

Mapeando o Campo da Comunicação no Brasil: desafios e descobertas metodológicas de uma metapesquisa

Paula Guimarães Simões

Doutora; Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
paulaguimaraessimoes@yahoo.com.br

Vera Regina Veiga França

Pós-Doutora; Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
veravfranca@yahoo.com.br

Ana Karina de Carvalho Oliveira

Doutora; Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
anakarina.akco@gmail.com

Laura Antônio Lima

Mestra; Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
lauraantoniolima@gmail.com

Lucas Afonso Sepulveda Alves

Doutorando; Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
afonsepuv@gmail.com

Maíra Lobato Bicalho Chagas Moura Campos

Mestra; Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
mairalobatomoura@gmail.com

Paulo Henrique Basílio Santana

Mestrando; Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, MG, Brasil
paulobasilio28@gmail.com

Resumo

O artigo se debruça sobre os desafios de uma pesquisa que mapeia as principais referências teóricas do campo da Comunicação no Brasil, a partir da coleta e do processamento de artigos apresentados entre 2006 e 2015, no evento da Associação Nacional dos Programas de Pós-Graduação em Comunicação. Descrevemos aqui os procedimentos de raspagem das referências teóricas de cada trabalho apresentado

em Grupos de Trabalho do evento e as formas encontradas para processar dados textuais através de ferramentas estatísticas, discutindo a articulação entre métodos quantitativos e qualitativos. Por fim, refletimos sobre as descobertas metodológicas de metapesquisa sobre campos científicos. O texto demonstra como a articulação entre procedimentos quantitativos e qualitativos pode ser profícua para o desenvolvimento das pesquisas nas Ciências Humanas e Sociais.

Palavras-chave

Metapesquisa. Metodologia quantitativa. Metodologia qualitativa. Campo científico.

1 Introdução

O presente artigo apresenta e discute etapas, desafios e descobertas metodológicas para a construção de um mapa das principais teorias acionadas por pesquisadores do campo da Comunicação no Brasil. Partimos do pressuposto de que “A relevância das questões metodológicas na pesquisa é inconteste.” (BRAGA, 2011, p. 7), de modo que o compartilhamento do percurso desenvolvido em uma investigação também o é. O texto é fruto de um trabalho¹ voltado para a identificação das principais referências teóricas utilizadas em nossas pesquisas, no sentido de mapear e evidenciar a pluralidade constitutiva de nossa área.

Considerando a extensão desse campo disciplinar – número de Programas de Pós-Graduação e associações científicas –, escolhemos, para a definição do nosso universo de pesquisa, os trabalhos apresentados nos Encontros Anuais da Associação Nacional dos Programas de Pós-Graduação em Comunicação (Compós). Trata-se de um evento de grande relevância e abrangência, exatamente por congregar os inúmeros Programas e contemplar diferentes áreas temáticas do campo comunicacional. Estabelecemos como recorte os trabalhos apresentados durante dez anos do evento, compreendendo o período 2006-2015. Até o momento foram analisadas seis áreas temáticas, totalizando 562 artigos coletados em diferentes GTs, conforme apresenta o quadro abaixo²:

¹ O projeto *As Novas Teorias da Comunicação: mapeamento de um campo científico* é coordenado pelas professoras Vera França e Paula Simões e investiga os alicerces teóricos das áreas temáticas que configuram o Campo da Comunicação no Brasil. Já foram tratadas seis áreas: Comunicação e Política, Estudos de Televisão, Novas Mídias, Estudos de Jornalismo, Epistemologia da Comunicação e Comunicação e Sociabilidade. Confira França e outros (2016, 2018), Simões e outros (2019a, 2019b).

² As áreas temáticas coincidiram com os GTs na maior parte da pesquisa, com exceção de Estudos de Televisão, que não contava com um GT homônimo desde 2006, quando a coleta foi iniciada.

Tabela 1 – Seis áreas temáticas pesquisadas (entre 2006 e 2015)

Área Temática	Número de artigos	GTs onde foram coletados
Comunicação e Política	99	Comunicação e Política
Estudos de Televisão	67	Fotografia, Cinema e vídeo; Mídia e entretenimento; Produção de sentido na mídia; Estudos de televisão.
Novas Mídias	100	Comunicação e Cibercultura
Estudos de Jornalismo	100	Estudos de Jornalismo
Comunicação e Sociabilidade	98	Comunicação e Sociabilidade
Epistemologia da Comunicação	98	Epistemologia da Comunicação
Total	562	

Fonte: Dados da pesquisa.

O objetivo deste artigo é apresentar desafios e descobertas metodológicas encontradas a partir do uso criativo de ferramentas ainda pouco utilizadas no campo da Comunicação, como a linguagem Python para raspagem de dados e o software Stata para o processamento desses dados. O caráter inovador desse tipo de exploração é que, a nosso ver, ainda são poucas as pesquisas que se utilizam dessas ferramentas de estatística para rodar dados não numéricos, isto é, dados textuais.

