

ISSN 2317-6644

bib

**Revista Brasileira de Informação Bibliográfica
em Ciências Sociais**

BIB – *Revista Brasileira de Informação Bibliográfica em Ciências Sociais* (ISSN 1516-8085 impresso e 2317-6644 a partir do nº 69 no formato digital) é uma publicação semestral da Associação Nacional de Pós-Graduação e Pesquisa em Ciências Sociais (ANPOCS) destinada a estimular o intercâmbio e a cooperação entre as instituições de ensino e pesquisa em Ciências Sociais no país. A BIB é editada sob orientação de um editor, uma comissão editorial e um conselho editorial composto de profissionais vinculados a várias instituições brasileiras.

Diretoria (gestão 2017-2018)

Presidente: Fabiano Guilherme Mendes Santos (IESP-UERJ); *Secretário executivo:* Ricardo Mariano (USP); *Secretária adjunta:* Emília Pietrafesa de Godoi (UNICAMP); *Diretora de publicações:* Adrian Gurza Lavalle (USP); *Diretores:* Carlos Guilherme do Vale (UFRN), Roberto Veras de Oliveira (UFPE) e Silvana Krause (UFRGS)

Conselho Fiscal: Ciméa Barbato Bevilacqua (UFPR), Cláudio Gonçalves Couto (FGV-SP) e Marcelo Domingos Sampaio Carneiro (UFMA)

Coordenação: Adrian Gurza Lavalle (USP)

Comissão Editorial

Editores-chefe: Marcia Consolim (Unifesp); *Editores de área:* Adriano Codato (UFPR), André Botelho (UFRJ), Deise Lucy Montardo (UFAM), Eliana dos Reis (UFMA), Igor Machado (UFSCar), Julie Cavignac (UFRN), Luiz Cláudio Lourenço (UFBA), Marcelo Carvalho Rosa (UnB), Mariana Batista (UFPE), Ricardo Fabrino Mendonça (UFMG), Soraya Fleischer (UnB); *Membros estrangeiros:* Frédéric Lebaron (École Normale Supérieure – Cachan, França), Frédéric Sawicki (Université Paris 1 – Panthéon-Sorbonne - França), Gustavo Sora – (CONICET / Universidad Nacional de Córdoba - Argentina), Virgílio Borges Pereira (Universidade do Porto - Portugal); *Editora de dossiês:* Márcia Lima (USP); *Editores(as) assistentes:* Raissa Wihby Ventura (PUC/SP), Mírian da Silveira Pavanelli (Anpocs)

Conselho Editorial: Alex Keyssar (Harvard Kennedy School), Brasília Sallum (USP), Charles Pessanha (UFF), Edna Castro (NAEA/UFPA), Elide Rugai Bastos (Unicamp), Letícia Calderón Chelius (Instituto Mora), Lúcia Bogus (PUC-SP), Marcos Costa Lima (UFPE), Maria Helena de Castro Santos (UnB), Par Engstrom (College of London) e Ruben George Oliven (UFRGS)

Edição

Assistente Editorial: Mírian da Silveira Pavanelli e Raissa Wihby Ventura (PUC/SP)

Preparação/revisão de texto/copidesque: Fernanda Corrêa e Tatiana Custódio | Tikinet

Versão/tradução de resumos: Laura Varanda (francês) e Giovanna Marchetti (inglês) | Tikinet

Edição eletrônica: Karina Vizeu Winkaler | Tikinet

Appropriate articles are abstracted/indexed in:

Hispanic American Periodicals Index

Dados Internacionais de Catalogação na Publicação (CIP)

Vagner Rodolfo CRB-8/9410

B581

BIB: Revista Brasileira de Informação Bibliográfica em Ciências Sociais / Associação Nacional de Pós-Graduação e Pesquisa em Ciências Sociais – n. 41 (1996), – São Paulo : ANPOCS, 1996-.

Anual

Resumos em português, inglês e francês.

Título até o n. 40, 1995: BIB: Boletim Informativo e Bibliográfico de Ciências Sociais.

ISSN: 1516-8085 (impresso até edição nº 68)

ISSN: 2317-6644 (digital a partir da edição nº 69)

1. Ciências humanas 2. Ciências sociais 3. Sociologia 4. Ciência política 5. Antropologia.
I. Associação Nacional de Pós-Graduação e Pesquisa em Ciências Sociais - ANPOCS.

2017-366

CDD 300

CDU 3

Índice para catálogo sistemático:

1. Ciências sociais 300
2. Ciências sociais 3

Associação Nacional de Pós-Graduação e Pesquisa em Ciências Sociais – Anpocs

Universidade de São Paulo – USP

Av. Prof. Luciano Gualberto, 315 – 1º andar

0558-010 São Paulo – SP

Telefax.: (11) 3091-4664 / 3091-5043

**Revista Brasileira de Informação Bibliográfica
em Ciências Sociais****Sumário**

Fronteiras do debate metodológico em Ciências Sociais 5
Equipe editorial

ARTIGOS

**Sociologia como ciência das populações: contribuições de
Carlos Hasenbalg e Nelson do Valle Silva no Brasil** 7
Carlos Antonio Costa Ribeiro

**Alusões biográficas e trajetórias: entre esquemas analíticos e
usos flexíveis** 36
Eliana Tavares dos Reis e Irllys Alencar F. Barreira

**Análise de correspondências múltiplas: fundamentos,
elaboração e interpretação** 68
Elisa Klüger

**O retorno da comparação etnográfica:
a epistemologia em momentos de mudança política** 98
Erik Bähre

**O papel da ciência política no discurso científico brasileiro:
institucionalização e avanços de uma área em construção** 116
Emerson Urizzi Cervi

**O texto como dado: desafios e oportunidades para as
ciências sociais** 138
Maurício Izumi e Davi Moreira

Fronteiras do debate metodológico em Ciências Sociais

Equipe editorial

Este número é resultado dos debates recentes que emergiram no seio do atual corpo editorial da *BIB* sobre os rumos de uma publicação especializada em divulgar debates bibliográficos contemporâneos. Via de regra, publicações com esse perfil privilegiam artigos que analisam o campo de debates em torno de temas de pesquisa consolidados. O desafio recente do nosso corpo editorial foi abrir espaço, primeiro no encontro da Anpocs em 2017 e, posteriormente, em um número específico para conhecermos melhor as questões metodológicas que condicionam a escolha e análise de objetos específicos de pesquisa. Foram fundamentais nesse processo as iniciativas e sugestões de Carlos Antonio Costa Ribeiro e Eduardo Marques.

Nos textos que compõem este número são apresentadas algumas fronteiras do debate metodológico que contribuíram para delinear a pesquisa recente, no Brasil e em algumas partes do mundo, nos diversos campos e disciplinas das ciências sociais.

O trabalho de Ribeiro recupera a trajetória e a estrutura dos argumentos construídos pioneiramente por Nelson do Valle Silva e Carlos Hasenbalg para refletir sobre a noção de sociologia como uma complexa ciência das populações, um estilo de pesquisa centrado nos estudos das relações raciais no Brasil que descreveu as populações não brancas observando regularidades em amostras de grandes bancos de dados. Levando em consideração desafios metodológicos fundamentais e lançando mão de probabilidades, tendências e mecanismos

que explicam a desigualdade social, o autor defende que esse modo de fazer pesquisa contribuiu para nutrir o debate político que temos hoje no país, sem abrir mão do rigor e da inovação metodológica.

Dois artigos trazem, por vias distintas, os rastros metodológicos da obra de Pierre Bourdieu na sociologia contemporânea. Se o arcabouço conceitual desse autor se tornou referência na formação de cientistas sociais no Brasil, contribuindo para inúmeros debates em várias áreas de pesquisa social, os textos aqui publicados reforçam sua centralidade também por seu impacto metodológico.

Eliana Tavares dos Reis e Irllys Alencar F. Barreira apresentam um estudo sobre biografias a partir do levantamento dos artigos que utilizam dados biográficos como foco de pesquisa publicados na plataforma SciELO. A hipótese central das autoras é de que o lugar epistemológico das biografias cedeu espaço às disputas sobre os usos de dados biográficos. Percebe-se, aqui, uma importante distinção metodológica, nem sempre considerada em estudos que trabalham com indivíduos. O estudo de biografias é epistemologicamente distinto do simples uso de dados biográficos para a construção de argumentos nas ciências sociais. Ao marcar essas distinções, o texto sugere que, apesar da flexibilidade e da diversidade de estudos sobre o tema, a força dos laços intelectuais com o legado de Pierre Bourdieu contribuiu para o domínio preponderante de seus conceitos e técnicas de pesquisa sobre o tema no Brasil.

A contribuição de Elisa Klüger, por sua vez, invoca a herança de Bourdieu para o desenvolvimento das análises de correspondências múltiplas. Sua contribuição apresenta os fundamentos, os elementos constitutivos e as etapas necessárias para elaborar esse tipo de análise. Ressaltando que se trata de um olhar que privilegia a dimensão estrutural, a vantagem seria a de apontar afinidades ao permitir a visualização do espaço das posições relacionais dos agentes investigados.

O artigo de Erik Bähre contribui para que leitores conheçam alguns dos caminhos recentes da antropologia, recuperando a noção de comparação que certos estilos de etnografia euro-americana aportam para o debate internacional. O ponto específico do texto é como a noção de reflexividade, vista dos ângulos disciplinar e interdisciplinar, contribuiu para limitar as ambições objetivistas desse tipo de pesquisa comparada. Associada a isso, ganha terreno também a aposta epistemológica no potencial baixo controle que os autores e as autoras de etnografia contemporânea têm sobre a dinâmica da vida de seus sujeitos e seu ambiente social.

A contribuição de Emerson Urizzi Cervi analisa o tipo de contribuição metodológica desenvolvida na ciência política brasileira.

A questão central é a relação entre a teoria que sustenta as técnicas de modelagem estatística e sua maior ou menor conexão com a realidade empírica. Descrevendo costumes usos inadequados de certas técnicas de predição, o texto alerta para os excessos teórico-metodológicos que tendem a enfraquecer ou limitar as contribuições dessa área do conhecimento.

Finalmente, Maurício Izumi e Davi Moreira apresentam, a partir de seus próprios trabalhos, estudos que usam ferramentas e técnicas de análise textual. As técnicas trabalhadas no texto são as que organizam a obtenção e pré-processamento dos dados, a fim de demonstrar a semelhança entre os textos analisados. São explorados também os possíveis métodos de classificação e escalonamento de fontes textuais, em particular de discursos políticos.

Longe de esgotarem e resumirem os debates teórico-metodológicos recentes, os textos aqui apresentados visam contribuir para manter aberta a reflexão sobre os estilos e formas de pesquisa que sustentam ou limitam os avanços empíricos das ciências sociais. Nosso compromisso será o de manter esta seara aberta em edições futuras.

Sociologia como ciência das populações: contribuições de Carlos Hasenbalg e Nelson do Valle Silva no Brasil

Carlos Antonio Costa Ribeiro¹

*“Devem exigir que eu procure a
verdade, não que a encontre.”*

Denis Diderot²

Introdução

O objetivo deste artigo é apresentar as contribuições dos sociólogos Carlos Hasenbalg e Nelson do Valle Silva para a constituição de uma perspectiva sociológica que privilegie o estudo das populações no Brasil. Argumentamos que as pesquisas desses autores foram fundamentais para se entender melhor como as desigualdades raciais permaneceram na sociedade brasileira ao longo de seu processo de modernização, justamente porque ambos adotaram uma abordagem que privilegia o estudo das regularidades populacionais que só são passíveis de investigação sociológica a partir de análises estatísticas usando grandes bancos de dados. Em outras palavras, a abordagem adotada possibilitou identificar padrões de regularidade e mudança impossíveis de serem apontados por estratégias heurísticas e metodológicas alheias às dinâmicas populacionais. Também foi fundamental o fato de as pesquisas se filiarem à área de

estudos sobre estratificação e mobilidade social, que privilegia a análise dos fatores relacionados à reprodução e/ou superação das desigualdades ao longo do ciclo de vida dos indivíduos.

Uma marca central da perspectiva sociológica no estudo de populações é a incorporação da variabilidade entre os indivíduos ou coletividades que as compõem. Essa é uma ideia muito simples, que fica evidente quando olhamos para as diferenças físicas, de personalidade ou na história de vida dos indivíduos, ou as distinções entre qualquer outra unidade de análise: famílias, escolas, empresas, cidades, países etc. É fácil chegar à conclusão de que não há dois indivíduos iguais no mundo, assim como não há, por exemplo, duas escolas ou empresas ou países exatamente iguais. Mesmo assim, somos capazes de descrever algumas regularidades nas populações compostas por esses indivíduos ou unidades de análise. As generalizações, no entanto, não são absolutas, na medida em que cada um pode ter uma combinação diferente de algumas dessas características, bem como diversas outras que o façam únicos e singulares. Partindo dessas concepções, o objeto da sociologia seria detectar e descrever

1 Professor e pesquisador do Instituto de Estudos Sociais e Políticos da Universidade do Estado do Rio de Janeiro (IESP-UERJ), e Ph.D. em Sociologia pela Columbia University. E-mail: carloscr@iesp.uerj.br

2 Esta frase foi citada por Nelson do Valle Silva em um seminário organizado em sua homenagem no IESP-UERJ no dia 6 de outubro de 2017. No dia 20 de junho de 2016 também houve um seminário em homenagem a Carlos Hasenbalg. Ambos os seminários podem ser vistos na íntegra no seguinte endereço eletrônico: <https://bit.ly/2IwctaC>

as regularidades, cada vez mais complexas, encontradas nas populações estudadas e, em seguida, explicar como estas se constituem. Diante da variabilidade fundamental entre os indivíduos ou as unidades, temos que reconhecer o caráter probabilístico dos fatos que descrevemos e explicamos. O raciocínio probabilístico é uma consequência direta do trabalho com a variabilidade e não é uma marca meramente técnica ou metodológica, mas epistemológica, uma vez que toda noção de causalidade implícita nos raciocínios sobre populações incorpora necessariamente a imprevisibilidade e rejeita explicitamente qualquer sugestão determinística como forma cientificamente viável de explicação no âmbito da sociologia e das ciências sociais.

Um exemplo prosaico auxilia na construção do argumento. Pensemos hipoteticamente na população dos sociólogos no Brasil. Provavelmente encontraremos algumas regularidades e muita variabilidade em suas características, tais como renda do trabalho, posição política, tipo de formação acadêmica, tipo de emprego que ocupam, classe social de origem etc. Digamos que haja uma correlação entre ser sociólogo e ter opiniões políticas de esquerda. Assim, haveria uma alta probabilidade de um indivíduo que é sociólogo ter opiniões de esquerda, mas certamente haveria também alguns com opiniões distintas. Enfim, poderíamos estudar regularidades na população de sociólogos, sem deixar de reconhecer a diversidade entre eles.

Adotar essa abordagem da sociologia como ciência das populações foi fundamental para que Carlos Hasenbalg e Nelson do Valle Silva pudessem alcançar explicações e interpretações inovadoras sobre as desigualdades e as relações raciais. Além disso, também foi importante o fato de suas pesquisas terem se concentrado na área de estratificação social, o que os levou a estudar temas como a mobilidade social, a

industrialização e as mudanças no mercado de trabalho, a desigualdade educacional, a pobreza, os mercados matrimoniais etc. Essa perspectiva populacional, probabilística e temática os levou a estabelecer alguns fatos importantes, como o de que há forte desigualdade racial no acesso e na progressão no sistema educacional brasileiro. Uma vez estabelecidos os fatos (ou regularidades) e as probabilidades relacionadas, torna-se fundamental explicar como esse estado de coisas se mantém. Quais seriam os mecanismos que explicam, por exemplo, essas desigualdades raciais no sistema educacional? A resposta a essa pergunta depende, em grande medida, da proposição de mecanismos causais que devem ser pensados em termos das ações sociais dos indivíduos envolvidos. Hasenbalg e Silva (1988) sugerem que, em termos da desigualdade racial de um modo geral, o principal mecanismo de reprodução é o de acumulação de desvantagens (que incluem discriminação racial, mas não se limitam a esse único mecanismo) por parte dos pretos e pardos ao longo de seus ciclos de vida. Essa acumulação aumenta a *probabilidade* de os indivíduos desses grupos raciais encontrarem mais dificuldades para, por exemplo, progredir no sistema educacional e alcançar melhores empregos.

Essa perspectiva sociológica parte da ideia de que primeiramente é fundamental estabelecer quais são as probabilidades relacionadas às regularidades estabelecidas em análises estatísticas usando grandes bancos de dados de alta qualidade. Em seguida, é necessário buscar explicações usando a ideia de mecanismos causais ou explicativos. Esses procedimentos sempre levam em conta a ideia de que há muita variabilidade entre os indivíduos ou entre quaisquer outras unidades de análise (família, empresa, cidade, país etc.), ou seja, as regularidades encontradas não devem jamais ser pensadas como absolutas,

dada a natureza probabilística de qualquer causalidade passível de ser identificada nas ciências sociais. Por exemplo, quando se diz que pretos tendem a estar sobrerrepresentados entre os pobres, não está se afirmando que ser pobre implica necessariamente ser preto. Há pretos ricos, bem como brancos pobres. No entanto, há uma maior probabilidade de que pretos sejam pobres e brancos sejam ricos. Estabelecer esses tipos de probabilidades, suas tendências e os mecanismos que as explicam seriam os principais objetivos da sociologia pensada como uma ciência das populações.

Este artigo está dividido em três partes. Na primeira, explicamos melhor a ideia da sociologia como ciência das populações. A descrição será baseada principalmente no trabalho do sociólogo inglês John Goldthorpe (2016), mas também traremos para o debate outras contribuições e exemplos relevantes. Em seguida, apresentamos resumidamente os trabalhos de Carlos Hasenbalg e Nelson do Valle Silva, que foram reunidos em quatro livros (HASENBALG; SILVA, 1988, 1992, 2003; HASENBALG; SILVA; LIMA, 1999). Ao final, fazemos um balanço da obra dos autores, enfatizando, por um lado, a importância da perspectiva sociológica que adotam e, por outro lado, propondo uma breve reflexão sobre a relação entre os aspectos políticos e os resultados substantivos de suas pesquisas.

Sociologia como ciência das populações

A abordagem adotada na sociologia, pensada como uma ciência das populações, assemelha-se a um tipo de pensamento que também é comum na teoria evolucionista ou darwinista (LIEBERSON, 1992). Essa

semelhança, no entanto, não está relacionada à ideia de “evolução”, tampouco ao que se convencionou chamar de fundamentos biológicos das atividades humanas (debate *nature versus nurture*). A ideia que se assemelha é justamente a da variabilidade das e nas populações e a necessidade de incorporá-la, principalmente porque, assim como a biologia, a sociologia é uma ciência eminentemente observacional e não experimental. No caso da biologia, o foco é a variabilidade biológica entre os indivíduos de cada espécie, bem como a pluralidade de espécies. Na sociologia tratamos da variabilidade das histórias dos indivíduos e de algumas coletividades, tais como escolas, empresas, países etc. Diante dessa diversidade fundamental, a atividade científica deveria se concentrar na identificação e na explicação das regularidades que surgem no nível populacional. Em outras palavras, a característica comum de tais populações é que, apesar de seus elementos individuais apresentarem incontáveis diferenças entre si e serem aparentemente indeterminados em seu comportamento, elas podem apresentar regularidades de tipo probabilístico em seus níveis agregados.

Em sua apresentação da biologia evolucionista, Ernst Mayr (2001) sugere que esse tipo de “pensamento populacional” se desenvolveu de forma muito explícita em oposição ao “pensamento tipológico”, mais comum na química e na física, caracterizado pelo foco nas propriedades determinísticas e homogêneas das partículas ou dos elementos. Em contraste, na biologia houve um reconhecimento crescente da variabilidade existente entre os indivíduos que compõem as populações de espécies estudadas³. Ao mesmo tempo, o interesse se concentrou, por um lado, nas

3 O “pensamento populacional” de Darwin foi em parte inspirado por Malthus, que era um cientista social (MAYHEW, 2014).

regularidades probabilísticas que ainda podem ser observadas em meio à diversidade e, por outro lado, nos mecanismos através dos quais essas regularidades foram criadas. Esse tipo de “pensamento populacional”, no caso da sociologia, foi veementemente defendido por Otis Dudley Duncan, um dos principais precursores, defensores e também críticos do uso da metodologia quantitativa em sociologia (DUNCAN, 1984, 1992; GOLDTHORPE, 2016; XIE, 2007).

Embora Dudley Duncan fosse um sociólogo estadunidense, não é correto dizer que essa abordagem seja tipicamente norte-americana, como se costuma afirmar no Brasil. Seria também um exagero dizer que o “pensamento populacional” é predominante na sociologia, mas não há dúvidas de que um número cada vez maior de sociólogos no mundo todo vem adotando e desenvolvendo essa perspectiva em suas pesquisas (GOLDTHORPE, 2016). Consequentemente, estruturas bastante complexas de regularidades populacionais estão sendo identificadas e vêm se tornando objeto de explicação sociológica. Por exemplo, há interesse em regularidades não apenas na estabilidade de formas particulares de ação individual e suas consequências em certas populações ou em diferenças entre populações e entre subpopulações, mas também nas mudanças ao longo do tempo – seja em termos de períodos históricos, sucessão de coortes de idade ou ciclo de vida dos indivíduos. Além disso, as regularidades estudadas por sociólogos não se limitam a populações de indivíduos, mas incluem também as localidades micro, meso e macro em que estes se encontram, como redes sociais (*networks*), associações, instituições, cidades, bairros, países etc. Em todos esses casos, no entanto, o estabelecimento das regularidades populacionais nas sociedades humanas depende da coleta e análise de grandes bancos de dados.

De fato, a principal contribuição científica da sociologia baseada em pesquisas populacionais (principalmente *surveys*) e na análise dos dados coletados é, sem dúvidas, sua comprovada capacidade de estabelecer e revelar regularidades que simplesmente não seriam observadas sem o uso da metodologia quantitativa. Por exemplo, usando dados do censo populacional de 1980, Nelson do Valle Silva (1987) revelou algo que não era conhecido sobre os casamentos inter-raciais no Brasil. Os modelos log-lineares usados indicaram que, apesar de haver maior probabilidade de casamentos racialmente homogênicos, os casamentos inter-raciais entre brancos e pardos eram mais prováveis do que os entre brancos e pretos e entre pardos e pretos. Esses resultados foram confirmados em um estudo que mostrou um aumento dos casamentos inter-raciais entre 1960 e 2000 (RIBEIRO; SILVA, 2009). Vale lembrar que esse tipo de regularidade é de cunho probabilístico. Em outras palavras, há uma maior probabilidade de casamentos entre brancos e pardos do que entre pardos e pretos, embora existam diversos casamentos individuais entre todos esses grupos raciais. O estabelecimento das regularidades não deixa de considerar a variabilidade marcada pela singularidade de cada casamento na sociedade brasileira.

De acordo com a abordagem do “pensamento populacional”, o principal objetivo da pesquisa sociológica seria descrever e explicar as regularidades probabilísticas encontradas nas populações, em vez de eventos singulares. Para explicar esse ponto central, Goldthorpe (2016, p. 46) distingue entre “chances essenciais”, que seriam o puro acaso ou a coincidência (a ideia, por exemplo, de que uma pessoa que está no lugar certo na hora certa pode ter algum tipo de vantagem), e “chances operacionais” (ou “probabilidade”), que ocorrem de acordo com os contextos sociais em que os indivíduos se encontram (por exemplo, indivíduos que ficam

mais tempo na rua têm mais chances de serem vítimas de alguns crimes do que indivíduos que raramente saem de casa). Nesse sentido, Goldthorpe (Ibidem, p. 46-47, tradução minha) afirma o seguinte:

Assim, em análises feitas no nível da população – isto é, cobrindo um número relativamente grande de indivíduos – diversos tipos de regularidades probabilísticas na vida social emergem apesar das forças da “chance ou acaso essencial” – embora frequentemente sejam tipos complexos de regularidades que não são imediatamente visíveis. E são então essas regularidades que, sob os auspícios das “chances operacionais”, podem ser tratadas como as *explananda* para as quais explicações sociológicas podem ser apropriadamente procuradas⁴.

Seguindo essa linha de pensamento, não há como negar que a estatística assume um papel fundamental, na medida em que é o meio pelo qual as regularidades populacionais podem ser estabelecidas como os “objetos de estudo” da sociologia. Vale lembrar que a estatística não nega a pluralidade essencial que existe entre os indivíduos ou outras unidades de análise, pelo contrário, ela é uma forma de mapear não só as tendências centrais (com diversas técnicas de análise das estruturas das medidas de tendência central), como também a variabilidade humana em suas diversas formas (o que de certa maneira é levado em conta pelas diversas medidas de dispersão)⁵.

Primordial para o estabelecimento de qualquer análise é uma cuidadosa construção de conceitos. Nesse sentido, nenhuma análise pode estar descolada de técnicas de

mensuração. Em outras palavras, é fundamental e plenamente factível a aplicação de conceitos sociológicos em análises estatísticas na medida em que podem ser expressos por instrumentos de mensuração – classificações, escalas etc. – que tenham algum grau de confiabilidade (*reliability*) e validade (*validity*) (BOHRNSTEDT, 2010). A ideia de confiabilidade define em que medida um instrumento de mensuração, pelo qual um conceito é operacionalizado em uma variável, pode ser usado consistentemente em diferentes situações levando aos resultados esperados. Há vários testes de confiabilidade disponíveis. A ideia de validade é mais complexa, já que envolve diferentes aspectos e existem vários tipos. No entanto, o mais importante, geralmente chamado de “validade do constructo” (*construct validity*), se refere à capacidade de um instrumento mensurar de modo empírico o que ele conceitualmente pretende capturar. O ponto central nessa discussão é que, ainda que se assumam as particularidades epistemológicas da sociologia e sua teorização (a natureza discursiva da teoria, a imprevisibilidade da ação humana), há em última análise uma realidade passível de ser capturada de forma consistente e o arbítrio sobre a “captura” dessa realidade pode ser racional e empiricamente estabelecido com base nos próprios resultados alcançados por diferentes pesquisadores.

Também podemos exemplificar esse ponto com uma discussão sobre classificação racial que, de certa forma, foi iniciada

4 No original: “Thus, in analyses made at the population level – that is, covering relatively large numbers of individuals – probabilistic regularities in social life of many kinds do still emerge despite the pervasiveness of essential chance – though often they are regularities of a complex and not readily visible [...] kind. And it is then these regularities that, under the auspices of operational chances, can be treated as the *explananda* for which sociological explanations may properly be sought”.

5 De certa forma, estou simplificando essa distinção ao afirmar que as regularidades são capturadas por medidas de tendência central, uma vez que medidas de dispersão também podem apresentar regularidade, como é o caso das medidas de desigualdade de renda, que são essencialmente medidas de dispersão.

nos trabalhos de Carlos Hasenbalg e Nelson do Valle Silva. Em vários artigos, ambos os estudiosos usaram a distinção entre brancos e não brancos (pretos mais pardos) para mensurar as desigualdades raciais. O uso do termo “não branco” foi adotado simplesmente porque na maior parte das análises estatísticas implementadas as diferenças entre pretos e pardos não eram estatisticamente significativas. Não havia qualquer implicação ideológica nessa junção, como nos fazem crer Bourdieu e Wacquant (1999) em um texto bastante equivocado sobre a influência da sociologia norte-americana nas ciências sociais e, em particular, no Brasil. Telles (2003), que estudou importantes aspectos das relações raciais no Brasil, critica veementemente as ideias de Bourdieu e Wacquant (1999), afirmando que não houve dualização ideológica das relações raciais nas pesquisas sociológicas sobre o país. Em outras palavras, ao juntar pretos e pardos, Hasenbalg e Silva não estavam dualizando as relações raciais brasileiras, como nos fazem crer, de forma implícita, os autores franceses – Hasenbalg e Silva não foram nominalmente citados no texto, embora seja feita uma menção a brasileiros que estudaram nos Estados Unidos (BOURDIEU; WACQUANT, 1999). Essa rejeição do dualismo fica clara em outros trabalhos dos autores brasileiros, não apenas nos textos que mostraram a diferença entre pretos e pardos em termos de casamentos inter-raciais, como mencionado (SILVA, 1987), e as tendências de miscigenação (HASENBALG; SILVA; BARCELOS, 1989), mas também pelo fato de terem feito as primeiras análises quantitativas sobre a ambiguidade da classificação de cor ou raça no Brasil (SILVA, 1994, 1996). Nesse último trabalho, Silva (1994) mostra que há uma classificação social da raça no Brasil e que um gradiente de cor poderia ser usado para definir as “raças sociais”, mas também que esse gradiente está fortemente

correlacionado à classificação tricotômica de brancos, pardos e pretos. Além disso, o autor mostra que a categoria “moreno” é intrinsecamente indeterminada, porque pode ser usada tanto para classificar brancos quanto pardos e pretos. Essa linha de análise iniciada por Silva (1994, 1996) recuperava uma longa tradição de estudos sociológicos iniciada com os trabalhos de Oracy Nogueira (1998), que pretendia definir de forma adequada o conceito de raça tal como se expressava no Brasil e que certamente era diferente do que ocorria nos Estados Unidos. Além disso, Silva (1994, 1996) antecipou diversos trabalhos que foram desenvolvidos mais recentemente sobre a classificação racial no Brasil e em outros lugares do mundo (MUNIZ, 2010, 2012; RIBEIRO, 2017; SCHWARTZMAN, 2007; TELLES, 2014). Apesar dessa tendência a uma classificação contínua da cor ou raça, Silva (1994) sugere que a variável tricotômica é “confiável” e “válida” na maioria dos casos, tendo em vista que está fortemente correlacionada à classificação contínua de cor da pele – Telles (2014) confirma essa ideia em estudo comparativo, assim como Loveman, Muniz e Bailey (2012) em análises recentes. Em suma, as análises quantitativas estão em pleno diálogo com os esforços sociológicos, buscando definir corretamente conceitos analíticos adequados para entender as relações sociais.

Tendo em vista que as análises estatísticas são fundamentais para estabelecer as regularidades populacionais, é necessário bastante rigor na escolha de métodos de coleta e análise de dados. Os sociólogos tiveram papel ativo e contribuíram de forma decisiva não só para a elaboração de várias técnicas de coleta de dados – como as amostras probabilísticas (KISH, 1965; PLATT, 1996) – e de construção de instrumentos de coleta em entrevistas e questionários (PLATT, 1996), como também na formulação de métodos estatísticos de análise de dados

(CLOGG, 1992) – como, por exemplo, os modelos log-lineares (GOODMAN, 1984). Normalmente os críticos ao uso de métodos estatísticos em sociologia, que costumam definir essa abordagem, pejorativamente como “positivista”, não são capazes de perceber que os principais críticos ao uso dessas técnicas são justamente aqueles que as usam. Em termos de construção de bancos de dados, há uma constante preocupação tanto com a coleta – que deveria ser feita principalmente por meio de técnicas de amostragem probabilística (frequentemente encontramos dados quantitativos que não são coletados dessa forma, por exemplo diversas pesquisas eleitorais que usam técnicas de amostragem por cotas) – quanto com a elaboração de instrumentos de coleta, que devem considerar aspectos desde a formulação de perguntas até a preocupação com a influência da interação entre entrevistadores e entrevistados na qualidade dos dados obtidos. Essas preocupações são constantes e certamente são levadas em conta por bons pesquisadores que adotam a perspectiva da metodologia quantitativa nas ciências sociais.

Os métodos de análise, por sua vez, devem ser capazes de expressar corretamente os padrões e as tendências das regularidades populacionais que a sociologia enquanto ciência das populações visa descrever e explicar; levar em conta a variabilidade essencial existente entre as unidades de análise (indivíduos, famílias, escolas etc.) e respeitar a métrica de mensuração das variáveis (que podem ser intervalares, nominais, ordinais). Cabe mencionar que a combinação entre métodos quantitativos e qualitativos pode ser feita justamente a partir das regularidades encontradas pelas análises quantitativas da seguinte forma: unidades próximas (valores

preditos) e distantes (*outliers*) das regularidades descritas pelos métodos estatísticos poderiam ser escolhidas como foco de investigação qualitativa, o que contribuiria para um entendimento mais amplo do fenômeno estudado e para a reformulação de análises quantitativas posteriores⁶. Assim, a crítica aos métodos quantitativos não deveria ser feita em termos de uma contraposição aos métodos qualitativos. Na realidade, ambos podem e devem ser complementares. Outro exemplo de estudo qualitativo que dialoga com o trabalho de Carlos Hasenbalg e Nelson do Valle Silva é o de Silva e Leão (2012) sobre a exploração das fronteiras simbólicas da identidade racial dos brasileiros pardos.

Uma objeção que me parece mais relevante foi levantada, entre outras, por Abbott (1992). Segundo esse autor, muitos trabalhos quantitativos perdem de vista as interações e ações entre os indivíduos que, em última instância, compõem a vida social. Ou seja, muitos estudos quantitativos passam a ideia de que são as variáveis que estão atuando e não os indivíduos. Por exemplo, ao dizer que o “grau de escolaridade” está correlacionado à “renda do trabalho” (quanto maior a educação maior a renda), o pesquisador passa incorretamente a ideia de que a escolaridade está “agindo” para aumentar a “renda”. Obviamente não é esse o raciocínio como mostraremos. É importante entender os mecanismos, no nível da ação e interação entre os indivíduos, que explicam a correlação entre educação e renda, bem como qualquer outra regularidade encontrada. A crítica, no entanto, é fundamental e deveria ser levada a sério por pesquisadores usando métodos quantitativos.

Como vimos insistindo, o principal objetivo do uso de métodos quantitativos

6 Para um exemplo, cf. Ribeiro (1995).

é descrever as regularidades probabilísticas no nível populacional, o que Robert Merton (1987, p. 2-6, tradução minha) chamou de “estabelecendo os fatos”⁷. Para tanto, uma série de métodos estatísticos cada vez mais complexos e sofisticados vem sendo utilizada na sociologia contemporânea. Podemos citar, por exemplo, o uso de modelos lineares hierárquicos que foram desenvolvidos para diferenciar o efeito das características das escolas do efeito das características dos alunos e de suas famílias em seu desempenho acadêmico (RAUDENBUSH; BRYK, 2002). Essa metodologia, assim como diversas outras, foi desenvolvida especificamente para tratar de um problema sociológico identificado em trabalhos anteriores na sociologia da educação (COLEMAN et al, 1966). Além disso, precisamos estar atentos para diferentes formas de usar e interpretar os resultados obtidos a partir de técnicas estatísticas.

Um exemplo importante é nada menos do que o uso da regressão linear múltipla – e, de modo mais geral, dos modelos lineares generalizados –, que é a técnica mais usada nas ciências sociais, e a partir da qual diversas outras são derivadas. O sociólogo Yu Xie (2007) explica que, embora a técnica seja exatamente a mesma, é possível utilizá-la de dois modos diferentes. O primeiro, que denomina concepção gaussiana da regressão, tem o objetivo de estabelecer relações causais – expressando leis gerais – entre variáveis independentes e dependentes, e os desvios das observações individuais dessa relação causal (o termo de erro da regressão) são tratados como erro de mensuração, ou simplesmente como ruído indesejado⁸. A segunda abordagem, chamada de concepção galtoniana da

regressão, é associada ao “pensamento populacional” desenvolvido, primordialmente, nos trabalhos de Otis Dudley Duncan. Nesse caso, o objetivo das regressões não é determinar relações causais, mas produzir descrições parcimoniosas da variabilidade populacional em termos das associações ou resultados em jogo nas análises. Nessa perspectiva, as análises se concentram no componente sistemático das regressões, ou seja, na variabilidade que ocorre entre os grupos de interesse sociológico, definidos pelas variáveis independentes das análises. Ao mesmo tempo, o termo de erro das equações é visto como um reflexo da variabilidade interna aos grupos, além de representar o erro de mensuração *stricto sensu*. Embora as análises devam diminuir essa variabilidade interna a partir do uso de outras variáveis independentes teoricamente relevantes, devemos aceitar que a variabilidade individual dentro dos grupos sempre será substancial.

Em um livro em que apresenta o uso de modelos de equações estruturais na sociologia, Duncan (1975, p. 166-167) afirma que embora seja importante identificar os aspectos institucionais e estruturais das sociedades que modificam a variabilidade em diversas características individuais, estas sempre serão mais numerosas. Ou seja, mesmo que sejamos capazes de identificar tendências gerais, as características que tornam singulares as histórias e personalidades de cada um sempre serão muito mais numerosas do que esses aspectos estruturais que diminuem as diferenças entre os indivíduos. Duncan vai adiante e afirma que seria horrível viver em uma sociedade na qual a vida das pessoas fosse totalmente determinada por características

7 No original: “establishing the phenomena”.

8 Em sua tese de doutorado, Barbosa (2018) argumenta que o modelo econômico do capital humano desenvolvido por Mincer é partidário dessa abordagem “gaussiana”.

estruturais e institucionais – para usar uma linguagem estatística, seria horrível viver em uma sociedade em que o R^2 das regressões fosse de 95% ao invés de 30% ou 35%. Por exemplo, pessoas cujos pais eram das classes baixas têm menos chances de mobilidade ascendente do que pessoas com origens de classes mais altas, mas tendo em vista que os indivíduos são diferentes em diversos outros aspectos além da classe social da família de origem, sempre encontraremos algumas pessoas mais pobres que experimentaram grande mobilidade ascendente. Felizmente, não há sociedades totalmente determinísticas, fato amplamente reconhecido pelo “pensamento populacional” e negligenciado pelo “pensamento tipológico”. Parece-nos que é exatamente por serem determinísticas que as interpretações “holísticas” ou “tipológicas” do mundo social tendam a ser mais sedutoras, embora de fato sejam limitadas e simplistas, para não dizer erradas. Em suma, o fundamental aqui é reconhecer que as técnicas de regressão linear, bem como diversos outros métodos estatísticos, interpretadas de acordo com a concepção galtoniana, são centrais para avançarem o conhecimento e a perspectiva do “pensamento populacional” na sociologia⁹.

Embora as regressões lineares (e outros métodos estatísticos relacionados) sejam úteis para estabelecer regularidades, elas não podem ser diretamente usadas para desenvolver explicações causais, porque se não levarem em conta todas as variáveis de alguma forma relacionadas, tanto às outras independentes quanto às dependentes, haverá vieses que impossibilitam a definição da causalidade. Além disso, frequentemente nem sabemos se há outros fatores importantes. Esse é apenas um dos diversos problemas que invalidam o uso

das regressões para a definição de relações causais. Apesar dessas limitações, há obviamente interesse sociológico em definir as causas das regularidades que venham a ser estabelecidas.

Do ponto de vista puramente estatístico, a definição da causalidade depende de um experimento em que dois grupos são comparados: um de tratamento (ou experimental) e outro de controle. Em sociologia também podemos definir experimentos em pesquisas especialmente desenhadas para tanto ou, alternativamente, podemos tirar vantagem de experimentos ou quase-experimentos que ocorrem “naturalmente” nas sociedades. No entanto, na maioria das pesquisas essas abordagens são inviáveis na medida em que os dados geralmente não são obtidos de forma “experimental”, mas sim de maneira “observacional”, ou seja, são amostras de populações em que o pesquisador não faz qualquer tipo de manipulação controlada das variáveis. Uma alternativa defendida por uma série de sociólogos (que será apresentada posteriormente) é formular mecanismos causais de explicação que sejam válidos ao nível dos indivíduos e/ou das interações sociais. A seguir, explicaremos brevemente as duas abordagens sobre causalidade: a experimental e a que utiliza a ideia de mecanismos.

Qualquer explicação causal, mesmo partindo de experimentos, implica um contrafactual, ou seja, faz referência a algo que não ocorreu. Por exemplo, quando dizemos que uma pessoa chegou atrasada porque perdeu o ônibus, estamos implicitamente dizendo que se não tivesse perdido o ônibus, não teria se atrasado, ou seja, referenciamos a algo que não ocorreu, um contrafactual. Essa lógica é válida para qualquer experimento, mesmo nas ciências

9 Para uma analogia entre os métodos de regressão e algumas formas de análise qualitativa em sociologia histórica, cf. Stinchcombe (2005).

naturais. Por isso, devemos falar em “efeito causal médio” (*average treatment effect*, entre outras variações do termo), que é a diferença média encontrada em uma variável dependente (ou *outcome*) entre as unidades selecionadas aleatoriamente para dois grupos, um de tratamento e outro de controle. São essas as diferenças de resultados quantificadas para que o cálculo do “efeito causal médio” – a causalidade é pensada em termos da média, tendo em vista que cada unidade foi ou não afetada pelo tratamento. O “efeito causal” é entendido como o resultado de uma intervenção que altera a variável dependente, ou como usualmente se diz na literatura, de um “tratamento” (*treatment*). Nas palavras de Holland (1986, p. 958, tradução minha): “não há causalidade sem manipulação”¹⁰.

Um enorme número de técnicas estatísticas vem sendo desenvolvido nesse campo de estudos (conhecido como *potential outcome approach*) não só no que diz respeito ao desenho de experimentos (denominados de *randomised controlled trials*, RTC), como também para dados observacionais que não foram coletados de forma adequada para experimentos (ROSENBAUM, 1987, 2002)¹¹.

Em sociologia, o principal trabalho que advoga em favor dessas técnicas é o livro de Morgan e Winship (2007)¹², que na verdade é útil para diversos tipos de pesquisa, mesmo

qualitativas, nas ciências sociais. Essa abordagem no entanto é mais relevante e útil para avaliar intervenções¹³ do que para explicar regularidades normalmente encontradas em nível populacional¹⁴. Além disso, mesmo que uma explicação causal sobre o efeito de determinada intervenção (uma nova técnica de ensino, por exemplo) seja comprovada, a abordagem dos experimentos controlados não dá informações sobre como a intervenção funcionou na ação e na interação (entre alunos e professores, por exemplo). Em outras palavras, não há uma explicação em relação aos mecanismos que geram o efeito causal. Finalmente, também devemos ficar atentos para o fato de que os efeitos causais detectados por vezes são limitados a pequenas partes da população que foram expostas a algum tipo de intervenção e que, do ponto de vista demográfico, permanecem de certa forma restritas. A despeito dessas limitações, não há dúvidas de que essa abordagem deveria ser mais estudada e usada em estudos sociológicos. No Brasil, em particular, há pouquíssimo uso de tais técnicas nas ciências sociais, embora a nosso ver esse tipo de pesquisa devesse ser estimulado em nossa comunidade acadêmica. Afinal de contas, saber que algum tipo de política pública, por exemplo, tem probabilidades altas de ser eficiente é melhor do que não saber ou do que apenas ter uma opinião sobre a eficiência.

10 No original: “no causation without manipulation”.

11 Para um exemplo sobre o efeito da “riqueza” na mobilidade social no Brasil, cf. Torche e Ribeiro (2012).

12 Para uma introdução a diversas destas técnicas em estudos econométricos, cf. Angrist e Pischke (2015).

13 Esse tipo de abordagem deveria ser usado mais frequentemente para avaliar políticas públicas, o que quase não é feito no Brasil. Sobre esses métodos, cf. o excelente livro de Cano (2002).

14 Outra limitação, que também é uma vantagem em alguns casos, é que ao focalizar intervenções específicas, não é possível desvendar as causas mais amplas. Por exemplo, Cohen e Dupas (2010) mostram, em um estudo experimental, que doar uma rede de proteção contra mosquitos é mais eficiente na prevenção de doenças tropicais do que vendê-la por preços módicos (como era defendido em estudos anteriores). No entanto, a causa maior da propagação dessas doenças pode ser a pobreza e o pouco desenvolvimento do país (no caso o Quênia), que são consequências de uma desigualdade entre países provavelmente oriunda do colonialismo e de outras formas de dominação mais amplas.

Essa linha de análises através de desenhos experimentais de pesquisa não é a única alternativa para explicações causais, embora seja a estatisticamente mais rigorosa. Há um grupo bastante diversificado de autores nas ciências sociais (GOLDTHORPE, 2016; HEDSTRÖM, 2005; HEDSTRÖM; BEARMAN, 2009; HEDSTRÖM; SWEDBERG, 1998) defendendo a explicação a partir do uso de mecanismos sociais. Na introdução de uma coletânea sobre sociologia analítica, Hedström e Bearman (2009) se filiam à tradição sociológica iniciada por Robert Merton, com sua *Theory of the Middle Range*, para afirmarem que o objetivo da explicação via mecanismos causais é tornar inteligíveis as regularidades observadas a partir da especificação detalhada de como surgiram. É importante ter em mente que essas explicações não envolvem necessariamente novas técnicas, e são de caráter predominantemente teórico e sociológico – embora também haja técnicas de simulação que podem ser empregadas para avaliar a plausibilidade dessas explicações (referimo-nos aqui a simulações do tipo *agent based*)¹⁵. De fato, podemos defini-las como “narrativas gerais das ações e interações” subjacentes às regularidades que queremos explicar. Essas narrativas devem ser adequadas às regularidades que pretendam explicar, e também devem ser passíveis de testes empíricos. Alguns autores defendem uma perspectiva que focalize principalmente as decisões racionais dos indivíduos em determinados contextos sociais (*bounded rationality*) – como Breen e Goldthorpe (1997) –, enquanto outros abrem espaço para explicações mais relacionais envolvendo redes sociais (*networks*) e outros mecanismos que possam ser compilados no que Elster chamou

de *toolbox of mechanisms – nuts and bolts, cogs and wheels* (ELSTER, 1989; HEDSTRÖM; BEARMAN, 2009). Em ambos os casos, no entanto, as explicações devem ser dadas a partir da ação e das interações entre indivíduos, o que Elster (1989) e Coleman (1990) chamam de microfundações das explicações sociológicas.

Uma vez identificadas as regularidades e formuladas as narrativas definindo os mecanismos causais a partir da ação e da interação dos indivíduos, Goldthorpe (2016, p. 119-121) sugere algumas estratégias para verificar se as explicações são realmente relevantes. A primeira seria a observação em estudos de caso bem elaborados e conduzidos, ou seja, pesquisas qualitativas que considerem as regularidades encontradas nas análises quantitativas. A segunda seria de uma espécie de observação indireta, em que o mecanismo descrito implicaria outras regularidades, que se fossem observadas nos dariam mais certeza de sua validade. A terceira, finalmente, seria a formulação de experimentos que colocassem em prova as hipóteses levantadas nas narrativas dos mecanismos que o pesquisador usou para explicar as regularidades populacionais. Uma quarta estratégia, que não foi levantada por Goldthorpe (2016), mas que também vem se mostrando interessante, é o uso de simulações com base nos mecanismos causais propostos (HEDSTRÖM, 2005).

Essa breve descrição da perspectiva sociológica pensada como uma ciência das populações teve o objetivo de apresentar aos leitores uma abordagem cada vez mais adotada por um grande número de pesquisadores em vários países. Não acreditamos que essa perspectiva tenha colorações nacionais, embora pareça ser mais comum em países europeus, na América

15 Um clássico nessa área é Schelling (1978), mas há uma imensa literatura seguindo essa linha. Para sugestões de literatura e de material na internet, cf. as entrevistas de Jerônimo Muniz e Nelson do Valle Silva em Bachini e Chicarino (2018).

do Norte e na Ásia. Também há, no entanto, vários grupos de pesquisa na América Latina seguindo essa linha, e no Brasil, em particular, há um crescente número de sociólogos fazendo pesquisas quantitativas altamente relevantes, bem como demógrafos com estudos de cunho sociológico – principalmente no Centro de Desenvolvimento e Planejamento Regional da Universidade Federal de Minas Gerais (Cedeplar/UFGM)¹⁶.

Dois dos precursores desse tipo de sociologia no Brasil foram Carlos Hasenbalg e Nelson do Valle Silva, cujas contribuições descrevemos a seguir.

A sociologia de Carlos Hasenbalg e Nelson do Valle Silva

Os sociólogos Carlos Hasenbalg e Nelson do Valle Silva trabalharam juntos durante 25 anos, desde o início da década de 1980 até meados da década de 2000. Argumentamos que, ao longo desse período, desenvolveram pesquisas que se filiam ao que chamamos de sociologia como ciência das populações.

