



# Análise da obra de Pierre Bourdieu pelo método da mineração de texto, utilizando R.

José Carlos Porto Arcoverde Jr  
UFPE / CESAR School  
Recife, Brasil  
mabuse@cesar.school

Leonardo Augusto Gomez Castillo  
UFPE  
Recife, Brasil  
leonardo.castillo@ufpe.br

**Resumo** — Nos últimos anos vimos os campos de ciência de dados e ciências sociais cada vez mais se aproximando. A estatística, que sempre teve um papel importante na sociologia, ganha espaço em outras disciplinas como antropologia e literatura. Tal expansão se dá por conta do desenvolvimento de um poder computacional associado à criação de novos algoritmos, em especial no campo da *Natural Language Programming*, promovendo, mais do que uma análise estatística, uma possibilidade de entendimento de contexto, aplicada à mineração de textos. A contribuição da comunidade de Software Livre é observada no presente artigo na produção de pacotes na linguagem R, que tem sido um meio de iniciação de pesquisadores não-programadores no uso desses recursos computacionais. No presente artigo é introduzido o uso dos métodos de mineração de texto, por meio do Sistema R, por usuário que não tem o conhecimento de programação, como exercício propomos uma análise de agrupamento para dar uma nova possibilidade de leitura de parte da obra do Sociólogo Pierre Bourdieu.

**Palavras-chave**—Design; Mineração de Texto; Linguagem R; Código Aberto.

**Abstract** - *In the last few years we have seen the fields of data science and social sciences getting closer and closer. Statistics, which has always played an important role in sociology, is gaining ground in other disciplines such as anthropology and literature. Such expansion is due to the development of a computational power associated with the creation of new algorithms, especially in the field of Natural Language Programming, promoting, more than a statistical analysis, a possibility of understanding the context, applied to text mining. The contribution of the open source community is observed in this article in the production of packages in the R language, which has been a means of initiating non-programmer researchers in the use of these computational resources. This article introduces the use of text mining methods, through the R System, by a user who does not have programming knowledge, as an exercise we propose a cluster analysis to give a new possibility of reading part of the Sociologist's work Pierre Bourdieu.*

**Keywords**—Design; Text Mining; R Language; Open Source.

## I. INTRODUÇÃO

O presente artigo trata de duas questões: a importância das ferramentas *open source* de estatística que potencialmente podem ser aplicadas à pesquisa qualitativa, e o papel das comunidades de desenvolvimento de software livre, em especial para o público de não-programadores originários das disciplinas das ciências sociais. Essa investigação se dá a partir de uma perspectiva do *design*, por entendermos que um novo objetivo da sua prática no século XXI é a investigação criativa [1] e que, como criatividade compreendemos a combinação de estruturas anteriormente

não relacionadas, para obtenção de um todo que seja maior que o ponto de partida [2], então podemos pensar na importância de instrumentos tecnológicos que contribuam para tal tarefa do *designer*, sendo esse um tema abordado no Brasil ao menos desde o final do século passado [3].

Assim é apresentado o uso de um procedimento técnico originário da linguística computacional, aplicado por meio da linguagem *Open Source R*, como instrumento que contribui para novas leituras de documentos textuais, e como exemplo utilizaremos um recorte da obra do sociólogo Pierre Bourdieu.

## II. PIERRE BOURDIEU, O PENSADOR DAS DESIGUALDADES.

A sociologia passou por grandes transformações em meados do século XX e em especial após a Segunda Guerra Mundial. Mudanças geopolíticas como a derrota das nações fascistas e a ascensão e consolidação econômica e militar dos Estados Unidos da América, aparentemente demonstraram uma superioridade do sistema democrático liberal ocidental [4] é nesse cenário que se forma aquele que foi considerado como um dos grandes pensadores das desigualdades no século XX [5].

Pierre Bourdieu nasceu e cresceu na cidade de Denguin, no sudeste da França, em 1930. No início dos anos 1950 mudou para Paris e ingressou na *École Normale Supérieure*. Ao iniciar sua vida acadêmica, decepcionou-se com a doutrina da Filosofia do Sujeito do existencialismo (predominante na época) e deu passos para a Filosofia do Conceito, entrando então em contato com epistemólogos como Gaston Bachelard e Georges Canguilhem ao mesmo tempo que se aprofundava na fenomenologia de Edmund Husserl e Merleau-Ponty [6].

