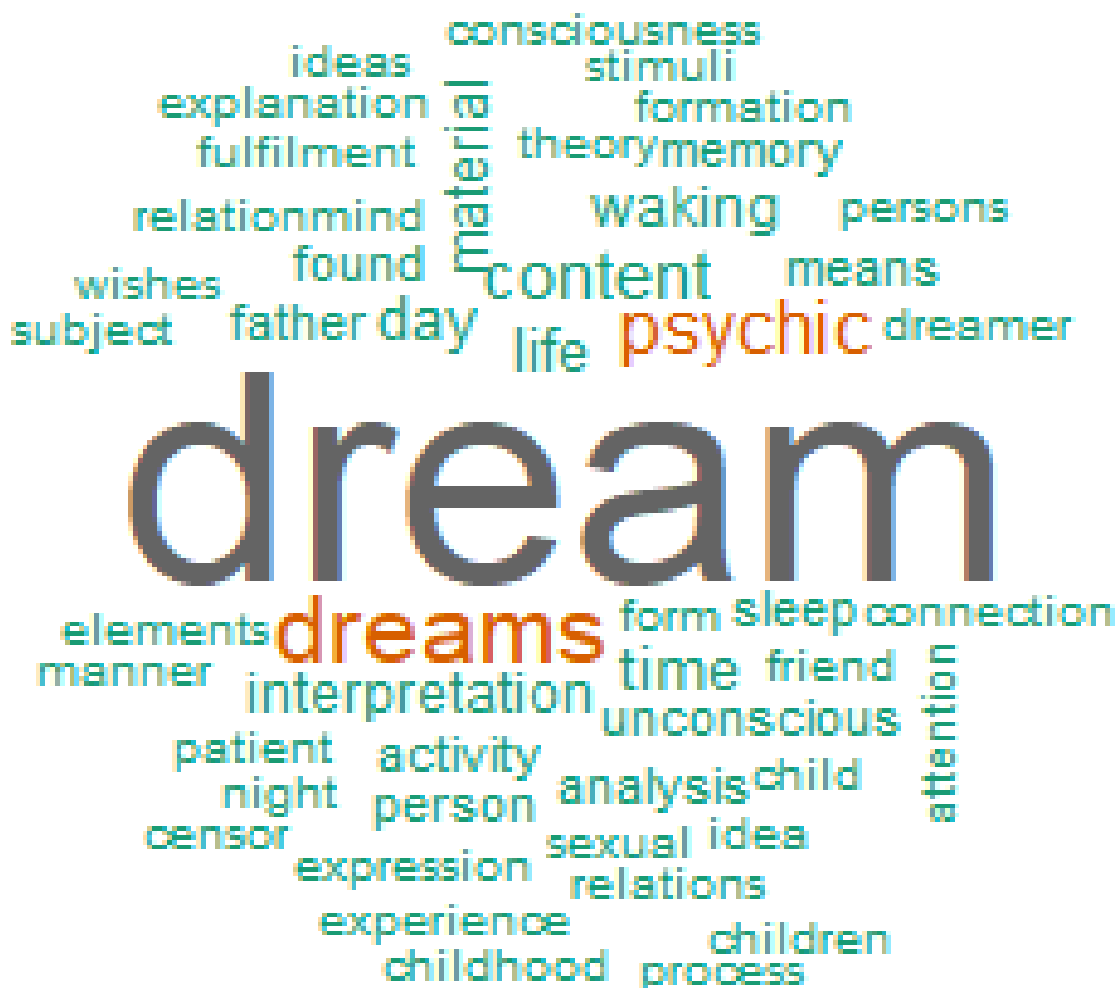


Figura 8. Gráfico de Nuvem de Palavras

Fonte: Autoria própria

Essa talvez seja a representação gráfica mais comum e de fácil interpretação ao abordar o Processamento de Linguagem Natural. Trata-se de uma análise lexical simples, onde as palavras são distribuídas de forma aleatória, porém seguindo algumas diretrizes. As palavras mais frequentes no texto são representadas em uma fonte maior, dispostas no centro do gráfico, enquanto as menos frequentes são apresentadas em uma fonte proporcionalmente menor, de acordo com sua taxa de frequência, sendo dispostas ao redor. Isso segue a regra de que palavras maiores e mais centralizadas indicam maior frequência, ao passo que palavras menores e mais distantes do centro possuem menor frequência no texto. Além disso, conforme estabelecido no código, o gráfico adota um padrão de cores