A seguir, apresentaremos o **desenho metodológico** da nossa investigação, bem como a forma como foi feita a **coleta de dados por meio da raspagem**, considerando o que de profícuo se revelou durante esse processo para a construção de um mapa teórico do campo da Comunicação.

2 Desenho da pesquisa

Etimologicamente, a palavra método vem do grego *methodos* e significa **caminho para** (meta: para; hodos: caminho). Metodologia, assim, pode ser vista, de forma mais ampla, como o estudo dos caminhos e estratégias adotadas em uma pesquisa científica. Não entendemos, contudo, que tais procedimentos já existem como receitas dadas *a priori* que

são simplesmente aplicadas a um fenômeno. Adotamos o posicionamento de Braga, para quem,

Diferente de uma perspectiva sobre metodologia como um conjunto de regras de encaminhamento apriorísticas; ou como mero relato posterior de componentes adjetivos à substância do tema; ou ainda como simples adesão a um quadro teórico estabelecido; eu gostaria de enfatizar tais questões como ações concretas e refletidas durante todo o desenvolvimento da pesquisa, desde as primeiras hipóteses até os resultados finais. (BRAGA, 2011, p. 7).

Nesse sentido, a metodologia é vista como o lugar das escolhas em uma pesquisa, como “[...] uma sabedoria na tomada de decisões em que o pesquisador se vê constantemente envolvido.” (BRAGA, 2011, p. 8). Existem vários procedimentos provenientes dessas escolhas e definidos ao longo de uma investigação e, no caso das “[...] Ciências Humanas e Sociais (CHS), é o forte consenso atual de que não é possível assumir abstratamente abordagens prévias e ‘fechadas’, a serem aplicadas a uma diversidade de pesquisas.” (BRAGA, 2011, p. 2)

Partindo dessa compreensão, apresentamos aqui os processos de tomada de decisões realizados na pesquisa e os métodos que buscamos entrecruzar, atentando para o modo como os caminhos foram sendo repensados ao longo do percurso de forma a constituir nossa metodologia.

Para elaborarmos um mapeamento dos principais autores, obras e conceitos utilizados em diferentes áreas do campo da Comunicação, como já adiantamos acima, optamos por trabalhar com os artigos disponíveis nos anais dos Grupos de Trabalho do evento anual da Compós, entre os anos 2006 e 2015, através do levantamento da bibliografia utilizada pelos trabalhos apresentados.

Há uma diversidade enorme de dados que podem ser coletados a partir dos artigos apresentados nos Grupos de Trabalho: primeiramente, cada trabalho está inserido em uma área temática (ou subárea da Comunicação) – como Comunicação e Política – e um Grupo de Trabalho (GT) que pode ou não ser homônimo à área. Há também, para cada publicação, a informação sobre o ano, o local onde o evento foi realizado, o nome do arquivo disponibilizado no site, o título, pesquisadores autores do artigo, instituições as quais esses pesquisadores estão vinculados, os autores usados como referência (e quais dessas referências são autocitação), a nacionalidade desses autores referenciados, e, por fim, as obras citadas.

Frente a essa diversidade, encontramos muitas dificuldades no início do projeto devido aos poucos recursos disponíveis naquele momento. Os bancos de dados das três primeiras áreas coletadas – Comunicação e Política, Televisão e Novas Mídias – foram feitos **à mão**, isto é, feitos a partir da coleta de referência por referência, sem nenhum procedimento automatizado, em cada artigo publicado de cada GT, para todos os anos selecionados. Com os bancos feitos, muitas revisões foram necessárias: vários títulos de obras referenciadas poderiam estar citados em diferentes idiomas, para além dos erros de digitação e inúmeras formas de referenciar os autores e as obras citadas.

Por fim, com os bancos de dados prontos, também tivemos que recorrer a um processamento **manual**. Quantos autores foram referenciados no total? Quantos deles eram de nacionalidade brasileira ou estrangeira? Quais eram os principais autores referenciados em cada GT? Quais dessas referências eram pesquisadores participantes do GT analisado? Quantas obras foram referenciadas e quais foram as mais referenciadas em cada área? Inicialmente, tentamos responder a essas perguntas sem o uso de ferramentas de processamento de dados, que poderiam ter nos fornecido respostas rápidas e com maior grau de certeza e confiabilidade. Consequentemente, a nossa análise desses dados **processados** também se mostrou pouco produtiva, devido às condições exaustivas de coleta e processamento **ao olho humano**. Com um grupo de cerca de dez pesquisadores, não conseguimos através do método **manual** um cruzamento possível dos dados obtidos para visualizações minimamente mais complexas.

Um passo importante e transformador foi a luta contra nossa resistência e o pouco hábito que tínhamos, em nossas pesquisas majoritariamente qualitativas, em utilizar métodos quantitativos e ferramentas estatísticas para nossa análise. Vale aqui uma breve discussão sobre a definição e o uso desses métodos nas pesquisas em Ciências Humanas e Sociais. Na abordagem qualitativa, a pesquisa é entendida como uma **prática social**, em que as escolhas e os caminhos parecem mais sensíveis e flexíveis ao contexto e aos sentidos (SPINK; MENEGON, 2004). A abordagem quantitativa busca a mensuração dos dados, a partir de contribuições da estatística, para chegar a uma generalização que responda ao problema de pesquisa.

