

18º Congresso Brasileiro de Sociologia

26 a 29 de Julho de 2017, Brasília (DF)

Grupo de Trabalho 39 – Sociologia Digital

Combinando *webscraping* em R e ATLAS.ti na pesquisa em ciências sociais: as possibilidades e desafios da sociologia digital

Leonardo F. Nascimento – UNILAB/PPGCS-UFBA



Brasília
2017

Combinando *webscraping* em R e ATLAS.ti na pesquisa em ciências sociais: as possibilidades e desafios da sociologia digital

Leonardo F. Nascimento¹

Resumo: A aquisição de habilidades computacionais na prática profissional do cientista social constitui um dos campos do que é denominado de Sociologia Digital. Atualmente estão disponíveis uma infinidade de ferramentas digitais que auxiliam na extração e montagem de banco de dados, na análise e interpretação desse material e, por fim, na organização e divulgação dos resultados. A comunicação apresentará um pequeno estudo de caso onde foram extraídos através de um script em R todos os artigos da Revista Brasileira de Ciências Sociais e depois realizadas análises automatizadas utilizando o software ATLAS.ti. O propósito é oferecer uma reflexão crítica sobre como essas ferramentas digitais podem vir a representar novas formas de estruturação da imaginação e da investigação sociológica exigindo avaliarmos “até onde é possível pensar sociologicamente” dentro das lógicas de programação que estruturam estes softwares.

A presente comunicação estará organizada da seguinte maneira. Eu inicio apresentando um estudo de caso onde foram extraídos, através de um script em R, todas as edições da Revista Brasileira de Ciências Sociais² (RBCS) e depois realizadas análises automatizadas utilizando o software ATLAS.ti. Em seguida, descrevo as potencialidades e problemas no uso destas ferramentas para os diferentes tipos de investigação da sociedade. Eu finalizo com algumas considerações sobre o futuro da pesquisa em sociologia em relação ao panorama proposto pela Sociologia Digital. O meu objetivo mais fundamental não é *apenas* apresentar um modo de processamento bibliométrico automatizado de dados digitais, mas de propor ao Grupo de Trabalho uma reflexão sobre a extração, análise e tratamento de dados sociais digitais onde estejam presentes as seguintes indagações: as TDIC's representam ou não novas formas de estruturação da imaginação e da investigação sociológica? Até onde é possível pensarmos sociologicamente dentro das lógicas de programação que estruturam estes softwares?

¹ Doutor em Sociologia, professor da Universidade da Integração Internacional da Lusofonia Afro-Brasileira – UNILAB e do Programa de Pós-graduação em Ciências Sociais da UFBA – PPGCS/UFBA. Email para contato: leofn3@gmail.com. Twitter: [@leofn3](https://twitter.com/leofn3). Website: <https://www.leofn.com>. Eu gostaria de agradecer aos estudantes de pós-graduação e de graduação que cursaram a minha disciplina Sociologia Digital e, em especial, a minha equipe de bolsistas PIBIC e orientandas: Ive Carvalho, Mylena Alecrim, Patrícia Cerqueira, Priscila Andreatta e Ludmila Silva.

² http://socialsciences.scielo.org/scielo.php?script=sci_serial&pid=0102-6909&lng=pt&nrm=iso acessado em 20 maio 2017.

1. Coletando os artigos da RBCS via *webscraping* em R

O primeiro passo foi montar um script³ na linguagem de programação R⁴ para fazermos o download de todos os artigos e resenhas da Revista Brasileira de Ciências Sociais (RBCS) que encontra-se disponível no meu GitHub pessoal⁵. O script foi feito utilizando um pacote de funções específicas para o R denominado XML⁶. Esse pacote permite ao R acessar websites e extrair os dados em páginas nos formatos HTML. O formato HTML é uma sigla para *HyperText Markup Language*, trata-se de um padrão oculto atrás de quase tudo o que vemos e fazemos ao navegar na Web (MUNZERT et al., 2015, p. 19). Podemos dizer que o script criado tem como principal função extrair grandes quantidades de dados na web. Tal procedimento vem se popularizando no meio científico em geral e é denominado de *webscraping* (literalmente, raspagem de dados na web), *web mining* ou ainda *web harvesting*. Em poucas palavras, podemos definir o *webscraping* como “uma técnica proeminente para a coleta automatizada de dados on-line” (MARRES; WELTEVREDE, 2013).

É importante perguntar porque utilizarmos um script para retirar informação de websites se podemos simplesmente copiar e colar estas mesmas informações? Em suma, por que fazer *webscraping*? Vejamos isso com calma. A maioria dos navegadores não oferecem a funcionalidade de salvar os dados que são apresentados nos sites para, digamos, um uso pessoal. Por exemplo, para salvar uma tabela de dados de algum site e, em seguida, usarmos em um editor de planilhas (Excel), nós nos habituamos a copiar e colar manualmente aquilo que queremos. Além disso, quando precisamos fazer o download de documentos em um site com dezenas de links, nós somos obrigados a fazer isso um por um, cada link de uma vez. Este trabalho de copiar e colar e fazer múltiplos downloads constitui uma tarefa repetitiva e tediosa que os pesquisadores geralmente – e sem nenhuma expressão de remorso – relegam aos heroicos bolsistas de pesquisa.