Para comparação, foi gerada uma segunda Nuvem de Palavras pelo software Iramuteq, a fim de comparar os resultados entregues por ambas as ferramentas. Nesta segunda nuvem de palavras não houve uma preocupação com cores e quantidade de palavras, pois pretendia-se apenas uma comparação de resultados ampla, afinal, por se tratar do mesmo corpus e da mesma técnica, esperava-se que o resultado fosse similar. O que se mostrou verdadeiro conforme observado na imagem, com um detalhe interessante a ser notado: o software Iramuteq aplica a lematização automática sobre o corpus, procedimento esse que não foi aplicado durante o tratamento dos dados na primeira parte enquanto se utilizava a linguagem R e RStudio propositalmente, o que significa maior controle sobre o pré processamento do corpus, a fim de evidenciar essa diferença entre as duas ferramentas. Visível na Figura 9.



12

Uma outra tentativa de visualizar a similitude entre as palavras do texto foi aplicada, porém selecionando as 20 palavras mais frequentes. O que demonstrou um parâmetro muito mais assertivo, trazendo um gráfico extremamente legível e interpretável. Conforme na Figura 11.

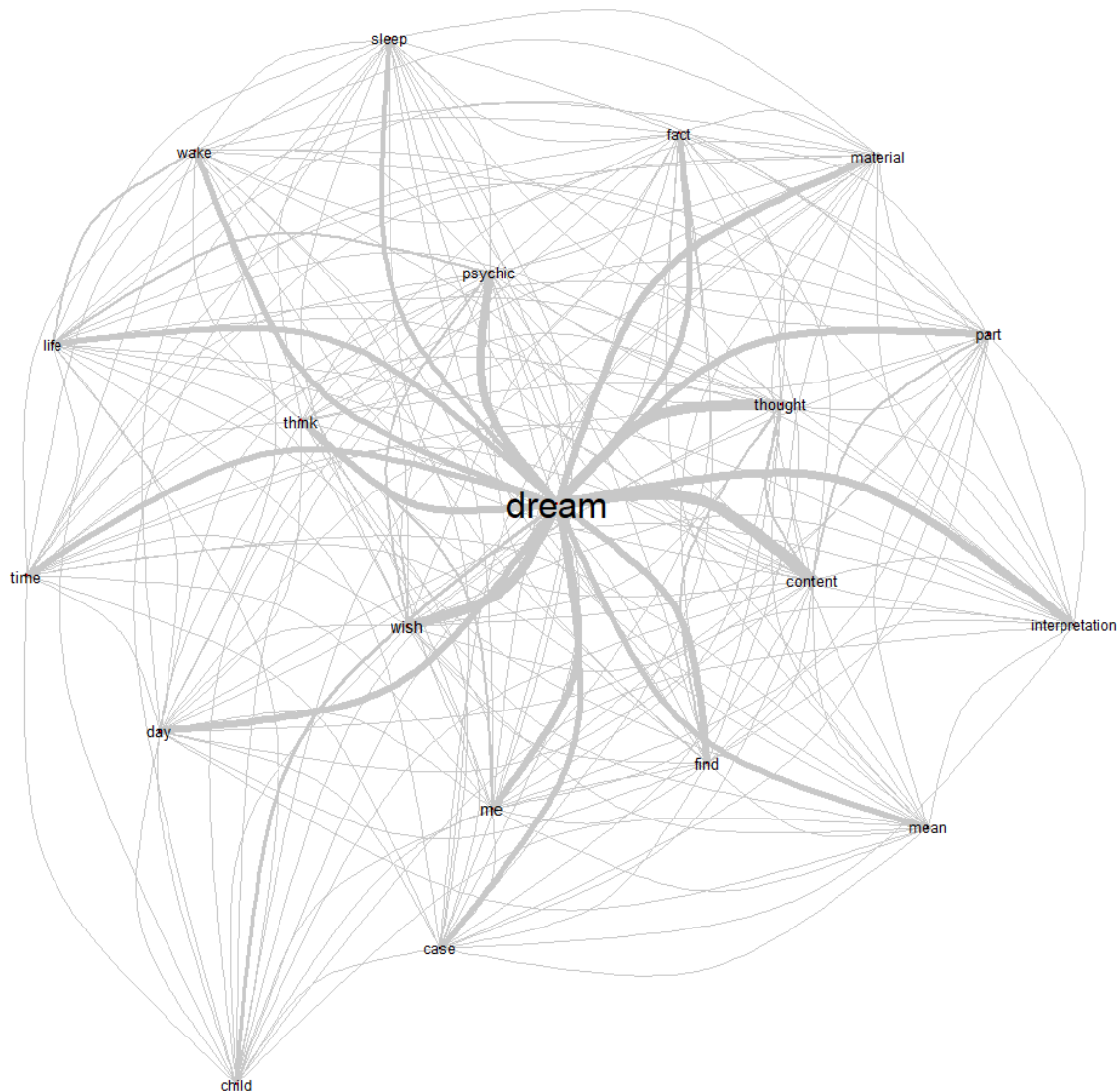


Figura 11. Gráfico da Análise de Similitude das 20 palavras mais frequentes do corpus
Fonte: Autoria própria