Há um amplo debate que procura apontar os méritos e os limites de cada uma dessas abordagens, mas não é nosso objetivo apresentar, aqui, as nuances dessa discussão. Importante apenas acentuar que elas não precisam ser vistas como dicotômicas, como bem sugeriram Spink e Menegon (2004). Ao contrário, elas podem ser vistas como

complementares na análise de um fenômeno e no desenvolvimento de uma pesquisa, como procuramos demonstrar neste trabalho.

Assim, foi superando a visão dicotômica entre as perspectivas que buscamos as contribuições oferecidas pelas ferramentas de uso quantitativo e estatístico. Com isso, mudamos o ritmo, a precisão e a produtividade dos nossos métodos. Optamos, então, pela coleta de dados por meio de raspagem, a partir da linguagem de programação Python, e pelo processamento de dados na ferramenta Stata.

Vemos o trabalho quantitativo como um ponto de partida, orientador de caminhos mais objetivos e proíferos para uma análise qualitativa dos dados, numa dinâmica de complementaridade entre os dois tipos de abordagem. Apesar da quantidade numericamente menor dos nossos bancos de dados em relação a outras pesquisas quantitativas que fazem uso de um volume maior de dados, o nosso desafio e inovação metodológica está em trabalhar com **dados textuais** diversos, utilizando métodos quantitativos. Dados de texto abrem espaço para muitas variáveis e erros e, assim, podem se revelar mais vulneráveis que números exatos. Fez-se necessário, portanto, a construção de um *script* próprio para as nossas necessidades e o uso do processamento de dados com o maior rigor possível, de forma a responder as perguntas que nos orientam na construção do mapa teórico das áreas selecionadas. São essas etapas metodológicas que apresentaremos a seguir: **a coleta de dados por raspagem** e o **processamento de dados** realizados na nossa pesquisa.

3 Coleta de dados por meio de raspagem

A partir do desafio para coletar, organizar e processar um volume considerável de dados, como já mencionado, recorreremos a ferramentas que pudessem automatizar, agilizar e aumentar a precisão da nossa pesquisa. Nesse sentido, é importante dizer que o uso dessas ferramentas também tornou possível estabelecer o cumprimento dos nossos objetivos dentro de um prazo menor.

Para realizar a raspagem dos dados, recorreremos a um *script*, desenvolvido na linguagem de programação Python, na versão 3.6.. Lançado em 1991, o Python atual tem um modelo de desenvolvimento comunitário, aberto e administrado pela organização sem fins lucrativos Python Software Foundation. O código aberto, colaborativo e gratuito garante um certo dinamismo para as soluções desenvolvidas a partir de seu uso. Seu criador foi o

programador de computadores holandês Guido van Rossum, que continua a supervisionar o seu processo de desenvolvimento. Uma das principais características dessa linguagem é proporcionar uma leitura fácil do código e exigir poucas linhas de comando, quando comparada a outras linguagens.

De acordo com a Wikipédia, o Python

Foi considerada pelo público a 3ª linguagem ‘mais amada’, de acordo com uma pesquisa conduzida pelo site Stack Overflow³ em 2018, e está entre as cinco linguagens mais populares, de acordo com uma pesquisa conduzida pela RedMonk.⁴ (PYTHON, 2018, doc. não paginado).

A técnica de raspagem de dados é algo relativamente simples no campo da computação, atualmente. O termo raspagem de dados (do inglês, *Data Scraping*) se refere ao processo automatizado de coletar um volume considerável de dados por meio de alguma ferramenta. “[...] é uma técnica computacional na qual um programa extrai dados de saída legível somente para humanos, proveniente de um serviço ou aplicativo. Os dados extraídos geralmente são minerados e estruturados em um formato padrão como CSV, XML ou JSON.” (RASPAGEM..., 2018, doc. não paginado).

Por se tratar de uma coleta de dados realizada em arquivos em *Portable Document Format* (PDF), enfrentamos um desafio maior. A técnica de raspagem é algo que requer certa regularidade na estrutura dos dados, e os arquivos com os quais trabalhamos apresentavam irregularidade: há, nos arquivos, muitas mudanças de templates ao longo dos anos da Compós, além do lançamento diferenciado das referências bibliográficas por parte de cada pesquisador. Isso significa que, quando se faz uma raspagem em um site, o código de programação da página oferece uma estrutura com repetições e padrões que facilitam o desenvolvimento de um *script* capaz de encontrar as informações desejadas e coletá-las. Porém, não é exatamente assim com PDFs. Dessa forma, uma raspagem de dados em arquivos PDF requer o uso de expressões regulares, que são “padrões de caracteres que associam sequências de caracteres no texto” (LINS, 2007, doc. não paginado). No caso do Python, a lógica de expressões regulares já é algo nativo desta linguagem de programação, o que contribuiu muito para o desenvolvimento do *script*. Também é importante ressaltar que a ferramenta PDFMiner⁵ foi essencial para que essa raspagem pudesse ser feita.