³ Um script é uma coleção de tarefas escritas em uma determinada linguagem de programação para serem executadas de maneira automatizada e mais veloz que a execução por um operador humano.

⁴ <https://www.r-project.org/>

⁵ https://github.com/leofn/Digital-Sociology/blob/master/SciELO_RBCS_SBS-2017.R

⁶ <https://cran.r-project.org/package=XML>

Na tentativa de responder à pergunta do parágrafo anterior, podemos dizer que as ferramentas de *webscraping* nos possibilitam poupar um tempo precioso que podemos dedicar a outros aspectos da pesquisa como, por exemplo, a análise dos dados e materiais coletados. Além disso, a automação dos downloads evita a possibilidade de erros por distração, cansaço ou possível sabotagem de algum bolsista precarizado. Brincadeiras à parte, eu gostaria de afirmar que *o uso do webscraping na pesquisa em sociologia constitui uma das mais poderosas ferramentas para a obtenção de dados acerca da sociedade pois cada vez mais partes significativas dos “materiais” de que somos feitos – predileções e aversões, felicidades e sofrimentos, atos de linguagem e expressões de afetos além de música, filmes, livros, etc. – estarão disponíveis online*. Vejamos, no exemplo descrito a seguir, se podemos concordar sobre a afirmação anterior.

A Revista Brasileira de Ciências Sociais é uma publicação da ANPOCS - Associação Nacional de Pós-Graduação e Pesquisa em Ciências Sociais e se intitula “o mais importante periódico acadêmico no campo das ciências sociais no Brasil, desde sua criação em junho de 1986”⁷. Apesar da relevância, a revista foi escolhida de maneira completamente arbitrária segundo o tempo disponível para a feitura da presente comunicação. Poderíamos ter feito o mesmo exercício com outro periódico científico da base Scielo ou ainda, sem nenhum esforço humano adicional, com toda a base Scielo para as Ciências Sociais, isto é, para dizer de modo claro, com todos os artigos em periódicos das ciências sociais publicados no portal Scielo. Esta última possibilidade demandaria apenas uma maior capacidade de processamento e armazenamento das dezenas de *terabytes* dos artigos, resenhas e ensaios que seriam baixados.

Se os leitores executarem o script⁸, serão obtidos por download o total de 877 documentos (entre resenhas e artigos) no formato pdf hospedados na plataforma Scielo de 1997 a 2017, especificamente do volume 12, número 35 até o volume

⁷ <http://submission.scielo.br/index.php/rbcsoc/about/history>

⁸ Para a execução eu recomendo o uso do RStudio uma IDE (do inglês *Integrated Development Environment* ou Ambiente de Desenvolvimento Integrado) para a linguagem R.

32, número 94⁹. Notem que, para o presente objetivo, não nos importou nomear os materiais baixados com exatidão, pelo título/autor ou mesmo pelo DOI¹⁰. O script disponibilizado poderá ser modificado posteriormente pelo autor ou pelos leitores a fim de extrair somente uma daquelas modalidade de documentos ou padronizar a nomenclatura dos arquivos. Uma vez que tenhamos todo o material intelectual em formato digital produzindo em 20 anos da RBCS (1997-2017), utilizaremos uma segunda ferramenta digital para a análise dos materiais: o software ATLAS.ti.

2. Analisando os dados da RBCS através do ATLAS.ti e da função *autocoding*

2.1. O software ATLAS.ti

O ATLAS.ti é um dos softwares comerciais de código fechado mais utilizados no mundo acadêmico para análises qualitativas de dados. O ATLAS.ti surgiu a partir de um problema de pesquisa. Em 1986, após o desastre nuclear de Chernobil, o prof. Heiner Legewie da *Technische Universität Berlin* passou a fazer entrevistas em profundidade com sobreviventes e técnicos. Foram realizadas 60 entrevistas resultando em mais de 4.000 páginas de transcrições (LEGEWIE, 2014). Os pesquisadores não queriam fazer uma análise automatizada através de algoritmos de análise linguística disponíveis àquela época. Segundo descreveu Legewie, “eles estavam profundamente convencidos de que apenas falantes humanos competentes seriam capazes de compreender e analisar as notas de campo e as transcrições de entrevistas.” (LEGEWIE, 2014, p. 2). Sendo assim, eles optaram por analisar o material através da *Grounded Theory* (Teoria Fundamentada nos Dados), teoria desenvolvida pelos pesquisadores Barney Glaser e Anselm Strauss (GLASER; STRAUSS, 1966, 2009). Um deles, o professor Anselm Strauss, ajudou no desenvolvimento do software e acabou se tornando amigo de Legewie.