Em outra configuração, foi solicitado ao software que trouxesse o mesmo gráfico de similitude das 20 palavras mais frequentes, porém trazendo o peso estatístico das relações. Em termos práticos para análise em primeiro momento, pode parecer pouco aproveitável, porém, visualmente a espessura de uma linha de relação pode ser muito similar a outra, e visualizar os pesos auxilia na interpretação da hierarquia das relações, livrando assim o

analista de falsas inferências e arbitrariedades sobre essas relações. Conforme na Figura 12.

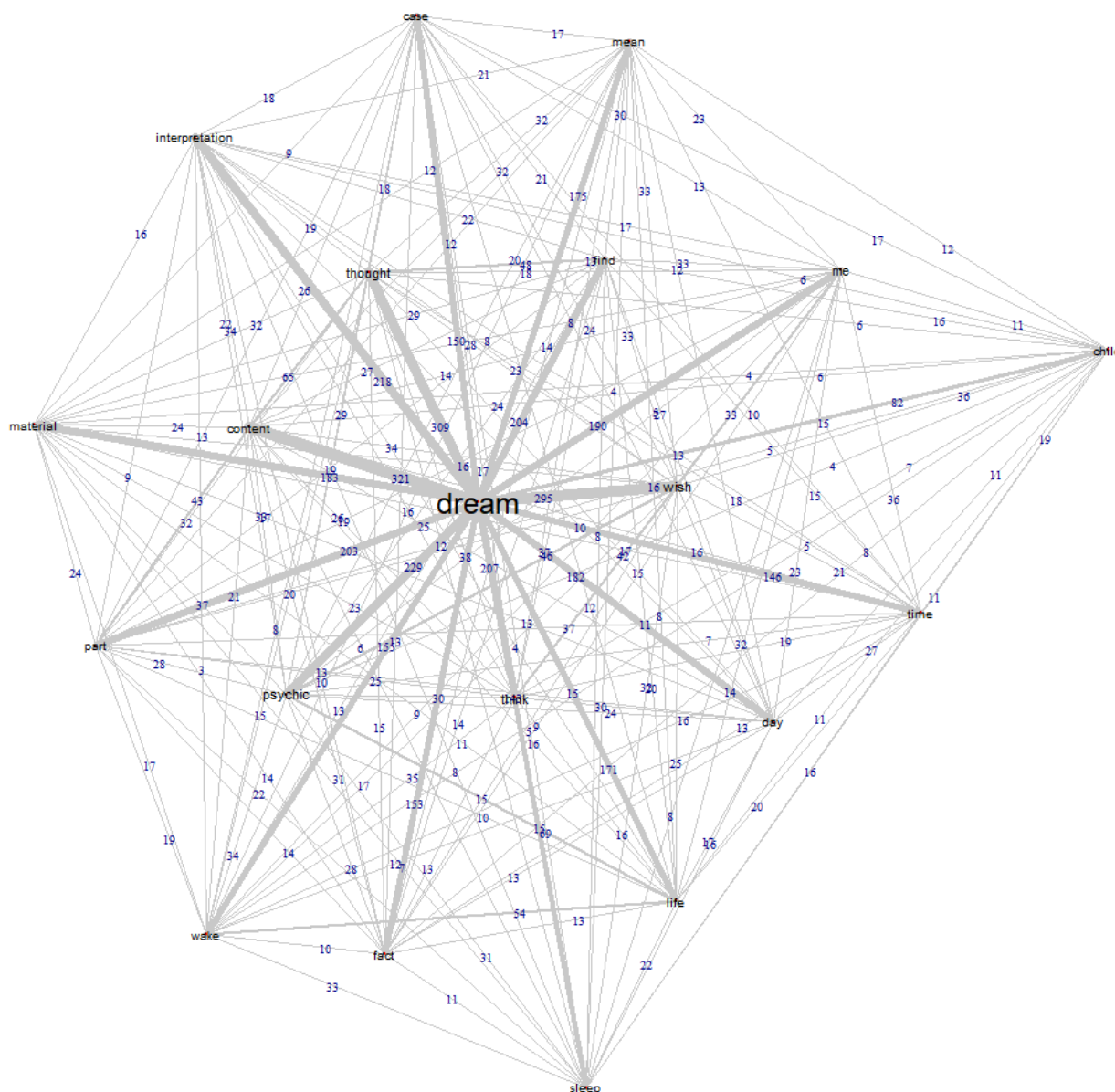


Figura 12. Gráfico da Análise de Similitude das 20 palavras mais frequentes do corpus com peso das relações
Fonte: Autoria própria