Seus artigos foram publicados em diversas revistas acadêmicas e coletâneas, e também foram reunidos em quatro livros (HASENBALG; SILVA, 1988, 1992, 2003; HASENBALG; SILVA; LIMA, 1999). Muitos desses trabalhos foram feitos em co-autoria, enquanto outros, escritos por apenas um dos dois, faziam parte de uma agenda de pesquisas comum, na qual parece ter havido uma divisão de trabalho – ou seja, trabalhavam em temas separados, mas complementares. É muito raro encontrar colaborações tão duradouras e, nesse caso, parece que isso foi possível tanto por interesses comuns e por pertencerem à mesma instituição (o Instituto Universitário de Pesquisas do Rio

de Janeiro – Iuperj, atualmente Instituto de Estudos Sociais e Políticos – Iesp) quanto pela amizade entre os dois.

De fato, a colaboração começou depois que ambos terminaram o doutorado e começaram a trabalhar no Iuperj. Carlos Hasenbalg defendeu sua tese em 1979 no departamento de sociologia da Universidade da Califórnia, em Berkeley, sendo orientado por Robert Blauner (que escreveu importantes trabalhos sobre as relações raciais nos EUA) e tendo Arthur Stinchcombe (que fez importantes contribuições à sociologia econômica e histórica) como membro do comitê de tese. O trabalho, publicado como livro no mesmo ano (HASENBALG, 1979), apresenta o resultado de uma pesquisa de fôlego sobre as interpretações históricas e sociológicas da escravidão no Brasil, bem como reúne análise de pesquisas de opinião (*survey*) e interpretações sobre a manutenção das desigualdades e da discriminação racial no Brasil no período pós-abolição, em que ocorreu a industrialização do país. Seu estudo se filia ao que se convencionou chamar de “sociologia histórica” – em fase de crescente ascensão na época –, mas também se concentra na análise da situação geográfica e econômica das populações brancas, pardas e pretas. A tese principal foi formulada em oposição à ideia de Florestan Fernandes (1965) de que a discriminação e as desigualdades raciais eram apenas uma reminiscência da escravidão e desapareceriam com a ampliação das desigualdades de classe na sociedade burguesa industrial. Hasenbalg (1979) mostra que essas disparidades raciais permaneceram sendo reafirmadas e perpetuadas na sociedade industrial moderna que se constituía gradativamente no Brasil. Essa ideia fundamental também inspirou a tese de doutorado de

16 Cf. a entrevista de Jerônimo Muniz em Bachini e Chicarino (2018).

Nelson do Valle Silva, que havia lido um artigo de Hasenbalg e resolveu analisar empiricamente a hipótese. Hasenbalg (1979) comprovou sua hipótese principalmente analisando a situação geográfica de pretos e pardos que se concentravam nas áreas mais pobres do país. Além disso, o autor afirma e mostra, a partir da análise de dados sobre mobilidade social, que a desigualdade racial se reproduzia na sociedade contemporânea.

Nelson do Valle Silva defendeu sua tese um ano antes, em 1978, na Universidade de Michigan, onde a princípio pretendia estudar com Otis Dudley Duncan, mas acabou sendo orientado por William Mason porque o primeiro se transferiu para outra universidade pouco depois da chegada do brasileiro. Mason é um sociólogo e metodólogo bastante conhecido nos Estados Unidos, que orientou alguns trabalhos importantes na área de estratificação social, como o doutorado de Robert Mare – que veio a desempenhar um papel central nas áreas de sociologia e demografia nos Estados Unidos (HAUSER, 2017). O trabalho de Valle Silva usou o Censo de 1960 e analisou as disparidades raciais em termos da desigualdade de renda mostrando, pela primeira vez com dados agregados de alta qualidade, que havia forte desigualdade racial.

O encontro dos pesquisadores no início da década de 1980, logo depois da defesa de suas teses, culminou no início de uma frutífera colaboração. Outro fato fundamental foi a disponibilidade de dados do Instituto Brasileiro de Geografia e Estatística (IBGE) contendo informações sobre cor/raça, bem como sobre diversos outros temas. Em particular, a Pesquisa Nacional por Amostra de Domicílios (PNAD) de 1976, que havia coletado em seu suplemento informações

relevantes sobre raça, mobilidade e outros aspectos do mercado de trabalho e da escolaridade dos respondentes, visava particularmente avaliar a relevância dos dados sobre raça que não haviam sido coletados no Censo de 1970¹⁷.

Como consequência dos resultados da PNAD de 1976, o Censo do IBGE passou a coletar as informações sobre raça a partir de 1980. Além disso, a PNAD de 1973 havia coletado dados de mobilidade social por sugestão do sociólogo paulista José Pastore, que havia concluído seu doutorado orientado por Archibald Haller, outro autor importante na área de estratificação social e professor da Universidade de Wisconsin, em Madison. Pastore publicou seu livro sobre mobilidade social em 1979 (PASTORE, 1979). Em suma, além do interesse em comum sobre o tema das relações raciais e da expertise metodológica, Valle Silva e Hasenbalg tiveram a sorte de ter acesso a grandes bancos de dados de altíssima qualidade assim que começaram a colaborar. Nas palavras de Merton (1987), os pesquisadores se encontravam diante de *strategic research materials*, informações que abrem fronteiras totalmente novas para a pesquisa em determinada ciência. Pela primeira vez, a sociologia brasileira tinha à disposição dados de tal qualidade, e podia portanto responder a perguntas importantes sem se valer apenas da interpretação ensaística ou de pesquisas qualitativas. Assim, iniciava-se a tradição da sociologia como ciência das populações no Brasil.

As contribuições de Hasenbalg e Valle Silva se deram principalmente por sua filiação aos estudos de estratificação e mobilidade social, que é internacionalmente reconhecida como uma das principais áreas

17 Para uma explicação desmistificadora dessa ausência dos dados sobre raça no Censo de 1970, cf. a entrevista de Nelson do Valle Silva em Barbosa et al. (2013).

de pesquisa empírica quantitativa na sociologia. Assim, os trabalhos que publicaram nos quatro livros mencionados (HASENBALG; SILVA, 1988, 1992, 2003; HASENBALG; SILVA; LIMA, 1999) seguem um esquema analítico desenhado para estudar a reprodução e/ou superação das desigualdades ao longo dos ciclos de vida dos indivíduos e dos períodos históricos. No que diz respeito aos ciclos de vida, o esquema parte das origens familiares dos indivíduos e os segue ao longo de diversas etapas da vida, desde a infância, passando pela escolaridade, pela entrada no mercado de trabalho, escolha conjugal, formação de uma nova família, culminando em sua situação econômica, social e cultural na vida adulta. Nesse processo, os capitais econômico, social e cultural desempenham papel central nas chances desiguais de mobilidade social. Em termos dos períodos históricos, a principal preocupação se volta aos efeitos do avanço da sociedade capitalista (industrialização, urbanização etc.) nas condições socioeconômicas das populações de homens, mulheres, brancos, pardos, pretos e outros grupos sociologicamente relevantes. Em todos os casos, as análises empíricas são os parâmetros que guiam a compreensão dos processos de mudança. Nesse sentido, classificamos os diversos artigos reunidos nos quatro livros (HASENBALG; SILVA, 1988, 1992, 2003; HASENBALG; SILVA; LIMA, 1999) em temas claramente centrais da área de estratificação e mobilidade social e que seguem o esquema analítico mencionado: (1) industrialização e mudança social; (2) origens sociais; (3) desigualdade de renda e pobreza; (4) seletividade marital e miscigenação; (5) estratificação educacional; e (6)

mobilidade social. A seguir descrevemos as contribuições em cada um deles.

Industrialização e mudança social

O primeiro artigo em que Hasenbalg e Valle Silva colaboraram foi “Industrialização e estrutura de emprego no Brasil: 1960-1980” (HASENBALG; SILVA, 1988), que foi preparado para um seminário realizado em 1983. Esse trabalho faz uso da PNAD de 1973 e dos Censos de 1960, 1970 e 1980 para analisar as enormes mudanças na estrutura de emprego e mobilidade social ocorridas no período de mais rápida industrialização na história brasileira. Seu foco é a distribuição da população em diferentes grupos ocupacionais e de renda, bem como a ampliação da participação feminina no mercado de trabalho e nas desigualdades, particularmente a racial, presentes nesse período. Os autores concluem que o período foi marcado por um aumento das relações de trabalho capitalistas, principalmente pela expansão do setor industrial e o encolhimento do rural; por forte mobilidade intergeracional estrutural, ou seja, a mobilidade induzida pelas mudanças na estrutura ocupacional entre as gerações de pais e filhos adultos indica que este processo talvez tenha sido um dos mais intensos em todo o mundo; pela expansão da participação feminina no mercado de trabalho; e pelo aumento considerável da desigualdade social. As conclusões desse trabalho foram confirmadas e ampliadas para novas décadas em Ribeiro (2007) e Silva (2008). Finalmente, cabe ressaltar que o foco foi principalmente nas mudanças na estrutura ocupacional, que sempre foram um tema caro à sociologia quantitativa.

Em outros dois textos, Carlos Hasenbalg¹⁸ analisou o tema da industrialização e, mais especificamente, as desvantagens que os negros sofreram ao longo do processo no Brasil. Nesses textos, Hasenbalg também voltou a desenvolver estudos históricos que marcaram sua formação e tese de doutorado. Ambos os artigos se valem de alguns trabalhos históricos – dos diversos Censos brasileiros, cujas tabulações estão disponíveis para o período de 1890 a 1950, e da PNAD de 1987 – para mostrar que a proletarianização da população negra foi tardia, e que houve manutenção da desigualdade racial ao longo de todo o período. A forte imigração europeia no início do século XX foi absorvida como mão de obra proletária, principalmente em São Paulo, mas também no Rio de Janeiro e na região Sul, enquanto os negros foram relegados a posições de trabalho doméstico ou em outros precários, além de terem se concentrado nas regiões mais pobres do país, no Norte e Nordeste. Assim, a dinâmica populacional e as políticas de Estado (em termos de imigração, entre outras) fizeram com que a população negra não fosse capaz de aproveitar as novas oportunidades que surgiam com a industrialização. Em suma, Hasenbalg formula sua famosa hipótese de que a desigualdade racial foi reproduzida ao longo do período de industrialização e não pode ser explicada simplesmente por uma herança da sociedade escravista, como sugeria Florestan Fernandes (1965).

O tema dos efeitos da industrialização na estratificação social sempre foi central nos estudos sociológicos. A teoria da modernidade preconizava uma crescente diminuição da desigualdade de condições e oportunidades com o avanço da sociedade industrial. O pressuposto de eficiência burocrática e

tecnológica desse tipo de sociedade tornaria necessária a alocação dos indivíduos às posições na estrutura de estratificação levando em conta apenas suas capacidades e não sua origem de classe, raça ou gênero (TREIMAN, 1970). Várias teorias e críticas posteriores mostraram a limitação dessa abordagem na medida em que as desigualdades foram mantidas mesmo nas sociedades industriais mais avançadas¹⁹.

Os trabalhos de Hasenbalg e Valle Silva revelam que essa continuidade nas desigualdades raciais foi mantida ao longo do processo de industrialização no Brasil.

Origens sociais e classificação da estrutura ocupacional

O tema do efeito das origens de classe social, mais especificamente das características das famílias de origem, nas chances de vida dos indivíduos está presente em vários trabalhos de Hasenbalg e Valle Silva. De fato, a perspectiva sociológica de estudos sobre desigualdades sociais sempre se preocupou com a reprodução intergeracional. Nesta seção, no entanto, apresentamos brevemente três trabalhos mais diretamente relacionados ao tema publicados pelos autores em 2003, embora diversos outros textos que discutiremos posteriormente sejam igualmente relevantes. Abordaremos também o esquema de classificação ocupacional geralmente usado nas pesquisas de Hasenbalg e Silva.

No primeiro capítulo do livro *Origens e destinos* (HASENBALG; SILVA, 2003), Nelson do Valle Silva descreve o os esquemas analítico e de classes usados na pesquisa. O primeiro é basicamente aquele descrito anteriormente, presente em diversos outros

18 “O negro na indústria: proletarianização tardia e desigual” em Hasenbalg e Silva (1992), e “Desigualdades raciais no Brasil” em Hasenbalg e Silva (1988), este originalmente publicado em 1977.

19 Cf. Goldthorpe (1985).

trabalhos de ambos, ou seja, é um esquema que visa estudar a desigualdade nas diferentes fases do ciclo de vida dos indivíduos. O segundo esquema também já havia sido utilizado – na verdade, sua primeira versão foi apresentada em um artigo não publicado (SILVA, 1992), e já se encontrava em outros trabalhos, como o de Scalón (1999) e, de forma ligeiramente modificada para permitir comparações internacionais, o de Ribeiro (2007). Recentemente o esquema também foi usado por Reis (2017). Nessa classificação ocupacional, 16 grupos são distinguidos: (1) profissionais liberais; (2) dirigentes; (3) proprietários empregadores; (4) outros profissionais universitários; (5) supervisores do trabalho manual; (6) ocupações técnicas e artísticas; (7) empresários por conta própria; (8) ocupações não manuais de rotina; (9) trabalhadores da indústria moderna; (10) trabalhadores dos serviços gerais; (11) vendedores ambulantes; (12) trabalhadores da indústria tradicional; (13) trabalhadores dos serviços manuais; (14) trabalhadores dos serviços domésticos; (15) proprietários empregadores rurais; e (16) trabalhadores rurais. Embora outros esquemas de classe ou ocupacionais tenham sido usados na sociologia brasileira, há forte correlação entre todos eles e o esquema proposto por Nelson do Valle Silva (CARDOSO; PRETECEILLE, 2017; SANTOS, 2002). A principal vantagem deste é que ele permite classificar os dados brasileiros tanto para as ocupações dos pais (obtidas a partir de questões retrospectivas sobre sua ocupação) quanto para as dos respondentes das pesquisas, ao passo que outros esquemas não podem classificar as ocupações dos pais por usarem outras variáveis dos bancos de dados (CARDOSO; PRETECEILLE, 2017; SANTOS, 2002). Pensamos que todos os esquemas são igualmente bons e ruins, sendo adequados para diferentes tipos de análises.

De qualquer forma, Carvalhaes (2015) fez o único trabalho que conhecemos testando

a *confiabilidade* e a *validade* do esquema de classificação ocupacional usado em Ribeiro (2007). A partir dessa metodologia, a sociologia brasileira mapeou algumas regularidades populacionais importantes, como a expansão das ocupações urbanas e o encolhimento das rurais, e a divisão de trabalho entre homens e mulheres em diferentes tipos de ocupação. É importante notar que essas regularidades não são específicas de nossa sociedade, embora as mudanças no Brasil pareçam ser mais recentes.

Nos dois capítulos seguintes do livro (HASENBALG; SILVA, 2003), Hasenbalg analisa as características das famílias brasileiras e a situação das crianças usando dados das PNAD de 1981, 1990 e 1999, descrevendo algumas tendências importantes das décadas de 1980 e 1990. Entre estas tendências se destacam a diminuição do número de filhos nas famílias e o aumento do nível educacional dos adultos ao longo das coortes de idade (que certamente afetaram positivamente as condições de vida das gerações seguintes), e a diminuição da desigualdade racial e de gênero (embora essas permaneçam importantes). Analisando a situação das crianças de 5 e 6 anos nas décadas de 1980 e 1990, Hasenbalg (Ibidem) indica algumas melhoras, como a expansão dos percentuais de crianças com acesso a creches, e situações mais difíceis, como o aumento da proporção de crianças em famílias contando apenas com a presença das mães.

Nesses três trabalhos os autores se limitam a descrever algumas regularidades e tendências da estratificação social acerca das estruturas ocupacional e familiar brasileiras, não buscando entretanto desvendar os mecanismos causais mais gerais. As pesquisas sobre a estrutura familiar e as condições de vida na primeira infância são cada vez mais proeminentes nos estudos de estratificação social (ALMOND; CURRIE, 2011), configurando-se como um campo fértil para trabalhos sociológicos futuros.

Desigualdade de renda e pobreza

Os temas da desigualdade de renda e da pobreza são frequentemente citados em trabalhos ensaísticos sobre o Brasil, embora a contribuição da sociologia brasileira seja limitada nessa área quando comparada, por exemplo, às da economia. Os sociólogos brasileiros frequentemente declaram que o Brasil tem muitos pobres e é muito desigual, para em seguida proporem suas interpretações ensaísticas mirabolantes sobre as causas dessa situação – interpretações em geral erradas na medida em que desprezam as análises empíricas. Um exemplo típico é o trabalho de Jessé Souza (2006), que insiste em desqualificar, sem argumento ou evidência razoável, as pesquisas sobre desigualdade de renda. Seu estudo é um exemplo de pensamento “tipológico” ou “holístico”, que pensa o mundo social de forma completamente determinística, ou seja, não reconhecendo a variabilidade fundamental que o caracteriza. Como os leitores podem perceber, não simpatizamos com esse tipo de abordagem, que certamente não é a de Carlos Hasenbalg e de Nelson do Valle Silva, tendo em vista que ambos os pesquisadores contribuíram com trabalhos empíricos importantes sobre esses temas e são representantes do pensamento populacional – que embora pareça menos sedutor que o tipológico, seguramente é mais preciso e sutil.

No texto “Os deserdados do milagre” (HASENBALG; SILVA, 1988), Silva cita pesquisas que indicam queda bastante acentuada da pobreza nas décadas de 1970 e 1980. Em seguida, parte para análise detalhada das características dos pobres e miseráveis em 1985. O trabalho revela que as condições ainda eram de muito desamparo a despeito das enormes mudanças das décadas anteriores. Alguns anos mais tarde, Silva (1992) escreveu outro artigo, usando então os dados da PNAD de 1988

sobre as chances de os indivíduos estarem na situação de pobreza, trabalho que foi pioneiro por ser uma das primeiras publicações brasileiras a mostrar que pardos e pretos tinham mais chance de estar em situação de pobreza do que brancos. Usando uma regressão logística e diversas variáveis de controle, as análises indicaram que pardos e principalmente pretos têm chances muito maiores do que brancos de se encontrarem em tal situação, regularidade essa que permanece na sociedade brasileira atual. De fato, Hasenbalg (HASENBALG; SILVA, 2003) mostrou que em 1999 cerca de 62,5% dos pobres ainda eram não brancos (pretos e pardos).

Em outro artigo, Silva (HASENBALG; SILVA; LIMA, 1999) mostrou que as desigualdades raciais estão presentes em todo o processo de estratificação, inclusive no acesso à escolaridade, a melhores ocupações e, finalmente, à renda do trabalho. Nesse artigo foram estimadas diversas regressões e a partir das regularidades encontradas foram sugeridos dois mecanismos explicativos para a desigualdade racial, que também estão presentes em outros trabalhos dos autores. O primeiro é a discriminação racial; diante de tantas evidências e controles, bem como de simulações usando o método de Oaxaca (SILVA, 1985), não há como negar que ela desempenha um importante papel na manutenção das desigualdades raciais, de renda, dentre outras. Ainda que todas as características observadas do grupo populacional sejam idênticas (região de moradia, idade, sexo, escolaridade e outras), a discriminação se evidencia porque pretos e pardos têm rendimentos sistematicamente inferiores a brancos.

O segundo é o mecanismo da acumulação de desvantagens, ou seja, o processo pelo qual pretos e pardos acumulam desvantagens ao longo de seus ciclos de vida que contribui para a na manutenção das desigualdades raciais. Na acumulação de (des)vantagens, certas posições

na estrutura social em determinado ponto no ciclo de vida são altamente associadas a certos tipos de (des)vantagens em períodos posteriores (CARVALHAES; CHOR, 2016). Como pretos e pardos têm maior probabilidade de nascerem em regiões menos economicamente desenvolvidas, em famílias com menor escolaridade e assim por diante, o estado de (des)vantagem observado para o grupo em determinado período deve ser interpretado como um processo histórico, interconectado e intergeracional, em que as condições do grupo em determinada época dependem necessariamente das condições anteriores.

Em 2003, Nelson do Valle Silva voltou a escrever sobre desigualdades de renda (HASENBALG; SILVA, 2003), analisando as tendências da desigualdade de renda entre indivíduos e entre grupos sociologicamente relevantes (grupos ocupacionais, raças e gêneros) e concluiu que houve aumento da desigualdade na década de 1980 e leve queda na de 1990. De certa forma, o texto antecipa achados sobre a queda das desigualdades, que ficariam mais evidentes em trabalhos publicados ao longo das décadas de 2000 e 2010. A queda identificada, no entanto, deve ser pensada com cautela, visto que os níveis de desigualdade ainda eram (e ainda são) extremamente elevados. As fortes disparidades entre grupos também indicam que as desigualdades de renda têm fundamentos estruturais e institucionais e são duráveis.

Esses trabalhos sobre pobreza e desigualdade de renda não só foram capazes de identificar diversas regularidades, como também levaram à formulação de alguns mecanismos causais que provavelmente contribuem para manter tal estado de coisas. São certamente excelentes exemplos da sociologia pensada como ciência das populações e um verdadeiro antídoto à tradição ensaísta e de “pensamento tipológico” ou “holístico” que ainda persiste na sociologia brasileira.

Seletividade marital e miscigenação

Já mencionamos o texto sobre seletividade marital de Valle Silva (1987), que mostra que a barreira ao casamento entre pretos e pardos é mais forte do que entre brancos e pardos. Esse estudo desempenha um papel importante na contribuição de Hasenbalg e Valle Silva para o entendimento das relações raciais no Brasil. Ao mostrarem que as barreiras ao casamento inter-racial não eram tão fortes e que havia muita miscigenação (HASENBALG; SILVA; BARCELOS, 1989), os autores sugerem que as relações de sociabilidade racial não são tão rígidas quanto as de estratificação social ou desigualdade em termos de educação, ocupação e renda.

Assim, eles antecipam uma ideia posteriormente formulada por Telles (2004) sobre a diferença entre relações verticais e horizontais nas relações raciais no Brasil, em que a desigualdade racial em termos econômicos e educacionais é extremamente rígida, em contraste à sociabilidade racial menos rígida em outras esferas, baseada talvez em diferenças de status, mas não de classe ou econômicas. Essa ideia é central na interpretação de Telles (2004), cujo livro foi premiado pela American Sociological Association, em 2006. Na verdade é uma tese que já estava presente nos trabalhos de Hasenbalg e Valle Silva, como podemos ver na seguinte passagem:

contrariamente ao que sabemos das distâncias socioeconômicas entre os grupos de cor, o grupo pardo ocupa uma posição realmente intermediária entre brancos e pretos. De fato, parece mesmo estar ligeiramente mais próximo do grupo branco do que do grupo preto. O padrão dicotômico observado nos estudos socioeconômicos, isto é, brancos claramente diferenciados de pardos e pretos, estes últimos ocupando posição muito próxima entre si, claramente não se reproduz nos padrões de distâncias sociais implícitas nos padrões de casamento inter-racial no Brasil. Estes resultados sugerem que as distâncias sociais no casamento não são de natureza

primariamente socioeconômica, seguindo talvez outras hierarquias como, por exemplo, hierarquias de status ou de prestígio (SILVA, 1987, p. 50).

Além dessa importante contribuição para a interpretação das relações raciais no Brasil, os estudos sobre seletividade marital, em termos dos níveis educacionais de maridos e esposas, indicam uma tendência de casamentos homogâmicos para pessoas com níveis educacionais mais altos, que poderia levar a uma concentração de recursos nas famílias mais ricas (SILVA, 1987).

Várias regularidades foram encontradas nesses trabalhos sobre seletividade marital, tema de grande importância por tratar da criação de novas famílias, que constituirão a estratificação social das próximas gerações. Falta, no entanto, definir os mecanismos que explicam essas regularidades, bem como análises que usem dados longitudinais e que nos ajudem a entender a dinâmica dos casamentos e das separações (tema este ainda pouco estudado no Brasil).

Estratificação educacional

O processo de estratificação educacional foi tratado em vários trabalhos de Hasenbalg e Silva (1992, 2003; HASENBALG; SILVA; LIMA, 1999). Esse tipo de análise visa estudar a desigualdade de oportunidades educacionais²⁰, ou seja, o efeito das características de origem de classes (entre outras) nas chances de progredir no sistema educacional. Para tanto são usados modelos logit condicionais que estimam as chances de transição ao longo do sistema educacional. Por exemplo, primeiro se estimam as chances de concluir o ensino elementar e depois, para aqueles que

concluíram, são estimadas as chances de entrar no ensino médio, e assim por diante. De fato, em diversos trabalhos Hasenbalg e Silva (1988, 1992, 2003; HASENBALG; SILVA; LIMA, 1999) sugerem que o acesso à educação é uma das principais barreiras à mobilidade social de pardos e pretos.

Usando dados da PNAD de 1982, Hasenbalg e Silva (1992, 1999) mostram que pretos e pardos tinham resultados educacionais muito inferiores aos alcançados por brancos. Além disso, indicam que embora houvesse análises sobre desigualdades educacionais no Brasil, praticamente não havia estudos que focalizassem as questões raciais relacionadas. Nesses trabalhos, os autores mostram, pela primeira vez, que os brancos, em particular, tinham chances muito maiores do que pardos e pretos de entrar na escola, de completar a primeira série do ensino primário (atual fundamental), e de concluí-lo. Em um estudo posterior, os sociólogos Hasenbalg e Silva (HASENBALG; SILVA; LIMA, 1999) ampliaram seu modelo de transições educacionais usando diversas variáveis de controle relacionadas à situação socioeconômica das famílias de origem dos estudantes, que, mesmo assim, confirmaram as desigualdades raciais. Essas análises foram mais uma vez ampliadas por Silva (HASENBALG; SILVA, 2003), que chegou a outras duas conclusões importantes: (1) as vantagens de brancos aumentavam nas transições mais altas do sistema educacional; e (2) as desigualdades raciais diminuíram entre 1990 e 1999, principalmente porque houve uma enorme expansão do ensino fundamental naquele período. Esses resultados, ou regularidades, foram confirmados por todos os estudos posteriores sobre o tema (MARTELETO; CARVALHAES; BRITO, 2017; RIBEIRO,

20 Robert Mare (1980, 1981), colega de doutorado de Nelson do Valle Silva em Michigan, foi quem propôs inicialmente essa abordagem.

2009, 2011) e constituem evidência fundamental para entendermos as relações raciais brasileiras.

O passo seguinte no processo de estratificação social é a entrada no mercado de trabalho, ou a transição da escola para o trabalho. Carlos Hasenbalg (HASENBALG; SILVA, 2003) mostrou que os brasileiros começam a trabalhar muito cedo e com pouca escolaridade (situação essa que se modificou lentamente entre 1981 e 1999), além de fortes desigualdades em termos de classe de origem nessa transição.

Em suma, os trabalhos de Hasenbalg e Silva sobre desigualdade de resultados e oportunidades educacionais foram os primeiros na sociologia brasileira, mostrando importantes regularidades desconhecidas até então. Certamente esses estudos influenciaram todo o debate sobre desigualdades raciais no sistema educacional – que se iniciava naquela época e hoje está muito em voga –, e inclusive contribuíram para levar adiante as políticas de ação afirmativa, já implementadas na agenda nacional. Na época em que escreveram ainda não havia evidências e o debate era praticamente inexistente.

Mobilidade social

Na sociologia, os estudos de mobilidade social seguem duas correntes, ambas exploradas por Hasenbalg e Silva. A primeira analisa tabelas de mobilidade cruzando grupos ocupacionais de pais e de filhos adultos, além de distinguir taxas absolutas e relativas de mobilidade social. A segunda enfoca o processo de realização socioeconômica usando escalas que hierarquizam todas as ocupações em uma variável contínua. Assim, tanto as ocupações dos pais quanto as dos filhos são classificadas em duas variáveis, aplicadas em modelos de regressão linear para verificar o

grau de correlação existente – quanto maior for a associação, menor a mobilidade.

Usando dados da PNAD de 1973 e analisando as taxas relativas de mobilidade social entre coortes de idade, Silva e Roditi – trabalho também publicado como um dos capítulos de Hasenbalg e Silva (1988) – concluíram não haver mudança ao longo das coortes. O estudo também investigou o efeito mediador que a educação dos indivíduos desempenhava na associação entre classes de origem e destino, chegando à mesma conclusão. As taxas relativas de mobilidade são uma forma de analisar a desigualdade de oportunidades, descontando o efeito das mudanças estruturais ou marginais. É importante lembrar que, naquela época, Nelson do Valle Silva estava usando o que havia de mais sofisticado na área de estratificação social em termos metodológicos, ou seja, importantes artigos nessa área estavam sendo publicados justamente nesse período e chegavam a conclusões semelhantes para países como Estados Unidos, França, Inglaterra etc. O artigo em questão não analisou a desigualdade racial, mas posteriormente Silva (SILVA; PASTORE, 1999) usou os dados da PNAD de 1996 para investigar a desigualdade de oportunidades de mobilidade social entre brancos, pardos e pretos e sugeriu que havia importantes vieses de raça, ou seja, havia desigualdade de oportunidades em termos raciais.

Em dois outros artigos, Hasenbalg e Silva (1988; HASENBALG; SILVA; LIMA, 1999) analisaram os diferenciais em termos de raça nas taxas absolutas de mobilidade social, mostrando que brancos tinham muito mais mobilidade ascendente do que pretos e pardos. Diante dessas evidências, e também a partir de análises do processo de realização socioeconômica, sobre o efeito das características dos pais (ocupação e educação do pai) e do local de moradia (região do país e área – rural ou urbana) no status ocupacional e na renda dos filhos, Nelson do Valle Silva resumiu e

concluiu a visão de ambos os sociólogos na seguinte passagem:

Mostrou-se que ao longo de todo o ciclo de vida socioeconômico, negros e mulatos sofrem desvantagens geradas por atitudes discriminatórias, desvantagens que se acumulam na geração de chances de vida profundamente inferiores àquelas desfrutadas por brancos. Negros e mulatos estão sujeitos à discriminação no processo escolar, à discriminação no emprego, bem como à discriminação salarial. Essas desvantagens competitivas agem cumulativamente, explicando a maior parte das diferenças monetárias entre brancos e não brancos chegando a atingir mais de 50% dos rendimentos percebidos por brancos e não brancos (SILVA, 1981).

Assim, os autores negam veementemente tanto o fato de que a desigualdade racial seja fruto do passado escravista quanto a ideia de “democracia racial”. Essas conclusões, bem como diversas outras apresentadas anteriormente, só foram alcançadas porque os autores adotaram a perspectiva da sociologia como ciência das populações. Ou seja, primeiro identificaram evidências sobre as desigualdades entre brancos e não brancos ao longo do ciclo de vida dos indivíduos e, em seguida, propuseram mecanismos causais para explicar os fatos que haviam estabelecido. Em particular, eles propõem dois tipos de mecanismos: discriminação racial e acumulação de desvantagens. Em suma, Carlos Hasenbalg e Nelson do Valle Silva não só revolucionaram os estudos sobre as relações raciais, como também iniciaram a tradição da sociologia como ciência das populações no Brasil. Devemos muito a ambos e a seu pioneirismo.

Conclusão

O principal argumento ao longo deste artigo que, ao adotarem as abordagens da sociologia como ciência das populações e dos estudos de estratificação social, Carlos Hasenbalg e Nelson do Valle Silva fizeram

contribuições importantes para explicar não só as relações raciais no Brasil, como também diversos temas relacionados à desigualdade social e a sua reprodução. Em vez de resumir os argumentos apresentados ao longo deste artigo, aproveitamos a conclusão para discutir a relação entre política e conhecimento nas pesquisas de Hasenbalg e Silva. Para tanto, é importante contextualizar os trabalhos desses autores no período em que começaram a colaborar de forma mais intensa, na década de 1980.

Naquela época era mais comum do que dizer que não havia discriminação e desigualdade racial no Brasil, mas apenas desigualdades econômicas e de classe. De fato, os trabalhos de Hasenbalg e Silva foram os primeiros a mostrar de forma contundente, e usando bancos de dados representativos da população brasileira como um todo, que havia forte desigualdade racial no Brasil se sobrepondo à de classe ou econômica. Aquele período também foi de grande inovação em outras áreas das ciências sociais e da história no Brasil. Em 1988, a abolição da escravidão completava 100 anos, e diversos cientistas sociais se voltavam para entender a situação dos negros na sociedade brasileira.

Como aluno de graduação em ciências sociais, vivenciei aquele momento através da pesquisa de iniciação científica realizada no Núcleo da Cor do Instituto de Filosofia e Ciências Sociais da Universidade Federal do Rio de Janeiro (IFCS-UFRJ). No Centro de Estudos Afro-Asiáticos (CEAA), da Universidade Candido Mendes, também havia muitos estudos sendo desenvolvidos sob a coordenação de Carlos Hasenbalg, diretor do Centro, e também de Nelson do Valle Silva, que era um colaborador constante. Em várias outras instituições acadêmicas o tema das relações raciais era estudado e discutido. No ano de 1988 foi elaborado um curso (organizado por diversos professores da UFRJ,

UERJ, PUC-Rio, Iuperj, entre outras instituições) sobre a literatura de relações raciais no Brasil, que incluía autores da história, antropologia e sociologia, ministrado em diversas instituições seguindo mais ou menos o mesmo programa.

Paralelamente, havia ONG como a Federação de Órgãos para Assistência Social e Educacional (Fase) coordenando campanhas para que as pessoas declarassem no Censo e nas pesquisas do IBGE sua cor de forma fidedigna – o lema da campanha era: “não deixe sua cor passar em branco”. Houve inclusive um momento em que o IBGE colocou em dúvida a validade da questão racial em suas pesquisas e Carlos Hasenbalg, que era muito próximo ao movimento negro, espalhou a notícia de forma que os militantes se mobilizaram junto ao IBGE a favor da manutenção da questão sobre raça ou cor. Em suma, havia muita agitação política e acadêmica em torno da data que marcava os 100 anos da abolição da escravidão.

Foi nesse contexto altamente politizado que os trabalhos de Hasenbalg e Silva foram produzidos. Embora tivessem opiniões diferentes sobre política partidária, ambos concordavam que os excessos de desigualdades de condições e oportunidades deveriam ser entendidos para que pudessem ser combatidos. Não há dúvida de que essa abordagem e esse contexto são altamente políticos, ou seja, o conhecimento que buscavam era inspirado por preocupações políticas e morais, como aliás é muito comum nas ciências sociais. Isso não significa, no entanto, que as preocupações políticas e os valores dos pesquisadores impeçam uma avaliação objetiva das evidências encontradas. De fato, o que torna o trabalho sociológico objetivo

é o rigor nos métodos e na avaliação das evidências.

Os trabalhos de Hasenbalg e Silva são um bom exemplo desse rigor metodológico sendo aplicado a problemas sociológicos de interesse público. Ao investigarem a superação e a reprodução das desigualdades, em particular as raciais, ao longo das gerações e dos ciclos de vida dos brasileiros, os autores adotaram uma abordagem (sociologia das populações e da estratificação social) e um método quantitativo, com grandes bancos de dados representativos da população que permitiam uma enorme precisão na avaliação das evidências. Em outras palavras, seus valores e opiniões políticas não garantiam a validade da pesquisa, mas sim o cuidado metodológico e a abordagem da sociologia como ciência das populações adotados. Se as análises empíricas dos dados não tivessem revelado regularidades na desigualdade racial ao longo do processo de modernização da sociedade brasileira, os autores não teriam como comprovar as hipóteses que defendiam.

Hoje em dia ouvimos nos corredores dos cursos de ciências sociais que o conhecimento é sempre político, o que parece uma obviedade por vezes afirmada com um tom de denúncia ou de partidarismo. Tudo se passa como se tivéssemos que afirmar nossos interesses políticos para validar nosso conhecimento e como se não fosse possível analisar objetivamente a realidade. Assim, muitos estudantes e profissionais passam a discussões intermináveis sobre as possibilidades do conhecimento, em vez de se dedicarem a análises rigorosas que procurem explicar e descrever as regularidades no mundo social.

Charles Tilly, em uma entrevista dada em 2007²¹, resume bem o imbróglio em que

21 Entrevista disponível em: <<https://www.youtube.com/watch?v=TEFWsJlBP2o>>. Acesso em: 20 ago. 2018.

se encontra o conhecimento nas ciências sociais: “Grande parte do debate hoje em dia na sociologia é sobre [...] divergências filosóficas. Ou seja, divergências sobre epistemologia e ontologia ao invés de serem sobre explicações. E eu não acho essa uma boa posição para se estar”²². Em outras palavras, Tilly argumenta que o debate incessante sobre as possibilidades de conhecimento nas ciências sociais, que inclui a relação entre política e conhecimento, não é frutífero. Deveríamos, ao contrário, focar na explicação das regularidades encontradas no mundo social. Os trabalhos de Hasenbalg e Silva visavam descrever e explicar as regularidades relacionadas às desigualdades, em particular as raciais, no Brasil. Essas descrições e explicações foram inspiradas por opiniões políticas, mas não contaminadas por elas. O conhecimento que produziram é válido porque eles avaliaram com rigor as evidências que foram capazes de encontrar.

Um outro aspecto importante a ser frisado nesta conclusão diz respeito à defesa da sociologia como ciência das populações realizada ao longo do artigo. Procuramos mostrar que os trabalhos de Hasenbalg e Silva se filiam a essa tradição sociológica, bem como que essa abordagem se caracteriza pelo reconhecimento da enorme variabilidade existente entre os indivíduos e as relações sociais que compõem a sociedade que estudamos. As regularidades

devem ser buscadas no nível populacional. A defesa dessa abordagem procurou mostrar a importância da metodologia estatística para a avaliação das evidências e para a descrição de regularidades. Essa defesa não implica negar outras metodologias; análises qualitativas, históricas e etnográficas de casos específicos são fundamentais e, em última instância, também levam em conta a variabilidade fundamental que caracteriza as unidades de análise (indivíduos, famílias, bairros, cidades etc.) do mundo social.

Hasenbalg e Silva também reconhecem em seus trabalhos, e nas colaborações com outros pesquisadores, sua importância para a pesquisa. Novamente ressaltamos o contexto em que os autores seu conhecimento, frutífero também para a antropologia, ciência política e história. Novas pesquisas nessas áreas também estavam sendo desenvolvidas e seus pesquisadores participavam dos debates em que Hasenbalg e Silva se inseriam. Foi um período de muita inovação não só para a ciência das populações, como também para outros campos do conhecimento nas ciências sociais, que também se voltavam para o estudo das relações raciais. Hasenbalg e Silva fizeram contribuições fundamentais e ajudaram a consolidar a sociologia quantitativa no Brasil. Devemos muito a nossos professores, e seguiremos ampliando esses caminhos que eles ajudaram a abrir.

Referências

- ABBOTT, A. What do cases do? Some notes on activity in sociological analysis. In: RAGIN, C. C.; BECKER, H. S. (Eds.). *What is a case?* Exploring the foundations of social inquiry. Cambridge, UK: Cambridge University Press, 1992. p. 53-82.

22 No original: “Too much of the debate in sociology today is about [...] philosophical disagreement. That is, disagreement on epistemology and ontology rather than about explanation. And I don't think this is a good position to be in”.

- ALMOND, D.; CURRIE, J. Human capital development before age five. In: ASHENFELTER, O.; CARD, D. (Eds.). *Handbook of labor economics*. Amsterdam: Elsevier, 2011. v. 4B, p. 1315-1486.
- ANGRIST, J. D.; PISCHKE, J.-S. *Mastering metrics: the path from cause to effect*. Princeton: Princeton University Press, 2015.
- BACHINI, N.; CHICARINO, T. S. Os métodos quantitativos, por cientistas sociais brasileiros: entrevistas com Nelson do Valle Silva e Jerônimo Muniz. *Sociedade e Estado*, Brasília, DF, v. 33, n. 1, p. 251-279, 2018.
- BARBOSA, R. J. *A educação e a desigualdade de renda do trabalho: um enfoque sociológico*. 2017. 279 f. Tese (Doutorado em Sociologia) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2018.
- BARBOSA, R. J. et al. Ciências sociais, censo e informação quantitativa no Brasil: entrevista com Elza Berquó e Nelson do Valle Silva. *Novos Estudos Cebrap*, São Paulo, n. 95, p. 143-155, 2013.
- BOHRNSTEDT, G. W. Measurement models for survey research. In: MARSDEN, P. V.; WRIGHT, J. D. (Eds.). *Handbook of survey research*. 2. ed. Bingley: Emerald, 2010. p. 347-404.
- BOURDIEU, P.; WACQUANT, L. On the cunning of imperialism. *Theory, Culture and Society*, Thousand Oaks, v. 16, n. 1, p. 41-58, 1999.
- BREEN, R.; GOLDTHORPE, J. H. Explaining educational differentials: towards a formal rational action theory. *Rationality and Society*, Thousand Oaks, v. 9, n. 3, p. 275-305, 1997.
- CANO, I. *Introdução à avaliação de programas sociais*. Rio de Janeiro: Editora FGV, 2002.
- CARDOSO, A.; PRETECEILLE, E. Classes médias no Brasil: do que se trata? Qual seu tamanho? Como vem mudando? *Dados*, Rio de Janeiro, v. 60, n. 4, p. 977-1023, 2017.
- CARVALHAES, F. A. O. A tipologia ocupacional Erikson-Goldthorpe-Portocarero (EGP): uma avaliação analítica e empírica. *Sociedade e Estado*, Brasília, DF, v. 30, n. 3, p. 673-703, 2015.
- CARVALHAES, F.; CHOR, D. Posição socioeconômica, idade e condição de Saúde no Brasil. *Revista Brasileira de Ciências Sociais*, São Paulo, v. 31, n. 92, e319027, 2016.
- CHARLES Tilly interview: big questions. 10'26". *Daniel Leite*, YouTube, 18 dez. 2007. Disponível em: <<http://bit.ly/2MoMozf>>. Acesso em: 20 ago. 2018.
- CLOGG, C. The impact of sociological methodology on statistical methodology. *Statistical Science*, Beachwood, v. 7, n. 2, p. 183-196, 1992.
- COHEN, J.; DUPAS, P. Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *The Quarterly Journal of Economics*, Oxford, v. 125, n. 1, p. 1-45, 2010.

- COLEMAN, J. S. *Foundations of social theory*. Cambridge, MA: Belknap Press of Harvard University, 1990.
- COLEMAN, J. S. et al. *Equality of educational opportunity*. Washington, DC: U. S. Government Printing Office, 1966.
- DUNCAN, O. D. *Introduction to structural equation models*. New York: Academic Press, 1975.
- _____. *Notes on social measurement*. New York: Russell Age, 1984.
- _____. What if? *Contemporary Sociology*, Washington, DC, v. 21, n. 5, p. 667-668, 1992.
- ELSTER, J. *Nuts and bolts for the social sciences*. Cambridge, UK: Cambridge University Press, 1989.
- FERNANDES, F. *A integração do negro na sociedade de classes*. São Paulo: Dominus, 1965.
- GOLDTHORPE, J. H. On economic development and social mobility. *The British Journal of Sociology*, Hoboken, v. 36, n. 4, p. 549-573, 1985.
- _____. *Sociology as a population science*. Cambridge, UK: Cambridge University Press, 2016.
- GOODMAN, L. *The analysis of cross-classified data having ordered categories*. Cambridge, MA: Harvard University Press, 1984.
- HASENBALG, C. *Discriminação e desigualdades raciais no Brasil*. Rio de Janeiro: Graal, 1979.
- HASENBALG, C.; SILVA, N. V. *Estrutura social, mobilidade e raça*. São Paulo: Vértice, 1988.
- _____. *Relações raciais no Brasil contemporâneo*. Rio de Janeiro: Rio Fundo, 1992.
- _____. Race, schooling and social mobility in Brazil. *Ciência e Cultura*, Campinas, v. 51, n. 5-6, p. 457-463, 1999.
- _____. *Origens e destinos: desigualdades sociais ao longo da vida*. Rio de Janeiro: Topbooks, 2003.
- HASENBALG, C.; SILVA, N. V.; BARCELOS, L. C. Notas sobre miscigenação racial no Brasil. *Estudos Afro-Asiáticos*, Rio de Janeiro, n. 16, p. 188-197, 1989.
- HASENBALG, C.; SILVA, N. V.; LIMA, M. *Cor e estratificação social*. Rio de Janeiro: Contra Capa, 1999.
- HAUSER, R. A life in sociology. *Annual Review of Sociology*, Palo Alto, v. 43, p. 1-18, 2017.
- HEDSTRÖM, P. *Dissecting the social: on the principles of analytical sociology*. Cambridge, UK: Cambridge University Press, 2005.
- HEDSTRÖM, P.; BEARMAN, P. *The Oxford handbook on analytical sociology*. Oxford: Oxford University Press, 2009.

- HEDSTRÖM, P.; SWEDBERG, R. *Social mechanisms: an analytical approach to social theory*. Cambridge, UK. Cambridge University Press, 1998.
- HOLLAND, P. Statistics and causal inference. *Journal of the American Statistical Association*, Abingdon, v. 81, n. 396, p. 945-960, 1986.
- KISH, L. *Survey sampling*. New York: Wiley, 1965.
- LIEBERSON, S. Einstein, Renoir, and Greeley: some thoughts about evidence in sociology: 1991 presidential address. *American Sociological Review*, Washington, DC, v. 57, n. 1, p. 1-15, 1992.
- LOVEMAN, M.; MUNIZ, J. O.; BAILEY, S. R. Brazil in black and white? Race categories, the census, and the study of inequality. *Ethnic and Racial Studies*, Abingdon, v. 35, n. 8, p. 1466-1483, 2012.
- MARE, R. D. Social background and school continuation decisions. *Journal of the American Statistical Association*, Abingdon, v. 75, n. 370, p. 295-305, 1980. *American Statistical Association* 75:295-305, 1980.
- _____. Change and stability in educational stratification. *American Sociological Review*, Washington, DC, v. 46, n. 1, p. 72-87, 1981.
- MAYHEW, R. J. *Malthus: the life and legacy of an untimely prophet*. Cambridge, MA, Belknap Press of Harvard University Press, 2014.
- MAYR, E. *What evolution is?* London: Phoenix, 2001.
- MERTON, R. Three fragments from a sociologist's notebook: establishing the phenomenon, specified ignorance and strategic research materials. *Annual Review of Sociology*, Palo Alto, v. 13, p. 1-29, 1987.
- MORGAN, S.; WINSHIP, C. *Counterfactuals and causal inference: methods and principles for social research*. Cambridge, UK. Cambridge University Press, 2007.
- MUNIZ, J. O. Sobre o uso da variável raça-cor em estudos quantitativos. *Revista de Sociologia e Política*, Curitiba, v. 18, n. 36, p. 277-291, 2010.
- _____. Preto no branco? Mensuração, relevância e concordância classificatória no país da incerteza racial. *Dados*, Rio de Janeiro, v. 55, n. 1, p. 251-282, 2012.
- NOGUEIRA, O. *Preconceito de marca: as relações raciais em Itapetininga*. São Paulo: Edusp, 1998.
- PASTORE, J. *Desigualdade e mobilidade social no Brasil*. São Paulo: Edusp, 1979.
- PLATT, J. *A history of sociological research methods in America, 1920-1960*. Cambridge, UK. Cambridge University Press, 1996.

- RAUDENBUSH, S. W.; BRYK, A. S. *Hierarchical linear models: applications and data analysis methods*. Thousand Oaks: Sage, 2002.
- REIS, N. I. *Mudanças nas famílias brasileiras (1976-2012): uma perspectiva de classe e gênero*. Rio de Janeiro: Garamond, 2017.
- RIBEIRO, C. A. C. *Cor e criminalidade: estudo e análise da justiça no Rio de Janeiro (1900-1930)*. Rio de Janeiro: Editora UFRJ, 1995.
- _____. *Estrutura de classe e mobilidade social no Brasil*. Bauru: Edusc, 2007.
- _____. Contínuo racial, mobilidade social e “embranchamento”. *Revista Brasileira de Ciências Sociais*, São Paulo, v. 32, n. 95, e329503, 2017.
- RIBEIRO, C. A. C.; SILVA, N. V. Cor, educação e casamento: tendências da seletividade marital no Brasil, 1960 a 2000. *Dados*, Rio de Janeiro, v. 52, n. 1, p. 7-51, 2009.
- ROSENBAUM, P. Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, Oxford, v. 74, n. 1, p. 13-26, 1987.
- _____. *Observational studies*. New York: Springer-Verlag, 2002.
- SANTOS, J. A. F. *Estrutura de posição de classe no Brasil: mapeamento, mudanças e efeitos na renda*. Belo Horizonte: Editora UFMG, 2002.
- SCALON, M. C. *Mobilidade social no Brasil: padrões e tendências*. Rio de Janeiro: Revan, 1999.
- SHELLING, T. C. *Micromotives and macrobehavior*. New York: Norton, 1978.
- SCHWARTZMAN, L. F. Does money whiten? Intergenerational changes in racial classification in Brazil. *American Sociological Review*, Washington, DC v. 72, n. 6, p. 940-963, 2007.
- SILVA, G. M.; LEÃO, L. T. S. O paradoxo da mistura: identidades, desigualdades e percepção de discriminação entre brasileiros pardos. *Revista Brasileira de Ciências Sociais*, São Paulo, v. 27, n. 80, p. 117-133, 2012.
- SILVA, N. V. Updating the cost of not being white in Brazil. In: FONTAINE, P.-M. (Ed.). *Race, class and power in Brazil*. Los Angeles: Center for Afro-American Studies, 1985. p. 42-55.
- _____. Distância social e casamento inter-racial no Brasil. *Estudos Afro-Asiáticos*, Rio de Janeiro, n. 14, p. 54-84, 1987.
- _____. Uma classificação ocupacional para o estudo da mobilidade e da situação de trabalho no Brasil. [S.l.], 1992. Mimeografado.