Dentro de sua vasta obra um de seus conceitos mais importantes é o de *Habitus*, criação de grande engenhosidade, que propõe uma relação dialética entre dois conceitos teóricos até então opostos: o objetivismo (no qual a realidade social é um conjunto de relações e forças impostas por agentes estruturais que não tem desejos ou consciência) e o subjetivismo (onde por sua vez a realidade social é a soma dos atos de interpretação no qual as pessoas constroem o sentido, por meio de suas interações). A partir da relação dialética entre agente e estrutura, o *habitus* então é definido como um sistema de disposições duráveis, e fonte da percepção, apreciação e ação [7]. O que sua teoria apresenta de solução para a questão é o entendimento de que o ator social age dentro de um campo socialmente pré-determinado.

Uma das questões que se colocam sobre a obra de Bourdieu é sobre a relação dele com o pensamento marxista.

Para a socióloga Bridget Fowler [8], mesmo considerando a grande quantidade de autores que são trabalhados por Bourdieu, entre eles Weber, Durkheim, Husserl, Mauss, Elias, e Pascal entre muitos outros, é na sua obra e em citações presentes a nomes como Lukács, Gramsci, Adorno, Benjamin, Williams, e E. P. Thompson, que é possível entender Bourdieu como um dos grandes herdeiros do marxismo ocidental no século XX.

### III. MÉTODO DE INVESTIGAÇÃO

Com o aumento de material potencialmente adequado para construção de um corpus de pesquisa sendo disponibilizado de forma digital, seja pelos usos de redes sociais pelas pessoas, seja na forma de documentos eletrônicos pelos governos, e em paralelo com o desenvolvimento de algoritmos de detecção de estruturas semânticas que permitem a inclusão de contexto dentro de modelos linguísticos computacionais, houve um crescimento do uso de computadores na análise de textos de forma mais sofisticada, com isso temos ferramentas computacionais que saem "de uma contagem de palavras, para a extração de significado" [9]. Esse fenômeno abre uma possibilidade para pesquisadores das ciências sociais, como um novo campo para análise de conteúdo [10].

Ao processo de extrair informações úteis e de qualidade, de textos não-estruturados, com o objetivo de entender padrões e tendências utilizando recursos computacionais, é dado o nome de Mineração de Texto [11], esse é um campo do conhecimento fruto do encontro de dois campos maiores: a Mineração de Dados e a Linguística Computacional, que trata da investigação quantitativa de dados linguísticos. Essa abordagem atende à necessidade de uma nova lente para análise de um volume grande de dados, com o objetivo de dar uma outra perspectiva para uma investigação ou a confirmação de outro procedimento anterior.

Todo exercício apresentado a seguir foi realizado na linguagem R versão 4.0.2 [12], dentro do ambiente de desenvolvimento RStudio na versão 1.3.1073[13]. Os pacotes utilizados além dos recursos nativos da linguagem, foram: tm[14], magrittr[15] e wordVectors[16].

#### A. O processo da Mineração do Texto

Mineração de Textos é um campo de conhecimento fortemente interdisciplinar, construído a partir de áreas tão diversas como linguística, estatística e ciência da computação. Dentro desse campo algumas das técnicas mais utilizadas são classificação de texto, agrupamento (clustering) de textos, criação de ontologia e taxonomias, sumarização documental e análise de corpus de texto..

Segundo Aranha [17] a mineração de textos se dá em cinco fases:

1. Coleta - É a construção do *Corpus* que será utilizado nesta análise de texto.
2. Pré-processamento - Etapa necessária de limpeza, formatação e representação do *corpus*. É onde se dá o esforço em transformar os texto não-estruturado em um formato de melhor acesso para os algoritmos de análise e consome um tempo considerável na atividade de análise.
3. Indexação - É a organização dos termos com a descrição de suas fontes de dados, importante para volumes massivos de informação.
4. Mineração - O conjunto de cálculos, inferências e algoritmos para encontrar padrões e comportamentos que

podem, dentro do contexto, construir estruturas de informações.

5. Análise - Por meio das informações oferecidas, em uma atividade humana, se constrói conhecimento sobre o corpus da pesquisa [18].