Nesse contexto, é possível interpretar o gráfico da figura 12 a partir das relações estabelecidas graficamente pelas linhas de ligação entre as palavras e os valores dos pesos dessas relações. O ponto de destaque se dá pela a forte ligação entre a palavra "dream" ao centro, sendo a palavra mais importante do texto e a que possui ligações com todas as demais palavras presentes no gráfico, e as palavras "content", "thought" e "wish", com pesos 321, 309 e 295 respectivamente. Dessa forma é possível estruturar linhas de diálogo entre o gráfico e o texto, que simplificam e norteiam uma pesquisa para o entendimento mais

aprofundado de seu conteúdo. Aqui, pode ser exemplificado como uma sugestão de interpretação a possível frase que resume o livro: "sonhos contêm/são pensamentos e desejos".

Considerações Finais

Este trabalho não buscou de forma alguma ser ou se tornar um guia da aplicação das técnicas de Processamento de Linguagem Natural, porém teve-se a intenção de demonstrar um pouco da sua utilização aplicada sobre uma das obras mais conhecidas de Sigmund Freud.

Diferentemente de outros trabalhos que aplicam técnicas de PLN sobre o conteúdo gerado durante as sessões de terapia conduzidas pelos psicoterapeutas de forma transcrita, esse trabalho se propôs a aplicar tais técnicas sobre a produção teórica de um dos principais autores da psicanálise, indo numa direção contrária ao que se tem como comum, demonstrando que é possível e de grande utilidade a aplicação de PLN sobre as obras que formam a base do conhecimento psicanalítico, em especial quando se busca compreender melhor o texto teórico. Além disso, trouxe também casos de aplicação das técnicas e sua interpretação, alcançando assim todos os objetivos propostos previamente.

Por ser um autor muito conhecido, e pela obra analisada também ser conhecida, algumas pressuposições podem surgir a partir da simples leitura do título da obra, mas ao ser observado o resultado das técnicas aplicadas é possível quebrar um pouco dessas pressuposições.

Evidentemente, a aplicação das técnicas de PLN e análise de resultado das mesmas não substitui a leitura da obra, mas pode ser de grande importância no estudo de seu conteúdo, demonstrando correlações entre palavras e temas que podem passar despercebidos.

Sobre os resultados obtidos, é importante notar que a palavra “dream” é sem dúvidas a palavra central da obra, comprovada a partir das figuras 8 e 9, e também a partir do gráfico de frequência de palavras (tanto Figura 5 quanto Figura 6), e que em parte isso era de ser pressuposto dado o título da obra.

Porém, a fortíssima relação entre a palavra “dream” e outras palavras que não são tão frequentes e que não aparecem entre as 10 palavras mais frequentes visualizadas na figura 6 é algo de grande valor para a interpretação dos resultados.

A palavra “dream” se relaciona fortemente com as palavras “content”, “thought” e “wish”, sendo que essa relação é demonstrada nas figuras resultantes da análise de similitude (Figura 11 e Figura 12) que demonstram correlações entre as 20 palavras mais

frequentes. Vale notar também que a relação dessas palavras entre si é baixíssima, porém existente, e que durante a leitura integral da obra essas relações podem acabar não sendo percebidas, dada a sutileza das mesmas.

Outras análises poderiam ter sido apresentadas nesse trabalho fazendo uso das mesmas ferramentas (Linguagem R, RStudio e Iramuteq), como a Classificação Hierárquica Descendente (CHD) e a Análise fatorial de Correspondência (AFC), que também seriam de grande valia para a discussão dos resultados. Porém priorizou-se as de maior facilidade de entendimento e que atualmente são de relativo conhecimento popular, o que de forma alguma desmerece a significância de tais técnicas (CHD e AFC) para a análise textual. O resultado da aplicação dessas técnicas pode ser observado no apêndice deste trabalho.