³ Veja em Stack Overflow (2018).

⁴ Veja em O'Grady (2018).

⁵ “O PDFMiner permite obter a localização exata do texto em uma página, bem como outras informações, como fontes ou linhas.” (PDFMINER, 2014, doc. não paginado, tradução nossa).

Em relação ao *script* utilizado para a nossa coleta dos dados, os passos programados foram: acesso ao site; navegação (*crawler*); download dos documentos em PDF; conversão dos documentos para texto; saneamento (limpeza) dos dados para uniformizar entradas de texto e corrigir possíveis e prováveis erros; raspagem dos dados; geração das tabelas.

As bibliotecas utilizadas no código de programação foram: *lxml* (utilizada para navegação nas páginas HTML do site da Compós); *cssselect* (utilizada para navegação nas páginas HTML do site); *pdfminer3k* (utilizada para conversão dos arquivos PDF para texto); *requests-cache* (utilizada para navegação nas páginas HTML do site); *unidecode* (utilizado para conversão dos caracteres PDF dos documentos para texto); *pyexcel* e *pyexcel-xlsx* (utilizado para produção das tabelas finais).

Segue, abaixo, o pseudocódigo⁶ do código desenvolvido⁷ no Python que possibilitou a automatização da coleta de dados para a nossa pesquisa:

Pseudocódigo principal:

```
trabalhos = VAZIO
encontros = OBTER_ENCONTROS()
PARA CADA encontro EM encontros:
    gts = OBTER_GTS(encontro)
    PARA CADA gt EM gts:
        trabalhos_do_gt = OBTER_TRABALHOS(gt)
        trabalhos = trabalhos + trabalhos_do_gt

dados = VAZIO
PARA CADA trabalho EM trabalhos:
    trabalho_pdf = BAIXAR(trabalho)
    trabalho_txt = CONVERTER(trabalho_pdf)
    trabalho_dados = RASPAR(trabalho_txt)
    dados = dados + trabalho_dados
```

⁶ “Pseudocódigo é uma forma genérica de escrever um algoritmo, utilizando uma linguagem simples [...], sem necessidade de conhecer a sintaxe de nenhuma linguagem de programação.” (PSEUDOCÓDIGO, 2018, doc. não paginado).

⁷ Desenvolvido por Jonatas Renan Alves (2018).

GERAR_PLANILHA(dados)

Pseudocódigo das funções auxiliares:

FUNCAO OBTER_ENCONTROS:

Acessa site, obtém todos os **encontros**
RETORNA **encontros**

FUNCAO OBTER_GTS(encontro):

Acessa site, navega até o encontro, obtém seus **gts**
RETORNA **gts**

FUNCAO OBTER_TRABALHOS(gt):

Acessa site, navega até o **gt**, obtém seus **trabalhos**
RETORNA **trabalhos**

FUNCAO BAIXAR(trabalho):

Faz o download do **pdf** do trabalho
RETORNA **pdf**

FUNCAO CONVERTER(pdf):

Aplica **pdfminer** em **pdf** para obter texto
RETORNA **texto**

FUNCAO RASPAR(txt):

Remoção de textos fixos, acentos, caracteres não imprimíveis.
Raspa campos utilizando expressões regulares e outras ferramentas de mineração.
RETORNA **campos**

FUNCAO GERAR_PLANILHA(dados):

Produz uma planilha **excel** a partir de dados.

Consideramos importante explicitar esses passos do nosso trabalho de raspagem, pois isso viabiliza que o código possa ser replicado para materiais semelhantes. O código acima nos permitiu, então, coletar e organizar em uma planilha as principais informações disponíveis nos artigos analisados, a saber: nome do grupo de trabalho da Compós no qual o artigo foi apresentado; ano em que o artigo foi apresentado; local do encontro da Compós em que o artigo foi apresentado; nome do arquivo PDF referente àquele artigo; título do artigo; palavras-chave definidas no artigo; nome e instituição do autor do artigo; autores e títulos das obras referenciadas no artigo. A partir das planilhas geradas com todos esses dados coletados, partimos para a etapa de processamento, que será detalhada no próximo tópico.

4 Processamento dos dados

Como já foi mencionado, o trabalho com os textos da Compós nos retornou um grande número de dados que, ao mesmo tempo, nos ofereceram a base para um trabalho potencialmente profícuo e nos colocaram o desafio de como processar tal material. Nossa inexperiência em lidar com bancos de dados e metodologias quantitativas nos situou diante de uma busca, iniciada quase às cegas, por procedimentos adequados para a realização de tal tarefa.