1. As 5 fases da mineração de texto (ARANHA, 2006).

#### A. Agrupamento (Clusterização)

O agrupamento (também conhecido pelo seu termo português clusterização) é o processo de organização de grupos para uma coleção. Para que isso aconteça, alguma métrica adotada define a similaridade entre esses elementos. Tais métricas são definidas numericamente. Para isso, há uma transformação no documento, que tem suas palavras mapeadas de forma vetorial [19].

Existem duas questões que definem o agrupamento como uma tarefa de alta complexidade: a quantidade de parâmetros exigidos para essa vetorização e a falta de uma resposta "correta", para o problema do agrupamento, em seu lugar o que há é o agrupamento mais adequado para a atividade desejada [20].

A melhor representação de agrupamentos para o contexto desse artigo é conhecida como Dendrograma [21], que representa visualmente o agrupamento das entidades. No caso do presente artigo as entidades são termos organizados de acordo com os conceitos relacionados ao marxismo.

#### Coleta

Considerando a limitação de tempo de escrita do artigo, foi definido um recorte dos textos de autoria de Pierre Bourdieu, como critérios de recorte, o primeiro foi dos textos disponíveis na língua inglesa e em formato eletrônico, que facilite a manipulação computacional. Diante de uma produção tão vasta quanto a de Pierre Bourdieu, também entrou como critério de recorte textos que foram citados como importantes diálogos com o marxismo pela pesquisadora Bridget Fowler [8]: *Masculine Domination* [23], *Algerian landing*[24], *Distinction*[25], *Pascalians Meditations*[26] e *Homo Academicvs*[27].

#### Pré-processamento e mineração de texto

Em razão da pequena quantidade de dados analisados para o presente artigo não houve necessidade de uma indexação para o aumento da velocidade de processamento, logo, as fases posteriores à coleta foram adaptadas do processo apresentado na figura 1, para conter exclusivamente o pré-processamento e a mineração de texto.

Toda fase de pré-processamento e mineração foi executada utilizando a linguagem R.

Desenvolvida para a construção de programas estatísticos, R é uma linguagem de programação interpretada [28], criada como uma versão similar a linguagem e ambiente S, desenvolvida pela Bell Laboratories por John Chambers [29], durante muito tempo S foi a linguagem mais utilizada para estatística aplicada às mais diversas áreas do

conhecimento. No seu site a R Foundation identifica que a linguagem pode ser considerada uma implementação diferente de S, uma das diferenças significativas sendo que R é um projeto de código aberto com o objetivo de fomentar o desenvolvimento de aplicações estatísticas nas mais diversas áreas[30], disponível dentro da Licença Pública Geral GNU (*GNU General Public License, GPL*), em sua segunda versão, de autoria da *Free Software Foundation*. [31].

O ambiente R é formado por uma coleção de softwares que facilitam o uso da linguagem de programação, além da linguagem também faz parte o ambiente de programação *RStudio*, o *Shiny* que conta com um conjunto de widgets que facilitam a publicação na web de projetos do R e o *R Packages*, um gerenciador para os pequenos programas específicos criados em R e chamados de pacotes (*packages*) para fácil uso. É na riqueza da comunidade open source e sua contribuição na forma de *packages* que está o maior atrativo de uso do ambiente [31], de *packages* para Mineração de Textos [11] até captura de dados do Datasus [32] Dentro das ciências sociais é onde se tem hoje uma grande e ativa comunidade.

### Análise

Dentro da análise de agrupamentos textuais, a representação visual adotada no presente artigo se dá por meio de Dendrogramas. Sua leitura se faz da base do diagrama para o topo, as conexões entre palavras representam os termos encontrados mais próximos.

A partir dessa leitura é possível analisar padrões que podem levar tanto à leituras alinhadas com a análise tradicional dos textos, quanto à interpretação contraintuitiva, ou seja, que à primeira vista podem parecer obscuras e até erradas, mas tem potencial de lançar uma nova perspectiva sobre o texto analisado por meio de um recurso visual.

O dendrograma foi adotado como ferramenta de *Visual Thinking*, uma categoria de técnicas que contribuem para a solução de problemas a partir de uso de representações visuais, com o objeto de diminuir o esforço cognitivo da pessoa engajada na tarefa [33].