- _____. Uma nota sobre “raça social” no Brasil. *Estudos Afro-Asiáticos*, Rio de Janeiro, n. 26, p. 79-96, 1994.
- _____. Morenidade: modo de usar. *Estudos Afro-Asiáticos*, Rio de Janeiro, n. 30, p. 79-96, 1996.
- _____. Brazilian society: continuity and change 1930-2000. In: BETHELL, L. (Ed.). *Cambridge history of Latin America*. New York: Cambridge University Press, 2008. v. 9, p. 290-360.
- SILVA, N. V.; PASTORE, J. *Mobilidade social no Brasil*. São Paulo: Markon, 1999.
- SOUZA, J. *Invisibilidade da desigualdade brasileira*. Belo Horizonte: Editora UFMG, 2006.
- STINCHCOMBE, A. *The logic of social research*. Chicago: University of Chicago Press, 2005.
- TELLES, E. US foundations and racial reasoning in Brazil. *Theory, Culture and Society*, Thousand Oaks, v. 20, n. 2, p. 31-47, 2003.
- _____. *Race in another America: the significance of skin color in Brazil*. Princeton: Princeton: Princeton University Press, 2004.
- _____. Ed. *Pigmentocracies: ethnicity, race, and color in Latin America*. Chapel Hill: University of North Carolina Press, 2014.
- TORCHE, F; RIBEIRO, C. C. Parental wealth and children’s outcomes over the life-course in Brazil: a propensity score matching analysis. *Research in Social Stratification and Mobility*, Amsterdam, v. 30, n. 1, p 79-96, 2012.
- TREIMAN, D. Industrialization and social stratification. *Sociological Inquiry*, Hoboken, v. 40, n. 2, p. 207-234, 1970.
- _____. Otis Dudley Duncan’s legacy: the demographic approach to quantitative reasoning in social science. *Research in Social Stratification and Mobility*, Amsterdam, v. 25, n. 2, p. 141-156, 2007.

Resumo

Sociologia como ciência das populações: contribuições de Carlos Hasenbalg e Nelson do Valle Silva no Brasil

Neste artigo é realizado um balanço da obra dos sociólogos Carlos Hasenbalg e Nelson do Valle Silva, mostrando como as inovações que trouxeram para o estudo das relações raciais e da desigualdade social no Brasil foram alcançadas ao partirem de uma perspectiva sociológica que privilegia o estudo das populações e da estratificação social. Nesse sentido, são apresentadas as principais características da sociologia pensada como uma ciência das populações. Essa abordagem, que ganha cada vez mais adeptos em todo o mundo, faz uso de análises estatísticas de grandes bancos de dados para descrever e explicar tendências gerais em nível populacional, ao mesmo tempo em que reconhece a singularidade de cada unidade de análise (indivíduos, família, país, ou qualquer outra unidade). Os trabalhos de Hasenbalg e Silva adotaram essa perspectiva e se voltaram para temas fundamentais dos estudos de estratificação e mobilidade social. Assim, suas pesquisas não só inovaram o estudo das relações raciais no Brasil, como também foram precursoras da sociologia como ciência das populações e dos estudos de estratificação social no Brasil.

Palavras-chave: Metodologia Quantitativa; Populações; Sociologia Brasileira; Estratificação Social; Epistemologia das Ciências Sociais.

Abstract

Sociology as a Science of populations: contributions of Carlos Hasenbalg and Nelson do Valle Silva in Brazil

The article presents the work of sociologists Carlos Hasenbalg and Nelson do Valle Silva, showing how the innovations they brought to the study of racial relations and social inequality in Brazil were achieved, as both started from a sociological perspective that privileges the study of populations and social stratification. Therefore, the main characteristics of Sociology are presented, thought as a science of populations. This approach, which has gained more and more followers around the world, uses statistical analysis from large databases to describe and explain general trends in the population level, while recognizing the uniqueness of each unit of analysis (individuals, family, country, or any other unit). The works of Hasenbalg and Silva adopted this perspective and employed it in fundamental topics of studies of social stratification and mobility. Thus, Hasenbalg's and Silva's researches not only innovated the study of racial relations in Brazil, but also pioneered Sociology as a science of populations and studies of social stratification in Brazil.

Keywords: Quantitative Methodology; Populations; Brazilian Sociology; Social Stratification; Social Sciences Epistemology.

Résumé

La Sociologie comme science des populations : contributions de Carlos Hasenbalg et Nelson do Vale Silva au Brésil

Cet article présente le travail des sociologues Carlos Hasenbalg et Nelson do Valle Silva et montre comment leurs innovations apportées à l'étude des relations raciales et de l'inégalité sociale au Brésil ont été réalisées, car ils partaient d'une perspective sociologique que privilégie l'étude des populations et de stratification sociale. Par conséquent, les principales caractéristiques de la sociologie sont présentées, pensées en tant que science des populations. Cette approche, qui a gagné des plus en plus d'adeptes à travers le monde, utilise l'analyse statistique de grandes bases de données pour décrire et expliquer les tendances générales au niveau de la population, tout en reconnaissant le caractère unique de chaque unité d'analyse (individu, famille, pays, ou autres unités). Les travaux de Hasenbalg et Silva ont adopté cette perspective et l'ont utilisé dans des sujets fondamentaux d'études de la stratification et mobilité sociales. Ainsi, les recherches de Hasenbalg et Silva ont non seulement innové l'étude des relations raciales au Brésil, mais elles ont également été pionnières de la Sociologie en tant que science des populations et des études de stratification sociale au Brésil.

Mots-clés: Méthodologies Quantitatifs ; Populations ; Sociologie Brésilienne ; Stratification Sociale ; Epistémologie des Sciences Sociales.

Alusões biográficas e trajetórias: entre esquemas analíticos e usos flexíveis

Eliana Tavares dos Reis¹
Irllys Alencar F. Barreira²

Uma mirada exploratória nas discussões em torno do uso e apropriações de “biografias” nas ciências sociais permite detectar o estatuto incerto, ambivalente e polissêmico que as caracterizam. E, antes disso, verificamos mesmo a raridade com que a noção ou categoria “biografia” aparece como referência de síntese nas discussões, pois, geralmente, elas são tomadas como sinônimo ou base para trabalhos orientados por diferentes tradições de estudos (das histórias de vida e oral aos estudos de trajetórias). De qualquer modo, a imersão na bibliografia direcionada à mobilização de dados biográficos traz à tona polêmicas suscitadas e flutuações de ênfases nas utilizações feitas ao longo do tempo, conforme as áreas específicas de conhecimento dos pesquisadores, suas preocupações e escalas de análise. Entre a biografia substancializada, em um polo, e a problematização das informações biográficas (que pode passar pela reflexão das autobiografias e memórias), em outro, passamos por empenhos na construção de biografias ou usos irrefletidos de dados e

fontes biográficas. No entanto, a constatação de François Dosse (2015, p. 11) permanece válida: “todas as gerações aceitaram a aposta biográfica” e “cada qual mobilizou o conjunto dos instrumentos que tinha à disposição”.

Situamos a perspectiva desse artigo no apanhado dessa pluralidade de ativações mais ou menos reflexivas da expressão biografia e seus correlatos conceituais, mapeando algumas tendências na produção recente das ciências sociais brasileiras e verificando pressupostos teóricos e metodológicos que mais se destacaram. Procuramos, ainda, apontar seus principais usuários e porta-vozes no Brasil.

Para tanto, em primeiro lugar, recorremos à pesquisa no Scientific Electronic Library Online (SciELO), cuja base de dados permitiu observar artigos publicados em periódicos de ciências sociais, sobre distintos universos de investigação e produzidos por autores e autoras com diferentes filiações a áreas do conhecimento. É claro que isso significa abranger somente os periódicos considerados mais importantes, justamente por estarem indexados

-
- 1 Eliana Tavares dos Reis é doutora em Ciência Política pela Universidade Federal do Rio Grande do Sul (UFRGS), professora e pesquisadora do Programa de Pós-Graduação em Ciências Sociais da Universidade Federal do Maranhão (UFMA). Coordena o Laboratório de Estudos sobre Elites Políticas e Sociais, é autora do livro *Trajetórias, espaços e repertórios de intervenção política* (Editora Zouk, 2015), coautora do livro *Elites parlamentares e a dupla arte de representar* (FGV Editora, 2016) e de outras coletâneas sobre estudos de elites políticas e culturais (EDUFMA, 2016, 2014).
 - 2 Irllys Alencar F. Barreira é doutora em Sociologia pela Universidade de São Paulo (USP) e professora titular de Sociologia do Programa de Pós-Graduação em Sociologia da Universidade Federal do Ceará (UFC). É líder do Grupo de Pesquisa Lideranças, Representações e Práticas Políticas e autora de vários livros, entre eles: *Chuva de papéis: ritos e símbolos de campanha eleitoral no Brasil* (Relume Dumará, 1998); *Imagens ritualizadas, apresentação de mulheres em cenários políticos* (Pontes Editores, 2008); e *O labor criativo na pesquisa: experiências de ensino e investigação em ciências sociais* (Imprensa Universitária, 2017).

nessa “biblioteca eletrônica” (como é definido), e comportar apenas as publicações mais recentes, não dispondo de trabalhos seminais – dificultando observar processos/modificações mais longínquas em uma pesquisa de natureza genealógica. Mesmo com essas restrições, o SciELO se constituiu em um bom ponto de partida, com conjunto representativo de estudos que remetem à questão em pauta. Além disso, essa pesquisa exploratória tornou possível identificar algumas das principais referências acionadas nas discussões, os autores consagrados, os instrumentais mobilizados, as dimensões de análise e temáticas priorizadas.

Com base nisso, em segundo lugar, propomos discutir alguns dos marcadores teóricos e metodológicos fundamentais que sustentam as discussões e análises. Sem a pretensão de sermos exaustivas – pois sabemos que a amplitude da questão abre outras possibilidades de exploração analítica – retomamos algumas tradições de análise persistentes nos estudos nacionais, que não ficaram imunes às suas orientações analíticas e aos debates teórico-metodológicos importados de outras configurações históricas. Antes disso, os empregaram de forma flexível, adequando-os aos variados universos de análise quando voltados a evidenciar a presença e as características de atores sociais e políticos em conjunturas históricas diversificadas.

O fio condutor da exposição se ancora na percepção de que a discussão sobre o lugar epistemológico das biografias cedeu espaço às disputas sobre os usos de dados biográficos. Desse modo, as alusões biográficas funcionam como “pretexto” (no duplo sentido) para a afirmação de noções como história de vida, história oral, carreira e trajetória. A rotinização constatada do emprego (mais ou menos rigoroso) dessa última noção coloca

em primeiro plano a centralidade da posição de Pierre Bourdieu e da sociologia francesa no debate, bem como dos circuitos de difusão da sua obra no país.

Distribuição da produção recente sobre estudos biográficos ou dos usos de biografias nas ciências sociais brasileiras

Para pensar nos usos e apropriações de “biografias” nas ciências sociais brasileiras, nossa porta de entrada foi o levantamento feito na base de dados SciELO. Começamos, então, pesquisando essa categoria. Porém, foram escassas as ocorrências de artigos que a registraram nos seus títulos, resumos e palavras-chave. Por isso, decidimos testar a localização de outras ocorrências que, de algum modo, são percebidas como estudos biográficos ou utilizam sistematicamente as informações biográficas nas suas investigações. Portanto, visando englobar um universo mais amplo de artigos, com referências analíticas e procedimentos metodológicos relativamente semelhantes, optamos por explorar, em todos os índices disponíveis, outras cinco orientações: “histórias de vida”, “história oral”, “carreiras”, “perfis” e “trajetórias”³.

Considerando os periódicos em geral indexados, a quantidade foi bastante alta: 5.592. Restringindo a pesquisa para periódicos brasileiros, a quantidade diminuiu para 3.724 casos. Aplicou-se, então, um filtro ainda mais forte, para melhor delimitação do universo, mais compatível com os objetivos da discussão que propomos: a seleção somente de periódicos de ciências sociais e história. Trata-se de uma definição de áreas relacionadas à já sabida tradição desses tipos de estudos biográficos voltados ao *indivíduo* e aos *eventos*.

3 Observamos uma variação quando os termos são colocados no singular (menor ocorrência) e quando são colocados no plural (maior ocorrência). Neste teste, abarcamos o maior número possível de casos.

Chegamos, assim, a uma distribuição na qual fica claro que os mais baixos índices, em todos os níveis, são de trabalhos que mobilizam a etiqueta “biografia”, em contraposição às altas frequências de discussões que se associam aos estudos de “trajetórias” (fundamentalmente baseadas no uso de dados biográficos) – com pequena variação quando tomamos o montante total de artigos armazenados no SciELO. Tanto quando consideramos a quantidade geral de artigos como quando visamos os periódicos brasileiros, a hierarquia dos mais frequentes se mantém: os “perfis” em primeiro lugar (respectivamente 2.164 e 1.833 ocorrências), as “trajetórias” em segundo (1.084 e 775, respectivamente) e a “história oral” (1.000 e 504) e “história de vida” (864 e 346) em terceira e quarta posições. Observamos que há diminuição dessas duas últimas, que é ainda maior

quando consideramos somente os periódicos de ciências sociais e história. E, nesse caso, mais acentuada é a queda da frequência de “perfis” e a ascensão das “trajetórias”. O Quadro 1 permite comparar os dois extremos dos filtros aplicados:

É interessante observar que, não raramente, os artigos associam seus estudos de “trajetórias” a outra das demais rubricas (ou vice-versa), ocorrendo um leve aumento das frequências de “biografias” e das “carreiras”, principalmente uma relativa ampliação em todas elas, quando atentamos às décadas das publicações. Os primeiros artigos localizados no SciELO foram publicados no final da década de 1980 (dois sobre “histórias de vida”, ambos na revista *Tempo Social*, sendo um sobre mulheres e o outro sobre identidade étnica e política⁴). A maior consistência de publicações sobre “trajetórias” acontece a partir de 2006 e 2007. O Quadro 2 é ilustrativo.

Quadro 1
Distribuição das palavras-chave

| | Biografias (%) | Histórias de vida (%) | História oral (%) | Trajetoórias (%) | Carreiras (%) | Perfis (%) |
|--|----------------|-----------------------|-------------------|------------------|---------------|------------|
| Quantidade geral | 3,5 | 15,4 | 17 | 19,4 | 5 | 39,7 |
| Periódicos das ciências sociais e história | 6,7 | 10,8 | 9,1 | 51,4 | 13,2 | 39,7 |

Fonte: Elaboração nossa com base nos dados do SciELO⁵.

Quadro 2
Distribuição das categorias por períodos

| Anos de referência | Biografias | Histórias de vida | História Oral | Trajetoórias | Carreiras | Perfis |
|--------------------|------------|-------------------|---------------|--------------|-----------|--------|
| 1989-1999 | 0 | 4 | 2 | 4 | 0 | 1 |
| 2000-2006 | 0 | 7 | 8 | 26 | 4 | 4 |
| 2007-2016 | 21 | 29 | 21 | 131 | 40 | 24 |
| 2017(parcial) | 1 | 2 | 0 | 8 | 1 | 1 |

Fonte: Elaboração nossa com base nos dados do SciELO⁶.

4 O primeiro é de Elizabeth Souza Lobo (1989), sobre “Experiências de mulheres: destinos de gênero”, e o segundo é de Eva Blay (1989), intitulado “Inquisição, inquisições: aspectos da participação dos judeus na vida sócio-política brasileira nos anos 30”.

5 Acesso entre setembro e outubro de 2017.

6 Acesso entre setembro e outubro de 2017.

Com base nessas informações, atentamos à distribuição desses artigos nos periódicos específicos das ciências sociais e da história nos quais tais registros apareceram. Chegamos a 26 revistas com a seguinte ordenação dos 341 artigos identificados: *Cadernos Pagu* (30), *Revista Estudos Feministas* (25), *Revista de Sociologia e Política* (24), *Horizontes Antropológicos* (23), *Tempo Social* (23), *Revista Brasileira de Ciências Sociais* (21), *História* (São Paulo) (18), *Caderno CRH* (17), *Mana* (15), *Revista Brasileira de Ciência Política* (15), *Sociologias* (15), *Lua Nova: Revista de Cultura e Política* (13), *Tempo* (13), *Revista Brasileira de História* (12), *Topoi* (Rio de Janeiro) (12), *Opinião Pública* (11), *Varia Historia* (11), *Sociedade e Estado* (10), *Sociologia e Antropologia* (9), *Sociologia* (2), *Dados* (7), *Estudos Históricos* (Rio de Janeiro) (7), *Novos Estudos Cebrap* (6), *Revista de Antropologia* (1) e *Etnográfica* (1). Considerando as áreas às quais os periódicos são ligados, 125 (37%) artigos estão em revistas voltadas às diferentes áreas das ciências sociais (ou seja, publicam artigos de sociologia, antropologia, ciência política e híbridos); 73 (21,4%) em periódicos de história; 63 (18,5%) de ciência política; 40 (11,7%) de antropologia; e 40 (11,7%) de sociologia.

Evidentemente, há uma gama imensa e diversificada de temáticas tratadas, mas estabelecemos algumas categorias por universos de análise privilegiados nos textos, e excetuamos os estudos que as aplicaram de forma mais geral ou metafórica, geralmente para falar de eventos, processos, condições, conjunturas e fases definidas como “outros”⁷. Suspendendo a análise desses casos, ficamos, então, com um universo de 292 artigos classificados em

dezessete categorias (por ordem de frequência): “mulheres”; “outros gêneros”; “intelectuais”; “políticos”; “profissionais”; “militantes”; “crianças/jovens”; “trabalhadores”; “étnicos”; “agricultores”; “juristas”; “velhos”; “teóricos/metodológicos”; “regimes/instituições”; “eclesiásticos/religiosos”; “escravos”; “militares”.

Levando em conta que dois dos periódicos que contemplam publicações em geral das diferentes áreas das ciências sociais são voltados a temáticas de gênero (*Cadernos Pagu* e *Revista Estudos Feministas*), não é de se estranhar que o universo mais frequentemente tratado seja o de “mulheres”, estudadas enquanto intelectuais, políticas, militantes, profissionais etc.⁸. São 66 (22,6%) artigos, cuja maior incidência (41 – 14%) está nos periódicos de ciências sociais, aparecendo de forma relativamente bem distribuída nas áreas específicas. Também há quatro (1,4%) trabalhos que chamamos de “outros gêneros”, pois versam sobre travestis, transexuais, sexualidade e corpo. O segundo universo mais frequente é dos “intelectuais”: são 38 (13%) e, mais recorrentemente, publicados em periódicos de ciências sociais (16); em segundo lugar, de história (9), a mais baixa é em revistas de antropologia (2). Estes são seguidos, empatados na terceira posição, pelos “políticos” e “profissionais”, com 32 (11%) ocorrências cada. Entre os primeiros há alta acumulação em periódicos propriamente de ciência política (23), seguidos de longe por artigos na história e em periódicos de ciências sociais. Os “profissionais” encontram-se distribuídos em todas as áreas, provavelmente devido à abrangência da categoria e da sua aplicação para autônomos, empresários, investigadores, *experts*, artistas etc.

7 Foram 49 (14,4%) artigos assim classificados, sendo que, destes, 16 (32%) foram publicados em periódicos de ciências sociais e 12 (25,5%), especificamente, em revistas de ciência política.

8 Tomamos primeiro o corte de gênero, pois a declaração nas bases de apresentação do texto informa o seu primado frente à qualificação que o acompanha.

Na categoria “militantes” (movimentos sociais, ativismo transnacional, luta pela terra, meio ambiente, direitos humanos, exílios, guerrilheiros etc.), encontramos a frequência de 24 (8,2%) textos, sendo mais usual em periódicos de história (11), seguindo com distribuição relativamente equilibrada nos demais. Em sexto lugar, há 21 (7,2%) artigos que investigam “crianças/jovens”, com a maior incidência (11) em periódicos de ciências sociais, geralmente abordados em relação ao mundo escolar e à formação educacional. Eles são seguidos pelos “trabalhadores” e “étnicos”, com 14 (4,8%) artigos cada. Os primeiros comportam aqueles que desenvolvem atividades no meio urbano (indústria, serviço público, construção civil, ferroviários, desempregados etc.) e são mais abordados em periódicos das ciências sociais e da sociologia (6 e 5, respectivamente), estando o restante distribuído nas demais áreas. Quanto aos étnicos (referentes a imigrantes em geral, bem como negros e africanos), a frequência é maior em periódicos de história (9) estando o restante relativamente distribuído em quase todos os demais, não havendo nenhuma ocorrência em revistas de ciência política.

As demais temáticas são as que têm menor incidência. Os “agricultores” (contemplando camponeses, lavradores, estancieiros, patronato) são nove (3,1%) casos, com concentração em periódicos de história (6) e inexistindo nos de ciência política e antropologia; os “juristas” são oito (2,7%) e não aparecem em periódicos de antropologia e história; “velhos” são sete (2,4%) e não se encontram referidos em periódicos de ciência política e sociologia; “teóricos/metodológicos” são seis (2%), distribuídos equilibradamente para sociologia, antropologia e história; “eclesiásticos/religiosos” são cinco (1,7%) e bem distribuídos em todas as áreas; “escravos” e “militares” são três (1%) cada, todos os primeiros em periódicos de história, e os segundos divididos em história e ciência política.

Conjugando essas temáticas aos tipos de estudos mais frequentemente acionados para cada um dos universos, como foi dito no início, as trajetórias predominam com 51,4% dos trabalhos, sendo majoritárias nos artigos sobre os seguintes universos: “mulheres” (estando em segundo lugar as carreiras, acompanhadas de perto das histórias de vida); “intelectuais” (para os quais o segundo tipo mais frequente são as biografias); “agricultores”, “étnicos” e “militantes” (têm a história oral na posição seguinte); “crianças/jovens” (também recorrentemente tratados pela história de vida); e “profissionais”, “teórico/metodológicos”; “eclesiásticos/religiosos” e “trabalhadores” (para os quais não há concentração em um tipo específico de estudos depois das trajetórias). Entre os artigos sobre “políticos” – e acontece o mesmo para os “militares” – é muito próxima a frequência de trajetórias e carreiras (aparecendo, para esses últimos, os perfis em terceiro lugar). Para “escravos” e “outros gêneros” não há acúmulo de estudos de trajetórias, mas uma distribuição relativamente equilibrada nos diferentes tipos. Já para os “juristas”, o primeiro lugar é das carreiras e, depois, os perfis e, para os “velhos”, as maiores incidências são da história de vida e história oral, com somente uma ocorrência para trajetórias.

O levantamento realizado nas bases do SciELO revelou alguns eixos a serem explorados, dos quais grifamos ao menos três. Em primeiro lugar, a ausência de um campo delimitado de estudos biográficos e a existência de uma multiplicidade de usos e apropriações alusivas das informações biográficas, favorecendo uma ampliação e diversificação da discussão. Em segundo lugar, foi possível perceber algumas regularidades entre os tipos de estudos empregados, as temáticas priorizadas e a confluência com as diferentes áreas das ciências sociais. Em terceiro lugar, a consideração dos anos de publicação possibilitou detectar a

preponderância progressiva dos estudos de trajetórias e, em menor grau, dos de carreiras, em detrimento das reflexões sobre história de vida e história oral (que são mais frequentemente mobilizadas entre pesquisadores da história e da antropologia, em trabalhos sobre uma variedade de universos, como “militantes”, “mulheres” e “velhos”).

Desse último ponto, quando examinamos os textos, tanto os de revisão teórica e metodológica encontrados como os demais, outros dois aspectos são relevantes. O primeiro é referente à noção de carreira, que aparece em diferentes tipos de estudos, para diversos universos e com distintas aplicações. *Grosso modo*, verificamos, de um lado, o recurso a autores da chama Escola de Chicago para estudos de determinados universos (como “profissionais”, “militantes” e “crianças/jovens”) nos quais as biografias dos atores são observadas de uma perspectiva mais situacionista, subjetivista, institucionalista e internalista, frequentemente apropriadas na antropologia e sociologia. De outro lado, observa-se a presença da mesma noção costurada ao modelo bourdieusiano, sobretudo baseado na sociologia política francesa, para o exame de agentes relativamente bem situados em diferentes esferas sociais (especialmente forte nos trabalhos sobre “mulheres”, “políticos”, “intelectuais”, “militares”, “profissionais”, entre outros), selecionados a partir de estudos sociográficos e prosopográficos. A tendência constatada é a de que essas pesquisas são realizadas principalmente por pesquisadores da ciência política e sociologia.

De qualquer modo, há claramente o domínio dos modelos europeus, notadamente franceses, na pauta de discussões e operacionalização dos trabalhos, e o aumento exponencial de estudos de trajetórias e das referências a Pierre Bourdieu. Ele é negavelmente um dos autores mais lembrados nas ciências sociais e em outras áreas do conhecimento. Conforme

levantamento feito por Bortoluci, Jackson e Pinheiro Filho (2005, p. 217), no Sistema Lattes, Bourdieu fica em terceiro lugar como o autor mais citado nas ciências humanas e sociais (766 ocorrências), assumindo a primeira posição quando se considera apenas a sociologia, e o segundo lugar quando se trata das citações mais frequentes nos artigos publicados no SciELO. Essa constatação vai ao encontro dos altos índices que verificamos, em levantamento feito na mesma base de dados, sobre as referências ao estudo de trajetórias. Nos artigos voltados a discussões específicas e mais conceituais sobre biografias, Bourdieu é mencionado, mesmo que seja para contestar suas críticas a determinados usos das “histórias de vida”.

Sendo assim, para melhor situar essas indicações, ter mais indicadores do lugar das “biografias” nas ciências sociais e história e tentar qualificar um pouco mais as bases que alicerçam os trabalhos no Brasil, propomos retomar alguns dos referenciais que mais frequentemente são mobilizados nas discussões, localizadas a partir do levantamento, sobretudo daquelas de cunho bibliográfico, teórico e metodológico. Somamos a isso a possibilidade de recolocar aspectos e desdobramentos consagrados nas discussões sobre o percurso dos estudos biográficos e do uso de dados biográficos nas ciências sociais em geral. Podemos observar a afirmação de modelos analíticos que trazem consigo não apenas a circunscrição de dimensões, instrumentais, ênfases de análise e de fronteiras disciplinares, como indicam os universos de investigação a serem privilegiados. E, ainda, pontuar elementos da importação dessas discussões e delimitações nas ciências sociais brasileiras. Mesmo que de forma bastante sumária e geral, indicamos alguns dos principais pesquisadores e tipos de pesquisas, que os aplicam e se constituem mesmo, nas últimas décadas, como referências nacionais primordiais.

História, biografias dos indivíduos e a força das representações

Neste momento, propomos uma breve discussão sobre a ligação genética entre história e biografia, que pode estar na base da associação espontânea entre os estudos biográficos e as abordagens da história oral e de história de vida. De um lado, a necessidade de constante retomada e sustentação da relevância desses tipos de estudos indica seu caráter “historicamente incerto”, de outro, o lugar atribuído aos indivíduos e às narrativas de suas experiências singulares oportuniza interlocuções com outras predileções disciplinares, principalmente com perspectivas vindas da filosofia e desdobramentos no campo da antropologia.

A partir do artigo publicado pela historiadora Marieta de Moraes Ferreira (2002) na revista de história *Topoi*, podemos seguir a trilha do processo de institucionalização da história como disciplina na França, sobretudo a partir do final do século XIX, quando a “elite republicana” teria se dedicado ao trabalho de construção da memória nacional. A garantia do caráter cientificista da produção dos “historiadores profissionais” (em oposição aos “ensaístas”) estava associada à distância temporal dos acontecimentos descritos retrospectivamente. Assim, o estudo do “passado”, amparado em documentos devidamente arquivados e catalogados, afirmou-se como competência necessária à distinção do amorosismo que reinaria entre aqueles que recorriam aos “depoimentos diretos”, aos relatos e acontecimentos do presente. No fundamento da chamada abordagem *événementielle* (dos acontecimentos “em si” ou factuais) estaria igualmente a consagração biográfica das figuras ilustres (políticas e militares), dos seus “feitos” e dos eventos marcantes. Conforme Levillain (1996), havia uma dificuldade histórica de separar o trabalho do historiador (feita de acontecimentos coletivos, usando o

gênero narrativo para reconstituir eventos e demonstrar mudanças) daquele do biógrafo (feita a partir dos “fatos e gestos de um indivíduo” para consagrá-lo, portanto originalmente ligada ao panegírico, não verificável e baseada no gênero descritivo); a diálise entre história e biografia colaborou mais decisivamente para a afirmação da segunda como gênero do que para a o reconhecimento da história como disciplina e ciência no século XIX.

A emergência de discussão mais sistemática sobre a relação entre biografia e história, depois da Primeira Guerra Mundial – sobretudo com a publicação, em 1929, dos *Aspects de la biographie*, de André Maurois, “em nome de um método positivista” e “envolto num estilo laudatório” (LEVILLAIN, 1996, p. 157) – foi obscurecida pela ascensão dos argumentos dos historiadores ligados à *École des Annales*. Diferentemente das preocupações anteriores e com outra forma de representação do “tempo histórico”, o alvo passava a ser as dimensões econômicas e sociais mais amplas, atenta às estruturas estáveis, verificadas em processos de longa duração e com atores coletivos. Para tanto, seus expoentes propunham a incorporação de técnicas de quantificação, aplicadas a partir de uma diversificação de fontes, como “listas de preços, de salários, séries de certidões de batismo, óbito, casamento, nascimento, fontes notariais, contratos, testamentos, inventários”, com base nas quais pretendiam verificar “os condicionamentos econômico-sociais das ações e decisões individuais, a sociedade global e as massas, as condições materiais, e não os projetos individuais, subjetivos e ideológicos” (REIS, 2000, p. 23). Suas abordagens seguiam priorizando o “passado” e as fontes escritas, porém tratava-se de evitar as intersecções (com a literatura) e se afastar das perspectivas laudatórias (de abordagem moral, ideológica e testemunhal), que favoreciam os personagens das elites no poder, delimitando, assim, os limites da autonomia de uma história

que seria propriamente científica ou de uma história como ciência social⁹.

Durante algumas décadas as discussões sobre biografias e histórias de vida arrefeceram-se tanto para os historiadores como para os cientistas sociais, a ponto de Le Goff e Nora (1974), na organização do seu *Faire de l'Histoire*, decretarem a “excomunhão da biografia” em nome da renovação da história tradicional, propondo, nos três volumes publicados, novas abordagens, técnicas, fundamentos epistemológicos e objetos de estudo. Entretanto, foi no final dessa mesma década e nos anos 1980 que ocorrera a reincorporação dos trabalhos sob essa rubrica (no campo de estudos das ciências humanas), revigorando debates sobre procedimentos e estratégias metodológicas, bem como dimensões de análise a serem privilegiadas. Como veremos subsequentemente, para as ciências sociais, tal reabilitação se relaciona com a centralidade que os relatos e informações biográficas conquistaram entre cientistas sociais localizados em diferentes contextos nacionais (principalmente na Alemanha, nos Estados Unidos e na França), com suas ressonâncias nas apropriações de pesquisadores brasileiros.

Ferreira (2002, p. 319) ressaltou como, a partir da década de 1980, os distintos domínios da pesquisa histórica passaram por modificações consistentes, com o revigoramento da análise qualitativa e o resgate da “importância das experiências individuais”, das “situações vividas” e “singulares”. Essa perspectiva redundou em “novo impulso à história cultural, um renascimento do estudo

do político que se incorporou à história o estudo do contemporâneo” (Ibidem).

Provavelmente, no Brasil, acompanhando movimentos globais, o resgate das experiências guardou relação com os processos de redemocratização do país, com a entrada em cena de protagonistas e porta-vozes (incluindo pesquisadores), formados por versões de marxismo e valorização da “sociedade civil”¹⁰, preocupados com a história de personalidades anônimas, protagonistas de uma “história da resistência”. Conforme situado em Barreira (2003), os movimentos sociais foram considerados agentes ativos de processos sociais, fazendo um contraponto entre estrutura e experiência. Em vista disso, as trajetórias de lideranças operárias, de movimentos de bairro e outras organizações coletivas ganharam notoriedade, sendo enfatizada a lógica de atores e processos subjacentes de mudança social. Uma lógica de “subversão das estruturas” reforçou a preocupação com determinados atores e grupos, fertilizando críticas a modelos teóricos considerados deterministas, transformando as trajetórias e biografias em peças vivas de manifestação da dramaturgia social.

Dessa maneira, os depoimentos foram particularmente utilizados nas ciências sociais como modo de recuperar não apenas trajetórias individuais, mas contextos históricos de referência, incluindo profissões, ideologias, processos de mudança etc. Aspásia Camargo (1982), influenciada pela tradição francesa, se situou no “estudo das elites” para retomar as biografias de políticos, sendo observadas as relações entre ciclos de vida e momentos históricos. Ela cotejou

9 Na disputa pela imposição dos sentidos e definições da história, dos limites legítimos do fazer historiográfico, e pela consolidação de reconhecimento intelectual, os historiadores da *École des Annales* aplicaram-se a rechaçar as exaltações biográficas dos indivíduos, reivindicando que a história só faz sentido quando desvelada pelos historiadores. Dessa forma, substituíam “as limitações da cronologia linear pela cronologia como pano de fundo”, visando testar hipóteses com base na “determinação de regularidades e de sistemas de relações” (LEVILLAIN, 1996, p. 158).

10 Sobre a trajetória de militantes que lutaram contra a ditadura, a afirmação das ciências humanas e sociais, bem como de porta-vozes formados em versões de marxismo, que conquistam reconhecimento indissociavelmente intelectual e político, ver Reis (2015).

igualmente as lideranças camponesas, pontuando que, embora as motivações de representação política fossem convergentes, ambas apontavam referências culturais específicas provenientes da condição social. Os termos “elite” e “massa” vinham acionados para identificar a diversidade de comportamentos associados a padrões culturais. Por essa via, Camargo criticou a concepção estatística por meio da história de vida como recurso de investigação, considerando que o que se perderia em “representatividade” se ganharia em “inteligibilidade”.

Maria Izaura Pereira de Queiroz (1988) também problematizou a ênfase nos métodos quantitativos para acentuar que as histórias de vida foram consideradas úteis para registrar aquilo que ainda não havia sido objetivado em documentos. E pontuou a dimensão da preservação da memória, muito embora fosse alvo de críticas, pois, supostamente, tais depoimentos poderiam ser considerados “infieis”, porque contaminados pela psique individual¹¹.

Tais formulações influenciaram na formação de uma geração de intelectuais. A partir da década de 1990 houve investimentos consistentes na realização de eventos para discutir história oral, em publicações¹², na formação de redes com investigadores estrangeiros e na criação, em 1994, da Associação Brasileira de História Oral, entre outros empreendimentos, em oposição à diminuição desse recurso de investigação que experimentava as ciências sociais.

Com esses movimentos, os usos da história oral e história de vida eram delineados não apenas como técnica, incursionando também

no debate teórico e na crítica dirigida às abordagens estruturais, que desconsideravam a primazia dos sentidos atribuídos às relações sociais no espaço das interpretações.

Uma ponderação, geralmente retomada nas revisões acerca das histórias de vida, é sobre a distinção proposta por Ferreira (2002) entre, de um lado, seu uso complementar a outras fontes, como forma de “preencher lacunas” por meio da realização de entrevistas com roteiros, de maneira a garantir alguma objetividade e controle da pesquisa e, de outro lado, os depoimentos de caráter mais espontâneo, valorizados como foco de “representações” que desvelam o lugar da “memória” e da “história” nas reconstituições do “passado”. Por esse motivo, as “subjetividades”, inconsistências e distorções não são consideradas “elementos negativos”, pois “a veracidade dos depoimentos não é a preocupação central” (FERREIRA, 2002, p. 327-328).

No entanto, destaca-se na linha de frente um conjunto diversificado de estudos sobre segmentos das elites políticas, cuja ênfase nos dados biográficos tem se amparado na seleção, classificação e divulgação feitas pelo Centro de Pesquisa e Documentação de História Contemporânea do Brasil (CPDOC), criado em 1973. Outro aporte dessa perspectiva realça os trabalhos sobre “memória de operários”, trabalhadores fabris, seringueiros amazônicos, migração de camponeses, educação de mulheres, histórias de vidas de presidiários, entre muitos outros” (SANTOS; OLIVEIRA; SUSIN, 2014). As principais colaboradoras do CPDOC foram responsáveis pela elaboração de dicionário (biográfico) sobre

11 Uma referência importante foi o trabalho, publicado originalmente em 1978, do sociólogo Paul Thompson (1992), para o qual a história de vida seria compreendida como parte singular de uma totalidade. Perguntando-se sobre como se efetivavam as relações entre empregados e empregadores, como a consciência de classe variava entre cidades, países e profissões, o autor postulou a necessidade de construir uma teoria de modo processual, passo a passo, junto com a descoberta de fatos, relacionando economia, ideologia e família, por exemplo.

12 Do final da década de 1980 há, por exemplo, as publicações de Alberti (1989) e von Simson (1988), e as várias coletâneas sobre o tema produzidas a partir dos eventos, como Ferreira (1994), Ferreira e Amado (1996) e Moraes, Fernandes e Alberti (2000). Muitos deles são feitos a partir do Centro de Pesquisa e Documentação de História contemporânea do Brasil (CPDOC).

a história do Brasil no período após 1930. A maioria delas – pesquisadoras e porta-vozes da aplicação da “história oral” e “história de vida” –, ambivalentemente, alimentou esse repositório das “figuras ilustres” (portanto inevitavelmente de caráter laudatório) para disponibilizá-lo aos variados usos, sobretudo políticos e acadêmicos. E buscaram informações a partir de entrevistas, pensadas como situações de interação e influência mútua entre entrevistador e o entrevistado. Nesse sentido, a construção biográfica é concebida como manifestação “coerente” e “substantiva” com a “narrativa de experiências de vida produzida no contexto de uma entrevista de história oral” (ALBERTI, 2012, p. 165).

No bojo desses movimentos, entre os artigos localizados no levantamento feito, podemos citar como exemplares de trabalhos de *história oral*: Araújo (2002), Cardoso (2004) e Adão (2010). Entre os de *história de vida*, destacamos os trabalhos de Farias (2009), Hartmann (2014) e Guedes (2014).

Também localizamos o artigo recente de Maria da Glória Oliveira (2017), com o título provocador: “Quem tem medo da ilusão biográfica? Indivíduo, tempo e histórias de vida”, portanto, tendo como eixo o artigo de Bourdieu, publicado justamente em coletânea organizada por Ferreira e Amado (1996), intitulada *Usos & abusos da história oral*. A historiadora contesta os argumentos de Bourdieu como interdições, que deveriam ser superadas para recolocar a centralidade hermenêutica do “indivíduo”, sobretudo por meio da ideia *identidade narrativa* cunhada pelo filósofo francês Paul Ricoeur. Ao contrário, Carvalho (2003), psicologia, com pós-graduação

em educação e pós-doutorado em antropologia, compatibiliza sem receio muitas das dimensões de análise disponíveis no modelo analítico de Bourdieu com as perspectivas de autores como Clifford Geertz, Jerome Bruner, Ricoeur e Hans-Georg Gadamer. Tendo estudado “trajetórias no campo ambiental”, a autora empenhou-se particularmente em sustentar a relevância das proposições de Ricoeur e Gadamer para os estudos que mobilizam métodos biográficos. Considera a pesquisadora que essa contribuição “reitera a fronteira entre sujeito e história como o ambiente epistêmico por excelência da pesquisa biográfica e rompe com uma possível orientação realista” (CARVALHO, 2003, p. 293). Particularmente, trata-se de considerar os relatos biográficos como tipos de narrativas (individuais) que amarram simbolicamente as inter-relações “sujeito” e “história”, assim como entre “acontecimentos e sua reconfiguração na tessitura de vidas narradas” (Ibidem).

O modelo proposto, de certo modo, se aplica à discussão de Gonçalves, Marques e Cardoso (2012), que reflete as várias combinações de matrizes, instrumentos e universos de análise privilegiados na antropologia, na filosofia e na história, para sustentar a intersecção entre etnografia e biografia. Para eles, as narrativas, em sua função de agenciamento, traduziriam experiências do mundo, levando a problematizar “como o conceito de etnobiografia afeta, necessariamente, não só o modo como tratam as histórias que os sujeitos etnografados contam, mas também como os pesquisadores contam suas histórias etnográficas sobre essas histórias e seus personagens-pessoas” (Ibidem, p. 10)¹³. Apoiando-se

13 Problematicando de modo mais radical os conceitos abrangentes e substantivados que opõem o individual e o coletivo, o sujeito e a cultura, abrindo, assim, espaço para a individualidade, Gonçalves, Marques e Cardoso (Idem) defendem que o exercício de “etnografar uma biografia” se constituiria em uma missão relevante, tendo em vista verificar com acuidade as condições de realizar uma biografia, assim como a importância da narrativa na vida cotidiana dos próprios contadores. Essa concepção, mais desvinculada da ideia de representatividade objetivista, permitiria relativizar o caráter de “veracidade” como foco da questão, recompondo com mais vigor a dimensão de singularidade fortalecida nos ideais de objetividade, também parte significativa da narrativa científica.

também em Geertz, Brune, Ricouer, Gadamer e Bourdieu, entre muitos outros, Gonçalves, Marques e Cardoso (Idem) fazem referência ao argumento de Queiroz (1988) sobre o lugar do narrador “que, em última instância, decide o que é relevante ou não para ser tomado como fundamentação de sua narrativa, sobretudo o modo como organiza os eventos e o modo com que se pronuncia ou silencia sobre determinados fatos” (GONÇALVES; MARQUES; CARDOSO, 2012, p. 23). Estabelecendo a necessidade de problematizar e “repensar a tensa relação entre subjetividade e objetividade, pessoa e cultura” (Ibidem, p. 20), os autores lembram que “este modo de pensar o biográfico e o social” é semelhante às conceitualizações de Simmel sobre os “dualismos contraditórios que, positivamente, constituem a vida social” (Ibidem), e as formulações de Elias acerca do peso dos constrangimentos sociais sobre os processos diferenciados de individualização.

Contribuições da tradição sociológica alemã e a ênfase da Escola de Chicago

Nas reflexões sobre biografias ou usos de informações biográficas é comum a demarcação da Escola de Chicago como veio da restituição temática no âmbito das ciências sociais, particularmente na sociologia e na antropologia. É igualmente persistente o assentamento das influências da tradição alemã sobre os pesquisadores envolvidos nessa reabilitação. Howard Becker (1996) assinalou que os escritos de Georg Simmel vinham sendo traduzidos em Chicago desde antes de 1900 e muitos deles eram publicados na *American Journal of Sociology*. Isso graças ao interesse dos sociólogos americanos, geralmente de formação protestante, “na reforma social, voltado para o equacionamento dos problemas sociais que afligiam as grandes cidades americanas” (Ibidem, p. 178).

A influência de Georg Simmel já havia sido assumida por seu contemporâneo, Max

Weber, ao pensar a *forma social* para tipos de socialização, combinando dimensões da organização social e das orientações das ações individuais, visando estabelecer a diferença e superar (conforme o estágio de desenvolvimento do mundo ocidental) a distinção que havia sido estabelecida por Ferdinand Tönnies entre a *Gemeinschaft* (comunidade) e a *Gesellschaft* (sociedade). Assim, a contribuição de Simmel residiria na possibilidade de lidar com diferentes níveis, individual e coletivo, de análise das formas possíveis de socialização e sociabilidades, das subordinações, subjetividades e conflitos emergentes em condições de pluralidade, diversificação, especialização, enfim, de complexificação crescente da vida em sociedade.

Do mesmo modo, é notória a centralidade das formulações de Alfred Schütz no mesmo campo epistêmico. Podemos sublinhar a importância da sua *teoria da relevância* para determinados estudos de biografias orientados pela ideia de que “qualquer narrativa é uma interpretação realizada a partir de uma situação biográfica específica, uma vez que uma biografia é uma descrição de processos e experiências vinculadas pelo próprio indivíduo” (SANTOS; OLIVEIRA; SUSIN, 2014, p. 374). Dessa forma, ganham relevo os “processos interpretativos” nos quais os indivíduos manipulam “estoques de conhecimento” para “atribuir sentido” às suas ações e experiências e, conseqüentemente, definir suas decisões (SCHÜTZ; LUCKMANN, 1973).

Em consonância, Peter Berger e Thomas Luckmann (2004) publicaram *A construção social da realidade: tratado de sociologia do conhecimento*, em 1966, para compreender como os atores individuais e coletivos constroem suas realidades históricas e cotidianas. Dubar e Nicourd (2017, p. 47-55) assinalaram a centralidade das interações na composição da linguagem, que funciona como “meio de interiorização das categorias que estruturam

a relação em sociedade”; e a distinção entre os efeitos dos processos plurais e permanentes de socialização primária e secundária (sendo esta última dos “submundos institucionais ou fundados nas instituições”) ao estudo das dinâmicas biográficas. A noção de *choque* (entre *provincias finitas de sentidos*) de Schütz está na base da definição *choques biográficos*, que Berger e Luckmann utilizaram para precisar os casos em que há mudanças bruscas nos percursos, exigindo conversões (no sentido religioso) desintegradoras das realidades interiorizadas.

As diretrizes de Alfred Schütz foram mobilizadas pelos cientistas sociais da Escola de Chicago preocupados em fundamentar as dimensões múltiplas das realidades e das identidades no quadro do seu modelo construtivista. Erving Goffman estabeleceu particularmente a interlocução com Schütz no seu *Frame Analysis* (1974), no qual buscou compreender os “processos de construção social de realidade” focalizando “rituais e estratégias de interação”, como ressaltou Gilberto Velho (2002, p. 14) em depoimento sobre essa tradição de estudos.

William Thomas, Robert Park e Herbert Blumer, Everett Hughes, Howard Becker, Erving Goffman, entre outros, compõem a genealogia de pesquisadores que mobilizaram as histórias de vida, os relatos orais, sobretudo por meio de observações e entrevistas, como procedimentos congruentes ao tratamento das transformações sociais decorrentes dos deslocamentos migratórios, de “delinquentes” e de *carreiras desviantes* na sociedade norte-americana, particularmente de Chicago¹⁴.

De um modo geral, segundo Becker (1996), eram estudos voltados às interações sociais entre pessoas agrupadas cotidiana e sucessivamente para “fazer coisas em comum”,

agindo de dado modo durante determinado tempo, ou modificando paulatinamente suas maneiras de agir, ou alterando-as significativamente em função de algum acontecimento externo. Outra dimensão importante seria a do “contexto em que se davam os fatos estudados” ou a “dimensão histórica” do trabalho. E, para tanto, a composição de recursos metodológicos poderia (ou deveria) ser ampla, incluindo também dados estatísticos e históricos complementares. De qualquer modo, o tipo de pesquisa priorizada indicava aspectos da existência não captáveis nas fontes oficiais, permitindo desnaturalizar a suposta inclinação a sublinhar as características negativas de certos grupos (como aqueles de origem polonesa) e, por esse intermédio, contestar as ações do Estado sobre eles (PENEFF, 1994). Logo, a perspectiva desse *interacionismo simbólico* não somente era favorável à apreensão dos dados biográficos, como a considera necessária para dar conta das dinâmicas examinadas.