Conforme apresentado em trabalho anterior (FRANÇA *et al.*, 2018), em um primeiro momento do projeto piloto sobre a área de Comunicação e Política, utilizamos o software *Statistical Package for the Social Sciences* (SPSS), contando com o auxílio de terceiros. A partir dele, fizemos processamentos básicos que nos informaram, por exemplo, frequências de citações a autores e obras, o que nos permitiu gerar os primeiros *rankings*. Contudo, esses *rankings* não eram precisos, o que exigia o cruzamento da frequência das referências com outras variáveis – como a autocitação, no caso do *ranking* de autores mais citados. Outras questões surgiram, como, por exemplo, o número de autores diferentes citados na área, que não coincide com o número de registros na planilha e requer a eliminação de duplicidades.

Essas questões aparentemente simples sinalizaram um problema mais complexo: ainda que munidos de um software capaz de processar nosso banco de dados, necessitávamos, primeiramente, conhecer seu escopo para que pudéssemos elaborar as perguntas que nos interessavam. Essa necessária adequação entre procedimentos e objetivos da pesquisa nos levou a rever e repensar o uso dos softwares estatísticos. Afinal, como aponta Braga (2011),

O processo metodológico básico não é o de definir uma regra de encaminhamento e depois segui-la estritamente; mas sim o de rever cada passo dado e refletir sobre a justeza de seu direcionamento, corrigindo-o no próprio andamento da pesquisa. Planejar é replanejar. (BRAGA, 2011, p. 10).

Nesse processo de replanejamento da pesquisa, buscamos a qualificação da própria equipe do projeto para o trabalho com tais softwares e nos foi apresentado o software Stata, com o qual passamos a trabalhar.

Para além do processamento complexo dos dados, dois outros objetivos podem ser destacados no aprendizado e na utilização do Stata. Primeiro, como já pontuado, a

compreensão das possibilidades do software e sua articulação com os objetivos da pesquisa, o que tornou possível uma leitura mais direcionada e objetiva do banco de dados, articulando variáveis relacionadas e tornando nossos cruzamentos mais complexos e precisos. O segundo foi criar, a partir de tal compreensão, um roteiro padrão das perguntas que desejaríamos responder em relação a cada uma das áreas pesquisadas. A partir daí, criamos um protocolo com tais perguntas e os códigos correspondentes para o processamento pelo Stata, de modo que potencialmente qualquer pessoa da equipe possa executar os mesmos passos e responder aos mesmos objetivos nas diferentes áreas.

O protocolo conta com onze perguntas, apresentadas a seguir, e cada pergunta envolve um conjunto de procedimentos e comandos para o processamento dos dados pelo Stata: (1) Quantos/as autores/as diferentes são citados/as como referência? (2) Quantos/as são os/as pesquisadores/as apresentadores/as? (3) Quais são os/as dez autores/as mais referenciados/as, sem eliminar autocitações? (4) Quais são os/as dez autores/as mais referenciados/as, eliminando as autocitações? (5) Qual é a frequência de autocitações entre os/as pesquisadores/as que são primeiros/as autores/as? (6) Qual é frequência de autocitação entre os/as pesquisadores/as que são coautores/as? (7) Quantos/as autores/as de referência são brasileiros/as e quantos/as são estrangeiros/as? (8) Quantos/as são membros do GT e quantos/as não são? (9) Quantas e quais são as obras mais citadas? (10) Dos/as dez autores/as mais citados/as, quais são suas obras citadas mais de uma vez, sem contar as autocitações? (11) Que pesquisadores/as citam esses/as autores/as e quantas vezes citam?

A criação desse protocolo não só nos permitiu chegar a dados mais precisos, através de métodos mais confiáveis, como dinamizou significativamente o trabalho. Entre 2015 e 2018, trabalhamos no projeto piloto da área de Comunicação e Política, no qual foram feitos os vários testes e encontrados os diversos desafios aqui apresentados. Entre 2018 e 2019, após todo o processo de criação, teste, adequação e aprimoramento do protocolo, foi possível concluir as seis áreas temáticas indicadas anteriormente, de modo que foram perceptíveis os resultados do aperfeiçoamento metodológico em termos de rapidez e rigor.

5 A importância de uma análise qualitativa

No que se refere à questão qualitativa da pesquisa, identificamos a necessidade de aprofundar o nosso olhar sobre os dados encontrados a partir do processamento nos

programas e avançar em sua análise. Compartilhamos da perspectiva de Minayo (2004), para quem “[...] qualquer pesquisa social que pretenda um aprofundamento maior da realidade não pode ficar restrita ao referencial apenas quantitativo.” (MINAYO, 2004, p. 28). Os autores evocados para determinadas teorias nos artigos, os conceitos abarcados nas obras dos autores mais citados em cada área e a maneira como os autores são mencionados nos artigos (endosso, contradição ou contraponto), são aspectos que mostram a necessidade de um aprofundamento e uma análise qualitativa.