De forma geral, tanto o uso da linguagem R e RStudio a partir de scripts, quanto a utilização do software Iramuteq foram bem sucedidos no papel de demonstrar a aplicação das técnicas, seus fundamentos e motivações de aplicação, e abrem espaço para maiores estudos sobre a utilidade da PLN não somente como parte da ciência de dados, mas também como forte aliada nos estudos de outras áreas, contribuindo significativamente para o enriquecimento de outras ciências que não necessariamente estão ligadas a tecnologia, e também como ferramenta para fixação de conteúdo e conhecimentos.

Em trabalhos futuros pode-se aplicar essas e outras técnicas de PLN a todas as obras de Sigmund Freud, seguindo um caminho para além de uma classificação de textos ou sumarização, mas também comparando os resultados das análises específicas de cada obra entre as obras analisadas e quem sabe, revivendo as ideias do autor em sistemas de chatbots.

Agradecimento

Agradeço a Deus, sem Ele nada seria. Agradeço aos meus pais, minha esposa, meu sogro, sogra e cunhada, que me apoiaram e auxiliaram em cada momento difícil que atravessei no decorrer da vida e do curso, me incentivando a seguir até o fim.

Referências

ALMEIDA JUNIOR, Oberdan Alves de. Beck: um chatbot baseado na terapia cognitivo-comportamental para apoiar adolescentes com depressão. 2017. Dissertação de Mestrado. Universidade Federal de Pernambuco.

CARDOSO, Marcelo de Oliveira. Indústria 4.0: a quarta revolução industrial. 2016. 43 f. Trabalho de Conclusão de Curso (Especialização em Automação Industrial) - Universidade Tecnológica Federal do Paraná, Curitiba, 2016.

Camargo, B. V., Justo, A. M. (2013). IRAMUTEQ: Um Software Gratuito para Análise de Dados Textuais. Temas em Psicologia, 21 (2), 513-518.

COUTINHO, Gustavo Leuzinger. A era dos smartphones: um estudo exploratório sobre o uso dos smartphones no Brasil. 2014. 60 f., il. Monografia (Bacharelado em Comunicação Social)—Universidade de Brasília, Brasília, 2014.

Dale, Robert; MOISL, Hermann; SOMERS, Harold (Ed.). Handbook of natural language processing. CRC press, 2000.

FARIA, Pedro Duarte; PARGA, João Pedro Figueira Amorim. Introdução à Linguagem R: seus fundamentos e sua prática. 2. ed. Belo Horizonte: [s.n.], 2021. ISBN 978-65-00-12606-8.

Disponível em: <https://pedro-faria.netlify.app/pt/publication/book/introducao_linguagem_r/>

Finatto, M. J. B. Lopes, L. Ciulla, A. (2015) “Processamento de Linguagem Natural, Linguística de Corpus e Estudos Linguísticos: uma parceria bem-sucedida.”. DOMÍNIOS DE LINGU@GEM, v. 9, n.5

GIL, Antônio Carlos. Metodologia da pesquisa. São Paulo: Atlas, 2002.

Indurkha, Nitin, and Fred J. Damerau. Handbook of natural language processing. Chapman and Hall/CRC, 2010.

Jurafsky, Daniel; MARTIN, James H. An introduction to natural language processing, computational linguistics, and speech recognition. 2000.

KASZUBOWSKI, Erikson. PROCESSO PRIMÁRIO E DOMÍNIO SEMÂNTICO: APROXIMAÇÕES ENTRE PSICANÁLISE E ANÁLISE DO DISCURSO POR MEIO DO PROCESSAMENTO DE LINGUAGEM NATURAL, 2013.

Manning, Christopher; Schutze, Hinrich. Foundations of statistical natural language processing. MIT press, 1999.

OLIVEIRA, Alana Lúcia; MATOS, Leonardo; DELABRIDA, Zenith; DA SILVA, Gilton; MACÊDO, Hendrik; PRADO, Bruno. Chatbot para Terapia Cognitivo Comportamental focada em Ruminação (RFCBT) em jovens universitários. In: ESCOLA REGIONAL DE COMPUTAÇÃO BAHIA, ALAGOAS E SERGIPE (ERBASE), 18. , 2018, Aracaju. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2018 . p. 338-347.

SCHULTZ, Viviane; CRUZ, Roberto Moraes. Avaliação de competências profissionais e formação de psicólogos. Arquivos Brasileiros de Psicologia, v. 61, n. 3, p. 117-127, 2009.

Silge, Julia; Robinson, David. Text mining with R: A tidy approach. " O'Reilly Media, Inc.", 2017.