⁹ http://www.scielo.br/scielo.php?script=sci_issues&pid=0102-6909&lng=pt&nrm=iso

¹⁰ https://pt.wikipedia.org/wiki/Digital_object_identifier

Com o ATLAS.ti é possível codificar¹¹ livros, artigos, entrevistas, músicas, filmes, quadros, fotos, websites, em suma, praticamente todo e qualquer material de pesquisa e/ou trabalho nos mais diversos campos ou áreas. Tudo isto é possível, obviamente, desde que estes materiais tenham sua origem em dispositivos digitais ou tenham sido convertidos para o formato digital. O ATLAS.ti faz parte de uma classe de aplicativos agrupados sob o nome de CAQDAS, acrônimo de língua inglesa *Computer Assisted Qualitative Data Analysis* ou Análises de Dados Qualitativos auxiliada por Computador. Embora esteja descrito deste modo, o ATLAS.ti não se restringe às “análises qualitativas”, podendo gerar planilhas com frequência de códigos e contagem de palavras por documentos, sendo mais adequado considerá-lo uma plataforma para os denominados métodos mistos de pesquisa (CRESWELL; CLARK, 2015).

É indispensável sublinhar que o ATLAS.ti não substitui, em hipótese alguma, o trabalho do pesquisador. E que quando falamos de codificação nós estamos diante de três elementos fundamentais: 1) um documento (nos formatos de áudio, texto, vídeo ou imagem) a ser analisado; 2) um rótulo ou código que é atribuído a trechos selecionados dos documentos; e 3) temos como resultado, ao final de muitas codificações, um conjunto de trechos ou excertos dos documentos que expressam ideias ou aspectos importantes do(s) documento(s).

Todo esse *procedimento de atribuição de códigos ou codificação constitui um procedimento fundamentalmente hermenêutico*. A codificação, portanto, está baseada em processos interpretativos derivados das teorias sociológicas e da progressiva análise e intimidade com os dados. Nesse *intermezzo* precisam ser consideradas a pergunta da pesquisa, a construção do objeto e todas as etapas consagradas à investigação científica de caráter sociológico. Apesar do aspecto hermenêutico da codificação, o ATLAS.ti nos oferece algumas funcionalidades automatizadas que eu gostaria de explorar aqui. Acredito que essas funções, ao serem utilizadas em *combinação* com os processos

¹¹ “Um código na pesquisa qualitativa é, na maioria das vezes, uma palavra ou frase curta que atribui simbolicamente um atributo que resume, salienta, captura a essência e/ou evoca de uma porção de dados visuais ou baseados na linguagem.” (SALDANA, 2015, p. 3)

interpretativos dos documentos, podem acelerar análises e descobrir padrões e/ou regularidades nos dados.

2.2. A função *autocoding*

Uma das funcionalidades automatizadas mais interessantes do ATLAS.ti é o *autocoding*. Ele pode ser definido como a capacidade de fazermos buscas por palavras-chave em todo o material em formato texto (txts, docs, docxs, pdf, etc.) desde, obviamente, que os caracteres estejam reconhecidos ou, para dizer de uma maneira mais técnica, que o(s) documento(s) (especialmente os pdfs) tenha(m) passado por um reconhecimento ótico de caracteres ou OCR.

Desde a primeira vez em que fui apresentado, ainda em 2010, ao *autocoding*, eu fiquei intrigado com as possibilidades e limites desta função. Sobretudo porque, como veremos mais adiante, ele se diferencia de uma busca simples presente nos editores de texto (Ctrl + L do Word) ou dos navegadores (Ctrl + F do Chrome ou Firefox). Vejamos quatro parâmetros do *autocoding* antes de apresentarmos os resultados da aplicação aos materiais da RBCS.

O primeiro deles é que o *autocoding* funciona *intra* e *inter* documentos, por exemplo: podemos buscar, em poucos segundos, todas as citações de Nietzsche nas Obras Completas de Sigmund Freud ou apenas em um dos livros da coleção e ainda, fazermos o rastreamento do conceito de Violência Simbólica em um livro ou em toda a obra Pierre Bourdieu. Os limites da busca dependem exclusivamente da quantidade de material que você tenha no formato digital (pois muitos livros não estão, de fato, no formato digital). Imaginemos, por exemplo, que um pesquisador esteja interessado na influência da fenomenologia em toda a Teoria Social contemporânea (e que ele disponha das obras de Norbert Elias, Pierre Bourdieu, Anthony Giddens, etc. em formato digital). Nós poderíamos buscar por todos os sobrenomes dos principais autores vinculados à fenomenologia como um ponto de partida para nossas análises.

O segundo parâmetro, bem conhecido entre os programadores, é a possibilidade de utilizarmos expressões regulares ou GREP (sigla para **G**lobally search a

Regular Expression and Print). De modo resumido, podemos dizer que o GREP faz buscas no conteúdo textual dos documentos através de caracteres especiais. Por exemplo, se eu quiser codificar todos os trechos de livros e artigos com o código “Autores da Fenomenologia” toda vez que aparecerem os nomes de “Husserl”, “Merleau-Ponty” e “Sartre”, eu posso, utilizando o *autocoding* com GREP, fazer uma busca utilizando o caracter “|” que significa *or* (ou) da seguinte forma: Husserl|Merleau-Ponty|Sartre. Ou ainda utilizar colchetes “[]” para as variações de grafias nos nomes dos autores, por exemplo, S[ø]ren Kierkegaard¹².