Neste processo, tomamos os *rankings* de autores mais citados, optando por um recorte que considera as dez primeiras posições. Em casos de autores com a mesma frequência de citações, adotamos como critério de desempate a quantidade de artigos diferentes do GT que citam cada autor, considerando que isso permite ver a difusão do uso do autor no tempo e/ou entre pesquisadores. O autor citado em mais artigos pode ser considerado, assim, com maior potencial de influência na área.

Com os *rankings* aprimorados e definidos, os autores mais citados foram divididos entre a equipe do projeto, para leitura de suas principais obras (registrando sua perspectiva de trabalho, conceitos e/ou abordagens metodológicas) e apreciação da forma como foram referenciados pelos pesquisadores em seus artigos. Dessa maneira, foi possível mapear as principais teorias contemporâneas da Comunicação e suas interfaces.

Nessa etapa, contudo, cada área nos fez deparar com novas questões e desafios. Primeiro, com a percepção de que alguns autores com expressiva frequência no *ranking* aparecem predominantemente ou apenas nas referências bibliográficas, e não no conteúdo dos textos que os citam. Tal ocorrência nos demandou uma leitura ainda mais minuciosa dos trabalhos em questão, a fim de identificar se e de que maneira a perspectiva daquele autor se fazia presente no artigo analisado.

Uma outra questão em relação à interpretação dos dados e da compreensão da relevância de cada frequência dentro de seu universo foi a disparidade no número de citações das obras de um mesmo autor, dificultando identificar quais obras deveriam ser analisadas. Por exemplo: no ranking de autores mais citados, dois autores aparecem com 18 citações cada, distribuídas entre três obras. No primeiro caso, a distribuição é equânime, cada obra com seis ocorrências (ou com pequenas variações), o que faz com que, na parte qualitativa, todas elas sejam analisadas. No segundo caso, a distribuição é desigual: obra (1) com doze menções; obra (2) com três; obra (3) com uma menção. Neste caso, não é possível que as obras (2) e (3) sejam consideradas como tendo o mesmo peso para a área em questão

que a obra (1). Contudo, há grande variação entre autores, o que torna difícil a adoção de um critério objetivo para identificar que intervalo pode ser considerado significativo para que uma obra seja ou não considerada nas diferentes áreas temáticas.

Essa e outras questões deixam claro que, ainda que métodos quantitativos sejam fundamentais para mapear e organizar nosso *corpus* e fornecer as bases para nossas análises, os métodos qualitativos continuam desempenhando um papel essencial no desenho da pesquisa. A articulação entre eles é o grande desafio, mas também – acreditamos – um dos grandes méritos da pesquisa aqui apresentada.

Dando prosseguimento à etapa, cada membro da equipe produziu um texto contendo uma pequena biografia do autor que lhe foi atribuído, uma breve explanação de suas obras mais citadas, apresentando sua perspectiva teórica e/ou metodológica e seus conceitos-chave, além do modo como tais conceitos e teorias foram referenciadas nos artigos do GT que citam o autor. Por fim, esses materiais foram compilados em um único texto, buscando fazer uma espécie de raio-x das principais referências utilizadas em cada área temática analisada pela nossa pesquisa. Até o momento, buscamos compilar estes textos em artigos que tracem um mapa teórico de cada uma das áreas temáticas do campo da Comunicação.

6 Considerações finais

Este artigo teve por objetivo apresentar alguns desafios e descobertas no percurso metodológico construído ao longo de uma pesquisa que conjuga métodos quantitativos e qualitativos para mapear as teorias acionadas nas pesquisas no Campo da Comunicação. Desde o início, quando optamos pela Compós, encontramos desafios, como as diferenças de formatação entre os artigos, que dificultaram a raspagem e desconfiguraram a planilha, os nomes grafados de maneira incorreta, a quantidade de referências coletadas manualmente, entre outros. Mais do que concluir ou apontar caminhos **certos**, quisemos apresentar os processos de tomada de decisão, de planejamento e replanejamento da pesquisa, bem como o que aprendemos ao longo desse processo. Afinal, se “[...] a prática de pesquisa [...] [é] a atividade fundamental na produção do conhecimento.” (MINAYO, 2004, p. 10), acreditamos que o compartilhamento do percurso investigativo que desenhamos pode contribuir na produção de novos conhecimentos no Campo da Comunicação.

Como grupo de pesquisa e pesquisadores da área, tivemos de nos debruçar em um território pouco explorado da metapesquisa em comunicação e, como apontam Mattos e Oliveira (2018), construir juntos um modelo:

A metapesquisa, para além de um estudo bibliométrico, quer apontar elementos teórico-metodológicos que avancem em direção a um modelo de pesquisa, [...] uma possibilidade heurística que leve em conta a Comunicação e sua encruzilhada como saber (ciência). (MATTOS; OLIVEIRA, 2018, p. 17).

Levando em conta o que queríamos descobrir e as possibilidades e limitações encontradas, não tivemos muitos exemplos de pesquisas que combinavam métodos tal como feito em nosso percurso. Além disso, foi preciso parar e reorganizar várias vezes **como** daríamos seguimento às atividades.