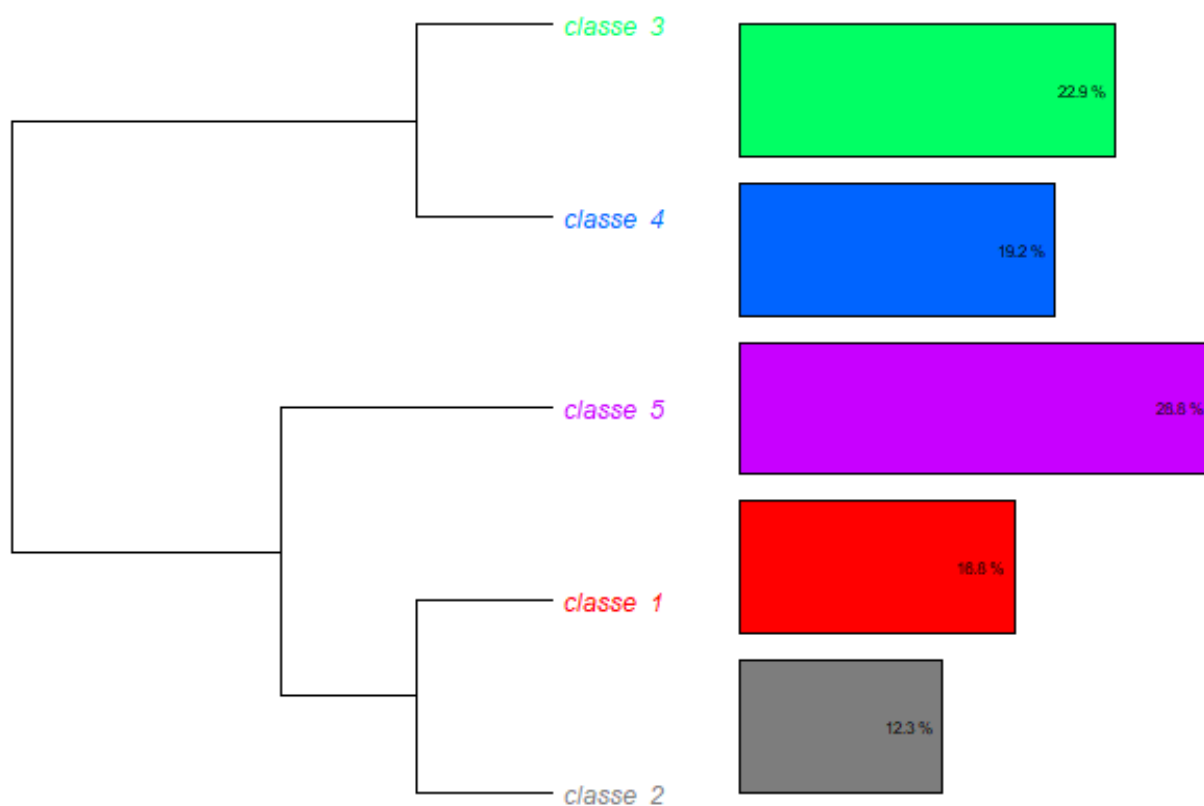
SILVA, B. J. F.; MALAGOLLI, G. A. UM ESTUDO SOBRE AS MUDANÇAS CAUSADAS PELO USO DE TECNOLOGIA DE INFORMAÇÃO NA ESTRUTURA ORGANIZACIONAL DE EMPRESAS. Revista Interface Tecnológica, [S. l.], v. 11, n. 1, p. 67–78, 2014. Disponível em: <https://revista.fatectq.edu.br/interfacetecnologica/article/view/91>. Acesso em: 4 nov. 2022.

VIEIRA, Renata; LOPES, Lucelene. PROCESSAMENTO DE LINGUAGEM NATURAL E O TRATAMENTO COMPUTACIONAL DE LINGUAGENS CIENTÍFICAS. Em corpora, p. 183, 2010.

Wickham, Hadley. "Data analysis." ggplot2. Springer, Cham, 2016. 189-201.