Everett Hughes, discípulo de Park (por sua vez, seguidor de Simmel), desenvolveu suas pesquisas em uma comunidade no Canadá francês, especialmente dedicadas às ocupações e às relações raciais. Becker estudou com ele e, nitidamente, a sua noção de *carreira* foi lapidada a partir desse encontro. Segundo Hughes (1937) podemos conjugar a dimensão mais objetiva ou coletiva da sucessão de papéis, status e cargos ocupados, com uma subjetiva das auto avaliações dos indivíduos, ou seja, uma “perspectiva dinâmica pela qual a pessoa concebe sua vida como um conjunto e interpreta o significado de suas diversas características, das ações e das coisas que lhe ocorrem” (HUGHES, 1937, p. 409-410). Essas dimensões poderiam entrar em conflito, produzir dilemas (devido à possibilidade dos indivíduos

14 Becker (1996) atribuiu a William Thomas a formulação original da ideia de “definição de situação”, inspirada na investigação coletiva sobre a relação entre comunidades de imigrantes e pobreza, estada na realização de entrevistas e histórias de vida tanto com residentes na Polônia como com pessoas que emigraram para os Estados Unidos (publicada com Florian Znaniecki em cinco volumes, intitulada *The Polish Peasant in Europe and America*).

transitarem entre diferentes *mundos*, com distintas expectativas e papéis a serem exercidos) e algumas posições, podendo se complexificar e se objetivar (em instituições, por exemplo), exigindo constantes ajustes.

Com o fim da Segunda Guerra Mundial, a produção da Escola de Chicago entrou em declínio e “a Escola de Chicago, de certo modo, deixou Chicago” e prosperaram os empreendimentos quantitativos, do tipo *surveys*, “menos abertos a estudos com abordagem antropológica” (BECKER, 1996, p. 187). Não por acaso, coincidiam com um momento chave de institucionalização da ciência política nos Estados Unidos que, desde os anos 1930 e 1940, fez confluír seus investimentos de pesquisa com o processo de burocratização do Estado americano (LOWI, 1994). Por exemplo, os “dirigentes” e membros da “classe política” procuravam entender as “aspirações de eleitores” para garantir a manutenção dos seus mandatos e os cientistas políticos produziam sondagens de opinião e forneciam tais respostas com bases “científicas”¹⁵. Ou, como provocou Wright Mills (1975), eram os tempos do “empirismo abstrato”.

Mas Goffman e Becker seguiram com uma produção consistente. O primeiro publicou, entre o final da década de 1950 e início de 1960, obras importantes como *A representação do eu na vida cotidiana* (1959) e *Estigma* (1963). Contudo, para o que interessa mais diretamente neste artigo, enfatizamos o livro *Outsiders*, de Becker, também de 1963, seguindo o modelo de Hughes. Nesse, a ideia de *desvio* é colocada como possibilidade de apreensão de regras coletivas, cujos fundamentos podem ser apreendidos por meio do estudo das *carreiras* de indivíduos inseridos

em processos de interação, com injunções e resistências, legitimidades, conflitos e moralidades, bem como articulando as dimensões objetiva e subjetiva na análise de grupos “que se encontram em baixo dos narizes” (BECKER, 2008, p. 194) dos sociólogos e que, portanto, podem ser atentamente observados.

Velho (2002), reconhecido importador dos pressupostos da Escola de Chicago, relatou que um dos obstáculos às apropriações de autores como Goffman no Brasil foi a preponderância de versões de marxismos e estruturalismos, que confluíam com as expectativas de resistência política ao regime militar. Mas, no final dos anos 1960, particularmente os antropólogos e “profissionais da área *psi*” começaram a utilizá-lo, e seus textos passaram a ser traduzidos, sobretudo na década de 1970. Assim, “a análise do cotidiano e das relações interpessoais, em uma perspectiva socioantropológica, estimulou o desenvolvimento de trabalhos e investigações com preocupação interdisciplinar” (VELHO, 2002, p. 10). O antropólogo, com estadas de estudo em instituições norte-americanas, investiu na apreensão, transmissão e produção de pesquisas, com a publicação de livros e coletâneas, fundamentados no uso desse referencial ou, em suas palavras, dessa “linhagem acadêmica” de Chicago, na qual não haveria “barreiras entre antropologia e sociologia” e seus autores. Em especial, Becker e Goffman representavam a “contribuição norte-americana para a temática indivíduo e sociedade” e aos “estudos urbanos” (VELHO, 2002, p. 10).

Seguindo essa trilha, Velho (2003) destacou que a noção de biografia é fundamental para a compreensão das sociedades complexas em escala individual, pois os indivíduos definem seus *projetos* e se amoldam (ou se

15 Desenhava-se, como destacou Theodore Lowi (1994), um acordo em via dupla: de um lado o compromisso do governo com o investimento em laboratórios e pesquisas em ciência política, e, de outro, o compromisso dos cientistas políticos em dar ao governo os instrumentos necessários às suas decisões com bases ou justificativas científicas.

metamorfoseiam) a uma diversidade de valores, inserções e trânsitos possíveis (não fixos no tempo e no espaço). Disso emerge o *potencial de metamorfose* detido e desigualmente distribuído entre indivíduos, que exercem um “repertório de papéis sociais” (Ibidem, p. 29), dispostos em múltiplos planos, e têm sua existência social condicionada por essa multiplicidade. A capacidade de circulação entre *mundos* (Becker) ou *províncias de significados* (Schütz) informa o potencial de *mediação* de determinados indivíduos, ou seja, de atuar como “negociador” habilitado (VELHO; KUSCHNIR, 1996). Com encontros e desencontros entre projetos individuais e coletivos traçados no decorrer de trajetórias, importa atentar “não [à] escolha isolada do indivíduo, mas [ao] seu processo para chegar até lá; percurso este que pode contemplar não apenas outros indivíduos, mas também instituições – entendendo-as como parte dos aspectos objetivos do real” (DE LUCA; OLIVEIRA; CHIESA, 2016, p. 469).

Do levantamento feito e exposto na primeira seção do artigo, podemos registrar como determinadas pesquisas seguiram esses parâmetros de análise: a de Barros (2010), sobre mobilidade e ciclos de vida entre jovens adultos; de Rial (2008), sobre circulação internacional de jogadores de futebol; e a de Pavão (2017), sobre universitários com dificuldades de usar a escrita – as três publicadas na revista *Horizontes Antropológicos*. Também a discussão de Maio e Lopes (2015), que trata da carreira de Donald Pierson e a Escola de Chicago. Considerando somente as ciências sociais, a noção de “carreira”¹⁶ é mais frequentemente utilizada de forma metafórica ou combinada com outras categorias, principalmente “perfis” e “trajetórias”, ou mesmo assume outras definições. Podemos grifar, entre os artigos da ciência política que

usam a ideia de carreira geralmente associada à noção de recrutamento político, o trabalho que discute gênero, de Miguel e Biroli (2010), a pesquisa sobre deputados, de Leoni, Pereira e Renno (2003), a análise sobre magistrados, de Da Ros (2012), e o texto sobre políticos de esquerda e direita, de Marengo e Serna (2007).

Destaca-se, também na Alemanha, a preponderância de uma perspectiva subjetivista nos anos 1980. Santos, Oliveira e Susin (2014) discutiram os estudos de Fritz Schütze que, seguindo a orientação teórica de Schütz, questionou, por exemplo, a realização de entrevistas controladas (mediante a elaboração de um roteiro, por exemplo) como mecanismo de suspensão da subjetividade do entrevistado. O depoimento do entrevistado não poderia ser “colonizado” em benefício da relevância sociológica.

Porém, podemos realçar o trabalho de Norbert Elias, que, na esteira do pensamento weberiano e freudiano, rompeu enfaticamente com a dicotomia conceitual entre “sociedade” e “indivíduo” e contribuiu às formulações a serem apropriadas nos estudos biográficos (inscrevendo-se no rol de autores alemães mobilizados nos estudos), propondo uma perspectiva ampliada – processual, relacional e concorrencial – da análise de como as biografias individuais somente podem ser apreendidas considerando a estrutura social e suas relações de poder (ELIAS, 1994).

De forma bastante sumária, é possível dizer que sua proposta foi de apreensão das transformações das estruturas sociais com impactos nas estruturas das personalidades dos indivíduos, considerando a configuração de cadeias de interdependências, amarrando e constringendo os comportamentos e percepções, tanto quanto sendo produzidas e

16 Há o debate sobre a necessidade de atualização da noção de *carreira*, constatando seu enfraquecimento e possibilidades de reabilitação. Ver Lawrence (2011) e Arthur (2008).

modificadas (com diferentes ritmos e graus de intensidade) como resultado (incerto) das relações de poder existentes entre seus componentes. Nessa linha, a importância da “biografia” pode ser verificada em várias direções para situar as inscrições dos indivíduos (com características sociais desiguais e cujos desejos, pulsões, expectativas, dilemas, comportamentos etc., são definidos nas relações de interdependência e concorrência com outros) nos processos ou para apreender processualmente as transformações históricas e sociais (que delineiam formas de Estado, instituições, governos, profissões, atividades em geral) conduzidas nas inserções práticas dos indivíduos.

Notadamente, grifamos o “Estudo sobre a gênese da profissão naval” (Idem, 2006). Dividido em duas partes, o trabalho investigou como a afirmação de uma carreira de oficiais da Marinha na Inglaterra do século XIX foi decorrente das interações, da convivência, dos conflitos entre dois grupos (marinheiros experientes e nobres), com origens sociais e atividades diferentes, que formaram uma configuração de interdependências, cujas lutas e convergências levaram à reconfiguração das suas atividades, mais especializadas e mais regradas. Destacamos, ainda, *Mozart: sociologia de um gênio* (ELIAS, 1995)¹⁷ – organizado a partir de diferentes manuscritos e não finalizado devido ao falecimento do autor. Na obra póstuma, Elias procura investigar (com maior ênfase psicogenética que sociogenética), mediante correspondências e materiais biográficos, as condições (e limitações) históricas, sociais e psicossociais de afirmação do “artista” Mozart. Elias considera que o traço mais singular de um indivíduo somente pode ser compreendido quando reconstituímos o tecido social no qual ele está ou esteve

inserido. Somente pela análise biográfica é possível entender o encontro entre o plano macro de características aparentemente estáticas e externas ao “indivíduo”, processos de interiorização e exteriorização de normas e libidos, em tentativas de ajustes nem sempre plácidas, antes disso, podendo ser bem conflitivas e dilemáticas. Assim, em *Mozart*, o autor oferece, como bem sintetizou Bernard Lahire (2010, p. 13), “o estudo preciso de um caso singular, variando as escalas de análise das macroestruturas fixando as relações entre os artistas-músicos e a aristocracia de corte às microestruturas familiares regrando a natureza das relações entre um pai e seu filho”.

As orientações teóricas de Elias, atinentes ao exame de dados biográficos, aparecem, principalmente, articuladas às referências de Pierre Bourdieu e relacionadas às pesquisas sobre intelectuais (o segundo universo de temas mais frequente, segundo mapeamento feito no SciELO), o que será mencionado na próxima seção.

Influência da sociologia francesa: a afirmação do estudo de trajetórias

De modo geral, na década de 1980, o recurso a fontes e informações biográficas – como entrevistas, autobiografias, diários, repertórios biográficos – ganharam força entre sociólogos de diferentes nacionalidades, que delas se apropriaram com diferentes ênfases de análise. O mesmo aconteceu na França, entretanto, com desdobramentos e redirecionamentos significativos nos períodos seguintes até a noção de trajetória assumir centralidade. As ressonâncias desses deslocamentos nas ciências sociais feitas no Brasil, constatadas anteriormente, justificam a breve reconstituição dos debates em torno da questão.

17 Para uma síntese e análise desse trabalho de Norbert Elias, ver Barreira (1997) e Lahire (1993).

Peneff (1994) acentuou a influência de Daniel Bertaux para uma geração de pesquisadores que adotaram a abordagem biográfica a partir dos anos 1970 na França, confluindo com o retorno à história oral ou de vida entre seus homólogos no resto da Europa e nos Estados Unidos. Originalmente reivindicando sua inserção na genealogia da Escola de Chicago e na etnometodologia, Bertaux (1980) enfatizou, no seu relatório “*L’approche biographique: sa validité méthodologique, ses potentialités*” (de 1976, publicado em 1980), o estudo das *narrativas de práticas* em detrimento do uso das biografias como relatos de vida. A coleta de relatos com essa ênfase serviria para extrair padrões de comportamentos e de relações sociais. Isso porque as formas de interação, que envolvem e se introduzem no indivíduo, bem como as lógicas de suas ações, são perceptíveis exatamente a partir das narrativas sobre as práticas, tratadas como fontes de informação cruzadas com outros dados objetivos (tipos de documentos) sobre os processos sociais (CHEVALIER, 1979; DUBAR; NICOURD, 2017).

Com essa “novidade” na forma de empreender uma análise objetivista de relatos de práticas, ou etnosociológica, Bertaux se afastou das perspectivas dominantes. Inverteu o ponto de vista marxista, no qual o indivíduo era explicado a partir das determinações da estrutura social, e questionou o alcance da abordagem acerca dos sentidos atribuídos pelos sujeitos, tornando-os base para explicar práticas e formas de associação. Nesse último caso, conforme Dubar e Nicourd (2017), o debate foi travado com Maurizio Catani, que defendia justamente uma abordagem de orientação declaradamente weberiana. Quer dizer, de buscar o significado subjetivo de uma prática, por meio do relato dado por uma pessoa, no qual ela própria produziria as vinculações a “eventos biográficos passados, inscritos em processos sociais” (Idem, p.13).

Por consequência, o pesquisador privilegiaria os sentidos conferidos às práticas e às relações, uma vez que são eles que fornecem o substrato sociológico à compreensão das ações e das interações sociais. A abordagem sociocompreensiva é utilizada não somente em estudos de casos (individualmente tratados, abordados de forma intensiva e dialógica), mas também na coleta de entrevistas com amostras de indivíduos que contam uma parte da sua vida (Idem).

Em resposta a críticas de Bertaux, Catani apontou a “pobreza” da ênfase nas práticas em detrimento da dimensão simbólica dos fatos estudados, o que teria levado Bertaux a renunciar a possibilidade de formulação de uma “hipótese forte”, conduzindo suas investigações e deduções em prol de uma postura mais indutivista (CHEVALIER, 1979, p. 99). Catani reivindicou a contribuição da sua abordagem às pesquisas concretas, com a introdução do simbólico, não como princípio explicativo, mas como princípio de elucidação, que não recusava a ordem econômica, os antagonismos e os processos históricos, mas os interpretaria de outra maneira (DUBAR; NICOURD, 2017).

Segundo Peneff (1994, p. 26), uma “mudança intelectual” aconteceu com a publicação da “Ilusão biográfica”, por Pierre Bourdieu, no número especial da revista *Actes de la Recherche en Sciences Sociales*, de 1986. Ainda hoje os argumentos do autor atraem várias avaliações que, sendo mais ou menos superficiais ou substantivas, contribuíram para colocá-lo no epicentro do debate acerca dos estudos biográficos e nas reconstituições consagradas ao tema.

Dubar e Nicourd (2017) lembram que o foco da crítica de Bourdieu dirigiu-se a uma “história de vida” trazida de contrabando da literatura para as ciências sociais, o que é diferente da suposta condenação do uso de “relatos de vida” e “biografias” nas pesquisas sociológicas. Esses autores recuperaram quatro aspectos

que informam, ao mesmo tempo, os limites e as potencialidades dos apontamentos de Bourdieu (principalmente pelos usos que sucederam à formulação original), sistematizados a partir de um colóquio específico (“Biografia e ciclos de vida”), realizado em 1988 para debater o artigo: (1) é preciso ter “vigilância com o subjetivismo sem refutar a pluralidade de racionalidades e crenças” (Ibidem, p. 32); (2) existem “zonas de opacidade, traços de construção e reconstrução através de percursos sinuosos”, dificilmente reconstruídos coerentemente nos relatos (Ibidem, p. 33); (3) “não há sentido preexistente à existência”, mesmo se aquele que a relata acredite às vezes nisso e tente apresentar essa coerência na sua narrativa; (4) a centralidade da análise de trajetórias, considerando, porém, os eventos biográficos como mais complexos que a fixidez dos “posicionamentos e deslocamentos no espaço dos campos sociais” (Ibidem, p. 34).

Pierre Bourdieu (1996a, 1996b) propõe a superação das dicotomias entre abordagens antitéticas (a exemplo da separação da dimensão prática e simbólica), questionando a validade da análise de relatos arbitrariamente acumulados, equiparados e dos quais se extraem regularidades, sem considerar as condições objetivas e específicas (inclusive a situação de entrevista) em que esses indivíduos são chamados a serem “ideólogos de suas próprias vidas”, bem como não atentando aos sentidos (relacionais, concorrenciais, constrangedores) que interferem nas suas seleções, conforme o estágio biográfico no qual se encontram (suas expectativas, frustrações, estímulos e estímulos etc.). Bourdieu aponta, ainda, (valendo para Bertaux e Catani) a necessidade de compreender os significados que os indivíduos atribuem para si, para sua vida, para os outros e para a situação em geral (no momento biográfico em que são chamados a “relatar”), levando em conta a localização dos agentes em um espaço de interdependências e inserções ou em uma

estrutura social, econômica e cultural que se impõem, construindo o consenso sobre sua imprescindibilidade e dos seus porta-vozes.

Sendo assim, a utilização de dados biográficos exige a problematização das condições sociais e institucionais de sua objetivação enquanto “dado”. Para tanto, é preciso levar em conta a dimensão simbólica, que garante que as classificações sociais sejam, ao mesmo tempo, produto e produtoras de imposições tanto das formas de auto apreciação como das percepções do mundo social. Da posição social e política que o agente ocupa no momento da sua enunciação e tendo em vista as demais posições correntes, os tipos e a distribuição dos recursos disponíveis, bem como os objetos de luta disputados, derivam duas construções que dificilmente podem ser decompostas. A primeira refere-se às expectativas quanto à maximização do depoimento para o próprio “jogo” do qual o depoente participa, a segunda – e talvez mais importante – é a busca daquele que se apresenta e apresenta sua vida de construir uma “constância de si mesmo como uma história bem construída” (BOURDIEU, 1996a, p. 81).

Assim, Bourdieu (1996a; 1996b) propõe que as *trajetórias sociais* sejam apreendidas levando em consideração um sistema dos traços pertinentes para a compreensão de uma biografia individual ou de um grupo de biografias, constituídas das posições que vão sendo sucessivamente ocupadas por determinado agente (individual ou coletivo) no âmbito de determinado campo de forças. Sendo assim, antes de tudo, é preciso construir esse *espaço de relações objetivas entre posições*, que se definem por *relações objetivas* entre indivíduos e entre *propriedades* (que convergem ou divergem no interior de um campo). A estrutura do *campo* é definida pelo estado das *relações de força* ou *lutas entre espécies de capital* (sobretudo social, econômico, cultural e simbólico). Essas *lutas* visam a imposição

dos *princípios legítimos de dominação*, que incidem sobre *representações e práticas* nos limites do campo e com efeitos na escala mais ampla de hierarquização do mundo social. O *princípio gerador e unificador* das *representações e práticas* é o *sistema disposicional* ou *habitus*, herdado (mormente na socialização familiar e escolar) e mais ou menos consoante com as *posições ocupadas* e a *trajetória* dos agentes (Idem, 1992, p. 191). Nesse sentido, as *trajetórias* são construídas em relação às propriedades e funcionamento do *campo* investigado, considerando as condições de origem ou de entrada e as *posições ocupadas* nesse *campo* e dele em relação a outros *campos* que compõem o espaço social mais amplo (em determinada fase ou em estágios sucessivos).

No sistema conceitual concebido pelo autor, a análise pode contemplar casos que ocorrem em um mesmo campo ou subcampo de forças, examinando seus movimentos relacionais e competitivos, conforme o acúmulo de certos capitais valorizados e a desvalorização de outros, acompanhando as sucessões de posições e destinos. E também são investigadas a partir do deslocamento de agentes no espaço social mais amplo, com origens mais ou menos dominantes ou dominadas em algum *campo*, para posições mais ou menos dominantes ou dominadas no mesmo campo ou em algum diferente. Com efeito, as *trajetórias* podem: assumir um caráter de ascensão ou declínio; decorrer de bloqueios de uma via, redefinindo a direção do percurso em outro rumo; configurar continuidades previsíveis, ou exigir *reconversões* significativas de capitais, sentidos e práticas, e assim por diante.

Apontamento específico pode ser feito sobre a necessidade de flexibilizar a metáfora do trajeto no metrô, que Bourdieu (1996b, p. 81) aciona para ponderar sobre o imperativo de tratar “os acontecimentos biográficos [...] como alocação e como deslocamentos no espaço social”.

Tentar compreender uma vida como uma série única e, por si só, suficiente de acontecimentos sucessivos, sem outra ligação que a vinculação a um ‘sujeito’ cuja única constância é a do nome próprio, é quase tão absurdo quanto tentar explicar um trajeto no metrô sem levar em conta a estrutura da rede, isto é, a matriz das relações objetivas entre as diversas estações. (Ibidem).

A crítica feita por Passeron (1995, p. 20) a Bourdieu a esse respeito parece elucidativa. Para começar, ele localizou dois polos igualmente vulneráveis de análise das biografias. O primeiro é o da “utopia biográfica”, que se orienta na convicção de que “nada é insignificante”, “baseia sua impressão de compreender na ilusão do imediato”, adere à “*ilusão da pan-pertinência* do descritível”, em detrimento da formulação de uma problemática de pesquisa orientando a análise dos “*traços pertinentes da descrição*” (Idem, p. 206-207, grifos do autor). Portanto, semelhante às críticas de Bourdieu. O outro polo é o da abordagem estruturalista, que deseja somente conhecer:

a realidade dos traços pertinentes e dos sistemas de relações que os constituem: realismo das posições e das oposições, das leis de transformação ou de reprodução, para que os devires individuais, deixando de ser convites à descrição do que, numa singularidade, presta-se à intelecção de generalidades, não podem mais ser percebidos como “portadores da estrutura”, ocupantes de um “sistema” ou de um “campo” – enfim, agregações de propriedades sincrônicas. (Ibidem, p. 208).

Dessa forma, o alerta de Passeron é quanto à necessidade de atentar não somente para a relação entre *trajetória* e *campo*, mas aos deslocamentos coletivos (com impactos diferenciados em cada indivíduo) e às transformações simultâneas e interdependentes que envolvem aparatos objetivados e expectativas subjetivas. Por isso, diz preferir a metáfora anterior de Schumpeter, pois, “antes que Bourdieu entrasse no metrô, Schumpeter já havia tomado o ônibus” (Ibidem, p. 210). A partir de determinado

“recorte”, definido como relevante ao estudo (classe, grupo, organização etc.), tomado como um *veículo* que produz efeitos duráveis: “assim como os ônibus cujo trajeto constitui um objeto específico de descrição, mesmo que na chegada os veículos já não contenham os mesmos viajantes que havia na partida e, no máximo, que no terminal não esteja mais nenhum dos que nele subiram” (Ibidem). Além disso, há a possibilidade de pensar nas “circunstancialidades” que podem agir no tempo do trajeto do coletivo (interferências no trânsito, por exemplo), que é distinta do “tempo dos viajantes”, mas mais ou menos sincronizadas. Tais reconfigurações levam a perguntar sobre essa diversidade, pois “mesmo portadores das *mesmas* propriedades os viajantes talvez jamais sejam os *mesmos*: não visando mais a mesma história” (Ibidem, p. 211).

Ainda mais recentemente, Bernard Lahire (2004, 2005, 2006), a partir de uma série de pesquisas realizadas em diferentes meios sociais (universo familiar, escola primária, estudantes, literatos etc.), estabeleceu uma interlocução direta com Bourdieu (sobretudo problematizando as noções de *campo* e a realocação da ideia de *habitus*) para propor seu modelo de análise disposicional e contextual. Há a defesa da pluralidade de possibilidades (não raro ambivalentes) de práticas, disposições e biografias dos indivíduos. Ele retorna à questão da complexificação e da diferenciação sociais, sublinhando que os indivíduos passam por múltiplas e concorrentes “instâncias socializadoras” (família, escola, profissão, cultura, política, religião etc.), assim como são plurais seus trânsitos, contatos, elos e os princípios e normas orientadores das ações. Portanto, a análise das *biografias sociológicas* seria uma forma de abarcar o impacto desses aspectos em escala individual, sem perder de vista as determinações sociais que os fabricam. Ou seja, é preciso apreender tanto os “*patrimônios individuais de disposições e de competências culturais incorporadas*”, que são “mais ou menos

homogêneas ou heterogêneas”, quanto as “*propriedades dos diversos contextos de práticas*, que entram mais ou menos em contradição com esses patrimônios individuais de disposições e competências” (LAHIRE, 2006, p. 21-22, grifos do autor).

Vale mencionar que, além da interlocução com o esquema bourdieusiano, no prólogo ao seu *Franz Kafka: elementos para uma teoria da criação literária* – no qual examina, principalmente por meio da autobiografia, correspondências e outros materiais, a “fabricação social do autor” – Lahire (2010) ressalta a inspiração contundente do *Mozart*, de Elias (já citado anteriormente), no seu trabalho. Principalmente, cabe realçar a distinção feita entre biografia sociológica e biografia literária ou histórica, na qual o estudo deve considerar, como citado, os universos sociais que contribuem na edificação do *estoque de disposições*, a inter-relação entre eles e em cada um deles. Inseparavelmente, os lugares ocupados, as variações diacrônicas e sincrônicas ou as reorganizações do patrimônio, as relações vivenciadas entre instituições, a influência das sociabilidades, as tensões ou crises enfrentadas e os contextos específicos de atuação (personagens, contatos, práticas etc.).

A gama de perspectivas revisitadas nas seções anteriores moldurou as apropriações e justaposições de referenciais feitos por cientistas sociais brasileiros, que se dedicaram à coleta e sistematização de informações biográficas. No momento seguinte, tentaremos esboçar, em grandes linhas, alguns circuitos de circulações dessas noções e estratégias metodológicas.

Uso de informações biográficas e estudos de trajetórias no Brasil

O uso de certos autores no Brasil é tributário da importação de pesquisadores (geralmente posicionados no centro da produção nacional), que incorporam tais discussões e referências quando têm contato com elas,

sobretudo em suas estadas de estudos no exterior. Não foi diferente com Bourdieu, que é um dos autores mais citados por cientistas sociais brasileiros. Segundo Bortoluci, Jackson e Pinheiro Filho (2015), as primeiras apropriações foram feitas no âmbito dos estudos rurais, e a primeira publicação do autor no país, do texto “Campo intelectual e projeto criador”, em 1968, foi viabilizada por Moacir Palmeira (que acompanhava, no período, seus seminários na École des Hautes Études en Sciences Sociales). O antropólogo teria sido o cientista social brasileiro pioneiro na utilização da obra do sociólogo francês em sua tese de doutorado (de 1971) e formou uma geração de disseminadores de Bourdieu no país, desde sua atuação no Museu Nacional, que se especializaram em pesquisas sobre transformações no mundo rural, disputas políticas e trajetórias intelectuais. Geralmente com força da abordagem mais qualitativa, das observações de campo, entrevistas ou outras fontes documentais, como suporte às reconstituições biográficas, inseridas em contextos de luta e mudança social, com a apreensão das representações e estratégias de reprodução social.

Com essa orientação, um conjunto de pesquisas sobre trajetórias políticas e intelectuais foram feitas. Destacamos os trabalhos de Afrânio Garcia Jr. (1993, 2004, 2011), Mário Grynszpan (1990, 1996), Gustavo Sorá (1996, 2011), Gabriela Scotto (1995), Federico Neiburg (1995) e Barreira (2002). Além das construções biográficas e os condicionantes do reconhecimento conquistado pelos agentes investigados, esses pesquisadores se debruçaram sobre reconfigurações mais amplas ocorridas em determinadas dinâmicas

históricas; com a produção de bens simbólicos, seu conteúdo e suas modalidades de recepção/circulação; com o desenho de redes de relações, o trabalho de mediação, de representação e as representações construídas/compartilhadas; entre outras dimensões de análise contempladas principalmente por meio da observação de campo, relatos e documentos produzidos nas dinâmicas de concorrência. Observa-se a preocupação com a descrição minuciosa das bases de formação e socialização dos agentes, dos seus vínculos pessoais, dos trânsitos, dos recursos acumulados e acionados nas relações de trocas em diferentes níveis.

Sergio Miceli contribuiu também decisivamente na tradução, afirmação e preponderância das ideias de Bourdieu no Brasil, sobretudo a partir de pesquisas no campo da sociologia da cultura. Ele assinou (MICELI, 1974) a edição da *Economia das Trocas Simbólicas*, “uma das primeiras coletâneas de Pierre Bourdieu publicadas no mundo”, e a apresentação da publicação, “que se tornaria uma espécie de guia de leitura para esse conjunto de textos – e, por extensão, para a obra –, com o condão de apresentar a novidade produzida no centro de modo a fazer sentido num sistema intelectual periférico” (BORTOLUCI; JACKSON; PINHEIRO FILHO, 2015, p. 228).

Em uma das suas pesquisas mais citadas (fruto das orientações de doutorado de Bourdieu e também de Leôncio Martins Rodrigues¹⁸), Miceli (1979) localiza os dados biográficos dos agentes estudados para a análise das trajetórias intelectuais com a preocupação de situá-los em espaços mais amplos de posições, seguindo de perto o modelo de análise bourdieusiano das condições e condicionantes

18 Cabe lembrar que Martins Rodrigues (2009) é sociólogo e cientista político, e atentou aos condicionantes sociais e culturais de atuação política de sindicalistas e parlamentares. Assim como Miceli (que escreveu sobre Mário de Andrade), ele compôs a publicação sobre “intérpretes” do Brasil (basicamente perfis biográficos), dedicando-se ao trajeto de Fernando Henrique Cardoso. Ver Botelho e Schwarcz (2009).

sociais e históricos de formação/transformação do campo cultural, da ocupação de posições e das tomadas de posição intelectuais e políticas dos agentes. É difícil recuperar a quantidade e diversidade de pesquisas publicadas pelo autor (MICELI, 1979, 1981, 1996, 2009a, 2009b, 2013, entre outras), que demonstram a mobilização das informações biográficas indissociadas dos condicionantes e das práticas sociais, das disposições e dos trunfos acumulados, das condições de inserção e das tomadas de tomada de posição dos agentes, ou seja, contemplando as dimensões de análise e de operacionalização propostas por Bourdieu para os estudos de trajetórias.

Podemos ainda mencionar a influência de Norbert Elias nas reflexões de Miceli sobre casos específicos, como a análise que fez de Jorge Luis Borges (MICELI, 2007a), recuperando informações sobre a configuração do campo intelectual argentino, as redes de socialização/sociabilidades, as oportunidades de acumulação de capital cultural, a relação com o pai, os conflitos internos, enfim, o conjunto de elementos que confluíram na construção do “escritor nato”. Na mesma configuração histórica, o exame das trajetórias *outsiders* (inclusive em relação a Borges) de Alfonsina Storni e Horacio Quiroga, localizados no polo mais heterodoxo do campo cultural em expansão (Idem, 2013). Para tanto, situou as informações biográficas, das origens sociais aos imponderáveis dos trajetos, analisou dos poemas e posicionamentos às representações em fotografias de ambos, das condições históricas aos dilemas psicogenéticos, entre tantas dimensões que favoreceram a apreensão da história social e dos destinos dos “personagens”.

Uma série de pesquisas foi realizada com essa orientação, com destaque para o estudo de trajetórias de mulheres artistas entre o final do século XIX e início do XX, feito por Simioni (2007a, 2007b), e de intelectuais franceses inscritos no campo de disputas por legitimidade no final do século XIX, em especial Gabriel

Tarde e Gustave Le Bon, empreendida por Marcia Consolim (2008). Com demarcadores semelhantes, salientamos as pesquisas desenvolvidas por Heloisa Pontes (2009, 2011, 2013, entre outros) e por ela junto com Miceli (2012). A autora tem se dedicado igualmente aos estudos intelectuais, em especial a artistas (principalmente de teatro) e com recorte de gênero, enfocados a partir das suas condições mais ou menos oportunas de profissionalização e de representação(s).

Com proximidade a esse tipo de análise – porém para investigar as intersecções entre domínios intelectuais e políticos – Grill (2015) analisou as *multinotabilidades* de Afonso Arinos, notadamente debruçando-se sobre suas “memórias” e utilizando repertórios biográficos (como CPDOC). A partir desse artigo queremos ponderar sobre uma última linha de estudos. Além do fortalecimento de uma sociologia da cultura, a influência de Bourdieu e de seus aplicadores no Brasil favoreceu igualmente o avanço dos estudos de trajetórias no âmbito de uma parte da ciência política brasileira, que se voltou à percepção das informações biográficas, em especial ao estudo de carreiras e dos perfis políticos.

Mais especificamente na linha das reflexões da chamada sociologia política francesa, os trabalhos de Daniel Gaxie (1977, 1980, entre outros), Michel Offerlé (1996) e, depois, de Frederic Sawicki (1997), são referências persistentes. Com eles, a compreensão da atividade política deve considerar padrões de recrutamento dos agentes políticos, seu *background*, origens sociais, “aquisição de um domínio prático das categorias e as disposições dos agentes” (GRILL, 2008a, p. 17), as multiposicionalidades desfrutadas, os capitais de notoriedade e de investitura acumulados, as empresas coletivas (organizações partidárias, midiáticas, jurídicas, econômicas etc.) empenhadas, o trabalho de mobilização eleitoral efetuado, as redes de relações acionadas etc. (Idem; GRILL, 2013).

O uso dessas discussões é tributário das apropriações de Odaci Luiz Coradini (2001, 2003, 2004, 2007, entre outros), que se doutorou em antropologia sob a orientação de Moacir Palmeira, porém alocado em um Programa de Pós-graduação em Ciência Política, da Universidade Federal do Rio do Sul. Ele contribuiu com as indicações à pesquisa de André Marengo dos Santos (2000), cientista político formado naquela instituição, então em fase de produção da sua tese de doutorado sobre recrutamento parlamentar. Os investimentos de inscrição e afirmação na disciplina desse último fizeram que suas apostas tenham sido dirigidas aos estudos de carreiras políticas (SANTOS, 1997, 2001), mediante os quais fusionou o jargão bourdieusiano (resquícios das leituras de Daniel Gaxie e Michel Offerlé, mas cada vez menos preponderante em seus textos) com outras influências como Angelo Panebianco, Pipa Norris e Richard Katz¹⁹. Ao priorizar as variáveis institucionais em detrimento das sociais, portanto afastando-se da abordagem da sociologia política francesa, Marengo foi acolhido como referência entre os pares e passou a ser considerado um dos porta-vozes do uso da noção de carreira na ciência política. Nessa linha, os levantamentos realizados com base na aplicação de modelos estatísticos permite cotejar o comportamento de variáveis, sobretudo políticas, com peso nos critérios institucionais mais formalizados. A formalização de padrões mais gerais de recrutamento ou de deslocamentos pode ser operada sem tanta preocupação com as vulnerabilidades das fronteiras entre os domínios sociais ou os limites da aplicação de modelos exógenos de análise.

Provavelmente por conta da sua formação original e dos universos de análise privilegiados, Coradini preocupa-se em utilizar de forma problematizada o esquema bourdieusiano (sem deixar de tomá-lo como referência principal), questionando certas fragilidades da importação e aplicação direta e propondo a combinação com outros referenciais de análise, igualmente pertinentes. Aproximando-se tanto do tipo de estudo estimulado por Palmeira, como daqueles empreendidos por Miceli, aplica a análise sociográfica como base para o estudo de trajetórias, a não restrição dos universos empíricos passíveis de serem investigados em múltiplos níveis, sobretudo, propõe a apreensão da multidimensionalidade e inter cruzamentos de lógicas entre eles.

Várias pesquisas foram feitas seguindo essa trilha. Recortaram-se os universos no sentido de apreendê-los de forma amalgamada, testando a hipótese da preponderância estrutural e estruturante dos princípios políticos (não no sentido estrito, mas amplo) nos múltiplos domínios sociais²⁰. Realçamos os trabalhos sobre recrutamento de elites e carreiras políticas de Bordignon (2017), sobre elites parlamentares e culturais de Grill e Reis (2016), sobre militares e eclesiásticos de Seidl (2008, 2017), sobre as inserções culturais de mulheres parlamentares de Reis (2014), sobre “famílias” e “memórias” de políticos, desenvolvidos por Grill (2008b, 2012), sobre intelectuais e as ciências sociais de Reis (2013), sobre elites jurídicas de Engelman (2006), entre outros. Afora coletâneas aglutinando trabalhos convergentes, como as publicações de Coradini (2008), de Seidl e Grill (2013), e de Reis e Grill (2016).

19 Perspectiva semelhante tem sido adotada por estudos produzidos por cientistas políticos da Universidade Federal do Paraná, que pesquisam carreiras e recrutamento político, mobilizando jargões da sociologia política e dos estudos de elites, assim como são responsáveis pela *Revista de Sociologia e Política*, localizada entre os periódicos que mais publicam artigos sobre a temática. Ver Renato Perissinotto, Costa e Tribess (2009) e Perissinotto et al. (2017) e Codato et al. (2016).

20 Para uma discussão um pouco mais detalha da ideia de “domínio”, ver Grill e Reis (2018) e Reis e Grill (2017).

Podemos ressaltar, nessa agenda de estudos, o uso das informações biográficas para a construção de quadros prosopográficos ou para a aplicação de análises de correspondência múltipla, que permitem o mapeamento relacional das características de um conjunto de agentes, inseridos em determinados domínios ou espaços de luta, problematizando os limites da ideia de campo para circunscrevê-los em configurações históricas como a brasileira. Essa perspectiva garantiria a representatividade do estudo de certas trajetórias “exemplares” ou “desviantes”, cujas reconstituições permitem perceber mais detidamente os posicionamentos assumidos e os recursos acumulados, combinando abordagem sincrônica e diacrônica.

Considerações finais

Neste artigo buscamos demonstrar um pouco da diversidade do uso de biografias mobilizadas em pesquisas das ciências sociais e história. Para tanto, apresentamos um levantamento nas bases SciELO, que permitiu evidenciar diferentes modalidades de estudos (particularmente, de história de vida, história oral, trajetórias e carreiras) relativamente aos periódicos e áreas das ciências sociais. Principalmente, observamos a tendência crescente das investigações sobre trajetórias (não raro adjetivadas ou em composição com outro dos enquadramentos), realizadas com representantes de diversas categorias sociais, com destaque para mulheres, intelectuais, políticos, profissionais e militantes.

Com base nesse mapeamento, localizamos artigos (sobretudo aqueles de cunho mais teórico e metodológico), que possibilitaram verificar aportes teóricos e metodológicos preponderantemente acionados nas discussões. Procuramos pontuá-los (muito sumariamente) seguindo uma lógica mais ou menos cronológica e geral de sua afirmação e apropriação

por cientistas sociais brasileiros. Dessa forma, aquelas regularidades, inicialmente apresentadas, ganharam elementos um pouco mais qualitativos e dinâmicos das suas condições de emergência e de consagração no âmbito nacional.

É possível dizer que a reflexão em torno das biografias caracteriza-se por uma flexibilidade e diversidade de aplicações em campos variados do conhecimento, que atualizam a tradicional oposição ou compatibilização das categorias “indivíduo” e “sociedade”, e que estão longe de se esgotarem, pois se referem a questões fundantes à investigação dos mecanismos e funcionamento da vida social. Versões de sociedade como fruto ou produto de interações se imiscuem na problemática, de forma direta ou indireta.

Os debates, polêmicas e ambivalências fundamentam-se na relevância no tratamento das características e percursos de indivíduos, por meio de informações/relatos que eles mesmos fornecem ou de fontes (incluindo publicações) institucionais que as armazenam, oscilando (no tempo e segundo as disciplinas) e se confrontando conforme: a adesão a métodos mais qualitativos ou quantitativos; visando o tratamento de dimensões mais subjetivistas (com peso relativo da dimensão das personalidades ou psicologizantes, ou das formas de representação do mundo social) ou objetivistas (operando dados mensuráveis dos quais é possível sistematizar padrões ou regularidades); voltados a diferentes escalas explicativas (mais micro ou macro); com foco em análises mais internalistas ou externalistas; priorizando categorias sociais de investigação alocadas em distintas posições “dominadas” ou “dominantes” – todos esses elementos podendo ser originalmente articulados nos trabalhos específicos.

Observamos que algumas tradições de conhecimento não apenas assumiram as biografias como objetos centrais da sua reflexão,

como seus porta-vozes as adotaram mesmo como causa a ser defendida, como forma de “visibilização” seja das elites ou das chamadas “minorias”, de qualquer modo, como condição de delimitação de fronteiras disciplinares e de manutenção de reconhecimentos intelectuais. Entram em cena as narrativas, singularidades, representações, subjetividades e tudo o que precisar ser exaltado contra a padronização ou abstrações das estruturas. Elas foram relativizadas ou ponderadas com abordagens que propõem a integração de níveis mais dinâmicos de análise para escapar das armadilhas que polarizam o estudo subjetivista do indivíduo (que parece atribuir uma essência ao biográfico) ou o estudo objetivista das coletividades abstratas. Noções como as de *mundo*, *configuração* e *campo*, respectivamente de Becker, Elias e Bourdieu, são basilares aos esforços de romper com a “forma dual, substancializada e reificante de falar, pensar e investigar” (GRILL; REIS, 2018, p. 169).

As importações das formulações de Pierre Bourdieu feitas por pesquisadores brasileiros, com estadas de estudos na França, influenciaram na gama de pesquisas sobre trajetórias produzidas nas diferentes áreas das ciências sociais brasileiras, com distintas ênfases, universos de análise e graus de rigor²¹. Elas têm diferentes desdobramentos, que não deixam de levar em conta a tentativa de não resistir à sedução da narrativa, na comoção do depoimento, na versão construída, examinando o material biográfico relativamente às condições particulares de sua produção e em relação a processos sociais mais amplos. Sobreretudo a questão principal é a problemática em pauta na pesquisa realizada, tentando, de preferência, encontrar o justo meio entre os extremos subjetivistas e objetivistas, que podem anulá-lo. Assim, a coleta e tratamento do material biográfico “depende da construção do olhar sociológico sobre esses dados e das modalidades de sua análise e interpretação” (DUBAR; NICOURD, 2017, p. 4).

Referências

- ADÃO, M. C. O. Aspectos da adesão feminina aos valores militares: o casamento e a família militar. *História*, Assis, v. 29, n. 2, p. 116-134, 2010.
- ALBERTI, V. *História oral: a experiência do CPDOC*. Rio de Janeiro: CPDOC, 1989.
- _____. De “versão” a “narrativa” no Manual de história oral. *História oral*, Rio de Janeiro, v. 15, n. 2, p. 159-166, 2012.
- ARAÚJO, M. P. N. Por uma história da esquerda brasileira. *Topoi*, Rio de Janeiro, v. 3, n. 5, p. 333-353, 2002.
- ARTHUR, M. B. Examining contemporary careers: a call for interdisciplinary inquiry. *Human Relations*, New Castle, v. 61, n. 2, p. 163-186, 2008.
- BARREIRA, C. Pistoleiro ou vingador: construção de trajetórias. *Sociologias*, Porto Alegre, ano 4, n. 8, p. 52-83, 2002.

21 Evidentemente, a referência às “trajetórias”, assim como acontece para a noção de *campo*, pode significar apenas a automatização do jargão, vinculado às possibilidades de apropriações cada vez mais amplas e frouxas (sem grandes pretensões de transposição de obstáculos teóricos), nas quais Bourdieu e outros autores são apenas instrumentalizados nas tomadas de posição que animam o/as cientistas sociais brasileiro/as.

- BARREIRA, I. “Mozart: sociologia de um gênio”, de Norbert Elias. *Revista Brasileira de Ciências Sociais*, São Paulo, v. I, p. 41-45, 1997.
- _____. O lugar do indivíduo na sociologia: sob o prisma da liberdade e dos contragimentos sociais. *Revista de Ciências Sociais*, Fortaleza, v. 34, n. 2, p. 51-63, 2003.
- BARROS, M. M. Trajetórias de jovens adultos: ciclo de vida e mobilidade social. *Horizontes Antropológicos*, Porto Alegre, ano 16, n. 34, p. 71-92, 2010.
- BECKER, H. Conferência: a Escola de Chicago. *Mana*, Rio de Janeiro, n. 2, v. 2, p. 177-188, 1996.
- _____. *Outsiders: estudos de sociologia do desvio*. Rio de Janeiro: Jorge Zahar, 2008.
- BERGER, P. L.; LUCKMANN, T. A construção social da realidade: tratado de sociologia do conhecimento. Petrópolis: Vozes, 2004.
- BERTAUX, D. L'approche biographique: sa validité méthodologique, ses potentialités. *Cahiers Internationaux de Sociologie*, Paris, v. LXIX, n. 2, p. 198-225, 1980.
- BLAY, E. A. Inquisição, inquisições: aspectos da participação dos judeus na vida sócio-política brasileira nos anos 30. *Tempo Social*, v. 1, n. 1, p. 105-130, 1989.
- BORDIGNON, R. R. Recrutamento e modalidades de entrada na carreira política: candidatos aos cargos legislativos no Rio Grande do Sul (1998-2006). *Política & Sociedade*, Florianópolis, v. 16, n. 35, p. 351-379, 2017.
- BORTOLUCI, J. H.; JACKSON, L. C.; PINHEIRO FILHO, F. Contemporâneo clássico: a recepção de Pierre Bourdieu no Brasil. *Lua Nova*, São Paulo, n. 94, p. 217-254, 2015.
- BOTELHO, A.; SCHWARCZ, L. M. (Orgs.). *Um enigma chamado Brasil: 29 intérpretes e um país*. São Paulo: Companhia das Letras, 2009.
- BOURDIEU, P. *A economia das trocas simbólicas*. São Paulo: Perspectiva, 1992.
- _____. *As regras da arte: gênese e estrutura do campo literário*. São Paulo: Companhia das Letras, 1996a.
- _____. A ilusão biográfica. In: BOURDIEU, P. *Razões práticas: sobre a teoria da ação*. Campinas: Papyrus, 1996b.
- CAMARGO, A. Os usos da história oral e da história de vida. In: CONGRESSO MUNDIAL DE SOCIOLOGIA, 10., 1982, Cidade do México. *Anais...* São Paulo: USP, 1982.
- CARDOSO, H. H. P. Narrativas de um candango em Brasília. *Revista Brasileira de História*, São Paulo, v. 24, n. 47, 2004.
- CARVALHO, I. C. M. Biografia, identidade e narrativa: elementos para uma análise hermenêutica. *Horizontes Antropológicos*, Porto Alegre, v. 9, n. 19, p. 283-302, 2003.