Foram diversas etapas e muito conhecimento acumulado em grupo (partindo da coleta manual para planilha do Excel até alcançar o aprendizado sobre linguagens de programação e utilização de softwares) para chegarmos aqui. Entendemos que iniciar o projeto com uma metodologia específica que dê conta de todo o percurso é uma pretensão ilusória, porque é na medida do surgimento dos entraves e dificuldades que novos recursos metodológicos se mostram necessários.

Entendemos que a divulgação de novos olhares para os métodos digitais é a principal contribuição para o campo da Comunicação e das Ciências Sociais/Humanas de maneira geral. A adesão aos métodos digitais para dados provenientes de redes sociais e internet vem crescendo (VIMIEIRO; BARGAS, 2018), mas a especificidade desta pesquisa é trabalhar com dados textuais – provenientes dos artigos da Compós – de uma maneira híbrida desde a coleta até a análise. Esse modelo de pesquisa ganha em dinamismo pela rapidez com a qual os softwares estatísticos conseguem processar os dados e aumenta a capacidade do grupo de se debruçar sobre outros enfoques.

Além disso, é interessante incorporar o olhar estatístico, desconstruindo o estereótipo da área de Ciências Humanas contrária aos números, e nos apropriando do que eles informam, à nossa maneira, sem abandonar a análise qualitativa. Esses dois olhares, quando combinados, enriquecem a análise, potencializando não só os objetos de pesquisa, mas a forma como os analisamos. Nesse sentido, a pesquisa se soma a outros esforços contemporâneos que vêm pensando nessa articulação de procedimentos quantitativos e qualitativos nas Ciências Humanas e Sociais:

A pesquisa científica em Ciências Sociais foi tão contaminada pela vertente administrativa, centrada em dados quantitativos, que durante décadas não se pensava em alternativa. Metodologia, para muitos, virou sinônimo de Estatística. A crítica ao positivismo e o surgimento de correntes teóricas críticas e construcionistas levaram muitos pesquisadores a questionar o viés quantitativo, considerado redutor, afastado do mundo das práticas sociais. Já a pesquisa qualitativa era encarada com desconfiança pelos pesquisadores das chamadas ciências duras, por ser supostamente pouco representativa da realidade dos fatos. Mas a ideia de que qualitativo e quantitativo são mutuamente excludentes não se sustenta na prática: todos nós, pesquisadores, prestamos tributos a um e a outro método, que, eventualmente, podem ser articulados e usados de forma combinada. (KISCHINHEVSKY, 2016, p. 288).

O desenho que apresentamos no texto é decorrência dessa articulação e combinação de procedimentos para a análise de seis áreas temáticas: Comunicação e Política, Estudos de Televisão, Novas Mídias, Estudos de Jornalismo, Epistemologia da Comunicação e Comunicação e Sociabilidade. As análises quantitativa e qualitativa dos dados permitiram chegar a um mapeamento de autores e teorias acionados pelos pesquisadores brasileiros.

De modo geral, observamos que não há uma homogeneidade no campo da Comunicação; pelo contrário, identificamos uma diversidade de autores de diferentes campos de conhecimento (Sociologia, Filosofia, Ciência Política). Sociólogos importantes no cenário internacional apareceram citados em várias subáreas, como, por exemplo, Pierre Bourdieu, um dos autores mais citados em Comunicação e Política, Estudos de Jornalismo, Comunicação e Sociabilidade e Epistemologia da Comunicação; ou o filósofo Michel Foucault, referenciado em Novas Mídias, Comunicação e Sociabilidade, Epistemologia da Comunicação. Isso mostra a diversidade constitutiva do campo da Comunicação e a necessidade de continuar refletindo sobre seu processo de constituição em pesquisas como a que nos propomos a realizar.

Financiamento

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

Agradecimentos

Agradecemos às agências financiadoras o apoio ao desenvolvimento de nossas pesquisas.

Referências

ALVES, Jonatas Renan Extração de dados de artigos em PDF raspados do site compos.org.br. *In*: ALVES, Jonatas Renan. **GitHub: compos-data**. [S. l.]: GitHub, 2018. Disponível em: <https://github.com/jonatasrenan/compos-data>. Acesso em: 25 jul. 2018.

BRAGA, J. L. A prática da pesquisa em Comunicação: abordagem metodológica como tomada de decisões. **E-Compós**, Brasília, v. 14, n. 1, p. 1-33, 2011.

FRANÇA, V. R. V. *et al.* Tendências das teorias da Comunicação: mapeamento de campos teóricos contemporâneos. **Questões Transversais**, São Leopoldo, v. 4, n. 8, p. 57-67, 2016.

FRANÇA, V. R. V. *et al.* Comunicação e Política: um mapeamento de autores/as e teorias que alicerçam essa área no Brasil. **Compolitica**, Rio de Janeiro, v. 8, n. 2, p. 5-40, 2018.