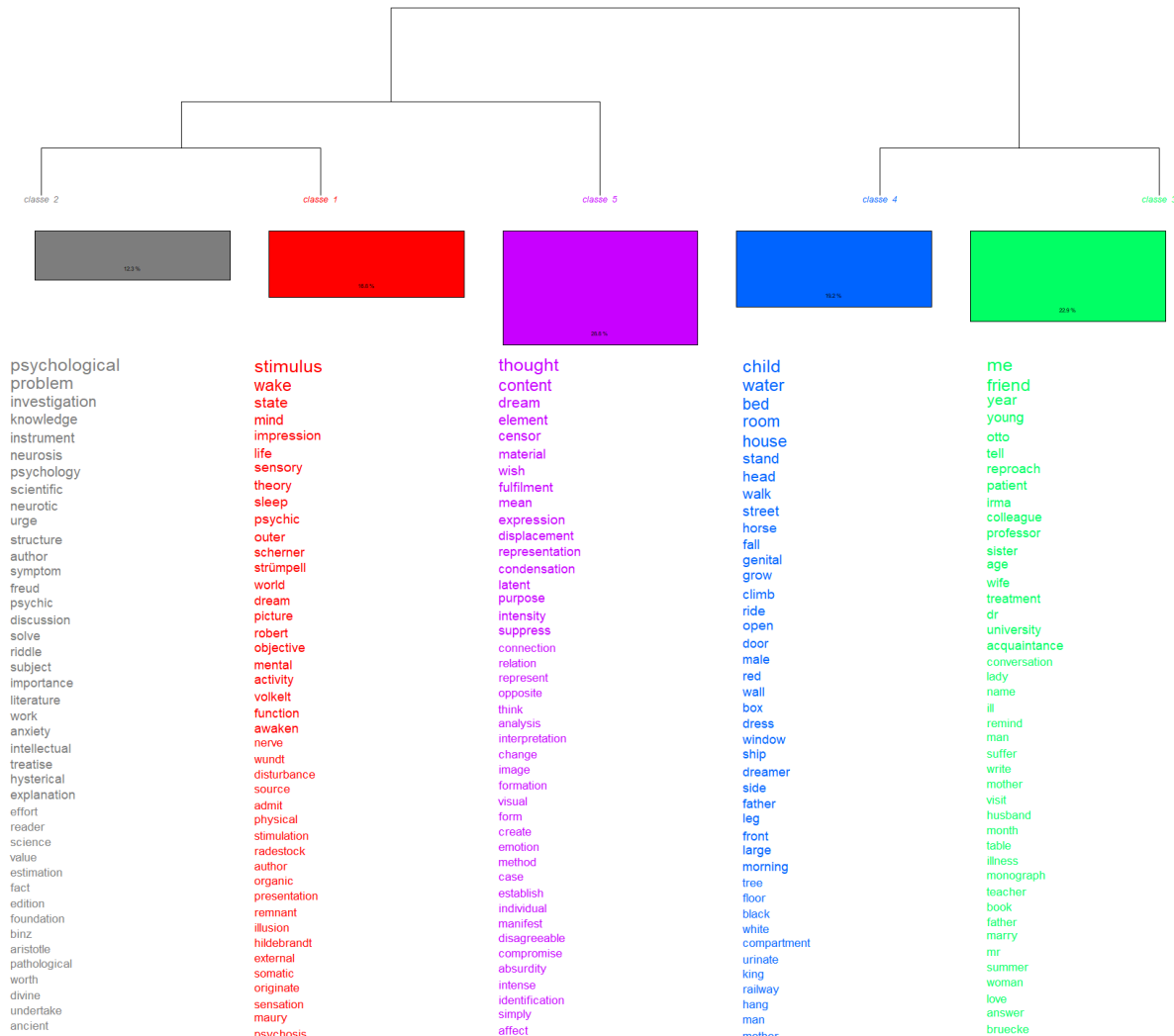
Yang, Jin, et al. "Brief introduction of medical database and data mining technology in big data era." Journal of Evidence-Based Medicine 13.1 (2020): 57-69.

Apêndice



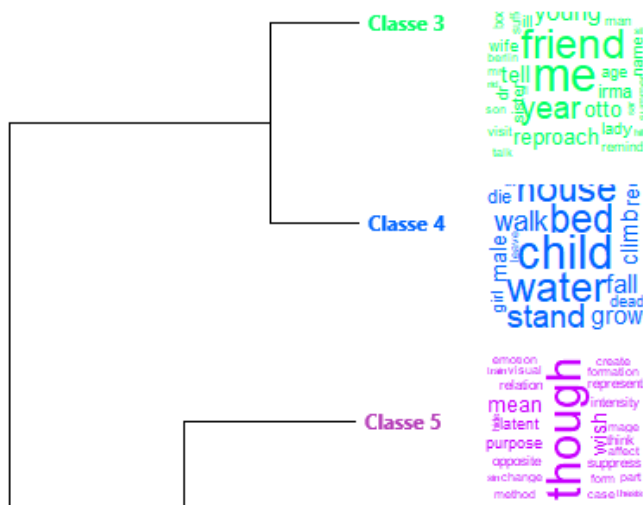
Apêndice 1. Gráfico de Classificação Hierárquica Descendente (CHD)