- CHEVALIER, Y. La biographie et son usage en sociologie. *Revue Française de Science Politique*, Paris, ano 29, n. 1, p. 83-101, 1979.
- CODATO, A. et al. Regime político e recrutamento parlamentar: um retrato coletivo dos senadores brasileiros antes e depois da ditadura. *Revista de Sociologia e Política*, Curitiba, v. 24, n. 60, p. 47-68, 2016.
- CONSOLIM, M. C. Gabriel Tarde e as ciências sociais francesas: afinidades eletivas. *Mana*, Rio de Janeiro, v. 14, n. 2, p. 269-298, 2008.
- CORADINI, O. L. *Em nome de quem? Recursos sociais no recrutamento de elites políticas*. Rio de Janeiro: Relume Dumará, 2001.
- _____. As missões da “cultura” e da “política”: confrontos e reconversões de elites culturais e políticas no Rio Grande do Sul (1920-1960). *Estudos Históricos*, Rio de Janeiro, n. 32, p. 125-144, 2003.
- _____. A formação da elite médica, a Academia Nacional de Medicina e a França como centro de importação. *Estudos Históricos*, Rio de Janeiro, n. 35, p. 3-22, 2004.
- _____. Engajamento associativo-sindical e recrutamento de elites políticas: tendências recentes no Brasil. *Revista de Sociologia e Política*, Curitiba, n. 28, p. 181-203, 2007.
- _____. (Org.). *Estudos de grupos dirigentes no Rio Grande do Sul: algumas contribuições recentes*. Porto Alegre: Editora da UFRGS, 2008.
- DA ROS, L. Juízes profissionais? Padrões de carreira dos integrantes das Supremas Cortes de Brasil (1829-2008) e Estados Unidos (1789-2008). *Revista de Sociologia e Política*, Curitiba, v. 20, n. 41, p. 149-169, 2012.
- DELUCA, G.; OLIVEIRA, S.; CHIESA, C. D. Projeto e metamorfose: contribuições de Gilberto Velho para os estudos sobre carreiras. *RAC*, Rio de Janeiro, v. 20, n. 4, art. 4, p. 458-476, 2016.
- DOSSE, F. *O desafio biográfico: escrever uma vida*. São Paulo: Edusp, 2015.
- DUBAR, C.; NICOURD, S. *Les biographies em sociologie*. Paris: La Découverte, 2017.
- ELIAS, N. *A sociedade dos indivíduos*. Rio de Janeiro: Jorge Zahar, 1994.
- _____. *Mozart, sociologia de um gênio*. Rio de Janeiro: Jorge Zahar, 1995.
- _____. Estudos sobre a gênese da profissão naval. In: NEIBURG, F.; WAIZBORT, L. (Orgs.). *Escritos & ensaios*. Rio de Janeiro: Jorge Zahar, 2006. (v. 1: Estado, processo, opinião pública).
- ENGELMANN, F. Internacionalização e ativismo judicial: as causas coletivas. *Lua Nova: Revista de Cultura e Política*, São Paulo, n. 69, p. 123-146, 2006.

- FARIAS, C. M. Entre lembranças e silêncios: reflexões sobre uma autobiografia feminina. *Estudos Históricos*, Rio de Janeiro, v. 22, p. 238-257, 2009.
- FERREIRA, M. M. (Coord.). *Entrevistas: abordagens e usos da história oral*. Rio de Janeiro: FGV Editora, 1994.
- _____. Apresentação. In: RÉMOND, R. (Org.). *Por um a história política*. Rio de Janeiro: UFRJ/FGV, 1996.
- _____. História, tempo presente e história oral. *Topoi*, Rio de Janeiro, v. 3, n. 5, p. 314-332, 2002.
- FERREIRA, M. M.; FERNANDES, T. M.; ALBERTI, V. (Org.). *História oral: desafios para o século XXI*. Rio de Janeiro: Editora Fiocruz/Casa de Oswaldo Cruz: Fundação Getúlio Vargas/CPDOC, 2000.
- FERREIRA, M. M.; AMADO, J. (Coords.). *Usos & abusos da história oral*. Rio de Janeiro: FGV Editora, 1996.
- GARCIA JUNIOR, A. Les Intellectuels et la conscience nationale au Brésil. *Actes de la Recherche en Sciences Sociales*, Paris, n. 98, p. 20-33, 1993.
- _____. A dependência da política: Fernando Henrique Cardoso e a sociologia no Brasil. *Tempo Social*, São Paulo, v. 16, n. 1, p. 285-300, 2004.
- _____. Meninos de engenho: tradições e dramas familiares feitos símbolos da brasilidade. *Revista Antropológica*, Niterói, n. 30, p. 21-47, 2011.
- GAXIE, D. Économie des partis et rétributions du militantisme. *Revue Française de Science Politique*, Paris, v. 27, n. 2, p. 123-154, 1977.
- _____. Les logiques du recrutement politique. *Revue Française de Science Politique*, Paris, v. 30, n. 3, p. 5-45, 1980.
- GONÇALVES, M. A.; MARQUES, R.; CARDOSO, V. *Etnobiografia: subjetivação e etnografia*. Rio de Janeiro: 7 Letras, 2012. (Coleção Sociologia e Antropologia).
- GRILL, I. G. Múltiplas dimensões de uma agenda comum de pesquisas: elites, profissionais e lideranças políticas. In: GRILL, I. G. et al. (Orgs.). *Elites, profissionais e lideranças políticas (RS e MA): pesquisas recentes*. São Luís: Edufma, 2008a.
- _____. "Heranças política" no Rio Grande do Sul. São Luís: Edufma, 2008b.
- _____. "Memórias" de políticos brasileiros: produção escrita, gestão de imagens e "teorizações" nativas do jogo político. *Revista Política & Sociedade*, Florianópolis, v. 11, n. 22, p. 11-40, 2012.
- _____. Especialização política: bases sociais, profissionalização e configurações de apoios. In: SEIDL, E.; GRILL, I. G. *As ciências sociais e os espaços da política no Brasil*. Rio de Janeiro: FGV Editora, 2013.
- _____. As múltiplas notabilidades de Afonso Arinos: biografias, memórias e a condição de elite no Brasil do século XX. *Revista de Sociologia e Política*, Curitiba, v. 23, n. 54, p. 21-42, 2015.

- GRILL, I. G.; REIS, E. T. Dos campos aos domínios das “elites” no Brasil. *Tomo*, São Cristóvão, n. 32, p. 163-210, 2018.
- _____. *Elites parlamentares e a dupla arte de representar: intersecções entre “política e “cultura” no Brasil*. Rio de Janeiro: FGV Editora, 2016.
- GRYNSZPAN, M. Os idiomas da patronagem: um estudo da trajetória de Tenório Cavalcanti. *Revista Brasileira de Ciências Sociais*, São Paulo, v. 5, n. 14, p. 73-90, 1990.
- _____. A teoria das elites e sua genealogia consagrada. *BIB: Revista Brasileira de Informação Bibliográfica em Ciências Sociais*, São Paulo, n. 41, p. 35-83, 1996.
- GUEDES, S. L. Vidas revisitadas: a reconstrução do passado de idosos em programas gerontológicos. *Revista Sociologia & Antropologia*, Niterói, v. 4, n. 1, p. 233-255, 2014.
- HARTMANN, L. “Me curei e hoje tô aqui sentadita contando a história.” Narrativa, performance e constituição da pessoa na fronteira entre Brasil, Argentina e Uruguai. *Revista Sociedade e Estado*, Brasília, v. 29, n. 3, p. 767-789, 2014.
- HUGHES, E. C. Institutional office and the person. *American Journal of Sociology*, Illinois, v. 43, n. 3, p. 404-413, 1937.
- LAHIRE, Bernard. NORBERT ÉLIAS, Mozart. Sociologie d’ungénie. *Annales: Économies, Sociétés, Civilisations*, Paris, année. 48, n. 6, p. 1629-1632, 1993.
- _____. *Retratos sociológicos*. Porto Alegre: Artmed, 2004.
- _____. *A cultura dos indivíduos*. Porto Alegre: Artmed, 2006.
- _____. *Franz Kafka: éléments pour une théorie de la création littéraire*. Paris: Éditions la Découvert, 2010.
- LAWRENCE, B. S. Careers, social context and interdisciplinary thinking. *Human Relations*, New York, v. 64, n. 1, p. 59-84, 2011.
- LE GOFF, Jacques; NORA, Pierre (dir.). *Faire de l’histoire* (volumes 1, 2 e 3). Paris: Gallimard, 1974.
- LEONI, E.; PEREIRA, C.; RENNÓ, L. Estratégias para sobreviver politicamente: escolhas de carreiras na Câmara de Deputados do Brasil. *Opinião Pública*, Campinas, v. 9, n. 1, p. 44-67, 2003.
- LEVILLAIN, P. Os protagonistas: da biografia. In: RÉMOND, R. (Org.). *Por uma história política*. Rio de Janeiro: Editora UFRJ, 1996.
- LOBO, E. S. Experiências de mulheres: destinos de gênero. *Tempo Social*, São Paulo, v. 1, n. 1, p. 169-182, 1989.
- LOWI, T. O Estado e a ciência política ou como nos convertemos naquilo que estudamos. *BIB: Revista Brasileira de Informação Bibliográfica em Ciências Sociais*, São Paulo, v. 38 p. 3-14, 1994.

- MAIO, M. C.; LOPES, T. C. For the establishment of the social discipline as sciences: Donald Pierson e as ciências sociais no Rio de Janeiro (1942-1949). *Sociologia & Antropologia*, Rio de Janeiro, v. 5, n. 2, p. 343-380, 2015.
- MARENCO, A.; SERNA, M. Por que carreiras políticas na esquerda e na direita não são iguais? Recrutamento legislativo em Brasil, Chile e Uruguai. *Revista Brasileira de Ciências Sociais*, São Paulo, v. 22, n. 64, p. 93-113, 2007.
- MICELI, S. *Pierre Bourdieu: a economia das trocas simbólicas*. São Paulo: Perspectiva, 1974.
- _____. *Intelectuais e classe dirigente no Brasil (1920-1945)*. São Paulo: Difel, 1979.
- _____. Carne e osso da elite política brasileira pós-1930. In: Fausto, B. (Org.). *III O Brasil Republicano 3: sociedade e política (1930-1964)*. São Paulo: Difel, 1981.
- _____. *Imagens negociadas: retratos da elite brasileira (1920-1940)*. São Paulo: Companhia das Letras, 1996.
- _____. Jorge Luis Borges: história social de um escritor nato. *Novos Estudos Cebrap*, São Paulo, n. 77, p. 155-182, 2007a.
- _____. "O chão e as nuvens: ensaios de Roberto Schwarz entre arte e ciência". In: Cevasco, M. E.; Ohata, M. (Orgs.). *Um crítico na periferia do capitalismo: reflexões sobre a obra de Roberto Schwarz*. São Paulo: Companhia das Letras, 2007b.
- _____. *A elite eclesiástica brasileira: 1890-1930*. 2. ed. São Paulo: Companhia das Letras, 2009a.
- _____. Mário de Andrade: a invenção do moderno intelectual brasileiro. In: Botelho, A.; Schwarz, L. (Orgs.). *Um enigma chamado Brasil: 29 intérpretes e 1 país*. São Paulo: Companhia das Letras, 2009b.
- _____. Voz, sexo e abismo: Alfonsina Storni e Horacio Quiroga. *Novos estudos Cebrap*, São Paulo, n. 97, p. 83-113, 2013.
- MIGUEL, L. F.; BIROLI, F. Práticas de gênero e carreiras políticas: vertentes explicativas. *Revista de Estudos Feministas*, Florianópolis, v. 18, n. 3, p. 653-679, 2010.
- NEIBURG, F. Ciências sociais e mitologias nacionais: constituição da sociologia na Argentina e a invenção do peronismo. *Anuário Antropológico*, Brasília, v. 94, p. 189-225, 1995.
- OFFERLÉ, M. Entrées en politique. *Politix*, Mont-Saint-Guibert, v. 35, p. 3-5, 1996.
- OLIVEIRA, M. G. Quem tem medo da ilusão biográfica? Indivíduo, tempo e histórias de vida. *Topoi*, Rio de Janeiro, v. 18, n. 35, p. 429-446, 2017.
- PASSERON, J.-C. A encenação e o corpus: biografia, fluxos, itinerários, trajetórias. In: PASSERON, J.-C. *O raciocínio sociológico: o espaço não popperiano do raciocínio natural*. Petrópolis: Vozes, 1995.
- PAVÃO, A. Carreira de usuários da cultura escrita e formação de professores: um ensaio etnográfico. *Horizontes Antropológicos*, Porto Alegre, v. 23, n. 49, p. 255-285, 2017.

- PENEFF, J. Les grandes tendances de l'usage des biographies dans la sociologie française. *Politix*, Mont-Saint-Guibert, v. 7, n. 27, 1994.
- PERISSINOTO, R.; COSTA, L. D.; TRIBESS, C. Origem social dos parlamentares paranaenses (1995-2006): alguns achados e algumas questões de pesquisa. *Sociologias*, Porto Alegre, n. 22, p. 280-313, 2009.
- PERISSINOTTO, R. et al. Redes sociais e recrutamento: o caso dos diretores e presidentes do Banco Central do Brasil (1994-2016)". *Tempo Social*, São Paulo, v. 29, n. 3, p. 61-82, 2017.
- PONTES, H. Beleza roubada: gênero, estética e corporalidade no teatro brasileiro. *Cadernos Pagu*, Campinas, n. 33, p. 139-166, 2009.
- _____. Corpo iluminado: Cacilda Becker e o teatro moderno. *Ciência Hoje*, Rio de Janeiro, v. 47, p. 30-35, 2011.
- _____. Mariazinha e Verônica: classe e gênero nos palcos da metrópole. *Novos Estudos Cebrap*, São Paulo, n. 97, p. 149-166, 2013.
- PONTES, H.; MICELI, S. Memória e utopia na cena teatral. *Sociologia & Antropologia*, Rio de Janeiro, v. 2, p. 241-263, 2012.
- QUEIROZ, M. I. P. Relatos orais: do "indizível" ao "dizível". In: VON SIMSON, O. M. *Experimentos com histórias de vida: Itália-Brasil*. São Paulo: Vértice, 1988. p. 14-43.
- REIS, E. T. Saberes em movimento: transações entre "intelectuais", definições de ciências sociais e a "política". In: SEIDL, E.; GRILL, I. G. *As ciências sociais e os espaços da política no Brasil*. Rio de Janeiro: FGV Editora, 2013.
- _____. Um tipo diferente de diferença na elite política: perfis políticos e inserções culturais de parlamentares brasileiras. *Cadernos Pagu*, Campinas, n. 43, p. 233-263, 2014.
- _____. Trajetórias, espaços e repertórios de intervenção política. Porto Alegre: Zouk, 2015.
- REIS, E. T.; GRILL, I. G. *Estudos sobre elites políticas e culturais: reflexões e aplicações não canônicas*. São Luís: Edufma, 2016. v. 2.
- _____. Estudos de elites políticas e as bases das multinotabilidades no Brasil. *Tempo Social*, São Paulo, v. 29, n. 3, p. 137-159, 2017.
- REIS, J. C. *Escola dos Annales: a inovação em história*. 2. ed. São Paulo: Paz e Terra, 2000.
- Rial, C. Rodar: a circulação internacional de jogadores de futebol brasileiro no exterior. *Horizontes Antropológicos*, Porto Alegre, ano 14, n. 30, p. 21-65, 2008.
- RODRIGUES, L. M. Fernando Henrique Cardoso: a ciência e a política como vocação. In: Botelho, A.; Schwarcz, L. M. (Orgs.). *Um enigma chamado Brasil: 29 intérpretes e um país*. São Paulo: Companhia das Letras, 2009.

- SANTOS, A. M. Nas fronteiras do campo político: raposas e outsiders no Congresso Nacional. *Revista Brasileira de Ciências Sociais*, São Paulo, v. 33, p. 87-101, 1997.
- _____. *Não se fazem mais oligarquias como antigamente*: recrutamento parlamentar, experiência política e vínculos partidários entre deputados brasileiros [1946-1998]. 2000. 259 f. Tese (Doutorado em Ciência Política) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.
- _____. Sedimentação de lealdades partidárias no Brasil: tendências e descompassos. *Revista Brasileira de Ciências Sociais*, São Paulo, v. 16, n. 45, p. 69-83, 2001.
- SANTOS, H.; OLIVEIRA, P.; SUSIN, P. Narrativas e pesquisa biográfica na sociologia brasileira. *Civitas*, Porto Alegre, v. 14, n. 2. p. 359-382, 2014.
- SAWICKI, F. *Les réseaux du Parti Socialiste*: sociologie d'un milieu partisan. Paris: Belin, 1997.
- SCHÜTZ, A.; LUCKMANN, T. The structures of the life-world. London: Heinemann, 1973. v. 1.
- SCOTTO, M. G. O visível e o invisível: considerações sobre uma candidatura. *Comunicações do PPGAS/MN*, Rio de Janeiro, v. 5, p. 21-38, 1995.
- SEIDL, E. Elites militares, trajetórias e redefinições político-institucionais (1850-1930). *Revista de Sociologia e Política*, Curitiba, v. 16, n. 30, p. 199-220, 2008.
- _____. Uma elite pouco (re)conhecida: o episcopado brasileiro. *Tempo Social*, v. 29, n. 3, p. 35-60, 2017.
- SEIDL, E.; GRILL, I. G. As ciências sociais e os espaços da política no Brasil. Rio de Janeiro: FGV Editora, 2013.
- SIMIONI, A. Paula Cavalcanti. As mulheres artistas e os silêncios da história: a história da arte e suas exclusões. *Estudos Feministas*, Florianópolis, v. 11, p. sn-si, 2007a.
- _____. Regina Gomide Graz: modernismo, arte têxtil e relações de gênero no Brasil. *Revista do Instituto de Estudos Brasileiros*, São Paulo, v. 45, p. 87-106, 2007.
- _____. *Profissão artista*: pintoras e escultoras acadêmicas brasileiras. São Paulo: Edusp, 2008.
- SORÁ, G. Os livros do Brasil entre o Rio de Janeiro e Frankfurt. *BIB: Revista Brasileira de Informação Bibliográfica em Ciências Sociais*, Rio de Janeiro, n. 41, p. 3-33, 1996.
- _____. A arte da amizade: José Olympio, o campo de poder e a publicação de livros autenticamente brasileiros. *Antropolítica*, Niterói, n. 30, p. 49-71, 2011.
- THOMPSON, P. *A voz do passado*: história oral. Rio de Janeiro: Paz e Terra, 1992.
- VELHO, G. Becker, Goffman e a antropologia no Brasil. *Sociologia: Problemas e Práticas*, Lisboa, n. 38, p. 9-17, 2002.

_____. *Projeto e metamorfose: antropologia das sociedades complexas*. 3. ed. Rio de Janeiro: Jorge Zahar, 2003.

VELHO, G.; KUSCHNIR, K. Mediação e metamorfose. *Mana*, Rio de Janeiro, v. 2, n. 1, p. 97-107, 1996.

VON SIMSON, O. M. *Experimentos com histórias de vida (Itália-Brasil)*. São Paulo: Vértice, 1988.

WHIGHT MILLS, C. *A imaginação sociológica*. Rio de Janeiro, Zahar, 1975.

Resumo

Alusões biográficas e afirmação dos estudos de trajetórias: reflexões sobre usos flexíveis

As “biografias” ocupam estatuto incerto, ambivalente e polissêmico no âmbito das ciências humanas e sociais. Neste artigo, mapeamos tendências na produção recente das ciências sociais brasileiras, retomamos alguns pressupostos teóricos e metodológicos preponderantes, bem como indicamos seus principais porta-vozes no Brasil. Tais procedimentos permitiram observar: as oscilações históricas, disciplinares e temáticas, que pesam nas alusões biográficas mobilizadas em diferentes empreendimentos metodológicos (como nas histórias de vida e orais; estudos de carreiras e de trajetórias); a oposição entre abordagens mais subjetivistas de construção biográfica e aquelas mais objetivistas de uso de dados biográficos; e a consagração da noção de trajetória e das referências a Pierre Bourdieu, com variados tipos de apropriação.

Palavras-chave: Biografia; Trajetória; Carreira; História de Vida; História Oral.

Abstract

Biographical allusions and affirmation of trajectory studies: reflections on flexible uses

The “biographies” occupy uncertain, ambivalent and polysemic status within the human and social sciences. In this article, we mapped trends in Brazilian social sciences’ recent production, resumed some prevailing theoretical and methodological assumptions, as well as pointed out its main spokesmen in Brazil. These procedures allowed us to observe: the historical, disciplinary and thematic oscillations that weigh in the biographical allusions mobilized in different methodological enterprises (as in the life history and oral history, studies of career and trajectories); the opposition between more subjectivist approaches of biographical construction and those more objectivist of using biographical data; and the establishment of the notion of trajectory and references to Pierre Bourdieu, with various types of appropriation.

Keywords: Biography; Trajectory; Career; Life history; Oral History.

Resumé

Allusions biographiques et affirmation des études de trajectoires : réflexions sur les usages flexibles

Les « biographies » ont un statut incertain, ambivalent et polysémique dans le contexte des sciences humaines et sociales. Cet article vise à dresser l’inventaire des tendances dans la production récente des sciences sociales brésiliennes, à reprendre certains présupposés théorico-méthodologiques prépondérants et à mettre en lumière leurs principaux porte-parole au Brésil. Partant de là, il a été possible d’observer : les oscillations historiques, disciplinaires et thématiques qui touchent les allusions biographiques présentes dans différentes entreprises méthodologiques (comme dans les histoires de vie et à l’histoire orale; les études de carrières et de trajectoires); l’opposition entre des approches plus subjectivistes de construction biographique et celles plus objectivistes d’utilisation de données biographiques; et la reconnaissance de la notion de trajectoire et des références à Pierre Bourdieu, avec plusieurs types d’appropriation.

Mots-clés: Biographie; Trajectoire; Carrière; Histoire de Vie; Histoire Orale.

Análise de correspondências múltiplas: fundamentos, elaboração e interpretação

Elisa Klüger¹

Introdução²

O primeiro contato de cientistas sociais com análises geométricas de dados ocorre, frequentemente, quando da leitura de obras de Pierre Bourdieu. Livros como *A distinção*, (2006), *Homo Academicus* (2011) e *La Noblesse d'État* (1998) – com suas primeiras edições lançadas em 1979, 1984 e 1989, respectivamente –, incluem quadros que representam espaços sociais em planos cartesianos demarcados por um par de eixos. Tais espaços foram construídos a partir de aplicações da técnica denominada análise de correspondências múltiplas (ACM), variante da análise geométrica de dados que utiliza dados categóricos (qualitativos) para diferenciar os agentes e posicioná-los no plano³. Em meio aos eixos, aparecem nuvens de pontos que representam ora agentes ora propriedades sociais – os ditos dados categóricos –, sendo a distância entre os pontos e seus padrões de

dispersão na nuvem dados centrais para observar as afinidades e polarizações existentes no universo social estudado.

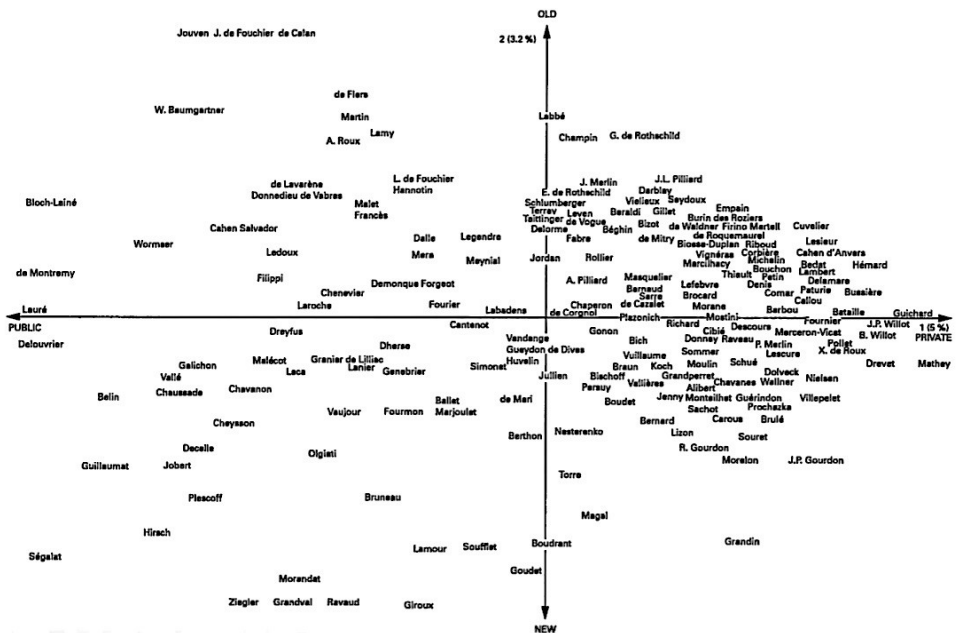
Bourdieu transpõe para as ciências sociais uma renovação no campo da estatística que ocorria na França, desde os anos 1960, ao redor do matemático e ex-colega de *École Normale Supérieure*, Jean-Paul Benzécri. Tratava-se da possibilidade de apropriação de uma técnica capaz de sintetizar quantitativamente dados qualitativos dispostos de maneira estrutural, multidimensional e relacional (GREENACRE; BLASIUS, 2006; LEBARON, 2010). A análise geométrica não é um dentre outros métodos mobilizados por Bourdieu e seus discípulos, mas aquele privilegiado em seus trabalhos justamente por ser entendido como uma representação filosoficamente afinada com a concepção relacional subjacente à teoria bourdieusiana dos campos (BOURDIEU; WACQUANT, 1992, p. 96). De acordo com Bourdieu e Loïc Wacquant (1992, p. 16-17, tradução e grifos meus),

-
- 1 Elisa Klüger é pós-doutoranda no Programa Internacional de Pós-Doutorado do Centro Brasileiro de Análise e Planejamento (Cebap) e Visiting Research Postdoctoral Associate na Princeton University. A autora conta, atualmente, com bolsa da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), processos 2017/13937-1 e 2018/09487-7. E-mail: elisa.kluger@gmail.com
 - 2 Expresso meus agradecimentos ao professor Frédéric Lebaron, que me apresentou às Análises de Correspondências Múltiplas e fez perceber a importância do domínio da técnica para a elaboração de trabalhos empíricos de matriz bourdieusiana. À Marcia Consolim agradeço o estímulo para transferir para o papel os conhecimentos sobre técnica com o objetivo de gerar material em português sobre o tema. Sou grata a Adriano Codato pelas recomendações e indicações bibliográficas. Possíveis erros são, evidentemente, de minha inteira responsabilidade.
 - 3 “Análises geométricas de dados (AGD) – nome sugerido por Patrick Suppes em 1996 – é a abordagem de estatísticas multivariadas iniciada por Jean-Paul Benzécri nos anos 1960, conhecida na literatura em língua francesa como *Analyse des Données*” (ROUANET, 2006, p. 138, tradução minha). A AGD inclui tanto as análises de correspondências múltiplas (ACM) – delineadas a partir de variáveis categóricas (categorias) – quanto análises de componente principal (ACP) – delineadas a partir de variáveis contínuas (números).

um campo consiste em um conjunto de relações objetivas e históricas entre posições ancoradas em certas formas de poder (ou capital) [...]. Cada campo prescreve seus valores particulares e possui seus próprios princípios reguladores. Esses princípios delimitam um espaço socialmente estruturado no qual agentes lutam, dependendo da posição

que ocupam no espaço, ou para transformar ou para preservar suas fronteiras e forma [...]. Um campo é um sistema padronizado de forças objetivas (bem à maneira de um *campo magnético*), uma configuração relacional dotada de uma *gravidade* específica que ele impõe a todos os objetos e agentes que o adentram.

Figura 1
O espaço dos chefes de grandes corporações francesas
nuvem dos agentes em *The State Nobility*



Fonte: Bourdieu (1998, p. 302)

A representação geométrica criada pela ACM é adequada à ideia de campo visto situar os agentes precisamente em um espaço objetiva e relacionalmente estruturado, no qual a distância entre uns e outros decorre de diferenças em suas propriedades sociais, inclusive dotações desiguais de capitais. Ao fazê-lo, a análise geométrica oferece elementos para a visualização dos fundamentos das polarizações e lutas travadas no espaço social

e para detecção de padrões de correlação entre as posições sociais dos agentes e suas práticas e tomadas de posição nos mais variados domínios da vida social.

Na obra de Bourdieu, as ACMs foram utilizadas pela primeira vez no texto “Anatomie du Gout” (BOURDIEU; SAINT MARTIN, 1976), escrito com Monique de Saint-Martin e publicado na revista *Actes de la Recherche en Sciences Sociales*, que versa sobre os espaços

dos estilos de vida na França. Ao longo da carreira, empregou a técnica para estudar os mais diversos objetos. Em *A distinção*, (2006), a empregou para mapear práticas culturais da população francesa, com dados gerados pelas grandes estatísticas nacionais. Na sequência, delineou o campo acadêmico francês, em *Homo academicus* (2011), indicando as posições na nuvem de conhecidos professores do sistema de ensino superior. Estudou o campo do poder em *The state nobility* (1998), posicionando relacionalmente o alto patronato francês; o campo das editoras francesas, em “Une révolution conservatrice dans l’édition” (1999); e o mercado habitacional francês, no livro *Les structures sociales de l’économie* (2000).

Seus discípulos estenderam os objetos explorados com a técnica. Por exemplo, Gisèle Sapiro (1999) a aplica ao campo dos escritores, Frédéric Lebaron (2000) aos economistas, Olivier Godechot (2001) aos financistas, François Denord (2003) ao campo do poder e Julien Duval (2006) ao campo do cinema. Diversos trabalhos sobre os desenvolvimentos da técnica são publicados por Brigitte Le Roux e Henry Rouanet (2010), parceiros de Pierre Bourdieu na área estatística, colaboradores no desenvolvimento de *softwares* capazes de executar a técnica e responsáveis por cursos de divulgação prática das ACMs em diversos países.

Com o tempo, a técnica ganha adeptos fora da França. Tony Bennett (2009), Mike Savage, Elizabeth Silva, Alan Warde, Modesto Gayo-Cal e David Wright conduzem projeto no qual mapeiam as práticas culturais dos ingleses aplicando ACMs à moda de *A distinção*. Virgílio Borges Pereira (2016), em

Portugal, combina estudo de classes sociais, práticas culturais e inserção geográfica na cidade do Porto. Felix Bühlmann, Thomas David e André Mach (2012) coordenam um projeto sobre as elites suíças que emprega a técnica, e Johs Hjellbrekke e Olav Korsnes (2013) dirigem projeto semelhante sobre as elites norueguesas. Marion Fourcade, Brian Lande e Evan Schofer (2016), pesquisadores afiliados a universidades estadunidenses⁴, comparam os espaços políticos de diferentes nações, entre tantos outros trabalhos responsáveis por multiplicar e difundir as ACMs.

No Brasil, a incorporação da obra de Pierre Bourdieu tem lugar no final da década de 1960, como efeito da circulação internacional de pesquisadores brasileiros que estabeleceu uma rede de cooperação entre bourdieusianos e pesquisadores brasileiros, primeiramente centrada na sociologia e antropologia rural e, contemporaneamente, na sociologia da cultura (BORTOLUCI; JACKSON; PINHEIRO FILHO, 2015). Ainda que seja um elemento central na obra de Bourdieu, a análise geométrica adentra as ciências sociais brasileiras de forma relativamente tardia. Em primeiro lugar, cabe notar que as principais obras a empregar ACMs foram traduzidas recentemente ou seguem sem traduções para o português. *A distinção* e *Homo academicus* foram traduzidas, respectivamente, em 2006 e 2011. As *Estruturas sociais da economia* têm tradução portuguesa de 2001. Os outros trabalhos mencionados seguem sem tradução. No que concerne à tradução de textos básicos sobre o uso da própria técnica, estão disponíveis em português o artigo “A análise geométrica de

4 Uma discussão corrente versa sobre a falta de permeabilidade da academia norte-americana a pesquisas desenvolvidas à dita estatística geométrica francesa, dada sua diferença no entendimento de causalidade quando comparada, por exemplo, com a estatística efetuada por meio de regressões. Tal resistência cria fronteiras para a disseminação da técnica nos Estados Unidos e para a publicação de artigos que a empregam em algumas das revistas acadêmicas de maior difusão e prestígio. Para um histórico das divergências entre as explicações sociológicas e metodologias adotadas nos dois países, ver Cibois (1981).

questionários: a lição de La Distinction de Bourdieu” (ROUANET; ACKERMAN; LE ROUX, 2000), publicado em Portugal em 2005, e o capítulo “Analisar um espaço social” (DUVAL, 2015), em livro sobre métodos nas ciências sociais organizado por Serge Paugam e publicado no Brasil.

A temporã tradução das obras que utilizam o método pode ser um dos elementos explicativos da decalagem na incorporação da ACM por pesquisadores brasileiros das ciências sociais⁵. Outro fator que deve ser considerado é a maior dificuldade de transposição de um saber que não é só teórico, mas implica aprendizados práticos. O estabelecimento de redes que propiciaram a adoção das ACMs por pesquisadores brasileiros ocorreu décadas depois do início da aproximação com os bourdieusianos franceses. Ademais, o centro dessa rede não está fincado na sociologia da cultura, e sim no entroncamento das sociologias econômica e das elites, ainda que haja trabalhos sobre cultura e educação que empreguem a técnica.

Maria Rita Loureiro (1997) – que esteve associada à *École des Hautes Études en Sciences Sociales* (EHESS), em Paris, de 1986 a 1988 – em seu livro sobre os economistas no governo, adota um esboço de representação, em plano cartesiano, do campo dos economistas sem propriamente construir uma ACM. O estudo dos economistas foi uma área fértil de disseminação das ACMs, contando com o mestrado de Rodrigo Cantu de Souza (2009), *A ciência dos economistas: entre dissensos científicos e clivagens morais*, sobre a produção em ciência econômica; e com o trabalho “Mapping the inflections in the

policies of the Brazilian National Economic and Social Development Bank during the nineties and two thousands within social spaces and networks”, de Elisa Klüger (no prelo), realizado a partir de dados coletados em entrevistas com dirigentes do Banco Nacional de Desenvolvimento Econômico e Social. Os dois foram alunos de Frédéric Lebaron, que, além de ser conhecido por seus escritos sobre economistas franceses, tem publicações sobre métodos e técnicas em ciências sociais, com ênfase nas análises geométricas. Ainda tratando de especialistas em economia, Eric Gil Dantas, Edilson Montrose de Aguiar Júnior, Felipe Calabrez e Adriano Codato (2017) apresentam ACMs relativas aos atributos sociopolíticos dos secretários do Ministério da Fazenda no *working paper* “Prosopografia dos ajudantes do Ministério da Fazenda: uma análise das lógicas de recrutamento, dos itinerários profissionais e do perfil social do segundo escalão”.

Um dos discípulos mais jovens de Pierre Bourdieu, Frédéric Lebaron é uma importante ponte para disseminação do método no Brasil. Seus laços com os professores Roberto Grün e Julio Donadone, especialistas em sociologia econômica e das finanças, favoreceram a realização de cursos sobre a teoria e os métodos de Pierre Bourdieu no Brasil e funcionaram como ponte para que diversos alunos da Universidade Federal de São Carlos estudassem a técnica com ele na França e a empregassem em suas teses. Foram os casos, por exemplo, de Thais Joi (2015), que produz a tese *Desejo, necessidade e realidade: os marcadores culturais e econômicos e suas implicações ocupacionais para o grupo profissional de engenheiros de produção*

5 Cabe enfatizar aqui que o artigo aborda o uso das ACMs nas ciências sociais, em especial em trabalhos associados à corrente bourdieusiana, visto haver uma temporalidade diferente na incorporação e nos tipos de uso da técnica em outras áreas do conhecimento. No Brasil as ACMs são mais disseminadas em outras disciplinas – notadamente na área médica –, incluindo saúde pública, psicologia e medicina, como indicam buscas simples no Google Scholar e na plataforma SciELO.

no Brasil; de Marcela Purini Belém (2015), que defende a tese *A lei Rouanet e a construção de um mercado de patrocínios culturais no Brasil* (2015); e de Leandro Garcez Targa (2017), que se doutora com *Os diplomatas brasileiros sob a perspectiva relacional: o campo dos diplomatas e o campo político*.

Na área de sociologia da educação, Graziela Seroni Perosa e Cristiane Kerches da Silva Leite escrevem, em colaboração direta com Lebaron (2015), o artigo “O espaço das desigualdades educativas no município de São Paulo”. Um estudo pioneiro na utilização de ACM nessa área é o de Ana Paula Hey, que foi contemporânea de Lebaron quando fez parte do seu doutorado na EHESS, no início dos anos 2000, trabalho publicado no livro *Esboço de uma sociologia do campo acadêmico: a educação superior no Brasil* (2008). É possível encontrar também um artigo sobre o espaço das práticas culturais de Lebaron e Philippe Bonnet (2016) no número sobre classes sociais e desigualdades da revista *Tempo Social*, organizado por Edison Bertonecelo (2016a), que também utiliza ACMs em suas pesquisas na área de classes sociais e cultura, como em “O espaço das classes sociais no Brasil”. Na área de elites políticas há outro vínculo com a França, costurado por Odaci Luiz Coradini (2006), que fez estágio de pesquisa na EHESS, em 2000, e logo utilizou ACMs em trabalhos, vide “Representação profissional e elites políticas no Brasil no período recente”. A disseminação da técnica promovida por Coradini pode ser vista, por exemplo, no trabalho de seu orientando Rodrigo Bourdignon (2017), que utiliza a técnica para falar de “Recrutamento e modalidades de entrada na carreira política: candidatos aos cargos legislativos no Rio Grande do Sul”.

O breve balanço bibliográfico sinaliza a proeminência da análise geométrica na produção bibliográfica de Pierre Bourdieu e, conforme indicado, a crescente difusão da técnica nas ciências sociais de inspiração bourdieusiana, inclusive no Brasil. A leitura e compreensão de parte significativa das obras associadas a essa corrente sociológica, ou por ela metodologicamente orientadas, requer, portanto, habilidade para reconhecer, decodificar e interpretar as análises de correspondências múltiplas mobilizadas para representação de espaços sociais e campos. Isso posto, este artigo consiste precisamente em uma introdução à ACM voltada a subsidiar a leitura de obras que empregam a técnica e, quiçá, a auxiliar na construção de pesquisas que explorem tal recurso. Posto tratar-se de uma apresentação primária do tema, não serão contemplados debates relativos à modelação matemática subjacente à técnica⁶, atendo-se o texto à discussão da lógica geométrica própria à ACM e à apresentação de procedimentos iniciais para elaboração e interpretação de ACMs.

O texto aparece dividido em três segmentos seguidos de uma conclusão. O primeiro deles é denominado “Campo magnético” por explorar a metáfora, presente na definição de campo citada, segundo a qual agentes podem ser entendidos como “partículas” em um “campo magnético” que imprime forças objetivas de atração e repulsão. O segmento contempla uma apresentação das tabelas agentes *versus* variáveis e discussões acerca do uso das fontes e dos princípios para a recodificação dos dados. Ademais, indica a importância da realização de estatísticas descritivas preliminares e cruzamentos simples entre variáveis antes de tomar decisões relativas à seleção de variáveis para construção das ACMs.

6 Para debates relativos aos fundamentos matemáticos, recomenda-se consultar Blasius e Greenacre (1998), Greenacre e Blasius (2006), Robson e Sanders (2009), e Le Roux e Rouanet (2010).

O segundo segmento é designado “Geometria” por se dedicar à apresentação dos procedimentos que dão origem à distribuição das “partículas” no plano cartesiano, tornando mensuráveis e representáveis visualmente as distâncias entre os integrantes das nuvens de agentes e de modalidades. O trecho contempla uma discussão acerca do estatuto da explicação e da predição na sociologia de Bourdieu que tem por objetivo informar as escolhas feitas na etapa de construção da ACM, principalmente aquelas relativas à classificação das variáveis em ativas e suplementares. Na sequência, é apresentada uma primeira representação da nuvem das modalidades, indicando como fazer a leitura do quadro eixo a eixo e descrevendo de que modo se diferenciam o comportamento das variáveis ativas e suplementares.

O terceiro segmento foi nomeado “Gravidade”, recorrendo à ideia já mencionada segundo a qual no campo “há uma gravidade específica que ele impõe a todos os objetos e agentes que o adentram” (BOURDIEU; WACQUANT, 1992, p. 16-17, tradução e grifos meus). A noção de gravidade faz-se pertinente na medida em que cada propriedade social possui uma massa e que esta varia em função do número de agentes que dela dispõe. Tal qual no cálculo da força de atração gravitacional, a massa e distância entre as partículas são os elementos variáveis. No caso, a massa e a distância das partículas em relação ao ponto central do plano cartesiano são os elementos que definem a contribuição das modalidades e variáveis para a variância dos eixos e, conseqüentemente, para a definição do desenho do espaço. Na sequência, será apresentada a tabela de contribuições e procedimentos indicados para reconhecer e prevenir distorções nas ACMs. Logo, define-se a noção de eixo, discute-se a variância observada ao longo de cada eixo e ressalta-se a importância da observação combinada dos espaços sociais construídos no plano cartesiano e da tabela

de contribuições para interpretações. Por fim, faz-se recurso a elipses de concentração para mapear o grau de dispersão/concentração de propriedades em meio aos agentes representados e comenta-se a possibilidade de combinar ACMs com outros métodos e técnicas, como as análises de classificação hierárquica e as análises de redes sociais.

Fundamentos da ACM: “campo magnético”, “geometria” e “gravidade”

A leitura do subtítulo apresentado poderia indicar uma inesperada aproximação do texto com as ciências exatas. Não se trata, entretanto, de um desvio, posto haver metáforas – não exatamente estritas – com noções da física e da geometria, tanto na produção teórica de Bourdieu quanto nos métodos por ele empregados. A seguir, propriedades dos campos magnéticos e da gravidade serão esboçadas para apresentação dos elementos essenciais e da lógica subjacente à construção das ACMs. A geometria, por sua vez, subsidia a representação visual dos espaços sociais e torna mensuráveis as distâncias entre os agentes ou entre as propriedades sociais representados nas nuvens.

O texto percorrerá as seguintes etapas pertinentes à construção e interpretação de ACMs: (1) elaboração das questões e hipóteses iniciais da pesquisa; (2) seleção das fontes e casos; (3) montagem das tabelas *agentes versus variáveis*; (4) descrição elementar dos dados da tabela; (5) realização de cruzamentos simples; (6) recodificação dos dados e busca de equilíbrio do número de modalidades por variável e de variáveis por rubrica; (7) seleção das variáveis ativas e suplementares; (8) montagem das ACMs; (9) observação inicial da contribuição dos eixos para a determinação da variância total das nuvens e da contribuição das variáveis e modalidades para a variância dos eixos; (10) avaliar se há alguma modalidade rara ou defectiva que tenha sido

demasiadamente contributiva na construção do espaço; (11) caso encontre alguma distorção, verificar os dados e decidir se tornará a modalidade passiva; (12) caso tenha alterado alguma modalidade, gerar novas ACMs e tabelas de contribuição; (13) decidir quais eixos reter de acordo com as questões formuladas e com o grau de variância total que se pretende contemplar; (14) interpretação das nuvens de modalidades e das nuvens agentes por meio da observação das polarizações nas configurações visuais resultantes, das tabelas de contribuições, do posicionamento das variáveis e modalidades suplementares, das elipses de concentração e da observação dos agentes, caso conhecidos; (15) combinação das análises de correspondências com outras técnicas e metodologias, como análises de classificação hierárquica e análises de redes sociais. Essa lista de etapas deve funcionar como uma via rápida para recapitulação e checagem dos procedimentos próprios à ACM.

“Campo magnético”

Ao explicar sua noção de campo, no livro *An invitation to reflexive sociology*, Bourdieu explora imagens provenientes do estudo do magnetismo, dizendo que:

Nesse campo eu encontrarei “partículas” (deixe-me fingir por um momento que estamos lidando com um campo físico) que estão sob a influência de forças de atração e repulsão, e assim por diante, como em um campo magnético. Tendo dito isso, ao falar sobre um campo, minha atenção recai sobre o primado desse sistema de relações objetivas entre as próprias partículas. (BOURDIEU; WACQUANT, 1992, p. 106, tradução minha)

A metáfora feita por ele é particularmente útil à apresentação dos fundamentos

subjacentes às ACMs, visto ser possível conceber os *agentes*⁷ como as ditas “partículas” sob influência do “campo magnético”. A proximidade de agentes no espaço social não depende da existência de interações entre eles, como seria o caso em uma rede social. Todos os agentes (conhecidos ou desconhecidos) aparecem objetivamente relacionados e hierarquicamente posicionados de acordo com sua localização na estrutura, determinada em função da distribuição desigual de *propriedades sociais* (incluindo atributos, capitais, práticas e tomadas de posição). Assim, a *proximidade* ou *distância* dos agentes no campo é determinada em função do grau de homogeneidade de suas propriedades sociais. Em termos bourdieusianos, a atração entre dois agentes será tanto maior quanto mais similares forem os seus *habitus* – “conjunto de relações históricas depositadas em corpos individuais na forma de esquemas mentais e corporais de percepção, apreciação e ação” (Ibidem, p. 16, tradução minha) –, e a repulsão tanto maior quanto menos parecidos do ponto de vista social.

No caso das propriedades sociais, aparecerão em posição próxima no espaço as comumente associadas nas respostas dos agentes analisados, indicando serem partilhadas por aqueles com perfis e preferências similares. Por exemplo, pode-se imaginar que pessoas que aprenderam a tocar instrumentos musicais tenham maior gosto por música, portanto, sua frequência a concertos seja maior do que aquela de indivíduos que não tiveram contato com instrumentos musicais. Assim, espera-se que o domínio de instrumentos musicais e a alta frequência a concertos apareçam em pontos próximos em um espaço social que retrate as preferências e práticas culturais de um conjunto de indivíduos, por serem partilhadas por um mesmo grupo de agentes.

7 Doravante, aparecem destacados em itálico os termos principais do vocabulário associado à ACM.

Antes de apresentar as etapas propriamente relativas à estruturação do “campo magnético”, é preciso definir o que se entende por agentes e propriedades sociais e discutir quais fontes podem ser empregadas para construir as bases de dados mobilizadas nesse tipo de análise. Em primeiro lugar, cabe dizer que o agente apresentado como uma “partícula magnética” pode ser tanto um indivíduo quanto uma instituição ou coletivo. As propriedades sociais – comumente denominadas variáveis –, por sua vez, podem compreender informações relativas aos capitais (culturais, econômicos, sociais, cosmopolitas etc.), além de dados relativos a práticas (em relação ao corpo, culturais, atividades sociais etc.) e indicadores de tomadas de posição (políticas, estéticas, morais etc.) dos agentes analisados. A montagem de quadros de ACM principia pela construção de tabelas *agentes versus variáveis*. Nelas há única resposta associada a cada agente para cada questão, sendo a resposta obrigatoriamente de natureza categórica (qualitativa), doravante denominada *modalidade*. Cada variável contempla, portanto, diversas modalidades. Um conjunto de variáveis relativas a uma mesma temática (por exemplo, preferências políticas, escolarização, hábitos alimentares etc.) é, por sua vez, denominado *rubrica*.

Não há uma regra fixa para predeterminação do número de casos que devem ser contemplados na montagem de uma ACM⁸. A ferramenta pode ser utilizada tanto para o estudo de largos espaços sociais quanto para pequenos grupos. O importante é obter exaustividade de material em relação ao universo que se optou por analisar, de modo a

construir um retrato complexo e apurado do espaço retratado. Os dados utilizados para preencher as tabelas, por sua vez, podem ter sido produzidos diretamente pelo pesquisador, em entrevistas ou *surveys*, ou ter sido extraídos de estatísticas nacionais, arquivos, dicionários biográficos, relatórios administrativos, entre tantas possíveis fontes. Em qualquer dos casos, os dados devem passar por um processo de verificação e ser recodificados de acordo com as questões centrais da pesquisa antes de passar à construção das ACMs, conforme exemplificado no Quadro 1.

Considere a seguinte tabulação de um *survey* imaginário relativo a práticas e preferências musicais dos participantes das quadrilhas da música “Flor da idade” de Chico Buarque (1975) e do poema “Quadrilha” de Carlos Drummond de Andrade (1930).

Múltiplas perguntas relativas às afinidades e repulsões entre propriedades sociais poderiam ser formuladas a partir do Quadro 1. É possível indagar, por exemplo, se os agentes que partilham da modalidade membros da quadrilha de Chico têm preferências musicais distintas daquelas dos membros da quadrilha de Drummond. Caberia avaliar, entre outras coisas, se a variável sexo seria associada a frequências distintas a concertos de música clássica ou se há conexões entre a modalidade de instrumento que dominam e as preferências musicais. Tais relações, que contrastam modalidades associadas a apenas duas variáveis, poderiam ser ilustradas por cruzamentos simples. Tomemos, por exemplo, a questão relativa à afinidade entre o domínio de instrumentos musicais e a frequência a concertos:

8 Ainda que não haja regra precisa para definir o número de casos e que pesquisas conhecidas tenham sido feitas com baixos efetivos – vide Bourdieu (1999) –, existem controvérsias a esse respeito, sobretudo dada a discussão acerca do efeito de baixos efetivos de respondentes por modalidades na criação de distorções, tema que será abordado adiante. Com o objetivo de minimizar tais distorções, Giovanni Di Franco (2016), por exemplo, propõe que o número mínimo de casos seja definido em função do número de modalidades ativas, com a sugestão de adoção de vinte casos para cada modalidade ativa.