O terceiro parâmetro diz respeito à extensão do trecho codificado. O *autocoding* do ATLAS.ti permite, uma vez que o aplicativo detecta o(s) termo(s) que estamos à procura, que escolhamos a extensão da codificação, podendo abarcar desde o trecho exato (*exact match*), a palavra (*word*), a linha (*sentence*), o parágrafo (*paragraph*) ou todo o texto (*all text*). Imagine que nós pretendemos fazer um *survey* com todos os autores da revista RBCS, para isso podemos mandar um link questionário do *Google Forms* para os e-mails de toda(o)s a(o)s autoras(es). Como consegui-los? Nós poderíamos buscar aquilo que se repete em todos os e-mails, e apenas neles, o arroba “@” e codificar com “Emails_Autores”. Agora se nós quisermos o trecho exato (*exact match*) o programa irá codificar apenas o arroba “@” e isso não nos serve de nada.

meuemail@email.com.br

Se escolhermos, por outro lado, a linha (*sentence*), nós corremos o risco de codificar por exemplo,

Unicamp, Campinas – SP, Brasil. E-mail: meuemail@email.com.br

O que vai nos gerar um trabalho tremendo de deletar a informação inútil. Agora, por fim, se escolhermos a palavra (*word*) ele pegará os caracteres para um lado

¹² Embora Kierkegaard não pertença à fenomenologia eu utilizei ele apenas para facilitar o exemplo.

e para outro até o espaço em branco e teremos exatamente a informação que queremos:

`meuemail@email.com.br`

É comum que “falsos positivos” sejam codificados, mas como podemos ver na base que foi disponibilizada, a taxa de acerto dos e-mails foi bastante elevada.

O quarto e último parâmetro representa algo extremamente útil para a pesquisa em toda e qualquer área, científica ou não. Uma vez feita a codificação – automatizada via *autocoding* ou através do modo tradicional – de trechos de texto, áudio ou vídeo ou de um detalhe de dados em imagem, é possível gerarmos um documento que contenha todos os trechos codificados. Um relatório das codificações indicando o nome documento, o autor, a data da criação e possíveis comentários feitos no trecho codificado. Isso constitui, do ponto de vista intelectual, uma ampliação da capacidade mnemônica sem precedentes. Eu costumo exemplificar em sala de aula: imaginem uma graduação em que todas as leituras, de todos os semestres e disciplinas tenham sido realizadas em livros e artigos digitais e dentro do ATLAS.ti. A nossa capacidade de sintetizar autores, genealogias e/ou arqueologias de conceitos não teria sido, em alguma medida, modificada? Isso não teria uma relação com a nossa capacidade intelectual? Vejamos agora alguns resultados que encontramos.

2.3. Alguns achados preliminares da base de dados da RBCS

De posse dos 877 documentos em pdf vamos agora fazer algumas análises dentro do ATLAS.ti utilizando o *autocoding*. Mesmos as análises automatizadas de palavras através do *autocoding* necessitam de alguma pergunta inicial que seja direcionada aos dados. Imaginemos, portanto, que um pesquisador tenha o interesse em avaliar a presença e/ou o uso dos conceitos da obra de Norbert Elias no pensamento sociológico brasileiro. E que, obviamente, ele deseje fazer isso não de modo hagiográfico baseado em opiniões pessoais constituídas muitas vezes de maneira arbitrária e errática, mas sim baseado em dados

concretos. Nada mais coerente que começar investigando o “mais importante periódico acadêmico no campo das ciências sociais no Brasil”, a RBCS. Vejamos o alcance dos resultados conjugando teoria social e metodologia digital que encontramos.

O primeiro passo foi inserir os 877 documentos no ATLAS.ti criando uma unidade hermenêutica (*Hermeneutical Unity*) intitulada sbs_rbc_s_2017.hpr7¹³. Criamos inicialmente um código “Menções_Norbert_Elias” e fizemos um *autocoding* em todos os documentos buscando todos os parágrafos que contivessem o termo “Elias”. Geramos um relatório com todas as citações e muito rapidamente eliminamos os falsos positivos (por exemplo, menções a “Elias Canetti”, etc.) e, por fim, obtivemos um total de 296 menções de Elias em um total de 72 documentos que compreendiam 62 artigos e 10 resenhas. Atentando apenas para os artigos, podemos perceber no gráfico 1 abaixo os artigos (nomeados segundo sua inserção no ATLAS.ti de P1 a P887) que mais fizeram menções a Norbert Elias:

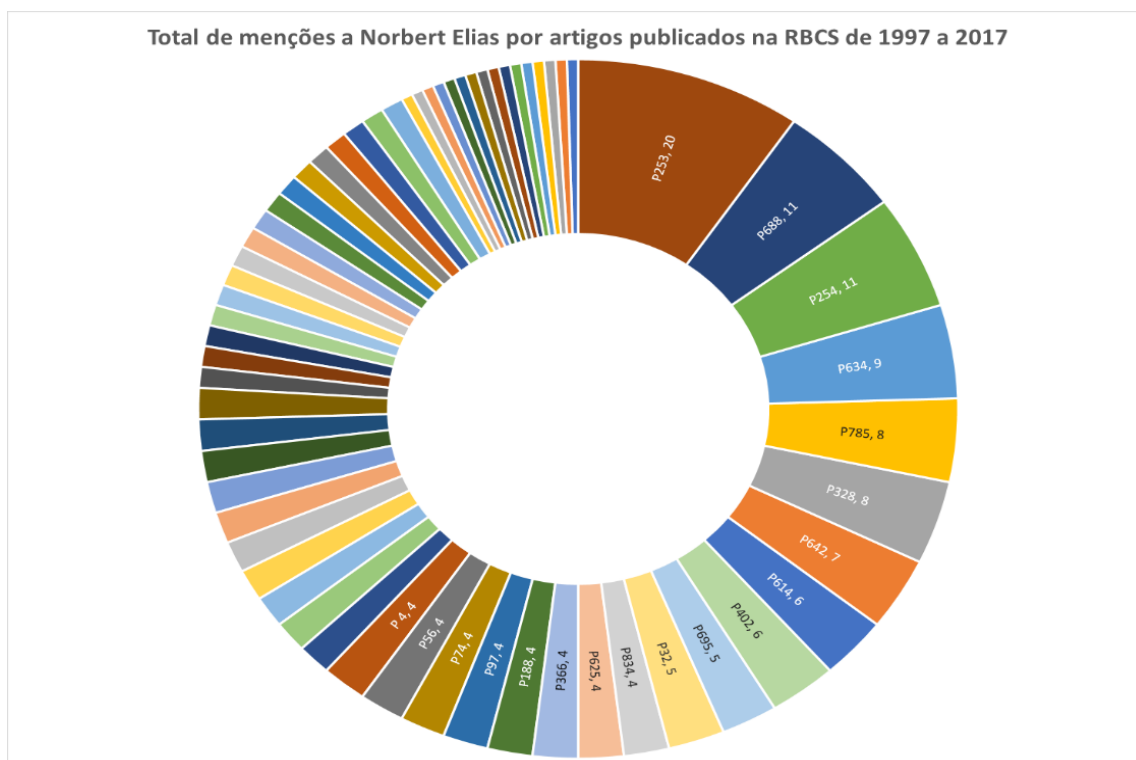


Gráfico 1

¹³ Os leitores poderão baixar uma “cópia pacote” (*copy bundle*) desta Unidade Hermenêutica com todos os documentos e codificações realizadas para este artigo no seguinte link: https://www.dropbox.com/s/hf77124l8mwbaur/sbs_rbc_s_2017.atlcb?dl=0

Uma vez que detectamos quais artigos com mais menções, todo um mundo de possibilidades analíticas se encontra a disposição. Vejamos, por exemplo, segundo o gráfico 1 acima, os quatro documentos que mais possuem menções a Elias: P253 (BRITO; MORAIS; BARRETO, 2011), P688 (SOUZA, 2001), P254 (COELHO et al., 2016) e P634 (PONTES, 2003).

O P253 é um artigo sobre sociologia do esporte em que os autores buscam uma “explicação sociológica da ideia de fair play”. Sabemos que Elias possui um trabalho que se tornou clássico sobre a sociologia do esporte (ELIAS, 1992), e, obviamente, o artigo faz diversas citações desta obra além de outras. O segundo, o P688, é um artigo sobre Roberto da Matta que lança mão de aspectos importantes da argumentação eliasiana, como a formação do estado, a economia dos afetos e o processo de construção do gosto entre as classes. Neste artigo, iremos encontrar nas referências apenas a obra clássica de Elias “O processo civilizador” (ELIAS, 2008) em sua versão alemã. O artigo P254 trata da autoridade policial nas situações de revista de bagagem e nas *blitz* da Lei Seca e faz um debate direto com a obra “Os estabelecidos e os outsiders” (ELIAS; SCOTSON, 2000) estudo clássico de sociologia urbana realizado por Norbert Elias e John Scotson. Por fim, o artigo da Heloísa Pontes (P634) também utiliza o modelo “estabelecidos-outsiders” para refletir sobre a posição da intelectualidade nova-iorquina entre 1930 e 1950.

O leitor criticamente inspirado poderia pensar com total razão: a quantidade de vezes que o termo “Elias” aparece nos artigos não me informa nada sobre o *uso teórico* da obra de Norbert Elias. Um autor poderia citar uma ou duas vezes Elias e, no entanto, ele ser o mote fundamental da argumentação do artigo. Peço aos possíveis leitores que sejam generosos em relação ao cerne do meu argumento: *o ATLAS.ti vai permitir que mergulhemos profundamente nos dados de modo qualitativo embora quantitativamente informado*. Acredito, por exemplo, que não se trata de algo ordinário que, entre 1997 e 2017, apenas 8,2 % das publicações (entre resenhas e artigos) da RBCS citam, ao menos uma vez, Norbert Elias. Das cinco resenhas, duas delas, P461(LANDINI, 2007) e P852 (WAIZBORT, 1998) eram sobre publicações de Norbert Elias. Além disso, a despeito da infestação da Teoria Sociológica nas graduações e pós-graduações, nenhum

artigo da RBCS tratou especificamente de Elias ou do desdobramento dos seus conceitos.