KISCHINHEVSKY, M. Métodos de pesquisa qualitativa aplicada à comunicação radiofônica. *In*: MOURA, C. P.; LOPES, M. I. V. (org.). **Pesquisa em comunicação: metodologias e práticas acadêmicas**. Porto Alegre: EDIPUCRS, 2016. p. 277-294.

LINS, K. **Iniciando expressões regulares**. Rio de Janeiro: DevMedia, 2007. Disponível em: <https://www.devmedia.com.br/iniciando-expressoes-regulares/6557>. Acesso em: 21 jul. 2018.

MATTOS, M. Â.; OLIVEIRA, M. E. Uma mirada no passado para projetar os novos rumos da metapesquisa (2011-2016). *In*: ENCONTRO ANUAL DA COMPÓS, 27., 2018, Belo Horizonte. **Anais [...]**. Belo Horizonte: Pontifícia Universidade Católica de Minas Gerais, 2018. p. 1-20.

MINAYO, M. C. de S. **O desafio do conhecimento: pesquisa qualitativa em saúde**. 8. ed. São Paulo: Hucitec, 2004.

O'GRADY, Stephen. **The RedMonk programming language rankings**: January 2018. [S. l.]: RedMonk, 7 Mar. 2018. Disponível em: <https://redmonk.com/sograde/2018/03/07/language-rankings-1-18/>. Acesso em: 21 jul. 2018.

PDFMINER. **Python PDF parser and analyzer**. [S. l.], 28 Sept. 2014. Disponível em: <https://euske.github.io/pdfminer/>. Acesso em: 23 jul. 2018.

PSEUDOCÓDIGO. *In*: WIKIPÉDIA. [S. l.]: Wikipédia, 2018. Disponível em: <https://pt.wikipedia.org/wiki/Pseudoc%C3%B3digo>. Acesso em: 23 jul. 2018.

PYTHON. *In*: WIKIPÉDIA. [S. l.]: Wikipédia, 2018. Disponível em: <https://pt.wikipedia.org/wiki/Python>. Acesso em: 21 jul. 2018.

RASPAGEM de dados. *In*: WIKIPEDIA. [S. l.]: Wikipédia, 2018. Disponível em: https://pt.wikipedia.org/wiki/Screen_scraping. Acesso em: 21 jul. 2018.

SIMÕES, P. G. *et al.*. Estudos de televisão no Brasil: uma abordagem de autores/as e teorias. **Contemporânea**, Salvador, v. 17, n. 2, p. 1-23, 2019a.

SIMÕES, P. G. *et al.* Mapeando as novas mídias no Brasil. **Eco-pós**, [s. l.], 2019b. No prelo.

SPINK, M. J.; MENEGON, V. M. A pesquisa como prática discursiva: superando os horrores metodológicos. *In*: SPINK, M. J. (org.). **Práticas discursivas e produção de sentidos no cotidiano**: aproximações teóricas e metodológicas. 3. ed. São Paulo: Cortez, 2004. cap. III, p. 42-70.

STACK OVERFLOW. **Developer survey results 2018**. [S. l.], 2018. Disponível em: <https://insights.stackoverflow.com/survey/2018#technology-most-loved-dreaded-and-wanted-languages>. Acesso em: 21 jul. 2018.

VIMIEIRO, A. C.; BARGAS, J. A virada computacional nas pesquisas em comunicação. *In*: ENCONTRO ANUAL DA COMPÓS, 27., 2018, Belo Horizonte. **Anais [...]**. Belo Horizonte: Pontifícia Universidade Católica de Minas Gerais, 2018. p. 1-22.

Mapping the Communication field in Brazil: methodological challenges and findings of a metaresearch

Abstract

This article leans over the challenges of a research that maps the main theoretical references of the Communication field in Brazil, from the gathering and processing of papers published in the Associação Nacional dos Programas de Pós-Graduação em Comunicação (Compós), between 2006 and 2015. We describe the procedures of mining of the theoretical references of each paper presented in the GTs and the ways we found to process textual data through statistical tools, discussing the articulation between quantitative and qualitative methods. Lastly, we reflect on the methodological findings of an metaresearch on scientific fields. The paper demonstrates how the articulation between quantitative and qualitative procedures can be useful for the development of research in the Humanities and Social Sciences.

Keywords

Metaresearch. Quantitative methodology. Qualitative methodology. Scientific field.

Recebido em 10/08/2018

Aceito em 04/11/2019

Copyright (c) 2020 Paula Guimarães Simões, Vera Regina Veiga França, Ana Karina de Carvalho Oliveira, Laura Antônio Lima, Lucas Afonso Sepulveda Alves, Maira Lobato Bicalho Chagas Moura Campos, Paulo Henrique Basílio Santana. Creative Commons License. Este trabalho está licenciado sob uma licença Creative Commons Attribution-NonCommercial 4.0 International License. Os Direitos Autorais dos artigos publicados neste periódico pertencem aos autores, e os direitos da primeira publicação são garantidos à revista. Por serem publicados em uma revista de acesso livre, os artigos são de uso gratuito, com atribuições próprias, em atividades educacionais e não-comerciais.