Quadro 1
Agentes e propriedades sociais dos participantes das quadrilhas
em survey imaginário sobre práticas e preferências musicais

| Agente | Instrumento musical | Frequência a concertos de música clássica | Tipo de música favorito | Sexo | Maior diploma | Quadrilha da qual faz parte |
|--------------------|-----------------------|---|-------------------------|-----------|---------------|-----------------------------|
| Carlos | Violão | Uma vez por semestre | Samba | Masculino | Graduação | Chico |
| Dora | Violino | Uma vez por mês | Clássica | Feminino | Doutorado | Chico |
| Lia | Flauta transversal | Uma vez a cada dois meses | Clássica | Feminino | Mestrado | Chico |
| Léa | Saxofone | Não frequenta | Jazz | Feminino | Doutorado | Chico |
| Paulo | Violino | Duas vezes por mês | Clássica | Masculino | Doutorado | Chico |
| Juca | Não toca instrumentos | Não frequenta | Samba | Masculino | Graduação | Chico |
| Rita | Violão | Uma vez por ano | Samba | Feminino | Graduação | Chico |
| Dito | Não toca instrumentos | Não frequenta | Samba | Masculino | Graduação | Chico |
| Pedro | Violão | Uma vez por semestre | Samba | Masculino | Mestrado | Chico |
| João | Guitarra | Uma vez a cada dois meses | Jazz | Masculino | Mestrado | Drummond |
| Teresa | Violoncelo | Duas vezes por mês | Clássica | Feminino | Doutorado | Drummond |
| Raimundo | Não toca instrumentos | Não frequenta | Jazz | Masculino | Mestrado | Drummond |
| Maria | Saxofone | Uma vez por semestre | Jazz | Feminino | Doutorado | Drummond |
| Joaquim | Clarinete | Uma vez a cada dois meses | Clássica | Masculino | Doutorado | Drummond |
| Lili | Não toca instrumentos | Uma vez por semestre | Clássica | Feminino | Graduação | Drummond |
| J. Pinto Fernandes | Não toca instrumentos | Não frequenta | Sem resposta | Masculino | Sem diploma | Drummond |

Tabela 1
Cruzamento simples das variáveis frequência a concertos e domínio de instrumentos musicais

| Frequência a concertos | Não toca instrumentos | Clarinete | Flauta transversal | Guitarra | Saxofone | Violão | Violino | Violoncelo | Total |
|---------------------------|-----------------------|-----------|--------------------|----------|----------|--------|---------|------------|-------|
| Duas vezes por mês | | | | | | 1 | 1 | | 2 |
| Uma vez por mês | | | | | | 1 | | | 1 |
| Uma vez a cada dois meses | | 1 | 1 | 1 | | | | | 3 |
| Uma vez por semestre | 1 | | | | 1 | 2 | | | 4 |
| Uma vez por ano | | | | | | 1 | | | 1 |
| Não frequenta | 4 | | | | 1 | | | | 5 |
| Total | 5 | 1 | 1 | 1 | 2 | 3 | 2 | 1 | 16 |

A Tabela 1 indica haver maior frequência a concertos daqueles que tocam instrumentos, conforme sugerido pela hipótese inicial. Evidentemente, a relação seria ainda mais visível caso as variáveis fossem recodificadas e reduzidas a um número menor de modalidades. A recodificação é uma etapa essencial na preparação para a construção da ACM. No caso, por exemplo, seria possível reunir a frequência a concertos entre alta, intermediária ou ausente. Um fator importante, ao recodificar, é garantir um razoável equilíbrio do volume de respondentes entre as faixas de frequência – para evitar a criação de modalidades raras, que, conforme será discutido adiante, pode ser problemática. Se as três modalidades que representam maior engajamento na atividade fossem reunidas como “alta frequência”, seis indivíduos seriam incluídos. Na modalidade

seguinte, reunindo as respostas uma vez por semestre e uma vez por ano como “frequência intermediária”, cinco indivíduos estariam representados. Finalmente cinco indivíduos aparecem como “não frequentadores”.

A recodificação da variável “domínio de instrumentos musicais”, por sua vez, é menos evidente, visto não se tratar apenas de um agrupamento guiado pela intensidade da prática, mas de uma agregação de natureza qualitativa. Seria possível, por exemplo, reunir instrumentos de sopro e instrumentos de corda ou reunir instrumentos associados com a música popular e instrumentos próprios às orquestras. A escolha, nesse caso, deve decorrer das perguntas formuladas. Caso interesse saber se o tipo de instrumento tem relação com as preferências musicais, seria coerente agrupar os instrumentos de acordo com o uso prevalente por gênero musical.

Tabela 2
Recodificação da variável frequência a concertos

| Frequência a concertos | Não toca instrumentos | Clarinete | Flauta transversal | Guitarra | Saxofone | Violão | Violino | Violoncelo | Total |
|--------------------------|-----------------------|-----------|--------------------|----------|----------|--------|---------|------------|-------|
| Alta frequência | | 1 | 1 | 1 | | | 2 | 1 | 6 |
| Frequência intermediária | 1 | | | | 1 | 3 | | | 5 |
| Não frequenta | 4 | | | | 1 | | | | 5 |
| Total | 5 | 1 | 1 | 1 | 2 | 3 | 2 | 1 | 16 |

Tabela 3
Recodificação da variável “domínio de instrumentos musicais”, exemplo com percentuais em linha

| Frequência a concertos | Não toca instrumentos | Instrumentos associados à música orquestral | Instrumentos associados à música popular | Total |
|--------------------------|-----------------------|---|--|------------|
| Alta frequência | | 5 83,3% | 1 16,7% | 6 100% |
| Frequência intermediária | 1 20% | | 4 80% | 5 100% |
| Não frequenta | 4 80% | | 1 20% | 5 100% |
| Total | 5 31,3% | 5 31,3% | 6 37,5% | 16 100% |

A apresentação das frequências e os cruzamentos simples entre variáveis ilustrados na Tabela 3, são parte dos procedimentos preparatórios para a elaboração de ACMs, auxiliando no conhecimento do universo estudado requerido para tomar decisões relativas à recodificação e apontando, desde o princípio, a existência de uma tendência de associação da alta frequência a concertos com o domínio de instrumentos orquestrais e de associação da

não frequência com o não domínio de instrumentos musicais.

O Quadro 2, resultante das recodificações propostas de acordo com as questões e hipóteses suscitadas, está pronta para análise, tornando possível passar às decisões propriamente voltadas à elaboração da ACM – vale destacar que retornos à etapa de recodificação e aos cruzamentos simples fazem-se frequentemente necessários para ajustes.

Quadro 2
Agentes e propriedades sociais dos participantes das quadrilhas em survey imaginário sobre práticas e preferências musicais – dados recodificados

| Agente | Instrumento musical | Frequência a concertos de música clássica | Tipo de música favorito | Sexo | Maior diploma | Quadrilha da qual faz parte |
|--------------------|-----------------------|---|-------------------------|-----------|---------------|-----------------------------|
| Carlos | Popular | Frequência intermediária | Samba | Masculino | Graduação | Chico |
| Dora | Orquestral | Alta frequência | Clássica | Feminino | Doutorado | Chico |
| Lia | Orquestral | Alta frequência | Clássica | Feminino | Mestrado | Chico |
| Léa | Popular | Não frequenta | Jazz | Feminino | Doutorado | Chico |
| Paulo | Orquestral | Alta frequência | Clássica | Masculino | Doutorado | Chico |
| Juca | Não toca instrumentos | Não frequenta | Samba | Masculino | Graduação | Chico |
| Rita | Popular | Frequência intermediária | Samba | Feminino | Graduação | Chico |
| Dito | Não toca instrumentos | Não frequenta | Samba | Masculino | Graduação | Chico |
| Pedro | Popular | Frequência intermediária | Samba | Masculino | Mestrado | Chico |
| João | Popular | Alta frequência | Jazz | Masculino | Mestrado | Drummond |
| Teresa | Orquestral | Alta frequência | Clássica | Feminino | Doutorado | Drummond |
| Raimundo | Não toca instrumentos | Não frequenta | Jazz | Masculino | Mestrado | Drummond |
| Maria | Popular | Frequência intermediária | Jazz | Feminino | Doutorado | Drummond |
| Joaquim | Orquestral | Alta frequência | Clássica | Masculino | Doutorado | Drummond |
| Lili | Não toca instrumentos | Frequência intermediária | Clássica | Feminino | Graduação | Drummond |
| J. Pinto Fernandes | Não toca instrumentos | Não frequenta | Sem resposta | Masculino | Sem diploma | Drummond |

“Geometria”

A principal propriedade da ACM é a capacidade de resumir graficamente quadros no formato agentes *versus* variáveis – como o

Quadro 2 – em planos cartesianos definidos por pares de eixos. Assim, permite explorar simultaneamente múltiplas correlações entre um largo número de variáveis retratando perfis, agrupamentos e polarizações que não

poderiam ser detectados a partir de cruzamentos simples de variáveis (CORADINI, 2006). A ACM dá origem a *nuvens de agentes e nuvens de modalidades*. Em uma nuvem, cada agente ou modalidade⁹ é um ponto e passa a poder ser referido de acordo com suas coordenadas nos eixos. Conforme mencionado, a distância, ao longo dos eixos, de cada “partícula magnética” decorre da proximidade ou da distância resultante dos padrões de afinidade ou de repulsão. Aparecem em posições vizinhas na nuvem de agentes aqueles que partilham respostas relativas às propriedades sociais em consideração e em posições próximas na nuvem de modalidades aquelas frequentemente coincidentes nas respostas fornecidas pelos agentes.

A fórmula que determina a distância entre os agentes no espaço é a seguinte: a distância entre os agentes A e B (d_{AB}) ao quadrado (devida à resposta fornecida por cada um deles para uma determinada questão) é igual a 1 dividido pelo número de pessoas que indicaram a modalidade X – escolhida por A para aquela questão – (n_X), dividida pelo total dos respondentes (n total), mais 1 dividido pelo número de pessoas que indicaram a modalidade y – escolhida por B para aquela questão – (n_Y), dividida pelo total dos respondentes (n total).

$$d_{AB}^2 = \frac{1}{\frac{n_X}{n \text{ total}}} + \frac{1}{\frac{n_Y}{n \text{ total}}}$$

A distância final entre os agentes será dada pela agregação das distâncias criadas por cada uma das questões para as quais os agentes

escolhem modalidades distintas e a dispersão total da nuvem se deve à distância entre todos os agentes (LE ROUX; ROUANET, 2010). O agente ficará, assim, localizado no baricentro das posições de todas as modalidades que representam propriedades sociais a ele associadas e as modalidades no baricentro de todos os agentes que delas partilharem (LE ROUX; ROUANET, 2010, p. 42). A geometria torna, assim, mensuráveis e representáveis visualmente as distâncias entre as “partículas” que integram as nuvens, como representado na Imagem II.

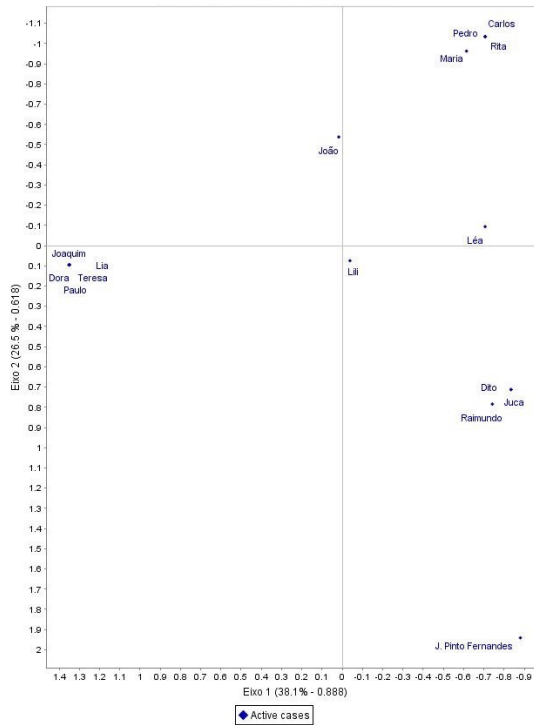
Na Figura 2, aparecem em posições próximas no espaço os agentes que partilham de propriedades sociais¹⁰. Estão reunidos, por exemplo, Dito, Raimundo e Juca, são os que não praticam música, nem frequentam concertos de música clássica e gostam de samba ou *jazz*. Observando verticalmente, no lado oposto do espaço estão Rita, Maria, Pedro e Carlos – os que têm frequência intermediária aos concertos, tocam instrumentos populares e gostam de samba ou *jazz*. Na horizontal, aparecem em posição oposta Dora, Lia, Teresa, Joaquim e Paulo, que são os que gostam de música clássica, tocam instrumentos orquestrais e têm alta frequência a concertos de música clássica. Em posição intermediária aparecem perfis com características mistas, como os de João, Lili e Léa. Finalmente, é possível encontrar J. Pinto Fernandes e sua não resposta, não frequência e não domínio de instrumentos, em posição bastante afastada, o que é significativo e será discutido em detalhes adiante.

9 As modalidades eventualmente aparecem designadas como categorias, adotando tradução do vocábulo dos textos em inglês.

10 Ao observar a nuvem de agentes vale atentar para a distribuição tendo em mente conhecimentos prévios sobre os participantes, o que ocorre sobretudo no caso de agentes conhecidos, ou seja, em estudos que retratam elites dirigentes, intelectuais, artistas, empresas (ROUANET; ACKERMANN; LE ROUX, 2000, p. 11). Neste caso vale inspecionar os agrupamentos resultantes e avaliar o sentido dos fracionamentos observados, se agentes que intuitivamente aproximariam resultam estar em posições semelhantes ou se aparecem afastados e investigar as divergências, caso detectadas.

Figura 2

Nuvem dos agentes – eixos 1 e 2, construída a partir da tabela agentes versus variáveis



As posições das “partículas” na nuvem dependem de definições feitas pelo pesquisador no momento de construção da ACM. Em primeiro lugar, nem todas as variáveis presentes em uma tabela são utilizadas para gerar a análise de correspondências, ou seja, nem todas as variáveis influenciam o estabelecimento dos eixos e a dispersão das *partículas*. Apenas *variáveis ativas* participam da construção do espaço. As *variáveis suplementares* – também ditas adicionais ou ilustrativas – por sua vez, são entendidas com “partículas” com massa zero e simplesmente projetadas no ponto médio de sua incidência no espaço tal qual gerado pelas variáveis ativas, sem alterar a configuração geométrica. As variáveis suplementares têm por função complementar a representação gerada pelas variáveis ativas,

oferecendo informações adicionais para que o pesquisador interprete os quadros gerados pela ACM (GREENACRE; BLASIUS, 2006).

A definição relativa ao caráter ativo ou suplementar de uma variável é feita diretamente pelo/a pesquisador/a, de acordo com o espaço ou campo que deseja representar, e é influenciada pelo tipo de questão formulada e pelo padrão de explicação que se pretende oferecer. Considerações acerca das noções de explicação pertinentes às ACM devem ser feitas para que se possa discutir quais critérios adotar ao classificar as variáveis.

Fundamentalmente, é preciso indicar a proximidade da técnica com duas ideias presentes na sociologia de Max Weber; a primeira centra-se na elaboração de tipologias. As representações gráficas providas pelas análises de correspondências

ajudam a evidenciar a presença de agregados de agentes e propriedades sociais que podem ajudar a compor tipos sociais à moda weberiana, na medida em que são entendidos como enquadramentos delimitados pelo pesquisador a partir de dados historicamente situados que enfatizam especificidades de cada um dos tipos (WEBER, 2006, p. 72-76) presentes no espaço, funcionando como um artefato para ajudar o pesquisador na tarefa de compreender os fenômenos estudados (CIBOIS, 1981).

A segunda aproximação com Weber deriva da noção de afinidades eletivas, havendo, para Weber, simultaneamente múltiplas direções possíveis para a causalidade entre variáveis (WEBER, 2006) em lugar de uma influência unidirecional de variáveis independentes sobre variáveis dependentes. Ao pensar em termos de afinidades eletivas, Weber contornava a causalidade unidirecional, não assumindo uma posição unilateral na querela sobre a primazia do material sobre o espiritual ou vice-versa. Ao analisar a relação entre *A* ética protestante e o espírito do capitalismo (Ibidem, 2004), ele faz notar a existência de afinidade entre certas formas de fé religiosa e certa ética profissional. Uma não teria sido causada pela outra, mas a afinidade entre as duas fez com que o movimento religioso pudesse exercer efeitos sobre o desenvolvimento da cultura material e o contrário (LÖWY, 2010). Nesse sentido, não se trata de uma relação de causalidade, mas de afinidades entre dois elementos que geram efeitos objetivos.

Bourdieu rejeita explicitamente a noção de causalidade unidirecional envolvendo variáveis que descrevem e variáveis a serem descritas, alegando que ela oculta a multiplicidade de fatores e a complexidade da interação entre as variáveis. A ideia de independência de variáveis é tida como problemática justamente por atribuir independência e constância a variáveis que são imbricadas e multideterminadas. Segundo ele, “a mais independente das variáveis ‘independentes’ esconde uma verdadeira rede de relações estatísticas que

estão presentes, subterraneamente, na relação que ela mantém com determinada opinião ou prática” (BOURDIEU, 2006, p. 98). Benzécri argumentava em direção similar ao rejeitar “os procedimentos estatísticos habitualmente utilizados no mundo anglo-saxão e os métodos de tipo hipotético-dedutivos. Para ele, a análise deve visar elaborar tipologias [...]. São os dados que servem de guia e não a teoria ou ideologia”. (1984 apud DURAND, 2012, p. 1)

Nas análises de correspondências, todas as variáveis ativas são mutuamente determinadas em uma complexa estrutura de inter-relações. Consequentemente, não cabe delinear explicações decorrendo de efeitos puros e mensuráveis de variáveis independentes sobre as dependentes (LEBARON, 2009). É possível, não obstante, estabelecer nexos explicativos ou preditivos em análises multicausais, a partir da detecção dos padrões de distribuição no espaço e da probabilidade de replicação de observações, notadamente em decorrência do tipo de aproximação entre variáveis ativas e suplementares. Henry Rouanet oferece o seguinte exemplo nesse sentido:

Consideremos um estudo educacional, por exemplo, no qual para cada estudante, variáveis relativas a diversos temas são adotadas como variáveis ativas para “criar distância” entre os estudantes e assim construir um espaço educacional. Suplementarmente a essas variáveis, fatores estruturantes, como características de identificação (gênero, idade etc.), poderiam ser registrados; sua relevância se reflete em questões como “como homens e mulheres se dispersam no espaço educacional?”. Em linguagem experimental, se poderia falar em termos do “efeito de gênero” ou, caso haja preferência por uma linguagem mais neutra, seria possível falar em *predição*, consequentemente a questão seria: sabendo o gênero de um estudante (“variável preditiva”), qual seria a posição desse estudante no espaço (variável geométrica a ser predita)? (ROUANET, 2006, p. 144, tradução minha)

Ao pensar em termos de variáveis geométricas a serem preditas, desestrutura-se a relação de causalidade simples de variáveis independentes e dependentes. A relação de

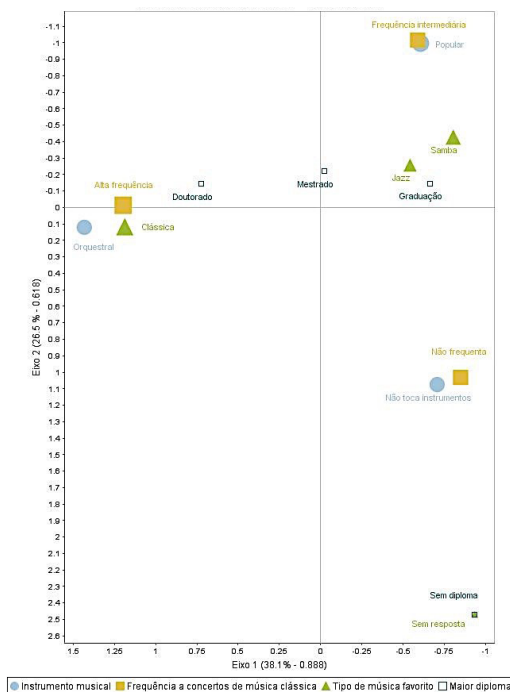
efeito de uma variável preditiva, nesse sentido, não se exerce sobre variáveis “independentes”, apagando as múltiplas relações entre variáveis subjacentes às relações. Ao construir primeiramente o espaço social, em função das múltiplas interações entre variáveis, é possível guardar a complexidade das influências mútuas entre elas; e só então buscar prever em função da probabilidade de reincidência de uma relação, como será discutido adiante.

Isto posto, a seleção das variáveis ativas e suplementares depende da questão que se pretende responder e do tipo de explicação buscada. Pode-se, dependendo da intenção sociológica, trabalhar apenas com variáveis ativas e dedicar-se predominantemente à interpretação das polaridades e aproximações detectadas na estrutura do espaço resultante.

Pode-se, alternativamente, empregar variáveis suplementares caso se pretenda visualizar a dispersão de alguma propriedade específica em um espaço social e fazer predições relativas ao ponto de ingresso, no espaço, de um agente adicional, dadas as suas propriedades sociais.

No caso do *survey* imaginário, seria possível, por exemplo, construir o espaço com base nas variáveis relativas às práticas e preferências musicais e, então, projetar a titulação dos agentes como variável suplementar. Tal construção seria pertinente, por exemplo, para investigar se existe efeito do maior diploma sobre a dispersão dos agentes no espaço das preferências musicais e indagar se seria possível prever a posição de um agente no espaço de acordo com seu diploma – correlação que será averiguada ao final do texto.

Figura 3
Nuvem das modalidades – eixos 1 e 2, construída a partir da tabela
agentes versus variáveis, com projeção das variáveis sobre o eixo 1



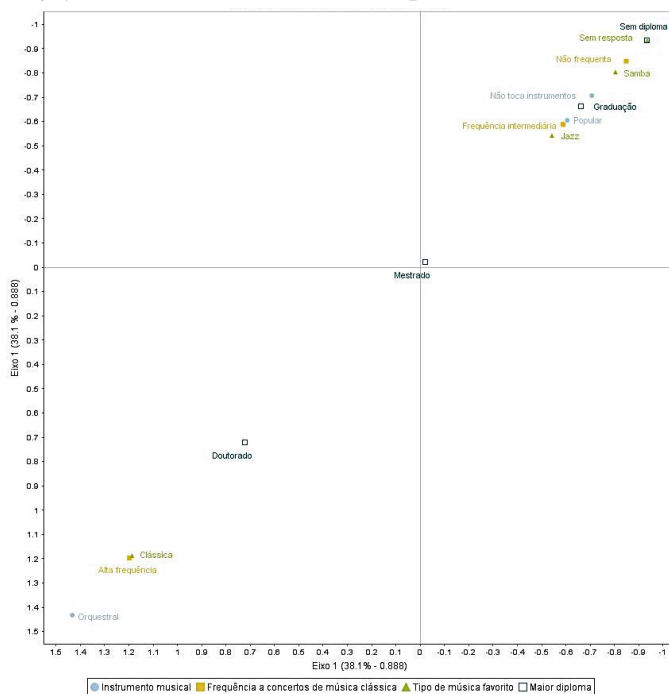
Na nuvem das modalidades representada na Figura 3¹¹ aparecem como variáveis ativas aquelas ilustradas por formas geométricas preenchidas: domínio de instrumento musical, frequência a concertos e tipo de música favorita. O tamanho das formas varia de acordo com o número de respondentes associado a cada modalidade, ou seja, de acordo com a massa da modalidade representada. As modalidades suplementares, por sua vez, são indicadas por quadrados transparentes. Posto que elas não têm massa, seu tamanho na representação não varia. Por razões didáticas, ao espaço social resultante foram acrescentadas linhas tracejadas que indicam a posição de cada modalidade

no eixo 1 – horizontal. A função das linhas é assinalar que a leitura do quadro deve ser feita a partir da projeção das modalidades em um dos eixos de cada vez. O resumo das distâncias resultante da projeção das modalidades no eixo 1 é apresentada na Figura 4.

Nas palavras de Jean-Paul Benzécri, principal desenvolvedor do método,

interpretar um eixo equivale a descobrir o que é similar, por um lado, entre todos os elementos figurando à direita do ponto central e, por outro lado, entre tudo aquilo que está escrito à esquerda; e expressar concisamente e com precisão o contraste (ou oposição) entre os dois extremos. (BENZECRI, 1992, p. 405, tradução minha)

Figura 4
Projeção das modalidades ativas e suplementar sobre o eixo 1



11 Todas as imagens foram construídas com o *software* SPAD, versão 9.0

A simples observação da Figura 4 – sem dispor ainda de ferramentas estatísticas para a análise – sugere que, no eixo 1, há maior proximidade do diploma de doutorado com os instrumentos orquestrais, a preferência pela música clássica e a alta frequência a concertos. No polo oposto, próximo ao diploma de graduação, encontram-se o *jazz* e o *samba*, a frequência intermediária a concertos e aqueles que não os frequentam, bem como a prática de instrumentos associados a gêneros musicais populares e o não domínio de instrumentos musicais. A observação livre da Figura 4, resultante da simples descrição das polaridades na nuvem de modalidades, deve ser complementada pela exploração das propriedades estatísticas da distribuição construída pela ACM.

“Gravidade”

Analisar a ACM requer a compreensão do princípio que rege a disposição espacial das “partículas” de acordo com suas massas e da ideia de contribuição das variáveis para a determinação dos eixos. Para tanto, faz-se útil o recurso a outra noção oriunda da física, a de “gravidade”, mencionada por Bourdieu e Wacquant quando afirmam que o campo é uma “configuração relacional dotada de uma *gravidade* específica que ele impõe a todos os objetos e agentes que o adentram” (BOURDIEU; WACQUANT, 1992, p. 16-17, tradução e grifos meus).

Ao observar um espaço social constituído ao redor de um par perpendicular de eixos, constata-se haver um *ponto central* no sistema. O ponto central encarnaria um hipotético agente médio da nuvem (DUVAL, 2015, p. 221) ou uma modalidade por todos partilhada e poderia ser pensado, inclusive, como um ponto com massa praticamente infinita em um sistema gravitacional. As propriedades sociais representadas no sistema têm tanto mais massa quanto mais disseminada for tal característica na população estudada. Supondo que a “força gravitacional” fosse o único determinante da posição dos corpos no sistema, quanto maior a massa de uma propriedade, maior a força de “atração gravitacional”¹² à qual estaria submetida. Em revés, uma propriedade pouco disseminada no grupo analisado teria menor massa e tenderia a aparecer em uma posição periférica por ser menos sujeita à “força gravitacional”. Lembrando que as variáveis suplementares seriam elementos de massa zero que não sofreriam, portanto, influência da “gravidade”.

A *massa* é um dos componentes da *contribuição* da propriedade social para a *variância* (ou *inércia*¹³) ao longo dos eixos que formam a nuvem. A variância de cada um deles é medida tanto em função do percentual da variância total da nuvem explicado por aquele eixo quanto em termos de seu valor próprio (*eigenvalue*). São esses os dois indicadores que aparecem nos quadros ao lado do número do eixo. O primeiro é o valor percentual, o segundo,

12 Uma aproximação com a metáfora poderia ser esboçada no seguinte sentido: dada a constante gravitacional, a força de atração gravitacional variará em função da multiplicação das massas dos dois objetos que se atraem, dividida pela distância entre eles ao quadrado. No caso, os dois corpos se atraem em direção ao centro combinado de suas massas, que fica mais próximo ao objeto de maior massa. Imaginemos agora que no centro do plano cartesiano haja um objeto de massa praticamente infinita, ou seja enormemente maior do que aquela das propriedades sociais ou dos agentes dispersos nas nuvens. No caso, a atração exercida sobre objeto de menor massa faria com que ele se movesse em direção ao ponto de maior massa e este quase não se deslocaria. Assim, as propriedades sociais e agentes seriam tanto mais atraídos quanto maior a sua massa, resultando ficarem mais próximos do centro do sistema.

13 A inércia dos eixos não é mais uma metáfora da importada da física por Bourdieu, e sim um termo aportado pelos estatísticos.

o *eigenvalue*¹⁴. Por exemplo, na Figura 3, o valor percentual explicado pelo eixo 1 é 38,1 e o valor próprio 0,888. Já o eixo 2 tem valor percentual 26,5 e valor próprio 0,618.

O outro componente considerado na medida da contribuição para a construção do eixo é a distância entre a modalidade e o ponto central em cada um deles. Nesse caso, a contribuição para a variância aumenta em função da distância da modalidade ao ponto central (LE ROUX; ROUANET, 2010). Assim, a contribuição resulta da combinação da massa da modalidade e de sua distância em relação ao ponto central, seguindo a seguinte fórmula: a contribuição de uma modalidade K para o eixo 1 ($CtrK_1$) será igual à massa relativa da modalidade (número de respondentes que escolheram aquela modalidade (nK) dividido pelo número total de respondentes (n total), dividido pelo número de questões ativas (Q) multiplicada pelo quadrado da distância da modalidade K até centro (G) no eixo 1 (dKG_1) – olhar a coordenada do ponto no eixo 1 – e dividida pelo valor próprio do eixo 1 (λ_1)¹⁵ (LE ROUX; ROUANET, 2010, p. 40).

$$CtrK_1 = \frac{\left[\frac{nK}{n \text{ total}} \right] \cdot (dKG_1)^2}{\lambda_1}$$

A contribuição é uma medida essencial para a interpretação da representação visual do espaço social pois indica a importância de determinada variável ou modalidade para

a definição da estrutura e configuração das polarizações observadas ao longo dos eixos. Na Tabela 4 estão tabuladas as contribuições das variáveis e modalidades ativas para as variâncias nos eixos 1 e 2. Figuram em destaque as contribuições superiores à média. No caso das modalidades, a contribuição será superior à média quando seu valor for maior que 100 dividido pelo número de modalidades ativas. No caso das variáveis, a contribuição será superior à média quando seu valor for maior que 100 dividido pelo número de variáveis ativas.

Uma vez destacadas as propriedades sociais que determinam as principais contribuições em cada eixo e antes de iniciar a interpretação, é preciso observar que a forte contribuição de modalidades raras (pouco frequentes e distantes do centro), não respondidas e agentes muito discrepantes do conjunto podem causar distorções no desenho do espaço. É interessante manter propriedades raras quando elas forem entendidas como elementos de distinção, mas a extrema raridade, os erros e as não respostas devem ser detectados e tratados de maneira especial. Uma possível solução está na recodificação e incorporação da modalidade em questão a uma categoria mais ampla. Isto posto, torna-se visível o quão crucial é a etapa da recodificação e passa a fazer sentido a preocupação com o equilíbrio entre o número de modalidades em uma variável e de variáveis em uma rubrica¹⁶, com o objetivo de evitar a extraordinária determinação dos eixos pelo excesso de massa ou pela excepcional raridade de alguma das modalidades (ROUANET; LE ROUX, 2010).

14 Uma terceira medida frequentemente utilizada para caracterizar o eixo é medida modificada da variância segundo a fórmula de Benzécri, desenvolvida pelo estatístico para corrigir casos em que é atribuído peso demasiado baixo à contribuição dos primeiros eixos para a variância total (ROUX; ROUANET, 2010, p. 41).

15 Na tabela de contribuições o resultado final da fórmula aparece multiplicado por 100.

16 Nessa pesquisa imaginária todas as variáveis concernem a um mesmo tema. Imagine que fosse um *survey* sobre práticas culturais em geral, no caso, poderia haver uma rubrica com questões relativas à prática/preferência em artes plásticas, outra concernente à música, outra sobre teatro etc., sendo então importante a manutenção do equilíbrio no número de variáveis e modalidades associadas a cada uma dessas rubricas.

Tabela 4
Contribuições das modalidades e variáveis para os eixos 1 e 2

| Rubrica (total 1) | Variável ativa (total 3 – contribuição superior à média > 33,3) | Modalidade (total de dez modalidades das quais dez são ativas – contribuição superior à média > 10,0) | Eixo 1 (38,1%) | Eixo 2 (26,5%) |
|----------------------------------|---|---|----------------|----------------|
| Práticas e preferências musicais | Instrumento musical | Não toca instrumentos | 5,9 | 19,5 |
| | | Instrumento orquestral | 24,1 | 0,2 |
| | | Instrumento popular | 5,2 | 20,1 |
| | | Total | 35,1 | 39,8 |
| | Tipo de música favorito | Jazz | 2,8 | 0,9 |
| | | Samba | 7,6 | 3,1 |
| | | Clássico | 19,8 | 0,3 |
| | | Sem resposta | 2 | 20,6 |
| | | Total | 32,2 | 24,8 |
| | | Alta frequência | 20,2 | 0,09 |
| Frequência a concertos | Frequência intermediária | 4,1 | 17,4 | |
| | Não frequenta | 8,4 | 18,0 | |
| | Total | 32,7 | 35,4 | |

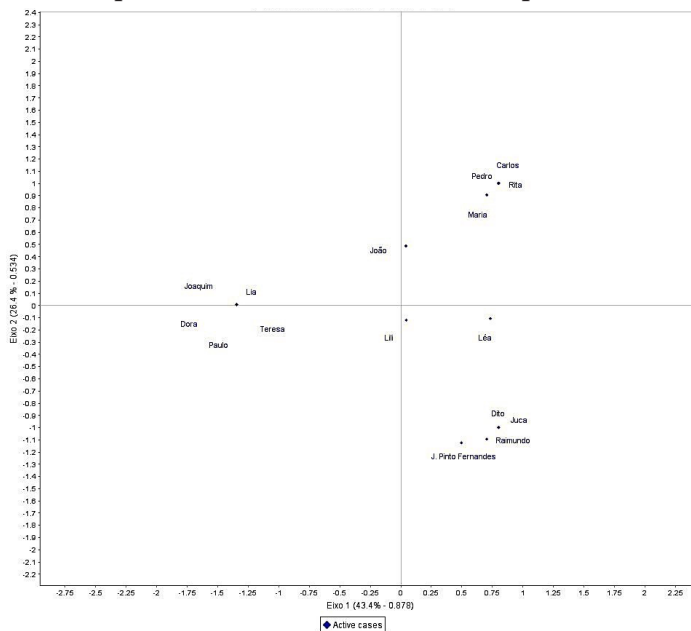
Em vermelho: modalidade com contribuição superior à média e orientação voltada para o polo positivo; azul: modalidade com contribuição superior à média e orientação voltada para o polo negativo; verde: variáveis com contribuição superior à média.

Outra solução para evitar problemas advindos de modalidades raríssimas ou defectivas está na realização de Análises de Correspondências Múltiplas com Seleção de Modalidades Ativas (Corem), técnica que torna passivas as modalidades responsáveis pela distorção do espaço. O caso das quadrilhas pode oferecer esclarecimentos a esse respeito. Nele, é preciso atentar à figura de J. Pinto Fernandes, pois seu padrão de respostas é discrepante daquele dos outros agentes representados na tabela. Por exemplo, a “não resposta” oferecida por Fernandes à questão relativa a preferências musicais figura como o elemento com a maior contribuição para a determinação do eixo 2 na Tabela 4. No caso da Figura 2, nota-se a larga distância entre J. Pinto Fernandes e o resto dos indivíduos e, na Figura 3, o grau de afastamento das propriedades a ele associadas, a saber, a ausência de diplomas e a não resposta relativa ao gênero musical favorito.

Poder-se-ia considerar a discrepância de Fernandes como decorrente de um erro

na coleta de dados. Seria importante, pois, verificar se não houve falha ao recolher ou tabular as informações. Caso seja descartada essa possibilidade, constata-se que se trata efetivamente de um caso diferente, que interessa, portanto, manter no estudo. A questão passaria, então, a ser como contemplar o caso sem distorcer o espaço por conta da introdução de propriedades raríssimas. Uma solução seria utilizar a técnica denominada Corem para transformar apenas a modalidade “sem resposta” em uma *modalidade passiva*, conservando ativas na construção do espaço as outras respostas relativas às preferências musicais. Além de neutralizar respostas muito discrepantes, o procedimento de seleção de modalidades é utilizado para anular, quando conveniente, o efeito desorganizador de não respostas, erros, modalidades “outros” e demais elementos sem interesse para a análise ou prejudiciais à leitura dos quadros (ROUANET; LE ROUX, 2010).

Figura 5
Nuvem dos agentes – eixos 1 e 2, após utilização de Corem para eliminar a modalidade “sem resposta”



Na Figura 5 – que representa a nova nuvem dos agentes, criada após a neutralização da modalidade “sem resposta” –, J. Pinto Fernandes aparece próximo ao grupo com o qual partilha mais propriedades, a saber, aquele sem frequência a concertos e sem domínio de instrumentos musicais. Ao eliminar o caráter ativo da “não resposta” fornecida por ele, logra-se tornar Fernandes um caso comparável a outros, sem que ele seja fundido com eles em um mesmo ponto. Outra vantagem decorrente da correção da distorção é a redução da distensão vertical resultante da existência da “não resposta”. Tal alteração modifica a relação de contribuição das modalidades para o eixo 2, permitindo avaliar com maior apuro quais os efetivos criadores de polarizações nessa dimensão do espaço.

Minimizadas as distorções, é possível passar à análise e interpretação detalhada da

nova tabela de contribuições (Tabela 5) e dos quadros a ela associados (Figuras 6 e 7). Em primeiro lugar, é preciso destacar que as medidas assinaladas tratam da contribuição das modalidades para cada um dos eixos, fazendo-se necessário definir, finalmente, o que são eixos e indicar como interpretar estatisticamente a variação ao longo dos eixos.

Existem múltiplos eixos que passam pelo ponto central. Cada plano cartesiano é formado por um par de eixos perpendiculares que se cruzam no ponto central, sendo eles enumerados. O primeiro (eixo 1) é aquele que contempla a maior parte da variância total da nuvem, ou seja, que apresenta o maior grau de estiramento e dispersão das propriedades sociais ao longo do eixo. O segundo (eixo 2) contempla a segunda maior variância, e assim sucessivamente. A dispersão das variáveis projetadas ressalta as oposições entre elas, assim, quanto maior a

parcela da variância representada por um eixo, mais importante ele será para exibir e interpretar as polarizações expressas no espaço. A cada eixo será atribuído um valor correspondente ao percentual da variância total da nuvem que é por ele explicado – indicado na tabela de contribuições abaixo do número do eixo (DUVAL, 2015; HJELLBREKKE; KORSNES, 2009). O que é possível observar no exemplo adotado é o crescimento da variância representada pelo eixo 1 decorrente da eliminação da “não resposta” (de 38,1% para 43,4%), aumentando a valia analítica da representação.

Uma vez construída a tabela de contribuições, cabe ao pesquisador decidir quantos e quais eixos pretende retratar, dadas as questões formuladas e tendo em mente contemplar a maior soma de variâncias pertinente. Passa-se, então, à etapa de *interpretação* da ACM, decorrente da descrição combinada da tabela de contribuições e das representações visuais das

nuvens de agentes e de modalidades dispostas ao redor dos eixos retidos.

Pode-se principiar pela leitura intuitiva das oposições em cada eixo, como feito anteriormente com a Figura 4, seguindo sempre a fórmula de Benzécri, segundo a qual é preciso entender qual é o sentido do agrupamento dos pontos que estão à esquerda do ponto central e o que une os pontos que estão à direita do ponto central. A tabela de contribuições ajuda a entender quais são as modalidades e as variáveis que devem ser observadas mais atentamente nesse processo, por serem as mais relevantes para a determinação dos eixos. Uma vez concluída essa etapa, Le Roux e Rouanet (2010) sugerem ser possível adotar rótulos para caracterizar cada um dos lados do eixo. Ainda que os rótulos possam ser simplificadores, eles ajudam o leitor a detectar de imediato o principal sentido da oposição expressa por cada eixo, conforme ilustrado nas Figuras 6 e 7.

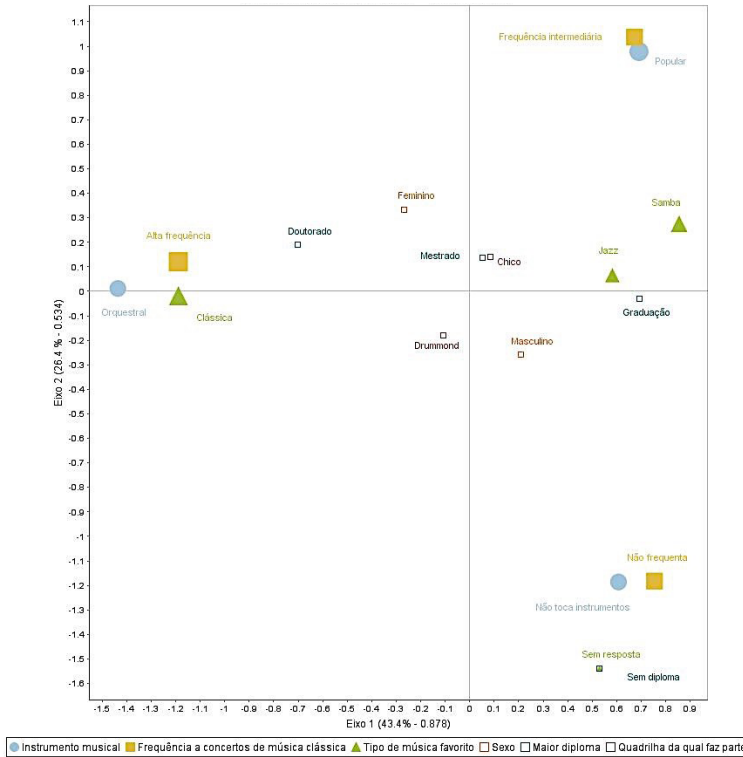
Tabela 5
Contribuições das modalidades e variáveis para os eixos 1, 2 e 3

| Rubrica (total 1) | Variável ativa (total 3 – contribuição superior à média > 33,3) | Modalidade (total de dez modalidades das quais nove são ativas – contribuição superior à média > 11,1) | Eixo 1 (43,4%) | Eixo 2 (26,4%) | Eixo 3 (18,8%) |
|----------------------------------|---|--|----------------|----------------|----------------|
| Instrumento musical | | Não toca instrumentos | 4,4 | 27,4 | 3,7 |
| | | Instrumento orquestral | 24,5 | 0,0 | 0,3 |
| | | Instrumento popular | 6,8 | 22,4 | 5 |
| | | Total | 35,7 | 49,8 | 8,9 |
| Práticas e preferências musicais | Tipo de música favorito | Jazz | 3,2 | 0,1 | 55 |
| | | Samba | 8,7 | 1,5 | 24,5 |
| | | Clássico | 20,1 | 0,0 | 2 |
| | | Sem resposta* | | | |
| | | Total | 32,0 | 1,6 | 81,4 |
| Frequência a concertos | | Alta frequência | 20,2 | 0,3 | 1,9 |
| | | Frequência intermediária | 5,4 | 21,00 | 6,6 |
| | | Não frequente | 6,8 | 27,3 | 1,2 |
| | | Total | 32,3 | 48,6 | 9,7 |

Em vermelho: modalidade com contribuição superior à média e orientação voltada para o polo positivo; azul: modalidade com contribuição superior à média e orientação voltada para o polo negativo; verde: variáveis com contribuição superior à média.

* Modalidades com baixa incidência que foram transformadas em suplementares para não distorcer o desenho do espaço, não contribuindo para a formação dos eixos.

Figura 6
Nuvem das modalidades – eixos 1 e 2, após utilização de
Corem para eliminar a modalidade “sem resposta”



Ao analisar os quadros, tendo em mente as variáveis com contribuição acima da média, é possível verificar que cada eixo retrata polarizações promovidas por elementos distintos. No eixo 1, a principal polarização resulta da oposição entre os adeptos da música clássica, incluindo a modalidade alta frequência a concertos e o domínio de instrumentos frequentemente utilizados para música orquestral, de um lado, e os adeptos da música popular, de outro. No eixo 2, a oposição predominante é entre aqueles sem engajamento musical (não frequentadores e não instrumentistas) e aqueles com engajamento musical (com destaque os frequentadores intermediários e praticantes de música popular). O eixo 3, por fim, expressa uma divisão interna ao

segmento da música popular, opondo os adeptos do *jazz* aos apreciadores do samba.

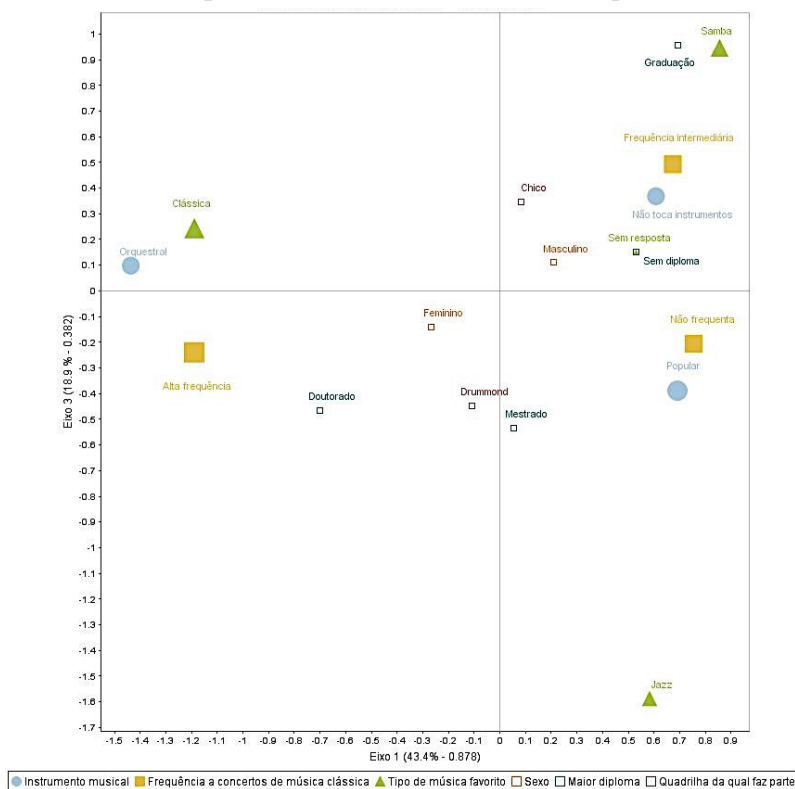
Uma vez analisada a estrutura da nuvem de modalidades e resumidas as oposições básicas dos eixos, torna-se possível passar à exploração da posição das variáveis e modalidades suplementares e à busca de regras para ingresso de novos agentes na estrutura, tal qual sugere Henry Rouanet (2006) em seu exemplo relativo ao espaço educacional.

Ao explorar as variáveis suplementares, é possível tecer as seguintes observações: os membros das quadrilhas de Chico e Drummond se repartem igualmente entre o clássico e o popular. Quanto à divisão entre os gêneros populares, faz-se visível, no eixo 3, uma distinção

entre as quadrilhas no que concerne ao tipo de popular com o qual têm afinidade – no caso da quadrilha do Chico, o samba, no caso da quadrilha de Drummond, o *jazz*. No eixo 2, o masculino aparece mais perto da baixa prática e baixa frequência musical do que o feminino. Le Roux e Rouanet (2010, p. 59) oferecem uma medida convencional para avaliar a importância da polarização de variáveis suplementares. Trata-se de uma classificação da distância entre elas, medida de acordo com as

coordenadas de cada um dos pontos em cada um dos eixos. Caso a distância seja menor que 0,5, não se trata de uma apreciável. Se estiver entre 0,5 e 1,0 é uma diferença possível de ser notada. Tratando-se de uma distância maior que 1,0, consiste em uma grande diferença. No caso, as únicas diferenças maiores que 0,5 são a quadrilha no eixo 3 (0,795) e o sexo no eixo 2 (0,593), assim, há dificuldade para a predição relativa ao ponto de entrada de um novo agente.

Figura 7
Nuvem das modalidades – eixos 1 e 3, após utilização de
Corem para eliminar a modalidade “sem resposta”



No que concerne ao efeito dos diplomas sobre as práticas e preferências musicais, o eixo 1 sinaliza uma grande diferenciação (1,395)

como efeito do diploma, decorrente da afinidade entre o doutorado e a adesão à música clássica e da proximidade entre a música

popular e os detentores do título de graduação. O eixo 2, por sua vez, só diferencia o sem diploma dos demais, não oferecendo pistas para posicionar no espaço um novo agente de acordo com o seu diploma. Já o eixo 3 permite sugerir que aqueles com mestrado e doutorado estão mais frequentemente associados ao *jazz*, e os graduados, ao samba (1,456).