Um último dado que certamente vai interessar ao leitor é o percurso temporal do uso de Norbert Elias expresso pela frequência de menções ao longo dos anos das publicações. Através do *autocoding* buscamos todos os anos de publicação dos artigos e fizemos uma contagem simples por ano de publicação que resultou no gráfico 2 abaixo:

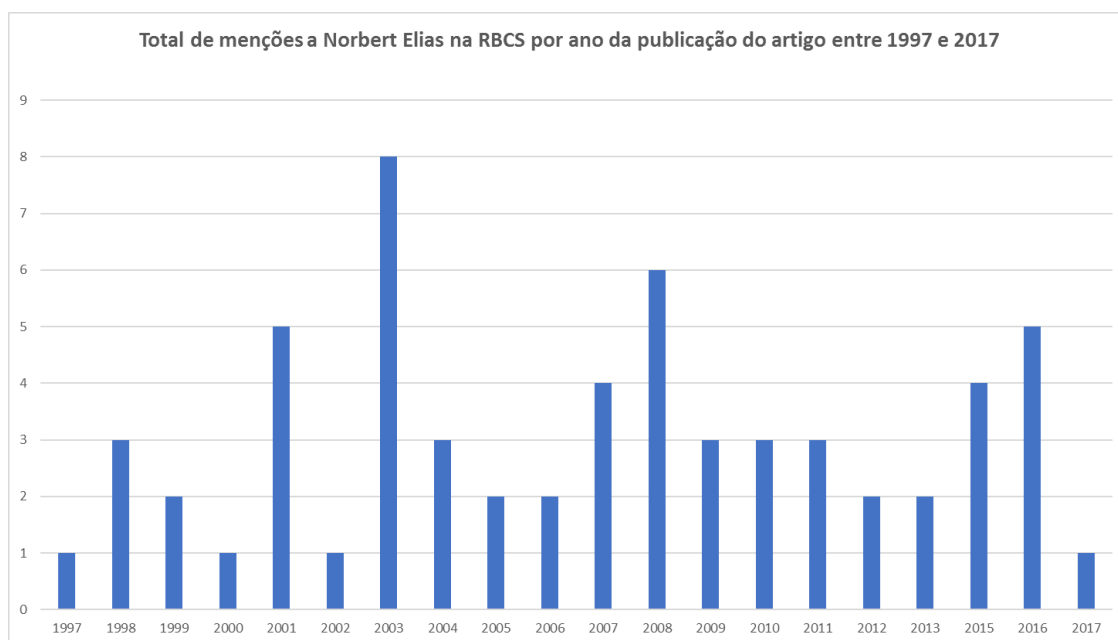


Gráfico 2

A título de curiosidade, nos idos de 1997, o primeiro artigo que cita Elias intitula-se “Exclusão e políticas públicas: dilemas teóricos e alternativas políticas” da profa. Alba Zaluar. Nele encontraremos a seguinte citação que se refere ao livro de “A busca da excitação” (ELIAS, 1992):

As gratificações no plano do simbólico, da auto-estima, da luta pela notoriedade, da disputa na qual se liberam as agressividades, a ostentação do poder e da riqueza, ou a busca da justificação, que chamam a atenção dos pesquisadores, continuam a aparecer, porém com regras explícitas que levam ao que Norbert Elias (1993) denominou de "equilíbrio de tensões" em lutas prolongadas, porém controladas por regras convencionadas. (ZALUAR, 1997, p. 10)

Para concluir, gostaria de dizer que as análises que foram realizadas utilizando o ATLAS.ti poderiam ser desdobradas em diversas direções. Seria possível, por

exemplo, através de co-ocorrência de códigos, rastrear os principais conceitos utilizados nos artigos, as obras mais citadas, os relacionamentos com outros autores, etc. E um aspecto fundamental é que – a despeito da intimidade do autor com o aplicativo – as análises aqui apresentadas foram feitas em uma única tarde com cinco horas ininterruptas de trabalho. Como foi mencionado acima, alguém interessado poderia expandir as análises realizadas para toda a base do Scielo o que proporcionaria um panorama amplo de como ocorre o debate entre a tradição sociológica brasileira e a obra de determinado autor (em nosso exemplo, Norbert Elias)

3. Possibilidades e desafios da Sociologia Digital

Temos mais de 15 anos de artigos e livros afirmando que a Sociologia estaria ficando obsoleta, necessitando, urgentemente, fazer uma atualização do modo como investiga e reflete acerca das dinâmicas sociais (ABBOTT, 2000; SAVAGE; BURROWS, 2007; WITTE, 2012; BURROWS; SAVAGE, 2014; LUPTON, 2015; MARRES, 2017). Trata-se de uma percepção que foi se formando gradativamente no pensamento de autores de distintas tradições e que alcança uma conjuntura dramática pois, segundo todos eles, o campo dos problemas propriamente sociológicos foi alterado substancialmente pelas tecnologias digitais.

Tudo parece indicar que as ferramentas tradicionais de pesquisa como *surveys* e entrevistas parecem não ser suficientes para acompanharmos a vastidão de dados e metadados sobre processos sociais que são produzidos na *Word Wide Web*. Todos os dias, voluntariamente e ininterruptamente, milhões de indivíduos deixam rastros digitais nas redes sociais e *microbloggings* (Facebook e Twitter) e nos *instant messengers* (Whatsapp, Telegram, etc.) (SAVAGE; BURROWS, 2007; BURROWS; SAVAGE, 2014). Por outro lado, a produção acadêmica está passando por uma mudança radical. Websites como *ResearchGate*, *Academia.edu*, *LinkedIn*, etc. e o forte emprego de bibliometria e indexação da produção, tal como vimos na análise dos artigos da RBCS, modificam o cenário de como pensamos a produção e divulgação de nossas pesquisas.