Fonte: Autoria própria



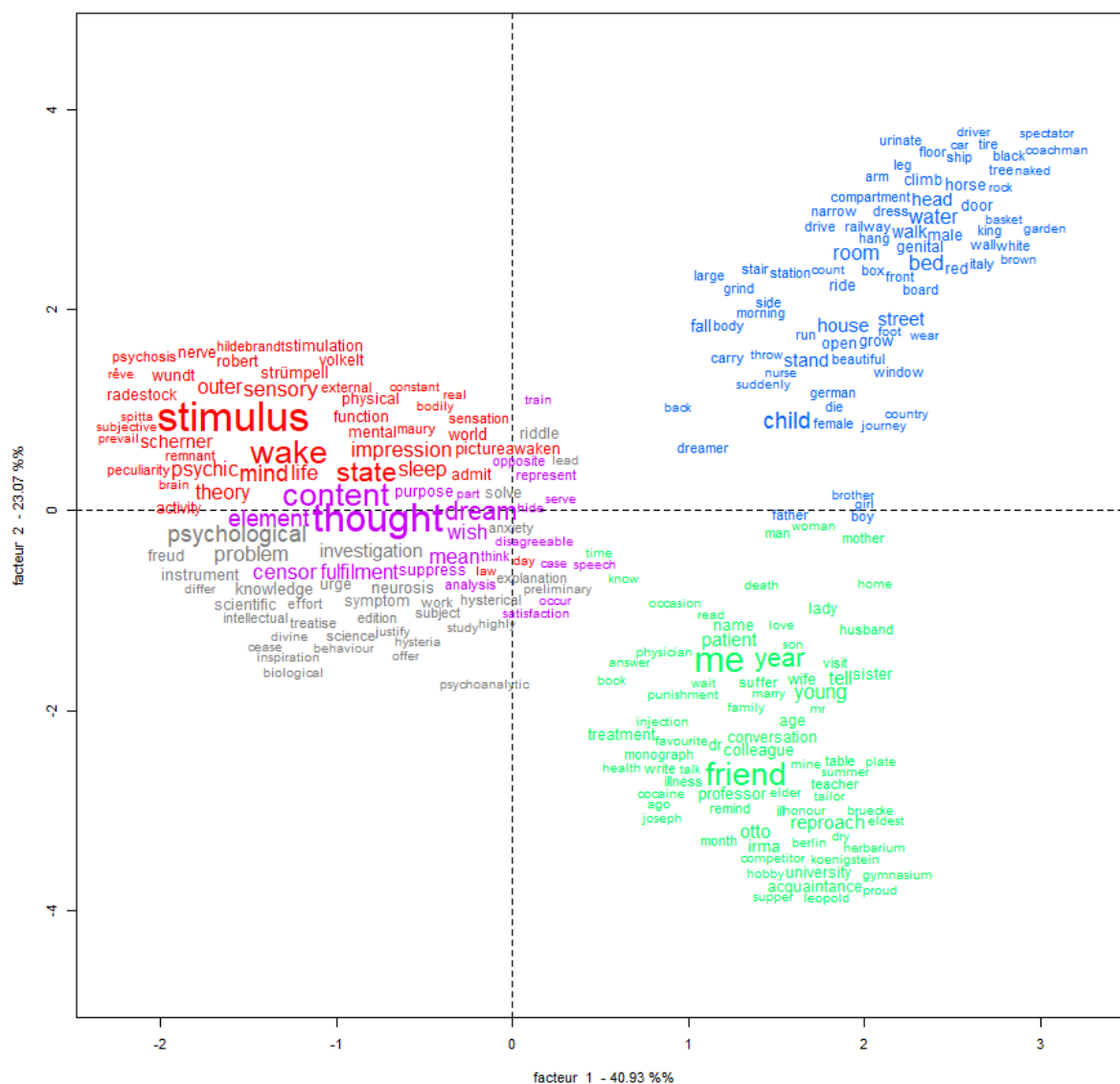
Apêndice 2. Gráfico de Classificação Hierárquica Descendente (CHD)

Fonte: Autoria própria



Apêndice 3. Gráfico de Classificação Hierárquica Descendente (CHD) com gráfico de Nuvem de Palavras de cada classe

Fonte: Autoria própria



Apêndice 4. Gráfico da Análise fatorial de correspondência (AFC)

Fonte: Autoria própria