Na busca por evidências relativas ao efeito de uma variável, a ideia de dispersão também deve ser considerada. Ao trabalhar com variáveis suplementares é possível – além de observar o ponto médio de incidência delas no espaço e as distâncias entre cada par de modalidades suplementares – fazer recurso a *elipses de concentração*. Essas são desenhadas a partir do ponto de incidência da modalidade suplementar selecionada e indicam o grau de dispersão no espaço daqueles que partilham de tal característica (HJELLBREKKE; KORSNES, 2009). Quanto maior a concentração de uma elipse, maior a proximidade, na nuvem dos agentes, daqueles que coincidem na propriedade social analisada.

No caso, torna-se possível avaliar se os portadores de cada diploma ocupam posições apartadas no espaço ou se há intersecções entre as áreas do diagrama ocupadas por cada um dos subgrupos, além de observar se estão concentrados em uma área pequena ou dispersos no espaço. Tais elementos ajudam a avaliar em que medida se pode prever onde um agente adicional teria maior probabilidade de figurar caso ingressasse no espaço.

Pode-se observar que, na nuvem dos agentes retratada na Figura 8, há maior concentração dos portadores de diplomas de doutorado e de graduação e uma grande dispersão dos detentores do título de mestre. O mestrado é, pois, um indicador com pouca efetividade para predição da posição que um novo agente ocuparia no espaço. Já os diplomas de graduação e doutorado aparecem concentrados em áreas mais restritas do espaço social.

Um novo agente com doutorado teria maior probabilidade de aparecer no polo positivo do eixo 1 e em posição central no eixo 2. Um novo agente com diploma de graduação teria posição altamente variável no eixo 2 e maior probabilidade de aparecer no polo negativo do eixo 1.

A interpretação primordial da ACM, na forma das suas nuvens de modalidades e nuvens agentes, deve estar desenvolvida após passar: pela observação e descrição das polarizações no espaço social, pela análise da tabela de contribuições para detecção dos elementos com maior peso na definição da estrutura do espaço e pela análise do posicionamento das variáveis e modalidades suplementares, incluindo observação das elipses de concentração e da posição de agentes em particular, quando se tratarem de figuras conhecidas. Nesse ponto, é possível avançar para análises que combinem ACMs com outros métodos e técnicas.

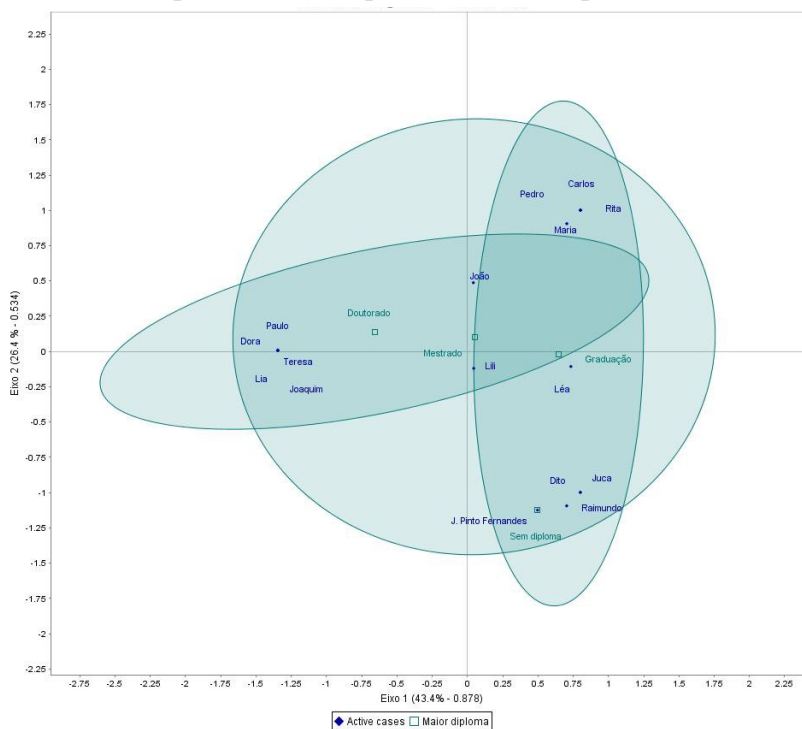
Uma combinação corrente é a de ACMs com análises de *classificação hierárquica* (ou classificação euclidiana). Trata-se de uma ferramenta que se apropria das árvores classificatórias (ou dendrogramas) das ciências naturais, agrupando sucessivamente aqueles que tem maior proximidade até chegar a um pequeno número de classes ou *clusters* tão diversos entre si quanto possível (BERTONCELO, 2016b; LE ROUX; ROUANET, 2004). Tal ferramenta foi utilizada por Bourdieu (1999) em sua pesquisa sobre o espaço editorial para delinear *clusters* que reuniam editoras com características similares, tratando-se de um instrumento útil para a construção de tipologias à moda weberiana.

Outra possibilidade é a realização de testes combinando as posições na nuvem de agentes da ACM com representações geradas por *análises de redes sociais*. Tal amalgama permite, por exemplo, averiguar se os agentes que ocupam posições próximas no espaço social têm maior chance de estabelecer conexões efetivas, dada

a similaridade de seus *habitus* e interesses. Uma forma de fazê-lo é projetar, no plano cartesiano da ACM, os laços entre agentes efetivamente conectados e verificar se a distância social entre eles é efetivamente menor do que a média das distâncias que resultariam das conexões hipotéticas entre quaisquer dois

agentes (KLÜGER, 2017). Outra forma de fazê-lo é observar as posições dos agentes nas redes sociais e nas ACMs e avaliar o grau de semelhança entre as duas representações, buscando explicações para as diferenças entre os posicionamentos delineados com o auxílio de cada um dos métodos (KLÜGER, no prelo).

Figura 8
Nuvem dos agentes – eixos 1 e 2 com adição de elipses para a variável suplementar maior diploma



As análises geométricas podem ser combinadas com outros métodos tanto sincronicamente – como nos casos da conciliação com as análises de classificação hierárquica e com as análises de redes – quanto diacronicamente, ou seja, ela pode ser uma etapa em pesquisas envolvendo mais de um método. Ressalta-se que o método pode contribuir largamente com incursões exploratórias, visto que ajuda

a descrever a estrutura social ao mapear as polarizações existentes em um conjunto de agentes e indicar as tendências expressas pela oposição entre suas propriedades sociais.

Conclusão

À guisa de conclusão, reforça-se a ideia de que há uma afinidade essencial entre uma

visão relacional da sociedade e a representação geométrica realizada com auxílio da ACM, dado que se trata de uma técnica capaz de posicionar os agentes e as propriedades sociais uns em relação aos outros, extraindo sentido das distâncias e proximidades observadas. Ao fazê-lo, o método geométrico oferece elementos para a visualização dos fundamentos das polarizações e lutas travadas no espaço social, objetivando a estrutura do campo e oferecendo elementos para compreender as

tomadas de posição dos agentes de acordo com suas posições na estrutura. Cabe ressaltar que a interpretação oferecida pelas ACMs é essencialmente decorrente da descrição e interpretação dos mapas e tem por princípio a não atribuição de independência a variáveis, por julgar os fenômenos da vida social como complexos e imbricados. Nesse sentido, privilegia as ideias de afinidade e predição e a construção de tipologias em lugar da busca de causas unidirecionais.

Referências

- BELÉM, M. P. *A lei Rouanet e a construção de um mercado de patrocínios culturais no Brasil*. 2015. Tese (Doutorado em Ciência Política) – Universidade Federal de São Carlos, São Carlos, 2015..
- BENNETT, T. et al. *Culture, class, distinction*. London: Taylor & Francis, 2009.
- BENZECRI, J.-P. *Correspondence analysis handbook*. New York: Dekker, 1992.
- BERTONCELO, E. O espaço das classes sociais no Brasil. *Tempo Social*, São Paulo, v. 28, n. 2, p. 73-104, 2016a.
- _____. O uso da Análise de correspondências múltiplas nas ciências sociais: possibilidades de aplicação e exemplos empíricos. In: ENCONTRO ANUAL DA ANPOCS, 40., 2016, Caxambu. *Anais...* Caxambu: Anpocs, 2016b. p. 1-25.
- BLASIUS, J.; GREENACRE, M. *Visualization of categorical data*. Boca Raton: Academic Press, 1998.
- BORDIGNON, R. R. Recrutamento e modalidades de entrada na carreira política: candidatos aos cargos legislativos no Rio Grande do Sul (1998 – 2006). *Política & Sociedade*, Trindade, v. 16, n. 35, p. 351, 9 jun. 2017.
- BORTOLUCI, J. H.; JACKSON, L. C.; PINHEIRO FILHO, F. A. Contemporâneo clássico: a recepção de Pierre Bourdieu no Brasil. *Lua Nova: Revista de Cultura e Política*, São Paulo, n. 94, p. 217-254, 2015.
- BOURDIEU, P. Une révolution conservatrice dans l'édition. *Actes de la Recherche en Sciences Sociales*, Paris, v. 126, n. 1, p. 3-28, 1999.
- _____. *Les structures sociales de l'économie*. Paris: Le Seuil, 2000.
- _____. *A distinção: crítica social do julgamento*. São Paulo: Zouk, 2006.
- _____. *Homo academicus*. Florianópolis: Editora da UFSC, 2011.

- _____. *The state nobility: elite schools in the field of power*. Palo Alto: Stanford University Press, 1998.
- BOURDIEU, P.; DE SAINT MARTIN, M. Anatomie du gout. *Actes de la Recherche en Sciences Sociales*, Paris, v. 2, n. 5, p. 2-81, 1976.
- BOURDIEU, P.; WACQUANT, L. J. D. *An invitation to reflexive sociology*. Chicago: University of Chicago Press, 1992.
- BÜHLMANN, F.; DAVID, T.; MACH, A. Political and economic elites in Switzerland: personal interchange, interactional relations and structural homology. *European Societies*, Abingdon, v. 14, n. 5, p. 727-754, dez. 2012.
- CIBOIS, P. Analyse des données et sociologie. *L'Année Sociologique*, Paris, v. 31, p. 333-348, 1981.
- CORADINI, O. L. Representação profissional e elites políticas no Brasil no período recente. *Política & Sociedade*, Trindade, v. 5, n. 9, p. 123-162, 1 dez. 2006.
- DANTAS, E. G. et al. Prosopografia dos ajudantes do Ministério da Fazenda: uma análise das lógicas de recrutamento, dos itinerários profissionais e do perfil social do segundo escalão. In: ENCONTRO ANUAL DA ANPOCS, 41., 2017, Caxambu. *Anais...* Caxambu: Anpocs, 2017.
- DENORD, F. *Genèse et institutionnalisation du néo-libéralisme en France (années 1930-années 1950)*. 2003. 517 f. Thèse (Doctorat en Sociologie) – École des Hautes Études en Sciences Sociales, Paris, 2003.
- DI FRANCO, G. Multiple correspondence analysis: one only or several techniques? *Quality & Quantity*, Heidelberg, v. 50, p. 1299-1315, 2016.
- DUVAL, J. L'art du réalisme: le champ du cinéma français au début des années 2000. *Actes de la Recherche En Sciences Sociales*, Paris, v. 1, n. 161-162, p. 96-115, 2006.
- _____. Analisar um espaço social. In: PAUGAM, S. *A pesquisa sociológica*. Rio de Janeiro: Vozes, 2015. p. 218-237.
- DURAND, C. L'analyse factorielle des correspondances. *Notes de cours et exemples* [online], [s.l.], p. 1-9, 2012. Disponível em: <<https://bit.ly/2Nbi1vM>>. Acesso em: 8 maio 2018.
- FOURCADE, M.; LANDE, B.; SCHOFER, E. Political space and the space of politics: doing politics across nations. *Poetics*, Amsterdam, v. 55, p. 1-18, 2016.
- GODECHOT, O. *Les traders: essai de sociologie des marchés financiers*. Paris: La Découverte, 2001.
- GREENACRE, M.; BLASIUS, J. *Multiple correspondence analysis and related methods*. London: CRC Press, 2006.
- HEY, A. P. *Esboço de uma sociologia do campo acadêmico: a educação superior no Brasil*. São Carlos: EdUFScar, 2008.
- HJELLBREKKE, J.; KORSNES, O. Héritiers et outsiders? Sur la noblesse d'État norvégienne. *Actes de la Recherche en Sciences Sociales*, v. 200, n. 5, p. 86-105, 2013.

- HJELLBREKKE, J.; KORSNES, O. Quantifying the field of power in Norway. In: ROBSON, K.; SANDERS, C. *Quantifying theory: Pierre Bourdieu*. Berlin: Springer Science & Business Media, 2009. p. 11-30.
- JOI, T. M. *Desejo, necessidade e realidade: os marcadores culturais e econômicos e suas implicações ocupacionais para o grupo profissional de engenheiros de produção no Brasil*. 2015. Tese (Doutorado em Ciência Política) – Universidade Federal de São Carlos, São Carlos, 2015.
- KLÜGER, E. Espaço social e redes: contribuições metodológicas à sociologia das elites. *Tempo Social*, São Paulo, v. 29, n. 3, p. 83-110, 2017.
- _____. Mapping the inflections in the policies of the Brazilian National Economic and Social Development Bank during the nineties and two thousands within social spaces and networks. *Historical Social Research/Historische Sozialforschung*, Köln, v. 43, n. 2, no prelo.
- LEBARON, F. *La croyance économique: les économistes entre science et politique*. Paris: Seuil, 2000.
- _____. How Bourdieu “quantified” Bourdieu: the geometric modelling of data. In: ROBSON, K.; SANDERS, C. *Quantifying theory: Pierre Bourdieu*. Berlin: Springer Science & Business Media, 2009. p. 11-30.
- _____. L’analyse géométrique des données dans un programme de recherche sociologique: le cas de la sociologie de Bourdieu. *Revue Modulad*, Paris, v. 102, n. 42, 2010.
- LEBARON, F.; BONNET, P. O espaço das práticas culturais: da construção do espaço social ao estudo dos subgrupos pela análise específica de classe. *Tempo Social*, São Paulo, v. 28, n. 2, p. 11-38, 2016
- LE ROUX, B.; ROUANET, H. *Geometric data analysis: from correspondence analysis to structured data analysis*. New York: Kluwer Academic Publishers, 2004.
- _____. *Multiple correspondence analysis*. London: Sage, 2010.
- LOUREIRO, M. R. G. *Os economistas no governo: gestão econômica e democracia*. Rio de Janeiro: FGV Editora, 1997.
- LÖWY, M. Sobre o conceito de “afinidade eletiva” em Max Weber. Tradução de Lucas Amaral de Oliveira e Mariana Toledo Ferreira. *Plural*, São Paulo, v. 17, n. 2, p. 129-142, 2010.
- PEREIRA, V. B. Classes sociais e simbolização na cidade do Porto: elementos teóricos e resultados de pesquisa empírica. *Tempo social*, São Paulo, v. 28, n. 2, p. 183-206, ago. 2016.
- PEROSA, G. S.; LEBARON, F.; LEITE, C. K. S. O espaço das desigualdades educativas no município de São Paulo. *Pro-Posições*, Campinas, v. 26, n. 2, p. 99-118, ago. 2015.
- ROBSON, K.; SANDERS, C. *Quantifying theory: Pierre Bourdieu*. Berlin: Springer Science & Business Media, 2009.

- ROUANET, H. The geometric analysis of structured individuals \times variables tables. In: GREENACRE, M.; BLASIUS, J. *Multiple correspondence analysis and related methods*. London: CRC Press, 2006. p. 137-160.
- ROUANET, H.; ACKERMANN, W.; LE ROUX, B. The Geometric Analysis of Questionnaires: the lesson of Bourdieu's La Distinction. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, Thousand Oaks, v. 65, n. 1, p. 5-18, jan. 2000.
- SAPIRO, G. *La guerre des écrivains: 1940-1953*. Paris: Fayard, 1999.
- SOUZA, R. C. *A ciência dos economistas: entre dissensos científicos e clivagens morais*. 2009. Dissertação (Mestrado em Sociologia) – Instituto Universitário de Pesquisas do Rio de Janeiro, Rio de Janeiro, 2009.
- TARGA, L. G. *Os diplomatas brasileiros sob a perspectiva relacional: o campo dos diplomatas e o campo político*. 2017. Tese (Doutorado em Ciência Política) – Universidade Federal de São Carlos, São Carlos, 2017.
- WEBER, M. *A ética protestante e o "espírito" do capitalismo*. São Paulo: Companhia das Letras, 2004.
- _____. *A "objetividade" do conhecimento nas ciências sociais*. São Paulo: Ática, 2006.

Resumo

Análise de Correspondências Múltiplas: fundamentos, elaboração e interpretação

As análises de correspondências múltiplas (ACM) foram popularizadas nas ciências sociais por Pierre Bourdieu e seus discípulos. A técnica foi privilegiada por eles por retratar o espaço social de maneira relacional, ao criar, em planos cartesianos, nuvens nas quais agentes são posicionados uns em relação aos outros, de acordo com suas propriedades sociais; e nuvens de modalidades nas quais propriedades sociais aparecem tanto mais próximas quanto mais frequentemente associadas aos mesmos agentes. Desse modo, a representação criada pela ACM permite observar a estrutura de polarizações própria ao espaço social, o que ajuda a compreender dinâmicas sociais decorrentes. O objetivo deste artigo é apresentar os fundamentos, os elementos constitutivos e as etapas necessárias para a elaboração de ACMs. Assim, pretende-se subsidiar a leitura de trabalhos que empreguem ACMs e, quiçá, auxiliar pesquisadores com intenção de utilizar a técnica em suas investigações.

Palavras-chave: Análise de Correspondências Múltiplas; Pierre Bourdieu; Campo Magnético; Geometria; Gravidade.

Abstract

Multiple Correspondence Analyses: fundamentals, elaboration and interpretation

Multiple Correspondence Analysis (MCA) was popularized in Social Sciences by Pierre Bourdieu and his followers. The technique was privileged by them since it allows representing the social space relationally, while creating, in Cartesian planes, clouds in which agents are positioned relative to each other according to their social properties; and clouds of social properties in which categories appear closer and more frequently when associated with the same agents. Thus, the representation created by the MCA allows detecting the structure of polarizations of social spaces, which helps understanding the resulting social dynamics. The goal of this article is to present fundamentals, basic elements and procedures to elaborate MCAs. Thereby, we intend to subsidize the reading of studies that use MCAs and, perhaps, to help researchers intending to use the technique in their work.

Keywords: Multiple Correspondence Analysis; Pierre Bourdieu; Magnetic Field; Geometry; Gravity.

Résumé

Analyses des Correspondances Multiples : fondements, élaboration et interprétation

L'Analyse des Correspondances Multiples (ACM) était popularisée dans les Sciences Sociales par Pierre Bourdieu et ses suivants. La technique en a été privilégiée car elle permet de représenter l'espace social relationnellement, tout en créant, dans des plans cartésiens, des nuages dans lesquels les agents sont positionnés les uns par rapport aux autres en fonction de leurs propriétés sociales ; et des nuages de propriétés sociales dans lesquels les catégories apparaissent plus proches et plus fréquemment lorsqu'elles sont associées aux mêmes agents. Ainsi, la représentation créée par le ACM permet de détecter la structure des polarisations des espaces sociaux, ce qui aide à comprendre les dynamiques sociales qui en résultent. Le but de cet article est de présenter les fondements, les éléments et procédures de base pour élaborer des ACMs. De ce fait, on a l'intention de subventionner la lecture des études qui utilisent des ACMs et, peut-être, d'aider les chercheurs qui ont l'intention d'utiliser cette technique dans leurs études.

Mots-clés: Analyse des Correspondances Multiples; Pierre Bourdieu; Champ Magnétique; Géométrie; Gravité.

O retorno da comparação etnográfica: a epistemologia em momentos de mudança política^{1,2}

Erik Bähre³

Introdução

A comparação é indispensável quando se tenta explicar alguma coisa. As explicações baseiam-se na identificação de diferenças que podem existir entre populações ou no interior delas, entre agora e depois ou aqui e ali, ou ainda entre experiências e interpretações.⁴ A comparação inevitavelmente levanta questões sobre a definição da diferença. Qual linguagem e quais conceitos podemos usar para comparar e como levar em conta que a linguagem e os conceitos estão sempre alicerçados em sociedades, culturas (acadêmicas) e momentos históricos específicos? Como a comparação produz ou reifica diferenças entre “nós” e “eles”, diferenças que podem facilmente assumir uma relação hierárquica? Este artigo examina abordagens relativas à comparação na etnografia à luz dos últimos debates. Contribuições recentes de Lazar (2012) e van der Veer (2016) e números especiais sobre comparação nas

revistas *Social Analysis* (publicado em 2013) e *HAU* (publicado em 2017), entre outros, colocaram a questão comparativa no centro do palco. Para ser franco, a comparação retornou à etnografia depois de estar ausente por quase quarenta anos.

Estudos comparativos tendem a envolver comparações entre países, o que reforça a ideia de que um Estado-nação é a escala de análise mais apropriada e obscurece o fato de que muitos fenômenos sociais não são organizados dentro desse arcabouço sociopolítico e legal. Algumas das principais categorias de comparação na antropologia, como culturas, comunidades ou etnias, têm seus próprios problemas e muitas vezes reificam essas identidades. A etnografia deu importantes contribuições para desconstruir tais categorias, identificando sua natureza contingente, suas muitas e frequentemente implícitas suposições subjacentes e as formas de violência simbólica ou estrutural que a pesquisa comparativa pode acarretar.⁵

1 Tradução de Sergio Lamarão.

2 Devido às conversas inspiradoras que lá tive, gostaria de agradecer aos participantes da oficina sobre comparação organizada pela Leiden Global, em 12 de maio de 2015, e da oficina Ethnographic Research and Comparison, realizada em 3 e 4 de maio de 2017, na Universidade de Leiden. Sou grato também a Nikkie Buskermolen, Maghriel van Crevel e Sian Lazar por seus comentários a respeito de uma versão anterior deste artigo. Gostaria de agradecer também aos editores e aos revisores anônimos da BIB por seus pertinentes comentários e a Sergio Lamarão pela tradução do artigo para português. Este projeto recebeu financiamento do Conselho Europeu de Investigação (CED) no âmbito do Programa de Investigação e Inovação Horizonte 2020 (Grant Agreement No. 682467).

3 Professor Associado da Universidade de Leiden. E-mail: ebaehre@fsw.leidenuniv.nl

4 Ver, entre outros, Mary Douglas (1966, 1994); Handler (2013); e van der Veer (2016).

5 Ver, por exemplo, Bourdieu et al. (2002), Bourdieu e Wacquant (1992), Fabian (1983), Das (1998), Das et al. (2000) e Hastrup (1995). A dominação pelo método é enfaticamente argumentada por Hacking (1990), que mostrou como o desenvolvimento da estatística e da matemática probabilística possibilitou comparar as populações de novas maneiras, e que essas comparações e medições poderiam servir para dominar os sujeitos comparados. Mais recentemente, o livro *Gross domestic problem* (Fioramonti, 2013) mostra como uma das medidas mais influentes, o Produto Interno Bruto, está longe de ser neutra, enfatizando, ao contrário, o resultado das ideologias e políticas temporais.

A comparação levanta questões epistemológicas fundamentais em relação ao controle. Partindo de um entendimento objetivista, a comparação necessita ter controle sobre o processo de investigação. Isso também implica considerar cuidadosamente se a população da pesquisa ou do estudo de caso é representativa de um todo maior e, portanto, precisa estar sob o controle de quem e do que se estuda. Fazer comparações de base objetivista requer o controle das categorias e das escalas usadas. Esse tipo de comparação deve ser capaz de destacar uma variável específica que controle outras variáveis. Por último, mas não menos importante, uma interpretação objetivista significa controlar a perspectiva do pesquisador. A influência do pesquisador precisa ser diminuída tanto quanto possível, a fim de possibilitar a comparação e garantir que os fatos estejam corretos. Chamarei a abordagem epistemológica objetivista – que determina que a comparação requer altos níveis de controle sobre o processo de pesquisa, incluindo o pesquisador e as categorias e a linguagem que permitem o confronto de dados – de “comparação controlada”.

A comparação que diverge dos altos níveis de controle sobre o processo de pesquisa, o pesquisador e a linguagem (conceitual) levanta inevitavelmente questões sobre o poder. As abordagens etnográficas contemporâneas são geralmente fundamentadas em uma epistemologia interpretativa e reflexiva que problematiza níveis elevados de controle. Elas consideram essas altas níveis de controle sobre o processo de pesquisa como uma forma de violência simbólica, na medida em que o pesquisador controla a linguagem e as categorias e decide quem e o que é representativo de populações e processos sociais maiores. Em vez de encarar o pesquisador como portador de uma perspectiva cujo efeito precisa ser mitigado, este tipo de abordagem encara a dimensão humana como aspecto fundamental do trabalho. A pesquisa

é entendida como um empreendimento humano, o que vale dizer que as interpretações e explicações são sempre conduzidas por pessoas de configurações sociais e históricas específicas.

Em vez de tentar filtrar esse dado, pode-se aceitá-lo, refletindo sobre a posição do pesquisador. As percepções são fortuitas. É possível comparar quando se tem menos controle sobre o processo de pesquisa, o material coletado, o pesquisador e a linguagem usada para entender as coisas? Chamo de “comparação fortuita” a comparação baseada numa epistemologia interpretativa, a qual, ao invés de destacar o controle e a representação, problematiza a reflexividade e a saliência. Uma comparação fortuita sugere níveis muito menores de controle e, portanto, é baseada em um ambiente de pesquisa que inclui relações e oportunidades específicas para encontros inesperados, a conciliação da linguagem e das categorias usadas para descrever e analisar e a relação entre o que é estudado (uma entrevista, um caso) e o todo maior que é baseado na saliência. Aqui, a saliência significa que se estuda um caso ou um evento específico porque ele oferece *insights* específicos sobre dinâmicas sociais mais amplas.

Como uma comparação fortuita pode ser formulada com mais clareza e como ela se relaciona à comparação controlada? A comparação controlada e a fortuita podem ser distinguidas ao longo das linhas de controle/reflexividade e de representatividade/saliência.

Se a pesquisa tem que refletir sobre as circunstâncias em que certas agendas, questões e abordagens de pesquisa se desenvolvem, então isso também vale para a questão comparativa na etnografia. A maneira como os etnógrafos se comparam, caso eles se comparem, caso adotem uma abordagem mais controlada ou fortuita, depende de acontecimentos sociais e políticos mais amplos. Para examinar esse aspecto, não restringirei minha análise ao recente debate sobre a comparação etnográfica. Primeiro, examinarei o trabalho de Mead e Bateson (1942),

realizado nas décadas de 1930 e 1940, e dois estudos-chave de Geertz (1962, 1973), todos eles importantes defensores das tradições de pesquisa que oferecem contribuições significativas aos debates sobre a comparação etnográfica. Mead e Bateson empregaram uma nova técnica de pesquisa, com o uso da fotografia e do filme, vista na época como um novo padrão para assegurar a objetividade e a comparação. Essas tecnologias tornaram-se fundamentais para uma comparação controlada, mediante a qual as pessoas podiam ver os fatos por elas mesmas (pelo menos é nisso que Mead e Bateson acreditavam). Todavia, em vez disso, o modo como usaram essas técnicas reproduziu hierarquias entre sociedades que se identificavam com as relações coloniais. O trabalho de Geertz revela uma importante mudança na antropologia, já bem estudada. Porém, para o propósito deste artigo, vale a pena comparar duas de suas publicações, que, embora examinem, ambas, a circulação do dinheiro, o fazem de maneiras muito diferentes.

Não argumento que a comparação feita por Mead e Bateson seja representativa da antropologia como um todo ou de uma escola

particular. Selecionei essas publicações por sua relevância. Mead e Bateson experimentaram um novo método “objetivo” que abriu caminho para novas formas de comparação na antropologia. Os textos selecionados ajudam a explicar o que quero dizer com os tipos de comparação mencionados: a comparação controlada e a comparação fortuita precisam ser entendidas nos moldes de controle/reflexividade e representatividade/saliência. Esses artigos são significativos, pois destacam como a comparação foi rejeitada pela mudança cultural nos anos 1970, mas que retornou recentemente. Esse retorno conduziu ao que chamo de comparação fortuita, ou seja, uma comparação baseada na epistemologia interpretativa, que reflete sobre o processo de pesquisa e define o objeto de estudo com base na saliência. Argumento que esse retorno da abordagem comparativa é resultado, pelo menos na Europa, de diferentes tipos de relações entre antropólogos, bem como de novas relações entre antropólogos e outros cientistas sociais. Essas novas posições revigoraram o debate sobre a comparação, que, na minha opinião, tem a ver, pelo menos em parte, com o fim da Guerra Fria na Europa.

Tabela 1
Epistemologias de comparação

| Tipo de comparação | Epistemologia | Conceitualização do processo de pesquisa | Relação entre objeto de estudo e sociedade |
|---------------------------|----------------------|--|--|
| Comparação controlada | Objetivista | Controle sobre o processo de pesquisa, a linguagem e o pesquisador. | Representatividade define a relação entre o estudo e a sociedade |
| Comparação fortuita | Interpretativa | Reflexiva sobre o processo de pesquisa, a linguagem e o pesquisador. | Saliência define a relação entre o estudo e a sociedade |

Uma comparação hierárquica

Ao afirmar que as culturas têm caráter, o estudo etnográfico intitulado *Balinese character*, de Mead e Bateson (1942), forneceu importante contribuição para a abordagem da cultura e da personalidade na antropologia. Como era comum na antropologia da época,

as culturas eram entendidas como entidades distintas, ou seja, havia limites claros entre as culturas e os valores, os hábitos e as pessoas que delas faziam parte. O caráter de uma cultura – argumentavam Mead e Bateson (1942) – correspondia à média das personalidades individuais existentes nessa cultura.

Sua abordagem de comparação tornou-se um exemplo emblemático dos motivos pelos quais a comparação pode ser problemática.

Em *Balinese character*,⁶ ancora-se firmemente a análise no uso da fotografia e do cinema que, na época, ainda eram métodos novos de pesquisa. A análise dos autores fundamenta-se amplamente em imagens de pessoas e suas interações. A fotografia e o filme permitem que Mead e Bateson examinem atentamente as posturas corporais, expressões faciais e emoções. Ao justificar o uso da fotografia para identificar o caráter balinês, eles escrevem:

O leitor é assim apresentado a declarações científicas de vários e diferentes graus de objetividade e generalidade. Cada fotografia pode ser considerada quase puramente objetiva [...] enquanto as legendas detalhadas contêm uma mistura de descrição objetiva e generalização científica. (Ibidem, 1942, p. 53)

De um ponto de vista objetivista, eles comparam Bali e os Estados Unidos de duas maneiras distintas, ambas com implicações teóricas de longo alcance. Primeiro, a pesquisa sobre o caráter balinês foi comparada à esquizofrenia (JACKNIS, 1988, p. 161), ou seja, uma condição médica grave e doença mental nos Estados Unidos foi usada como critério para medir o caráter balinês. Mead e Bateson (1942) argumentam que a natureza contraditória dos balineses é evidenciada em *rame* (barulhento, feliz e superpovoado) e *njepi* (silêncio, vazio, distanciamento), no sentido de que eles podem passar rapidamente de um a outro desses dois estados mentais. Além disso, o *rame* e o *njepi* são fundamentais para as interações sociais, a vida nas ruas, os rituais (como o transe), as emoções das pessoas, os estilos de conversa e a criação de filhos. Em segundo lugar, Mead e Bateson (Ibidem) comparam os EUA e Bali

com relação à aprendizagem e à socialização. Eles argumentam que os balineses não aprendem por meio da linguagem, mas pelo movimento e pelo corpo; trata-se, em outras palavras, de uma aprendizagem cinestésica. Essa distinção entre os norte-americanos e os balineses é posicionada hierarquicamente: “Enquanto os americanos rabiscam num pedaço de papel para lidar com as palavras de uma palestra, os balineses rabiscam com palavras, ao mesmo tempo em que seus corpos lidam, impecável e rapidamente, com as tarefas que lhes cabem” (MEAD; BATESON, 1942, p. 15). Na mesma seção, eles escrevem: “Os balineses não aprendem virtualmente nada com a instrução verbal e a maioria dos adultos [...] é incapaz de atender a três ordens consecutivas, o que consideramos como sinal de uma inteligência normal aos três anos de idade” (Ibidem).

Essas comparações produzem uma hierarquia em que certa doença mental americana inspirou a análise do caráter balinês e na qual as habilidades de aprendizagem das crianças americanas foram usadas para descrever como os balineses lidam com a linguagem. Assim, a comparação objetivista que Mead e Bateson adotam produz uma hierarquia clara, sendo a sociedade norte-americana apresentada como superior. Eles acreditavam que a fotografia e o cinema permitiam graus mais altos de objetividade, aumentando o controle sobre os resultados da pesquisa, eliminando a influência do pesquisador e permitindo que os colegas verificassem os fatos e fizessem suas próprias interpretações. Pelo menos é esse o argumento dos autores. No entanto, alguns estudiosos questionaram se a pesquisa dos dois é tão objetiva quanto sustentavam, afirmando que as imagens produzidas por eles foram mais manipuladas do que gostariam de admitir.⁶ Além disso, Nordholt (1993) revelou que o descrito por Mead e Bateson

6 Ver Jacknis (1988), entre outros, sobre as instruções dadas por Mead durante as filmagens. Ver também as críticas expressas por Freeman (1983) sobre a pesquisa de Mead em Samoa e o debate que se seguiu (EMBER, 1985; MARSHALL, 1993).

como caráter balinês reproduziu, na prática, um estereótipo que se perpetuou durante os séculos XIX e XX, período em que Bali esteve sob o domínio colonial holandês.

A análise de Mead e Bateson (1942) mostra como a fotografia e o cinema geram epistemologia objetivista, que enfatiza o controle de descobertas e interpretações. Essa epistemologia, por sua vez, cria uma comparação hierárquica, sendo as hierarquias estabelecidas entre a sociedade do pesquisador e a do pesquisado, bem como entre o poder colonial (Holanda) e a colônia (as Índias Orientais Holandesas das quais Bali fazia parte). Esse tipo de comparação hierárquica e objetivista representou uma abordagem desenvolvimentista das sociedades e uma característica central da teoria da modernização. Essas abordagens pressupunham que algumas sociedades eram mais desenvolvidas do que outras e que as sociedades evoluiriam de acordo com linhas mais ou menos semelhantes. As sociedades ocidentais eram vistas como o padrão de avaliação das outras sociedades, o que vale dizer que as diferenças humanas eram colocadas em relações hierárquicas (FABIAN, 1983).

Reflexividade pós-colonial

A relação epistemológica entre objetivismo e hierarquia é particularmente digna de nota quando se comparam duas publicações de Geertz (1962, 1973). Ambas examinam o uso do dinheiro nas relações pessoais e são relevantes porque mostram como uma comparação desenvolvimentista e hierárquica das sociedades deu lugar a uma descrição reflexiva e construtivista, deixando pouco espaço para comparação. Permitam-me demonstrar isso comparando “The Rotating Credit Associations: a ‘middle-rung’ in development”

(Idem, 1962) com *A interpretação das culturas* (Idem, 1973), especialmente o capítulo, muito conhecido, intitulado “Um jogo absorvente: notas sobre briga de galo balinesa”. A razão para selecionar esses dois trabalhos é que eles sugerem como as abordagens comparativas estão relacionadas a abordagens epistemológicas. Em suma, os dois estudos mostram que, quando a epistemologia objetivista deixou a etnografia, o mesmo ocorreu com a comparação.

Embora ambos os estudos examinem como o dinheiro é parte das relações sociais, a análise é muito diferente. No artigo “Middle-Rung”, Geertz (1962) defende uma visão desenvolvimentista, ao comparar a vida rural tradicional, caracterizada pela troca de presentes, com a vida urbana moderna, caracterizada por mercados monetários.⁷ Ele alega que as Associações de Crédito Rotativo facilitam a transição rural-urbana porque combinam elementos tradicionais e elementos modernos da sociedade, contribuindo, assim, para que a população rural, em rápida urbanização, se ajuste à sociedade urbana e monetarizada:

Há um movimento em direção a uma instituição cada vez mais formalista, impessoal e especificamente econômica, fato que se reflete na importância decrescente dos elementos ritualísticos de reforço da solidariedade, na crescente preocupação com a probabilidade financeira dos membros e líderes e na aplicabilidade legal das obrigações, e no desenvolvimento de padrões mais complexos de organização e cálculo comercial. (Ibidem, p. 261)

A Associação de Crédito Rotativo, argumenta Geertz, “é, portanto, um mecanismo de socialização” (GEERTZ, 1962; p. 260).

Onze anos depois, Geertz publicou “Um jogo absorvente” (Idem, 1973) – contribuição seminal para a mudança cultural na antropologia. Nesse capítulo de *A interpretação das*

7 Ver também Geertz (1963).

culturas, o dinheiro e as relações sociais foram abordados a partir de uma posição epistemológica muito diferente, segundo a qual, argumenta ele, o dinheiro não é característico de cálculos utilitaristas, referindo-se antes ao estabelecimento de relações entre as pessoas e os grupos aos quais pertencem. Geertz mostrou como as apostas feitas pelos homens nas brigas de galo precisavam ser interpretadas simbolicamente, ou seja, o dinheiro e as brigas de galo eram símbolos a serem interpretados. Com isso, o autor demonstrava que o dinheiro desempenha papel importante na articulação da masculinidade e na criação de hierarquias entre os homens e os grupos aos quais pertencem.⁸

Geertz introduziu o termo “descrição densa” como característica-chave de uma epistemologia interpretativa e construtivista. Segundo ele, o conhecimento é produzido dentro de redes específicas de interações das quais o etnógrafo faz parte. Nesse estudo, Geertz refletiu cuidadosamente sobre seu papel como pesquisador, a respeito da posição que manteve em relação às pessoas com quem se envolveu durante o trabalho de campo e das situações que serviram de base às interpretações. Em vez de comparar a “sua” sociedade com a sociedade “deles”, Geertz afirmou que as pessoas constroem significado juntas. Em vez de apresentar a si mesmo e sua sociedade como superiores, descreveu como era visto como um convidado engraçado, que se tornou parte de uma rede de interpretações: ele era um visitante que interpretava o que os balineses estavam interpretando.

A questão do controle é igualmente importante, embora não seja abordada de forma tão explícita quanto à reflexividade. Uma das características da pesquisa etnográfica é que o antropólogo não está realmente no controle

da situação de pesquisa, da linguagem usada, das pessoas atendidas ou das conversas. O baixo nível de controle sobre as situações de pesquisa significa que a etnografia se torna um espaço no qual encontros, conversas e ideias inesperados e fortuitos podem emergir.

A natureza aleatória do trabalho de campo é muito presente no estudo de Geertz quando ele descreve sua reunião com os balineses com os quais realiza seu trabalho de campo. O fato de Bali ter se tornado independente em 1949 parece ter colocado o pesquisador em uma posicionalidade diferente. Geertz começa o estudo mostrando sua chegada na condição de visitante. Descreve como tenta, sem muito sucesso, estabelecer relações com os balineses. Percebe que ele e sua esposa são ignorados: “Para eles, e até certo ponto para nós mesmos, éramos não-pessoas, espectros, homens invisíveis” (GEERTZ, 1973, p. 56). O caráter casual de sua pesquisa etnográfica se torna pungente ao relatar como sua relação com os balineses – e, portanto, sua compreensão – muda muito drasticamente. Geertz assistia a uma briga de galos ilegal quando, inesperadamente, a polícia chegou. Os homens fugiram da polícia, e o autor, espontaneamente, decidiu fugir também e se esconder junto com o resto da multidão. Esse momento foi crucial – ainda que imprevisto – e moldou suas relações de trabalho de campo. Geertz (Idem, p. 57) escreve: “Estávamos literalmente ‘dentro’”. Em nítido contraste com Mead e Bateson (1942), Geertz não sugere que estava no controle da pesquisa, das pessoas que estudou ou dos eventos de que tomava parte. Tampouco se propôs a demonstrar que a briga de galos era um caso representativo que caracterizava a sociedade balinesa. Em vez disso, mostra como a briga de galo, especialmente

8 Ver Ortner (1984) sobre como esta abordagem simbólica difere da de Turner. Geralmente, Geertz é considerado o pai fundador da reflexividade na etnografia, mas a questão da posicionalidade do pesquisador também foi desenvolvida por Gluckman (1950) e pela Escola de Manchester, da qual Turner fazia parte.

as apostas entre os homens, ofereceu visões importantes sobre a sociabilidade do dinheiro e o estabelecimento de hierarquias e limites.

Posteriormente, Geertz (1988, p. 131-132) explicou como essa mudança epistemológica na antropologia se relacionava à descolonização:

O fim do colonialismo alterou radicalmente a natureza da relação social entre aqueles que perguntam e aqueles que são inquiridos e observados. O declínio da fé no fato bruto, o estabelecimento de procedimentos e o conhecimento não situado nas ciências humanas e, na realidade, na erudição em geral, alteraram não menos radicalmente a concepção dos que perguntam e dos que observam sobre o que eles estavam tentando fazer. O imperialismo em sua forma clássica, metrópoles e possessões, e o cientificismo em seus impulsos e bolas de bilhar, caíram mais ou menos na mesma época.

Essa “dupla queda do cientificismo e do imperialismo” (CONQUERGOOD, 1991, p. 179) impactou a comparação etnográfica, que havia sido tão firmemente incorporada a uma epistemologia objetivista. Kuper (1994, p. 541) escreve: “A nova antropologia cultural não aspirou comparar nem explicar”. Ao contrário, examinou a posicionalidade, a reflexividade e o poder e a política da representação na escrita etnográfica.⁹ A epistemologia interpretativa tornou a comparação mais problemática porque desafiou a ideia de que culturas, sociedades e significados pudessem ser claramente distinguidos uns dos outros e colocados em uma relação hierárquica.¹⁰ Quando é controlada, a comparação sugere controle sobre o processo de pesquisa e a relação entre o que é estudado e a sociedade

em geral. A pesquisa de Mead e Bateson e os primeiros textos de Geertz sobre dinheiro também foram produzidos como parte da política do colonialismo destinada a controlar as pessoas estudadas pelos etnógrafos.¹¹

Essa reflexividade também é fundamental para uma abordagem ontológica de comparação que desestabiliza ainda mais as categorias usadas para comparação. *O gênero da dívida*, de Strathern (1988) compara gênero e poder na Melanésia com “o Ocidente” e, ao fazê-lo, tenta expor a parcialidade e a natureza fragmentada presentes quando se comparam lugares e momentos diferentes:

Tudo o que fiz foi explicitar essas comparações culturais implícitas que estão implicadas na justaposição acidental de usar uma língua como meio para revelar a forma que outra, comparável, poderia assumir. No entanto, isso também faz com que a comparabilidade desapareça. (Ibidem, p. 343)

O estudo foi criticado por reificar as diferenças entre “eles” e “nós” e por reconhecer insuficientemente como o mundo dos habitantes das Terras Altas da Papua-Nova Guiné se integrava a uma economia capitalista global.¹² Ao mesmo tempo, o livro se tornou base importante para uma abordagem ontológica de comparação que estava enraizada na epistemologia fenomenológica. O trabalho de Starthern centrou-se nos povos indígenas que estavam muito distantes da sociedade “ocidental”, pelo menos da forma como ela os examinou. Porém, em vez de comparar duas sociedades, seu objetivo era revelar a multiplicidade e a pluralidade da comparação e elaborar sobre a

9 Sobre a mudança cultural, ver, entre outros, Clifford e Marcus (1986) e Ortner (1984).

10 Ver também Kuper (1994), Holy e Stuchlik (1987) e Herzfeld (2001).

11 Said (1978), Bhabha (1990) e Spivak (1988) demonstraram como é fundamental examinar criticamente a produção de conhecimento nas relações de poder colonial. Não sugiro aqui que todas as abordagens comparativas tenham levado às hierarquias descritas neste artigo: Franz Boas, para mencionar apenas um exemplo, decidiu comparar sem criar tais hierarquias.

12 Ver, entre outros, Keesing (1992).

natureza fragmentada e instável das categorias usadas para a confrontação. A abordagem da comparação também é central no trabalho de Viveiros de Castro (2004, p. 4), para quem “toda cultura é um processo gigantesco e multidimensional de comparação”.

A abordagem ontológica de comparação levanta questões um pouco semelhantes às tratadas pela contribuição de Geertz para a mudança cultural. Assim como no caso de Geertz, há reflexão e destaque da posicionalidade. Além disso, do mesmo modo que o estudo de Geertz, a abordagem não se propôs a explicar a diferença ou, como Viveiros de Castro (2004, p. 5) formulou, “a antropologia compara no intuito de traduzir e não de explicar, justificar, generalizar, interpretar, contextualizar, revelar o inconsciente, dizer o que é óbvio, e assim por diante”.

Essa abordagem desestabiliza o próprio conceito de comparação por meio do processo de identificação de contrastes e perspectivas e de reconhecimento da pluralidade – Strathern (1992) usa o termo pós-pluralidade – do mundo. A exploração de contratos e perspectivas revela a multiplicidade da própria realidade. Em vez de colocar a questão explicativa “por que” no centro do palco, ela se propõe a entender o que Mol (1999, 2014) chamou de “a política do que”. Mol (2014) explica essa abordagem ontológica da seguinte maneira:

A ideia era de que não há apenas muitas maneiras de conhecer “um objeto”, mas sim muitas maneiras de praticá-lo. Cada forma de praticar estágios realiza, faz, encena uma versão diferente do “objeto”. Portanto, não se trata de “um objeto”, e sim de mais de um. Um objeto múltiplo... Para sublinhar nossa ruptura com essa herança monorealista do monoteísmo, importamos um termo filosófico antiquado, ontologia, e o colocamos no plural. Ontologias. Isso foi, na época, um oxímoro inédito.

Ainda mais do que Geertz e a abordagem construtivista para a qual ele contribuiu de forma tão notável, a abordagem ontológica enfraquece a autoridade que está implícita em identificar diferenças e contrastes e, ao fazê-lo, se afasta da comparação como uma forma de explicação e se aproxima da comparação como um método para tradução.¹³

Explicando reflexividades disciplinares e transdisciplinares

O debate recente leva a comparação a novas direções que colocam, mais uma vez, a pergunta “por que” no centro das atenções, voltando a *explicar* as diferenças, não apenas para traduzi-las ou revelá-las. Ao mesmo tempo, essas abordagens não fazem o caminho de volta a uma epistemologia objetivista ingênua, e sim continuam a explicar e a comparar reflexiva e posicionalmente. Lazar (2012), van der Veer (2016) e outros autores analisam reflexivamente as categorias usadas para comparação, mas, mesmo quando usam essas categorias, estabelecem, simultaneamente, as diferenças entre países, sociedades, culturas e eventos.

Antes de explorar as recentes contribuições para a antropologia comparada, convém apontar de que maneira as novas formas de reflexividade tornaram-se essenciais para a antropologia, especialmente – mas não exclusivamente – para os antropólogos que trabalham na Europa. Parece que a reflexividade entre os antropólogos estimulou as novas interpretações comparativas. Ela ocorre dentro da disciplina quando os antropólogos se comparam (chamo isso de reflexividade disciplinar) e entre disciplinas quando os antropólogos comparam seu trabalho com o de outros cientistas sociais

13 Ver também Strathern (1992) e Holbraad e Pedersen (2009). Para uma interpretação crítica do método ontológico, ver Bessire e Bond (2014).

(chamo isso de reflexividade transdisciplinar). Essas duas reflexões não são completamente novas, mas parecem ter se tornado mais importantes. Permitam que eu explique primeiro como e por que as reflexividades disciplinar e transdisciplinar tornaram-se mais cruciais para a disciplina e que, em seguida, examine as recentes contribuições intelectuais trazidas por elas à antropologia comparada. Aqui, o foco principal será nos desenvolvimentos ocorridos na Europa, o que obviamente não significa dizer que eles são exclusivamente europeus, mas que essas recentes contribuições para a antropologia comparada preocupam sobretudo os antropólogos que trabalham – mas não necessariamente – na Europa.