Se a produção do conhecimento depende, entre outras coisas, da capacidade de “fazermos referência” à informação anteriormente produzida, as mudanças no modo como acessamos e recuperamos a informação não constituiria, como sugere o historiador Roger Chartier, “uma mutação epistemológica fundamental” (CHARTIER, 2002, p. 108)? Se este for o caso, por quais mudanças estariam passando o modo como pensamos as sociedades em vista da progressiva “smartificação” e/ou “dataficação” do eu na vida cotidiana? Poderia a sociologia ficar incólume a todo esse processo?

Tendo em vista este panorama de questões, acredito que não seria suficiente apenas indagar se “a sociologia está preparada em termos teóricos e metodológicos para investigar e compreender as mudanças proporcionadas pelas tecnologias digitais?”(NASCIMENTO, 2016). Já está suficientemente claro que não. Talvez o que precisamos – enquanto um campo que reivindica alguma especificidade na compreensão interpretativa e na explicação causal dos problemas da sociedade – é encarar os desafios de um *upgrade* dos nossos debates e técnicas de pesquisa. Tal atualização vem conduzindo pesquisadores ao novo campo da sociologia digital.

Segundo a proposta da Sociologia Digital (PRIOR; ORTON-JOHNSON, 2013; CARRIGAN, 2014; LUPTON, 2015; DANIELS; GREGORY; COTTOM, 2016; MARRES, 2017), esse upgrade deve ser feito atentando para um duplo aspecto: a) todos os temas clássicos da pesquisa em sociologia tiveram sua dinâmica impactada, em alguma medida, pelas TDIC's (LUPTON, 2015, p. 10); b) os procedimentos metodológicos da sociologia precisam ser digitalizados sem perdermos de vista toda a crítica incessante que sempre devotamos aos nossos instrumentos de pesquisa. Por este motivo, acredito que temos a urgente necessidade de disciplinas e espaços de discussão que tenham como objetivo instrumentalizar digitalmente estudantes, professores e pesquisadores. Isso incluiria o desenvolvimento de habilidades em estatística, em manejo de bases de dados, em softwares de análise qualitativa e, sobretudo, em programação.

Deste modo, acredito que seja urgente debatermos a flexibilidade interpretativa (BIJKER; HUGHES; PINCH, 1987; BIJKER; LAW, 1992; BIJKER, 1997) do uso das ferramentas digitais de análise para compreendermos as sociedades. Muitas pesquisas já usam algumas das ferramentas descritas nesta comunicação e muitos pesquisadores, propositalmente ou não, omitem tal utilização. Eles acreditam que a extração dos dados e os processos automatizados constituem um assunto de programadores nada relacionado com o pensamento sociológico. Por este motivo eles deixam de descrever como sua reflexão estava suportada por mecanismos digitais.

No exemplo que vimos acima, os possíveis erros na extração por técnicas de *webscraping* em R e os possíveis *bugs* do ATLAS.ti na análise dos dados digitais coletados seriam erros tecnológicos *stricto sensu* ou seriam erros metodológicos do pesquisador em sociologia? E se essa última pergunta nos induzisse a erro, pois talvez não pudéssemos mais pensar em reflexão sociológica dissociada da tecnologia que proporcionou essa mesma reflexão? E se o nosso debate, por conta do uso de dados sociais e tecnologias de análise ambos digitais, passou de sociológico a *sócio-tecno-lógico*? Estas são algumas das indagações que gostaria de deixar em aberto.

Para finalizar gostaria de dizer que não se trata aqui, em hipótese alguma, de eliminar os métodos ou as “habilidades analógicas” (LÖFGREN, 2014) e tradicionais da sociologia em favor de uma cornucópia de ferramentas e métodos digitais. Mas sim de combiná-los ou, antes disso, de perguntar diante de nossas pesquisas: por que não os combinar? Por outro lado, acho que o debate sobre a digitalização das práticas de pesquisa passa necessariamente pela questão mais ampla da digitalização do cotidiano que nos referimos mais acima. Talvez o “digital”, da “sociologia digital” seja mais que um novo subcampo que busca angariar novos adeptos. A capilaridade semântica do termo digital – a amplitude e a velocidade de sua difusão – parece ser diretamente proporcional aos efeitos da digitalização na organização e na estrutura das sociedades contemporâneas.

Nesse contexto a digitalização da sociologia, mais que uma escolha, poderá a vir se tornar um verdadeiro imperativo.

Referências Bibliográficas

ABBOTT, A. Reflections on the Future of Sociology. **Contemporary Sociology**, v. 29, n. 2, p. 296, mar. 2000.

BIJKER, W. E. **Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change**. [s.l.] MIT Press, 1997.

BIJKER, W. E.; HUGHES, T. P.; PINCH, T. J. **The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology**. [s.l.] MIT Press, 1987.