Assim, como em outros procedimentos técnicos utilizados pelo *design* como mapas mentais, por não haver uma resposta "certa" em relação ao agrupamento, tem resultados bastante variados de acordo com o processo de mineração de dados utilizado, o que abre possibilidades para análises mais ampliadas [34].

### Dendrograma dos textos de Bourdieu

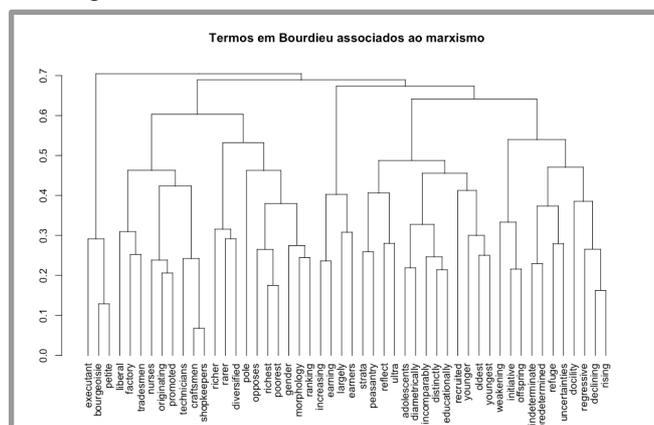


Figura 2. Dendrograma dos termos de Bourdieu em relação ao marxismo.

No presente artigo será apresentada a seguir uma seleção de alguns agrupamentos representados no dendrograma, e que chamaram a atenção na análise dos termos usados por Bourdieu. Utilizando o sistema da linguagem R é possível retornar ao texto original e encontrar em que momentos esses termos foram utilizados, o que pode contribuir para uma leitura dirigida sobre o texto.



Figura 3. Detalhe do Dendrograma dos termos de Bourdieu em relação ao marxismo.

O agrupamento representado na Figura 3 foi o primeiro no dendrograma, ligado diretamente ao subconjunto de todos outros, pode ser considerado um bom exemplo de um agrupamento que encontra uma referência explícita no texto, nos dando um caminho inicial de investigação. Nele a combinação de termos que em português se traduzem por "executantes" e "pequena burguesia", são o ponto de partida. Com o uso das próprias ferramentas de pesquisa do R, podemos encontrar rapidamente a referência do título do capítulo 6 do livro *A Distinção: "A Pequena Burguesia de Execução"* [35] e, a partir daí investigar outras referências nos outros livros que passaram pelo processo de mineração.

Podendo ser considerado um conceito secundário da obra de Bourdieu, a pequena burguesia de execução é uma subdivisão dos trabalhadores de colarinho branco. Nesse caso, a abordagem adotada colocou um ponto de entrada na obra do autor que não é necessariamente o caminho tomado frequentemente pelos seus analistas, mas tem um potencial interessante junto à relação com o marxismo.

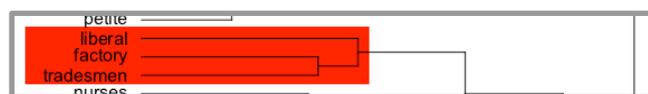


Figura 4. Detalhe do Dendrograma dos termos de Bourdieu em relação ao marxismo.

Outras combinações parecem levantar agrupamentos que se apresentam com clara relação com os conceitos marxistas, mas com pouca capacidade de articulação na obra analisada, de tão ampla interpretação, os conceitos pedem uma revisão no procedimento técnico para chegar a resultados mais aproximados de referências dentro do texto. Um exemplo encontra-se na Figura 4.

Uma das formas propostas de refino durante a fase de mineração seria uma melhor definição de palavras-chave que delimitam o que é o marxismo.

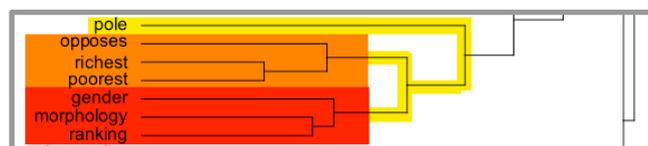


Figura 5. Detalhe do Dendrograma dos termos de Bourdieu em relação ao marxismo.