BIJKER, W. E.; LAW, J. **Shaping Technology/building Society: Studies in Sociotechnical Change**. [s.l.] Brooks/Cole, 1992.

BRITO, S. M.; MORAIS, J. V. de; BARRETO, T. V. Regras de jogo versus regras morais: para uma teoria sociológica do fair play. **Revista Brasileira de Ciências Sociais**, v. 26, n. 75, p. 133–146, fev. 2011.

BURROWS, R.; SAVAGE, M. After the Crisis? Big Data and the Methodological Challenges of Empirical Sociology. **Big Data & Society**, v. 1, n. 1, p. 2053951714540280, 1 abr. 2014.

CARRIGAN, M. **An Agenda for Digital Sociology**The **Sociological Imagination**, 30 dez. 2014. Disponível em: <<http://sociologicalimagination.org/archives/16568>>. Acesso em: 14 abr. 2015.

CHARTIER, R. **Os desafios da escrita**. [s.l.] UNESP, 2002.

COELHO, M. C.; SENTO-SÉ, J. T.; FERNANDES, R. B.; RIOS, F.; COELHO, M. C.; SENTO-SÉ, J. T.; FERNANDES, R. B.; RIOS, F. A EXPERIÊNCIA DA SUJEIÇÃO À AUTORIDADE POLICIAL: NOTAS SOBRE A ARTICULAÇÃO ENTRE COGNIÇÃO E EMOÇÃO NA VIDA PÚBLICA. **Revista Brasileira de Ciências Sociais**, v. 31, n. 90, p. 151–166, fev. 2016.

CRESWELL, J. W.; CLARK, V. L. P. **Pesquisa de Métodos Mistos - 2.ed.: Série Métodos de Pesquisa**. [s.l.] Penso Editora, 2015.

DANIELS, J.; GREGORY, K.; COTTOM, T. M. (ed.). **Digital Sociologies**. Reprint edition ed. [s.l.] Policy Press, 2016.

ELIAS, N. **A busca da excitação**. [s.l.] Difel, 1992.

ELIAS, N. **O Processo Civilizador 1**. [s.l.] Jorge Zahar, 2008.

ELIAS, N.; SCOTSON, J. **Os Estabelecidos e os Outsiders**. [s.l.] Zahar, 2000.

LANDINI, T. S. Escritos & ensaios: Norbert Elias em perspectiva. **Revista Brasileira de Ciências Sociais**, v. 22, n. 64, p. 169–173, jun. 2007.

LÖFGREN, O. Routinising research: academic skills in analogue and digital worlds. **International Journal of Social Research Methodology**, v. 17, n. 1, p. 73–86, 2 jan. 2014.

LUPTON, D. **Digital Sociology**. [s.l.] Routledge, 2015.

MARRES, N. **Digital Sociology: The Reinvention of Social Research**. [s.l.] John Wiley & Sons, 2017.

MARRES, N.; WELTEVREDE, E. Scraping the Social? **Journal of Cultural Economy**, v. 6, n. 3, p. 313–335, 1 ago. 2013.

MUNZERT, S.; RUBBA, C.; MEIßNER, P.; NYHUIS, D. **Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining**. [s.l.] John Wiley & Sons, 2015.

NASCIMENTO, L. F. A Sociologia Digital: um desafio para o século XXI. **Sociologias**, v. 18, n. 41, p. 216–241, abr. 2016.

PONTES, H. Cidades e intelectuais: os “nova-iorquinos” da Partisan Review e os “paulistas” de Clima entre 1930 e 1950. **Revista Brasileira de Ciências Sociais**, v. 18, n. 53, p. 33–52, out. 2003.

PRIOR, N.; ORTON-JOHNSON, K. **Digital Sociology: Critical Perspectives**. [s.l.] Palgrave Macmillan, 2013.

SALDANA, J. **The Coding Manual for Qualitative Researchers**. [s.l.] SAGE, 2015.

SAVAGE, M.; BURROWS, R. The Coming Crisis of Empirical Sociology. **Sociology**, v. 41, n. 5, p. 885–899, 1 out. 2007.

SOUZA, J. A SOCIOLOGIA DUAL DE ROBERTO DA MATTA: Descobrimos nossos mistérios ou sistematizando nossos auto-enganos? **Revista Brasileira de Ciências Sociais**, v. 16, n. 45, p. 47–67, fev. 2001.

WAIZBORT, L. Questões não só alemãs. **Revista Brasileira de Ciências Sociais**, v. 13, n. 37, p. 185–190, jun. 1998.

WITTE, J. C. A Ciência Social digitalizada: avanços, oportunidades e desafios. **Sociologias**, v. 14, n. 31, 2012. Disponível em: <<http://seer.ufrgs.br/index.php/sociologias/article/view/34927>>. Acesso em: 27 jul. 2014.

ZALUAR, A. EXCLUSÃO E POLÍTICAS PÚBLICAS: DILEMAS TEÓRICOS E ALTERNATIVAS POLÍTICAS. **Revista Brasileira de Ciências Sociais**, v. 12, n. 35, out. 1997. Disponível em: <http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0102-69091997000300003&lng=en&nrm=iso&tlng=pt>. Acesso em: 4 jun. 2017.