Por fim outros agrupamentos apresentam combinações de ramos que levantam possibilidades de investigação bastante promissoras, como no conjunto formado pela palavra "Polo" e ligada aos dois agrupamentos "oposição" e "gênero" que por sua vez são ligados, respectivamente à "mais ricos"/"mais pobres" e "morfologia"/"ranking". A análise desse conjunto complexo pode levar a leituras bastante interessantes, considerando os ramos do primeiro grupo com

termos mais relacionados ao discurso marxista e os ramos do segundo grupo à procedimentos técnicos relacionados à teoria de Bourdieu, sem necessariamente ter uma ligação explícita ao marxismo.

#### IV. CONCLUSÕES

A aplicação de procedimentos técnicos associados à mineração de textos para fazer emergir temas relacionados ao objeto de pesquisa é bastante promissor para a pesquisa em *design*. Principalmente considerando a possibilidade de uso de grandes volumes de informação. Para o presente artigo foi feito um recorte da obra de Bourdieu que resultou em cinco livros, mas os procedimentos técnicos utilizados foram desenvolvidos pensando na ordem de milhares de textos.

É importante ressaltar quatro pontos: a) o *corpus* da pesquisa tratado aqui é formado por textos de livros e artigos, mas pode ser aplicado o mesmo método para qualquer produção textual de posts em redes sociais à um conjunto de transcrições de entrevistas; b) que do ponto de vista dos métodos de pesquisa em *design*, esse recurso é entendido como uma rica contribuição estatística inicial para uma pesquisa qualitativa posterior; c) como da mesma forma nas análises semiótica, um conhecimento aprofundado no tema, no caso a obra de Pierre Bourdieu, leva à uma análise mais fácil dos diagramas gerados; e d) foi confirmado o grande potencial de uso por pesquisadores que tenham um perfil de não-programadores, em grande medida graças ao desenvolvimento de pacotes produzido pela comunidade de *software* livre.

#### V. TRABALHOS FUTUROS

O presente artigo se refere primeiro ensaio, com o recorte da obra de Pierre Bourdieu, que tinha como objetivo iniciar uma compreensão das possibilidades técnicas de estatística aplicadas ao *corpus* como uma das fases de uma pesquisa qualitativa, os próximos passos compreendem a análise do discurso, com os recursos da mineração de texto, dos artigos científicos no escopo de revistas acadêmicas de *design*.

Fazem parte desses trabalhos futuros: o estudo de outras possibilidades de pacotes para agrupamentos e representações visuais, o uso de corpus de textos mais robustos, assim como a investigação desse procedimento dentro de um projeto de *design*, como "gerador de *insights*" para o corpus da pesquisa que será aplicada no projeto.

#### REFERÊNCIAS

- [1] B. A. Lorenz, "Pesquisa através do *design* e prática crítica: uma investigação sobre o desenvolvimento de artefatos críticos no processo de construção de problemas de pesquisas acadêmicos," Universidade do Vale do Rio dos Sinos, 2018.
- [2] A. Koestler, *The Act of Creation*. London: Hutchinson & Co., 1966.
- [3] L. A. de Saboya and V. Iório, "Computadores e *design*: uma questão atual," *Estud. em Des.*, vol. 1, no. 1, pp. 28–30, 1993.
- [4] B. Fowler, "Pierre Bourdieu: Unorthodox Marxist?," in *The Legacy of Pierre Bourdieu: Critical Essays*, London: Anthem Press, 2011, pp. 33–58.
- [5] N. Nunes, "Desigualdades sociais e ação coletiva nas sociedades contemporâneas: a fecundidade teórica de Pierre Bourdieu e de Nicos Mouzelis," *Sociologia*, vol. 25, pp. 33–53, 2013.
- [6] L. J. D. Wacquant, "Pierre Bourdieu," in *Key Sociological Thinkers*, London: Macmillan Press, 1998, pp. 215–229.
- [7] R. Ortiz, "A procura de uma sociologia da prática," in *Pierre Bourdieu - Grandes Cientistas Sociais*, São Paulo: Ática, 1983, pp. 7–37.
- [8] B. Fowler, "Pierre Bourdieu: Unorthodox marxist?," in *The Legacy of Pierre Bourdieu: Critical Essays*, 2011.
- [9] G. Wiedemann, "Opening up to big data: Computer-assisted analysis of textual data in social sciences," *Hist. Soc. Res.*, vol. 38, no. 4, pp. 332–357, 2013.
- [10] M. Lemke, A. Niekler, G. S. Schaal, and G. Wiedemann, "Content Analysis between Quality and Quantity," *Datenbank-Spektrum*, vol. 15, no. 1, pp. 7–14, 2015, doi: 10.1007/s13222-014-0174-x.
- [11] A. Kumar and A. Paul, "Mastering Text Mining with R." Birmingham - Mumbai: Packt, 2016.
- [12] R Core Team, "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2020.
- [13] RStudio Team, "RStudio: Integrated Development for R." RStudio, PBC, Boston, 2020, [Online]. Available: <http://www.rstudio.com/>.
- [14] I. Feinerer and K. Hornik, "tm: Text Mining Package." 2020.
- [15] S. M. Bache and H. Wickham, "magrittr: A Forward-Pipe Operator for R." 2020.
- [16] B. Schmidt and J. Li, "wordVectors: Tools for creating and analyzing vector-space models of texts." 2020.
- [17] C. Aranha and E. Passos, "A Tecnologia de Mineração de Textos," *Rev. Eletrônica Sist. Informação*, vol. 5, no. 2, pp. 1–8, 2006, doi: 10.21529/resi.2006.0502001.
- [18] J. R. Carrilho Junior, "Desenvolvimento de uma Metodologia para Mineração de Textos." Rio de Janeiro, 2007.
- [19] G. Wiedemann, "Opening up to big data: Computer-assisted analysis of textual data in social sciences," *Hist. Soc. Res.*, vol. 38, no. 4, pp. 332–357, 2013.
- [20] L. S. Ochi, C. R. Dias, and S. S. F. Soares, "Clusterização em Mineração de Dados," *Programa de Pós Graduação em Computação (IC - UFF)*, no. May. 2004.
- [21] M. Drouot and L. Smith, "Reading dendrograms," *Ribotyping*, vol. 3, no. Slide 1, p. 8, 2012, [Online]. Available: <https://wheatoncollege.edu/wp-content/uploads/2012/08/How-to-Read-a-Dendrogram-Web-Ready.pdf>.
- [23] P. Bourdieu, "Masculine Domination". Stanford: Stanford University Press, 2002.

- [24] P. Bourdieu, "Algerian landing" *Ethnography*, vol. 5, no. 4, pp. 415–443, 2004, doi: 10.1177/1466138104048826.
- [25] P. Bourdieu, "Distinction: A Social Critique of the Judgement of Taste". London: Routledge, 2010.
- [26] P. Bourdieu, "Pascalians Meditations". Stanford: Stanford University Press, 2000.
- [27] P. Bourdieu, "Homo Academicus". Stanford: Stanford University Press, 1990.
- [28] G. Sawitzki, "R as a Programming Language and Environment." *Comput. Stat.*, pp. 193–231, 2009, doi: 10.1201/b17179-6.
- [29] J. M. Chambers, "Statistical computing: History and trends," *Am. Stat.*, vol. 34, no. 4, pp. 238–243, 1980, doi: 10.1080/00031305.1980.10483038.
- [30] J. M. Chambers, "S, R, and Data Science," *Proc. Acmprogr.*, vol. 4, no. June, 2020.
- [31] The R Foundation, "What is R?," 2020. <https://www.r-project.org/about.html> (accessed Nov. 12, 2020).
- [32] D. Petruzalek, "READ. DBC: um pacote para importação de dados do datasus na linguagem R," *J. Heal. Inf.*, vol. 8, no. supl.1, pp. 601–605, 2016, [Online]. Available: <http://pesquisa.bvsalud.org/brasil/resource/pt/biblio-906543>.
- [33] P. Pu and D. Lalanne, "Design visual thinking tools for mixed initiative systems," *Int. Conf. Intell. User Interfaces, Proc. IUI*, pp. 119–126, 2002, doi: 10.1145/502721.502736.
- [34] R. H. Baayen, *Analyzing Linguistic Data - A Practical Introduction to Statistics Using R*. Cambridge and London: Cambridge University Press, 2008.
- [35] P. Bourdieu, *A Distinção*. Porto Alegre: Zouk Editora, 2006.