Parece que as reflexividades disciplinar e transdisciplinar se tornaram mais relevantes após o fim da Guerra Fria. Durante esse período, os etnógrafos geralmente se reuniam nacionalmente ou em conferências internacionais realizadas nos Estados Unidos. O fim da Guerra Fria em 1989 alterou essa realidade, acarretando consequências significativas para as redes e os debates entre os etnógrafos que moldaram o que a etnografia é ou deveria ser. Isso estimulou novos tipos de reflexividade. A Associação Europeia de Antropologia Social (European Association of Social Anthropology – EASA) foi criada em 1989. Suas conferências ofereceram nova plataforma ao intercâmbio intelectual, principalmente entre os antropólogos que atuam na Europa. Dois anos depois, criou-se o Grupo África-Europa para Estudos Interdisciplinares (Africa-Europe Group for Interdisciplinary Studies – AEGIS), oferecendo uma nova arena para discussões entre antropólogos, sociólogos, cientistas políticos e outros. A fundação da Universidade Centro-Europeia, em Budapeste (Hungria), em 1991, com um departamento internacional

de antropologia e sociologia, e o estabelecimento do respeitado Max-Planck-Institut für Ethnologische Forschung, em 1999, em Halle, na antiga Alemanha Oriental, também foram iniciativas importantes.

Essas novas associações e instituições europeias fortaleceram as redes entre os antropólogos que trabalham na Europa e estimularam novos tipos de reflexividade. Os acadêmicos estavam envolvidos de forma mais intensa com diferentes abordagens etnográficas nacionais em toda a Europa, o que incentivou os antropólogos a refletirem sobre como as tradições nacionais influenciavam aquilo que constitui a etnografia. Essas instituições e eventos aumentaram a consciência da diversidade na antropologia na Europa e no Sul global. Essas diferenças são exploradas no volume *Other people's anthropologies: ethnographic practice on the margins* (BOŠKOVIĆ, 2008). O livro traz contribuições sobre a antropologia na Europa (Rússia, Holanda, Bulgária, Iugoslávia e Noruega) e em outras partes do mundo (Quênia, Turquia, Argentina, Camarões, Japão e Brasil), revelando uma riqueza de abordagens que não têm a ver com as dominantes francesa e anglo-americana, mostrando que essas últimas não podem reivindicar uma universalidade sobre o que definiria uma boa etnografia. Bošković e Eriksen (2008) enfatizam a importância de examinar a diversidade não apenas entre os povos que os etnógrafos estudam, mas a diversidade dentro de sua própria disciplina, incluindo definições e abordagens da etnografia.¹⁴ Em vez de encarar as disciplinas antropológicas de forma hierárquica – que colocam os EUA, o Reino Unido e, até certo ponto, a França na liderança –, as abordagens antropológicas foram vistas como construções históricas e sociopolíticas. Discutiram-se as diferenças nacionais de forma reflexiva pelas

14 Uma recente contribuição ao debate foi dada por Barrera-González, Heintz e Horolets (2017).

associações europeias, pelos institutos de pesquisa e por publicações como a de Bošković (2008), evitando a criação de hierarquias entre o Norte e o Sul, o Oriente e o Ocidente, e o francês/anglo-americano *versus* o resto. A comparação disciplinar implicou em refletir sobre as diversas abordagens etnográficas nacionais e tentar explicar as diferenças entre elas.

As reflexividades disciplinar e transdisciplinar foram estimuladas pelo avanço de uma agenda de pesquisa europeia destinada a fortalecer os desenvolvimentos científicos e incentivar a cooperação por meio de fronteiras nacionais e disciplinares. O Conselho de Pesquisa Europeu (European Research Council – ERC), fundado em 2007, também contribuiu para esse tipo de reflexividade. Os instrumentos de financiamento do ERC exigem, em geral, que os pesquisadores criem e liderem uma equipe de pesquisa para receber financiamento. Em vez de os etnógrafos formularem projetos individuais de pesquisa, a equipe tinha de trabalhar de forma colaborativa, e o valor agregado dessa colaboração precisava ser especificado. Qual foi o valor agregado de se elaborar um projeto envolvendo, por exemplo, quatro etnografias em quatro projetos etnográficos menores, financiados individualmente? Isso inevitavelmente levantou a questão de se comparar as descobertas etnográficas e refletir sobre como os antropólogos individuais realizam pesquisas de maneiras diferentes. O financiamento do ERC, assim como outros financiamentos para grandes projetos, encorajou os antropólogos a levantar a questão: como comparar as etnografias e refletir sobre as práticas etnográficas?¹⁵

A questão comparativa também se tornou mais relevante como resultado do processo de seleção de bolsas de pesquisa com

verbas elevadas. As propostas são normalmente revisadas e selecionadas por painéis interdisciplinares, nos quais muitas ciências sociais estavam representadas (LAZAR, 2013; MEYER, 2017; VAN DER VEER, 2014, p. 11). Os acadêmicos que conduziam as revisões e os processos de seleção nem sempre estavam familiarizados com a epistemologia em que a etnografia vinha se baseando desde a década de 1970 ou, às vezes, eram até hostis a uma epistemologia que não estava enraizada na tradição objetivista. Por causa disso, para haver êxito na obtenção de financiamento e garantir que a etnografia continuasse sendo um método legítimo, os antropólogos tiveram de dar respostas a perguntas do tipo “como você compara a evidência etnográfica?” ou “como você pode comparar diferentes países sem padronizar métodos?”. Os requisitos de financiamento implicavam que os antropólogos tinham de defender sua epistemologia, tornando sua abordagem comparativa mais explícita e convincente. É por esse motivo que a pergunta “por que”, que destaca as explicações, tornou-se mais importante.

Comparação fortuita

As reflexividades disciplinar e transdisciplinar estimularam a formulação da seguinte pergunta: como comparar as evidências etnográficas para explicar e discutir a questão “por que”, que não era central na mudança cultural e ontológica, mas de maneira reflexiva?

Lazar (2012) indaga como pode comparar seu trabalho de campo a respeito dos sindicatos na Bolívia e na Argentina, e que contribuiu essa comparação seria capaz de oferecer. Como os sindicatos e a relação entre sindicatos

15 Ao mesmo tempo, seria excessivamente determinista sugerir que as mudanças políticas europeias ou nacionais foram a única razão pela qual foi conferida mais atenção à etnografia comparada. Interesses, contribuições individuais, assim como tradições e redes de pesquisa também foram extremamente relevantes.

e política nacional podem ser mais bem compreendidos quando se equiparam esses dois países? A autora propõe distinguir dois tipos de comparação: a representativa e a disjuntiva. Ela destaca que ambos os tipos têm valor, mas produzem percepções diferentes que precisam ser avaliadas, recorrendo-se a critérios diferentes. Lazar define a comparação representativa como aquela mais tradicional, que confronta casos, pessoas e eventos representativos. De acordo com essa modalidade, as descobertas podem ser analisadas lado a lado. Já a comparação disjuntiva reflete sobre o papel do pesquisador que faz o trabalho comparativo e, portanto, reconhece que a compreensão é inevitavelmente um processo humano. Segundo ela, o critério não é a representatividade de casos ou pessoas, mas, antes, a oportunidade de encontrar diferenças inesperadas que exortam os pesquisadores a fazer novos tipos de perguntas e estabelecer um novo vocabulário para analisar essas diferenças. O valor da comparação disjuntiva fica evidente quando Lazar examina seu próprio material etnográfico, revelando claramente como uma comparação disjuntiva entre a Bolívia e a Argentina oferece *insights* sobre a relação entre organização coletiva e cidadania. Dessa forma, esse método abriu novos caminhos para explorar teoricamente a construção da individualidade política (LAZAR, 2012, p. 361). Portanto, a comparação disjuntiva situa-se nas experiências, ideias e surpresas da fortuidade etnográfica.

Inspiradas na abordagem ontológica de Strathern, as contribuições da edição especial da *Social Analysis*, intitulada “Cortando e conectando: perspectivas ‘afrinesianas’ sobre redes, relacionalidade e intercâmbio”, examinam como a etnografia melanésia e a antropologia africana influenciaram-se mutuamente e como os *insights* desenvolvidos numa área podem lançar luz sobre as práticas em outra área (MYHRE, 2013a). Elas abordam temas diversos, que variam desde o destrinchamento da carne e de como ela é compartilhada

(MYHRE, 2013b) até a comparação de doenças potencialmente fatais transmitidas pelo corpo humano (NIEHAUS, 2013), a socialidade, a personalidade (RODIMA-TAYLOR, 2013; VOKES, 2013) e, ainda, aos conceitos de direitos humanos e de redes (ENGLUND; YARROW, 2013)

Englund e Yarrow (2013) propõem uma etnografia comparativa que reconstitua o que eles definem como “produção de conhecimento relacional”, segundo a qual os etnógrafos têm de refletir sobre como suas teorias e conceitos são estabelecidos dentro de contextos específicos. Para eles, comparar significa incluir o etnógrafo na comparação. Sua abordagem comparativa leva em conta a natureza fortuita do trabalho de campo e os *insights* que ele gera. As comparações etnográficas, argumentam os autores, são valiosas porque “desestabilizam e complicam os termos [usados para comparação], o que por sua vez fornece a base para uma linguagem descritiva mais afinada às especificidades de contextos particulares” (ENGLUND; YARROW, 2013, p. 144).

Candea (2016) distingue entre o que chama de comparação frontal e de comparação lateral. Aqui também a reflexividade guia a distinção conceitual. Ele resume a frontal como aquela que compara “nós” e “eles”, enquanto a comparação lateral confronta “eles” e “eles”. O autor entende que ambas desempenham papéis importantes e destaca que eles podem se reforçar mutuamente e fortalecer a teoria etnográfica. Tanto a comparação frontal quanto a lateral podem ser reflexivas, argumenta Candea, mas são reflexividades diferentes, pois o pesquisador está posicionado de forma distinta na comparação frontal e na lateral: “Na comparação frontal, nós nos colocamos à prova. Na comparação lateral, testamos um ao outro. Os dois movimentos são mutuamente constitutivos” (Ibidem, p. 218).

O livro de van der Veer, *The Value of Comparison* (2016), critica a epistemologia

dominante na maioria das ciências sociais (economia, sociologia, ciência política). Ele contesta as alegações universais frequentes de cientistas sociais, afirmando que essas assertivas são baseadas em modelos ocidentais específicos de cognição, estrutura e racionalidade. Quando sociólogos, economistas e outros cientistas sociais fazem afirmações universais sobre como a sociedade, a cultura e o comportamento humano funcionam, ignoram o fato de que seus estudos são de pessoas que vivem no Ocidente, principalmente nos Estados Unidos. Acreditam que esses modelos sejam universais porque ignoram as diferenças que existem no mundo e assumem que as sociedades e as pessoas do Ocidente são representativas de todo o mundo. Isso acarreta consequências de longo alcance para comparação. Van der Veer salienta que “uma das maiores falhas no desenvolvimento de uma perspectiva comparativa parece ser a comparação quase universal de qualquer sociedade existente com uma modernidade euro-americana ideal-típica e totalmente autossuficiente” (VAN DER VEER, 2016, p. 28). Isso não quer dizer que os conceitos desenvolvidos no Ocidente não possam ser usados em um método comparativo, e o próprio van der Veer se compromete criticamente com os conceitos weberianos e maussianos. No entanto, isso implica que essas ideias e modelos precisam ser tratados de forma reflexiva, reconhecendo que são produzidos dentro de um contexto específico. A comparação só pode funcionar quando se percebe que uma afirmação universal é problemática e que sempre envolve uma compreensão histórica de como as ideias e os modelos ocidentais fazem parte de um processo de globalização, carregados, portanto, de relações globais de poder.

Van der Veer (2016) destaca os modelos universais usados na ciência cognitiva, entre os quais a metodologia individualista da escolha racional, que define uma abordagem

a-histórica e quantitativa da sociologia, da economia e da ciência política. Essas não refletem sobre o fato de que o vocabulário de comparação está enraizado em um terreno de pesquisa muito específico – ou seja, as sociedades ocidentais, particularmente os Estados Unidos. Ele argumenta que a adoção de um método comparativo é crucial para a compreensão completa de como e por que as sociedades diferem umas das outras. Contudo, a diferença só pode ser entendida se refletirmos sobre os valores da sociedade em que esses conceitos emergiram.

O livro apresenta uma agenda detalhada para a comparação reflexiva, formulando quatro vantagens da abordagem antropológica para a comparação, sendo elas: a compreensão completa do problema da tradução cultural da diferença; a capacidade de estudar um fragmento que não é modelo para um todo maior, mas que lança luz sobre ele; um generalismo que reconhece como as nações e civilizações são historicamente integradas; e uma compreensão do corpo e de seu disciplinamento que vai além da noção cognitiva limitada (Ibidem, p. 10). Essas quatro vantagens das exigências antropológicas representam uma crítica a certa comparação universalista que não considera suficientemente o problema da tradução cultural, que falha ao problematizar o individualismo metodológico e o universalismo cognitivo, e falha igualmente ao examinar como as nações e civilizações são integradas. Esses quatro temas percorrem as três seções do livro sobre “O fragmento e o todo”, “Civilização e comparação” e “Comparando a exclusão”.

Embora não considere a etnografia como o único método de comparação, van der Veer (Ibidem) argumenta que a etnografia pode dar contribuições específicas, porque se trata de um método bem adequado para examinar e compreender a relação entre o fragmento e o todo. Para ele, o fragmento, o cotidiano ou o “micro” não é estudado na etnografia porque

se acredita que representa o todo. O autor considera o fragmento importante porque só pode ser compreendido em relação a mudanças históricas maiores, locais, nacionais, regionais e, às vezes, globais. O oposto também é verdadeiro: a mudança histórica em grande escala só pode ser entendida por meio de evidências etnográficas detalhadas de situações particulares (VAN DER VEER, 2016, p. 25).¹⁶ Ao mesmo tempo, o autor enfoca a reflexividade durante o trabalho de campo etnográfico de forma menos exclusiva, pois é de opinião que uma abordagem comparativa não preocupa tanto a posicionalidade do etnógrafo quanto à posição das ideias e conceitos que os acadêmicos desenvolvem. Sua reflexão sobre o processo de pesquisa representa uma importante contribuição para a comparação, por incluir a posicionalidade do pesquisador e o uso de linguagem conceitual, e por mostrar como a relação entre o que é estudado e o que isso significa é definida pela saliência.

Conclusão

Com o tempo, surgiram diferentes tipos de reflexividade, que levaram a questões específicas sobre o que e como comparar e se há, realmente, necessidade de se comparar. A questão comparativa nunca desapareceu de todo, mas voltou com mais vigor na antropologia. Por várias décadas, a mudança cultural na comparação antropológica permaneceu sob suspeita, sendo dada relativamente pouca atenção à comparação etnográfica. A comparação remetia à falta de reflexividade e à incapacidade em entender como conceitos e ideias nela usados estavam situados e como, portanto, precisavam ser examinados. Também

significava assumir que a comparação havia levado, inevitavelmente, a uma hierarquia de culturas que fazia parte do colonialismo. Desde a década de 1990, a abordagem ontológica implicava que a comparação era uma maneira de traduzir e entender “a política do que”. A comparação revelou a pós-pluralidade de mundos, mas não representou uma tentativa de explicação, colocando a questão do “porquê” no centro da análise.

Dois tipos de reflexividade provocaram o retorno da questão comparativa, a qual se propunha a explicar, em vez de desconstruir. Tanto a reflexividade disciplinar, na qual os antropólogos comparam-se entre si, quanto a reflexividade transdisciplinar, na qual os antropólogos comparam-se com outros cientistas sociais, os estimularam a abordar a questão comparativa de novas maneiras. Na Europa, essas formas de reflexividade foram, em certa medida, resultado da Guerra Fria e promoveram novas conversações entre antropólogos e os forçaram a se engajarem ou a se defenderem de uma epistemologia positivista que caracteriza grande parte das ciências sociais.

Publicações recentes sobre a comparação na etnografia revelam as mudanças observadas nos contextos políticos em que os etnógrafos realizam suas pesquisas, a reflexividade que acompanha esses contextos e os objetivos que os movem para explicar as diferenças. Tais comparações reconhecem que a linguagem usada para descrever e analisar depende de relações de poder, que são globais e ocorrem entre pesquisadores e seus “sujeitos”. Ela reconhece que uma comparação objetivista é baseada no controle do processo de pesquisa, na nossa posição como pesquisadores na sociedade, nas perguntas que formulamos, nas respostas

16 Ver também Kapferer (2015) sobre o estudo de eventos. Ele se envolve criticamente com o trabalho de Gluckman, Deleuze e Guatarri para examinar como a etnografia pode nos ajudar a entender a ruptura, o conflito e a fragmentação na vida cotidiana, que constituem, por conseguinte, fontes de mudança social que transcendem esse evento específico.

que recebemos e, ainda, na seleção de quem estudamos. A comparação controlada implica que, embora o pesquisador esteja no comando, há menos espaço para entender a diferença, pois cabe aos pesquisadores e à linguagem conceitual que eles usam definir quais diferenças são percebidas e definidas como relevantes. Além disso, às vezes parece que o pesquisador não é tão responsável quanto as metodologias sugerem, o que leva a comparação controlada a tornar-se um véu que esconde o fato de que as práticas reais de pesquisa são sempre casuais, caóticas e aleatórias.

O baixo nível de controle sobre os “sujeitos” da pesquisa (meu desconforto com o termo “sujeitos” é porque ele sugere subordinação), a linguagem e as situações sociais são vitais para a nova teoria. Ao mesmo tempo, conforme o debate mostra, a falta de controle é contraintuitiva para um discurso metodológico que sugere que a comparação só pode ser feita quando se está no comando.

A comparação controlada tenta dirigir não só o ambiente de pesquisa, mas também

a linguagem e as categorias usadas para reunir e analisar os resultados, e a relação entre o que se estuda (uma amostra, um caso representativo) e o todo maior para aquilo que está sendo considerado. Esse tipo de comparação não explora suficientemente a falta de controle porque uma epistemologia objetivista o equipara a uma boa pesquisa. A mudança cultural e ontológica na antropologia removeu ou transformou a abordagem comparativa a tal ponto que a etnografia não mais se propôs a explicar a diferença. Hoje vemos o surgimento de uma comparação fortuita na etnografia, em que *insights* e explicações são possibilitados por um controle limitado sobre o ambiente de pesquisa, a linguagem e as categorias usadas para reunir e analisar os resultados e a relação entre o que se estuda (o fragmento, o caso representativo) e o todo maior que se constituem mutuamente. Ao mesmo tempo, o objetivo, agora, é explicar a diferença. A comparação fortuita coloca, assim, o inesperado e o caótico no centro de como o etnógrafo explica as diferenças.

Referências

- BARRERA-GONZÁLEZ, A.; HEINTZ, M.; HOROLETS, A. *European anthropologies*. New York: Berghahn, 2017.
- BESSIRE, L.; BOND, D. Ontological anthropology and the deferral of critique. *American Ethnologist*, Hoboken, v. 41, n. 3, p. 440-456, 2014.
- BHABA, H. K. *Nation and narration*. London: Routledge, 1990.
- BOŠKOVIĆ, A. (Ed.). *Other people's anthropologies: ethnographic practice on the margins*. New York: Berghahn, 2008.
- BOŠKOVIĆ, A.; ERIKSEN, T. H. Introduction: other people's anthropologies. In: BOŠKOVIĆ, A. (Ed.). *Other people's anthropologies: ethnographic practice on the margins*. New York: Berghahn Books, 2008. p. 1-19.
- BOURDIEU, P. et al. *The weight of the world: social suffering in contemporary society*. Cambridge: Polity, 2002.
- BOURDIEU, P.; WACQUANT, L. *An invitation to reflexive sociology*. Chicago: University of Chicago Press, 1992.
- CANDEA, M. On two modalities of comparison in social anthropology. *L'Homme*, Paris, v. 218, p. 183-218, 2016.

- CLIFFORD, J.; MARCUS, G. (Eds.). *Writing culture: the poetics and politics of ethnography*. Berkeley: University of California Press, 1986.
- CONQUERGOOD, D. Rethinking ethnography: towards a critical cultural politics. *Communication Monographs*, Abingdon, v. 58, n. 2, p. 179-194, 1991.
- DAS, V. Wittgenstein and anthropology. *Annual Review of Anthropology*, Palo Alto, v. 27, p. 171-195, 1998.
- DAS, V. et al. *Violence and subjectivity*. Berkeley: University of California Press, 2000.
- DOUGLAS, M. *Purity and danger: an analysis of concepts of pollution and taboo*. London: Routledge; Boston: Kegan Paull, 1966.
- _____. *Risk and blame: essays in cultural theory*. London: Routledge, 1994.
- EMBER, M. Evidence and science in ethnography: reflections on the freeman-mead controversy. *American Anthropologist*, Hoboken, v. 87, n. 4, p. 906-910, 1985.
- ENGLUND, H.; YARROW, T. The place of theory: rights, networks, and ethnographic comparison. *Social Analysis*, New York, v. 57, n. 3, p. 132-149, 2013.
- FABIAN, J. *Time and the other: how anthropology made its object*. New York: Columbia University Press, 1983.
- FIORAMONTI, L. *Gross domestic problem: the politics behind the world's most powerful number*. London: Zed Books, 2013.
- FREEMAN, D. *Margaret Mead and Samoa: the making and unmaking of an anthropological myth*. Canberra: Australian National University Press, 1983.
- GEERTZ, C. The Rotating Credit Association: a "middle rung" in development. *Economic Development and Cultural Change*, Chicago, v. 10, p. 241-263, 1962.
- _____. (Ed.). *Old societies and new states: the quest for modernity in Asia and Africa*. New York: Free Press, 1963.
- _____. *The interpretation of cultures: selected essays*. New York: Basic Books, 1973.
- _____. *Works and lives: the anthropologist as author*. New York: Basic Books, 1988.
- GLUCKMAN, M. The kingdom of the Zulu of South Africa. In: FORTES, M.; EVANS-PRITCHARD, E. E. (Eds.). *African political systems*. Oxford: Oxford University Press, 1950. p. 25-55.
- HACKING, I. *The taming of change*. Cambridge: Cambridge University Press, 1990.
- HANDLER, R. The uses of incommensurability in anthropology. In: FELSKI, R.; FRIEDMAN, S. S. (Eds.). *Comparison: theories, approaches, uses*. Baltimore: The John Hopkins University Press, 2013. p. 271-191.

- HASTRUP, K. *A passage to anthropology: between experience and theory*. London: Routledge, 1995.
- HERZFELD, M. Performing comparisons: ethnography, globetrotting, and the spaces of social knowledge. *Journal of Anthropological Research*, Chicago, v. 57, n. 3, p. 259-276, 2001.
- HOLBRAAD, M.; PEDERSEN, M. A. Planet M: the intense abstraction of Marilyn Strathern. *Anthropological Theory*, Thousand Oaks, v. 9, n. 4, p. 371-394, 2009.
- HOLY, L.; STUHLICK, M. *Actions, norms and representations: foundations of anthropological enquiry*. Cambridge: Cambridge University Press, 1983.
- JACKNIS, I. Margaret Mead and Gregory Bateson in Bali: their use of photography and film. *Cultural Anthropology*, Washington, v. 3, n. 2, p. 160-177, 1988.
- KAPFERER, B. Introduction. In: MEINERT, L.; KAPFERER, B. (Eds.). *In the event: toward an anthropology of generic moments*. New York: Berghahn Books, 2015. p. 1-28.
- KEESING, R. M. Book review of Marilyn Strathern, "The Gender of the Gift: Problems with Women and Problems with Society in Melanesia". *Pacific Studies*, Laie, v. 15, n. 1, p. 129-137, 1992.
- KUPER, A. Culture, identity and the project of a cosmopolitan anthropology. *Royal Anthropological Institute of Great Britain and Ireland*, London, v. 29, n. 3, p. 537-554, 1994.
- LAZAR, S. Disjunctive comparison: citizenship and trade unionism in Bolivia and Argentina. *Journal of the Royal Anthropological Institute*, London, v. 18, n. 2, p. 349-368, 2012.
- MARSHALL, M. The wizard from Oz meets the wicked witch of the East: Freeman, Mead, and ethnographic authority. *American Ethnologist*, Hoboken, v. 20, n. 3, p. 604-615, 1993.
- MEAD, M.; BATESON, G. *Balinese character: a photographic analysis*. New York: New York Academy of Sciences, 1942.
- MEYER, B. Comparison as critique. *Hau: Journal of Ethnographic Theory*, London, v. 7, n. 1, p. 509-515, 2017.
- MOL, A. A reader's guide to the 'ontological turn': part 4. *Somatosphere*, 19 mar. 2014.
- MYRHE, K. C. Introduction. *Social Analysis*, New York, v. 62, n. 4, p. 1-24, 2013a.
- _____. Membering and dismembering: the poetry and relationality of animal bodies in Kilimanjaro. *Social Analysis*, New York, v. 62, n. 4, p. 114-131, 2013b.
- NIEHAUS, I. Kuru, aids, and witchcraft: reconfiguring culpability in Melanesia and Africa. *Social Analysis*, New York, v. 62, n. 4, p. 25-41, 2013.
- NORDHOLT, H. S. Leadership and the limits of political control: a Balinese "response" to Clifford Geertz. *Social Anthropology*, Hoboken, v. 13, p. 291-307, 1993.
- ORTNER, S. Theory in anthropology since the sixties. *Comparative Studies in Society and History*, Cambridge, v. 26, p. 126-166, 1984.

- RODIMA-TAYLOR, D. Gathering up mutual help: relational freedoms of Tanzanian market-women. *Social Analysis*, New York, v. 57, n. 3, p. 76-94, 2013.
- SAID, E. *Orientalism*. New York: Pantheon, 1978.
- SPIVAK, G. C. *Can the subaltern speak?* Basingstoke: Macmillan, 1988.
- STRATHERN, M. *The gender of the gift: problems with women and problems with society in Melanesia*. Berkeley: University of California Press, 1988.
- _____. Parts and wholes: refiguring relationships in a postplural world. In: KUPER, A. (Ed.). *Conceptualizing society*. London: Routledge, 1992. p. 75-106.
- VAN DER VEER, P. *The value of comparison*. Durham: Duke University Press, 2016.
- VIVEIROS DE CASTRO, E. Perspectival anthropology and the method of controlled equivocation in Tipití. *Tipití: Journal of the Society for the Anthropology of Lowland South America*, [S.l.], v. 2, n. 1, p. 1-19, 2014.
- VOKES, R. New Guinean models in the East African Highlands. *Social Analysis*, New York, v. 57, n. 3, p. 95-113, 2013.

Resumo

O retorno da comparação etnográfica: a epistemologia em momentos de mudança política

Esse estudo examina diferentes abordagens das comparações etnográficas e analisa debates recentes e antigos sobre como comparar evidências etnográficas, bem como mostra de que modo epistemologias comparativas mudaram durante momentos políticos específicos (colonialismo, descolonização e fim da Guerra Fria). Recentemente emergiram novas formas reflexivas de comparação com raízes na epistemologia interpretativa. O fim da Guerra Fria estimulou formas de comparação e reflexividade que deram surgimento ao que chamo de comparação por serendipidade: uma abordagem comparativa baseada em uma epistemologia interpretativa que abraça serendipidade, reflexividade e relevância como mais importantes que controle.

Palavras-chave: Etnografia; Método; Comparação; Epistemologia; Política; Reflexividade; Serendipidade; Guerra Fria; Pós-colonialismo.

Abstract

The return of ethnographic comparison: epistemology at moments of political change

This study examines different approaches to ethnographic comparisons and analyses recent as well as older debates on how to compare ethnographic evidence. It shows how comparative epistemologies have changed during specific political moments (colonialism, decolonization, and the end of the Cold War). Recently, new forms of reflexive comparison have emerged that are rooted in an interpretative epistemology. The end of the Cold War stimulated new forms of comparison and reflexivity that gave rise to what I call serendipitous comparison: a comparative approach that is based on an interpretative epistemology that embraces serendipity, reflexivity, and saliency as being more important than control.

Keywords: Ethnography; Method; Comparison; Epistemology; Politics; Reflexivity; Serendipity; Cold War; Postcolonialism.

Résumé

Le retour de la comparaison ethnographique: l'épistémologie dans les moments de changement politique

Cette étude examine différentes approches des comparaisons ethnographiques et analyse les débats récents et anciens sur la manière de comparer les preuves ethnographiques et montre comment les épistémologies comparatives ont changé au cours de certains moments politiques (colonialisme, décolonisation et fin de la Guerre Froide). De nouvelles formes réflexives de comparaison avec des racines en épistémologie interprétative ont récemment vu le jour. La fin de la Guerre Froide a stimulé des formes de comparaison et de réflexivité qui ont donné lieu à ce que j'appelle une comparaison de sérendipité: une approche comparative basée sur une épistémologie interprétative qui considère la sérendipité, la réflexivité et la pertinence comme plus importantes que le contrôle.

Mots-clés: Ethnographie; Méthode; Comparaison; Épistémologie; Politique; Réflexivité; Sérendipité; Guerre froide; Postcolonialisme.

O papel da ciência política no discurso científico brasileiro: institucionalização e avanços de uma área em construção

Emerson Urizzi Cervi¹

Introdução

O que distingue um campo científico cujo objeto é a sociedade e suas relações das demais formas de produção de conhecimento ou meio social, como o jornalismo, por exemplo, é a linguagem própria para transmissão do saber e o público a que se destina², que, por sua vez, é resultado das diferenças entre os métodos utilizados por jornalistas e por cientistas políticos na retratação da realidade (RAGIN, 1994). Este artigo tem o objetivo de discutir as principais características do processo de autonomização acadêmica e institucionalização da área de ciência política dentro do grande campo das ciências sociais³. O objeto de análise principal da ciência política – as relações de poder em sociedade – não é exclusivo dela, porém, a especificidade da linguagem que os politólogos usam para tratar da sociedade está na metodologia, nos métodos e nas distintas técnicas de abordagem e interpretação do objeto social. Nossa hipótese é que a identificação de estratégias metodológicas cada vez mais usadas por politólogos é o que permitiu a conformação do novo campo científico cerca de um século atrás.

O que distingue o discurso científico sobre a sociedade e suas relações de poder não é o seu objeto, mas sim a linguagem (a forma) e o público a que se destina o discurso científico. Da mesma maneira, existem diferenças linguístico-metodológicas internas aos campos científicos. Assim, o artigo busca indicar os limites da discussão sobre técnicas de análise e o quanto o avanço de um campo de conhecimento científico está vinculado, primeiramente, à capacidade de interpretar fenômenos estudados para os próprios integrantes da comunidade. Para isso, a abordagem é predominantemente cronológica, indicando as etapas pelas quais o discurso científico da política passou, no último século, que é quando se dá a institucionalização do campo autônomo de conhecimento da ciência política.

Os estudos sobre política nasceram na antiguidade tendo como objeto de análise as disputas pelo poder, se considerarmos os textos de Platão. Portanto, a identificação de um campo científico chamado de ciência política não constrange os estudos sobre política a esse campo. Assim como não se pode defender que todo discurso sobre as disputas de poder é dotado de características científicas. Ao longo

-
- 1 Professor associado do Departamento de Ciência Política da Universidade Federal do Paraná e professor permanente do Programa de Pós-Graduação em Ciência Política (PPGCP) e do Programa de Pós-Graduação em Comunicação da mesma universidade.
 - 2 Versão inicial do texto foi apresentada na mesa “Fronteiras do Debate Metodológico em Ciências Sociais”, no Encontro Anual da Associação Nacional de Programas de Pós-Graduação em Ciências Sociais (Anpocs), em 2017.
 - 3 O autor agradece as críticas, correções e contribuições dos pareceristas anônimos, que colaboraram decisivamente para a versão final do texto apontando melhorias possíveis.

de mais de 25 séculos tem sido discutido o papel dos atores políticos, das instituições e dos processos de produção, concentração ou distribuição de poder político nas sociedades, portanto, muito antes de surgir a ciência tal como a conhecemos hoje. A maior parte da contribuição ao conhecimento humano sobre as relações de poder é teórico-filosófica, ou seja, são contribuições não científicas, pois a ciência como um todo, como a conhecemos hoje, tem pouco mais de 400 anos. Já a ciência política conta com um século, no máximo, de existência como disciplina científica autônoma. Ela surge apenas quando os pesquisadores começam a dispor de instrumentos metodológicos e informações sobre a realidade empírica de diferentes fontes para interpretar relações de poder entre indivíduos e por meio de instituições (MONROE, 1997). Portanto, não se pretende, aqui, limitar todo o conhecimento sobre as disputas de poder social ao discurso científico politológico, mas demarcar cronológica e institucionalmente a estruturação de uma nova forma de tratar das relações de poder: a do discurso científico com ênfase na empiria.

Partimos do princípio de que, na origem da disciplina, há uma inversão na forma de produzir conhecimento sobre as disputas de poder. Ao invés de, primeiro, formular teorias e depois validá-las na realidade, a ciência política, desde sua institucionalização como campo de conhecimento, começa pelos dados empíricos, analisa-os e, a partir dos resultados obtidos, formula suas teorias. Daí a centralidade da descrição, comparação empírica, estudos de caso e validação técnica dos dados para a formulação de postulados científicos pelos politólogos. Vale ressaltar que a ênfase na relação entre ciência política e empiria não exclui outros campos do conhecimento. A ciência política não é a primeira a fazer esse caminho, existiram outras antes, inclusive em campos de conhecimento que forneceram à

própria área os instrumentos metodológicos interdisciplinares necessários para a análise empírica dos fenômenos políticos no início do século XX.

De toda forma, a inversão do processo de produção do conhecimento pela ciência política teve dois grandes efeitos. O primeiro é que a disciplina abre mão da formulação de teorias generalizantes, como se dá em outros campos do conhecimento, pois a realidade fática que ela observa em seus estudos está em constante transformação. Não é que a Ciência Política não possa produzir generalizações, mas que isso não está entre seus objetivos principais. Uma explicação sobre determinado fenômeno em um lugar pode ser diferente sobre o mesmo fenômeno em outro espaço geográfico ou, até mesmo, no mesmo espaço, mas em diferentes momentos do tempo. O que o cientista político busca é a identificação de padrões e comportamentos empíricos em determinado momento e espaço para reformar ou ajustar teorias já existentes. Em outras palavras, a ciência política tem como ponto de partida que os fatos existem na realidade independentemente da presença do pesquisador e que, ao serem explicados nos estudos científicos, produzem-se novos conhecimentos sobre como se dão as disputas de poder político. Trata-se de uma ciência que objetiva os fatos da realidade como princípio, mas que não é constrangida à simples descrição.

O outro efeito – e mais importante para o que nos interessa aqui – é o vínculo de origem entre ciência política e pesquisa empírica (HATEMI; BYRNE; MCDERMOTT, 2012). A inversão da ordem no processo de produção do conhecimento, partindo da realidade objetivada para as teorias, só é possível quando há centralidade de técnicas de pesquisa empírica, tanto pelas estratégias quantitativas quanto as qualitativas. No caso da ciência política isso aconteceu a partir de dois processos independentes e paralelos.

O primeiro foi a importação de métodos de pesquisa quantitativos e qualitativos⁴ de outros campos do conhecimento entre o final do século XIX e início do século XX. Sem as técnicas interdisciplinares, de disciplina fim – como psicologia social – ou disciplina meio – como estatística, por exemplo – seria mais lenta a consolidação da ciência política como campo próprio de conhecimento. Por outro lado, teria pouca validade a aplicação de técnicas interdisciplinares de pesquisa empírica se não houvesse, no campo da política, uma disponibilidade crescente de informações. Este segundo processo, acelerado no período pós-Segunda Guerra Mundial, é consequência da disponibilização de dados oficiais de governos e institutos privados sobre relações políticas. As técnicas avançadas de pesquisa empírica sobre fatos políticos teriam pouca importância se houvesse à disposição dos pesquisadores informações sobre a realidade desses fatos, os estudos sobre política continuariam sendo filosóficos e ensaístas conceituais sobre o tema. A discussão do processo de consolidação e do papel do debate metodológico interno ao campo de pesquisas será aprofundada a partir do próximo tópico. A seguir, o texto trata da institucionalização formal da ciência política no Brasil no primeiro quarto do século XXI e, depois, aborda-se o papel da metodologia na conformação da especificidade da politologia diante de outros campos de conhecimento sobre a sociedade.

Caracterização da área

A ciência política se tornou um campo de conhecimento e de pesquisas próprio, a partir

do momento que conseguiu instituir uma metodologia predominante para o tratamento dos objetos de pesquisa comuns a outras áreas, formou uma comunidade que compartilha uma linguagem própria para discutir e avançar nas pesquisas e conhecimentos produzidos (GOODIN; KLINGEMANN, 1996). É impossível localizar temporal e espacialmente o ponto seminal desse processo⁵. O que se pode dizer é que, desde o fim do século XIX, economistas, filósofos, sociólogos, juristas e psicólogos sociais já vinham estudando fenômenos típicos de suas áreas com novo olhar e uma tentativa de configuração de uma nova gramática – a da ciência política. Mas ainda faltavam dois elementos: a institucionalização formal e uma justificativa moral para a autonomização da ciência política como campo próprio de conhecimento. Os dois elementos coincidem com o final da primeira metade do século XX e são consequência direta da Segunda Guerra Mundial (ROSANVALLON, 2010). A Organização das Nações Unidas (ONU), logo em seu início, resolveu financiar pesquisas acadêmicas para identificar padrões culturais e o comportamento político do cidadão comum (tipo de estudo pouco usual à época), o que começou a promover a institucionalização do novo campo científico. Isso tinha como justificativa a necessidade de entender o cidadão comum nas sociedades contemporâneas a fim de prevenir uma nova onda de guerra total. A justificativa moral para a institucionalização da ciência política está ligada a uma militância em favor de valores democráticos ao mesmo tempo que se defende uma neutralidade em relação a valores normativos (ALCÁNTARA, 2017).

4 Não é objetivo, aqui, discutir especificidades de cada um desses conjuntos de técnicas de pesquisas. Para uma apresentação detalhada das características próprias e do potencial de complementaridade entre técnicas quantitativas e qualitativas, ver Günter (2006) e Mahoney (2006).

5 Um indicador da conformação de uma área autônoma institucionalizada é o surgimento de manuais específicos da área, como pode ser visto em Pinto (1995), Caminal (1996) e Vallés e Martí (2015).

São representantes dessa primeira geração David Easton e Gabriel Almond, que recebem financiamentos da ONU para iniciar suas pesquisas no novo campo de conhecimento. O primeiro produz um estudo sobre a organização inicial da ciência política (EASTON, 1953), enquanto o segundo dá início às estratégias metodológicas de uso de *surveys* para identificação de padrões de comportamento e cultura política (ALMOND, 1990).

Com a ciência política, as afirmações sobre a disputa de poder político passaram a ser sustentadas por dados do mundo empírico. Isso gera uma primeira necessidade – essencial para a autonomização do campo de conhecimento – que é a incorporação de diferentes técnicas de pesquisa empírica originadas em outras áreas, principalmente a estatística, com todos os conjuntos de técnicas produzidos por elas para análise de dados micro (individuais ou de grupos, como elites) ou macrodados (agregados geográficos, comportamento de massa ou efeito de ordem temporal, por exemplo). Assim, a ciência política nasce tendo como objeto central de análise o mundo empírico das disputas de poder (BARREDA; RODRÍGUEZ, 2016).

Mas, para que a autonomização do campo da ciência política se concretizasse no século XX, foi necessária a ocorrência de dois fatores, um interno e outro externo ao próprio campo, sem os quais não seria possível definir essa nova disciplina de estudos empíricos sobre disputas de poder político. O fator externo é a disponibilização de dados sobre governos e sociedade. Um campo de conhecimento que se define como originalmente ligado a estudos empíricos de determinada realidade só pode existir se houver informações disponíveis. No início do século

XX, agências públicas e instituições privadas começam a produzir e tornar públicas informações sobre governos, sociedades e países de maneira sistemática. Tal prática se intensifica após a Segunda Guerra Mundial, quando países democráticos estabeleceram políticas de maior visibilidade da ação de seus governantes visando evitar o surgimento de novos líderes demagogos. O objetivo desse tipo de política inédita, até então, na história era informar as populações para torná-las menos vulneráveis aos discursos de lideranças individuais (CASTELLS, 2009).

O segundo aspecto, de ordem interna, é a ideia de que os fatos existem na realidade, independentemente do que pensam os pesquisadores. Essa ideia permitiu que cientistas políticos assumissem a responsabilidade de acessar, descrever, relacionar e fazer inferências a partir de dados do mundo empírico, usando métodos de pesquisa até então pouco comuns nos demais campos de conhecimento anteriores à ciência política (COLOMER, 2009). Essa centralidade nos objetos empíricos rendeu aos cientistas políticos a “acusação” de serem positivistas, como se fosse uma deficiência intelectual (GERRING, 2006). Para o desenvolvimento da ciência política, mais relevante do que se aproximar ou não do positivismo⁶ clássico tem sido o uso indevido ou inadequado das técnicas de análise importadas de outros campos de conhecimento, sem as devidas ressalvas ou adaptações necessárias ao objeto da política (ALCÁNTARA, 1993). A excessiva especificação de modelos estatísticos impróprios para os objetos de análise empírica da política podem afastar a ciência política de sua origem, que é a realidade política e não os resultados matemáticos dos modelos explicativos. A transformação de dados da

6 Para mais informações sobre a crítica ao positivismo da pesquisa empírica na ciência política do século XX, ver debate a respeito do Círculo de Viena, a Thomas Kuhn e Karl Popper, em McEwen (1963) e Schlick (1978).

realidade em números para modelos estatísticos de análises não é um fim em si mesma, mas um meio para se descrever com mais precisão o mundo empírico que move os estudos de ciência política.

O objeto específico de atenção dos estudos sobre a política é o poder e o processo de tomada de decisões em função do poder. Trata-se de um campo de pesquisas que analisa principalmente o exercício, a distribuição e organização do poder em dada sociedade, país ou qualquer grupo humano (AVILÉS, 2010). Desde o final do século XIX, quando a ciência política começou a se distinguir das demais disciplinas, um dos pontos principais dessa diferenciação foi a adoção de técnicas de análise do mundo real pelos cientistas. Na ciência política do século XXI, se passou a fazer cada vez mais estudos a partir de hipóteses iniciais com objetivo de formular teorias, e não o contrário – partir das teorias já formuladas para testá-las na realidade. Assim, é possível simplesmente analisar determinados fatos, descrevendo-os à exaustão, para buscar respostas concretas sobre a realidade. Essa forma de produção de conhecimento é muito distinta do que conhecemos como filosofia ou teoria política (SKINNER, 2013).

Um dos pontos de partida da filosofia política é a chamada *clássica*, representada pelos greco-romanos, em especial por Platão (427 a 348 a.C.), Aristóteles (334 a 322 a.C.), Cícero (106 a 43 a.C.) e Plutarco (46 a 119 d.C.). No intervalo de 500 anos entre Platão e Plutarco, os pensadores tinham como principal objeto de atenção os atores políticos. Dava-se importância às características dos líderes, que eram capazes de moldar, reformar ou até mesmo refundar instituições políticas. Os períodos ditatoriais e as fases democráticas das cidades-estados da antiguidade seriam resultado de traços característicos dos líderes de cada momento. Pensar na formação de bons líderes era fundamental para manter

um sistema político saudável, defendia Platão, por exemplo. Alcántara (1993) lembra que os autores da modernidade, que recuperaram o pensamento antigo, deram continuidade na atenção aos indivíduos como agente político. É o caso de Erasmo de Roterdã (1466 a 1536) e Maquiavel (1469 a 1527), entre outros europeus da modernidade. Em comum eles têm o fato de considerarem que as instituições se adaptam aos líderes ou que elas podem ser usadas em favor dos interesses dos líderes. Mais importante do que isso é o fato de que nesse período da filosofia política, os líderes adaptam as experiências que têm na esfera privada para as práticas em espaços públicos ou estatais. A única formação específica para o líder, segundo os filósofos políticos desses dois períodos (da antiguidade à modernidade), é a oratória (ALCÁNTARA, 2012).

Para a filosofia política da primeira fase da modernidade o que importa é entender como determinadas características individuais conseguem promover o acúmulo de poder político. O líder é quem mobiliza as pessoas, e não as instituições. Isso vale tanto para os realistas, como Erasmo de Roterdã, quanto para os normativos, como Maquiavel. Cerca de um século depois os filósofos políticos começam a se preocupar com o papel das instituições nas explicações sobre o acúmulo de poder. É nesse período que ganha importância a análise do direito, das formas de alterar e estabilizar as regras institucionais e, pela primeira vez, discute-se o equilíbrio entre diferentes poderes como forma de garantir a manutenção das sociedades políticas (ALCÁNTARA, 1993). Destacam-se Thomas Hobbes (1588 a 1679), John Locke (1632 a 1704), Montesquieu (1689 a 1755), Jean Jacques Rousseau (1712 a 1778) e, mais tarde, Tocqueville (1805 a 1859). Ainda que com diferentes abordagens sobre o papel das instituições, todos esses pensadores deram mais importância à configuração das regras

e estruturas políticas⁷ do que aos próprios atores. Caberia aos líderes políticos aprenderem a atuar politicamente dentro das instituições para conseguirem representar seus interesses. Em comum entre os autores está a predominância das abordagens normativas/prescritivas em detrimento da descrição da realidade. A exceção é Tocqueville (1998), que se propõe a explicar o funcionamento da democracia após conhecer instituições democráticas em diferentes países. Mas isso pode ser explicado pela distância temporal. Tocqueville está mais próximo dos autores que começam a se preocupar com a descrição da realidade no século XIX do que dos clássicos do século XVII.

No final do século XIX e início do XX, já inseridos no processo de descolamento entre filosofia e ciência política, os autores voltam a colocar o agente político no centro das atenções. Destacam-se Gustave Le Bon (1841 a 1931), Vilfredo Pareto (1848 a 1923), Gaetano Mosca (1858 a 1941), Robert Michels (1876 a 1936), Ortega y Gasset (1883 a 1955), Walter Lippmann (1889 a 1974) e Charles Wright Mills (1916 a 1962) nesse tipo de abordagem. Para eles, no geral, a elite política é um grupo específico, porém, nem muito exclusivo, nem muito acessível. Os integrantes dessa elite precisam usar experiência, tradições e um código profissional próprio para se manter no grupo que consegue controlar as instituições e concentrar o poder político (ALCÁNTARA, 2012). Um ponto comum das pesquisas sobre política nesse período é a crescente atenção dada pelos autores aos aspectos do mundo empírico e a prática cada vez mais corrente de só propor uma teoria depois de ter testado as hipóteses em uma realidade objetiva.

O reconhecimento da centralidade do subcampo da metodologia para a ciência política é tão importante quanto a identificação de excessos gerados pelo uso de técnicas de pesquisa empírica importadas de outros campos para a descrição ou predição de objetos específicos da política. Uma crítica contundente a esses excessos foi feita por Giovanni Sartori em conferência à Universidade de Columbia, em 2004, sob o título “Where is political science going?”, publicada posteriormente pela revista *Política y Gobierno*. Para ele, enquanto a ciência política se sustenta apenas em modelos estatísticos, cada vez mais abstratos e distantes da realidade fática, será um gigante crescendo sobre pés de barro (SARTORI, 2004). No Brasil, uma boa crítica sobre o distanciamento da ciência política de seu objeto empírico específico em função de pouco conhecimento metodológico específico pode ser encontrada em Reis (1991).

Assim, não apenas o objeto inicial da ciência, mas também sua justificativa como campo autônomo de conhecimento, passam a ser, a partir de 1950, os avanços na democracia, a reconfiguração do público, a revalorização da sociedade como agente político e dos direitos universais como liberdade e igualdade (HUNTINGTON, 1996). Mais especificamente, análises de mecanismos de representação tais como eleições, partidos, instituições e comportamento político são os primeiros objetos da gramática própria da ciência política (REIS, 1978; SIMÃO, 1956; SOARES, 1961). Como disciplina acadêmica, a ciência política é tributária de disciplinas clássicas como filosofia, economia, direito, história, sociologia e psicologia, principalmente. O que a torna autônoma em relação às áreas

7 Aqui usamos o termo “estruturas políticas” para representar o conjunto de instituições específicas da arena política, tais como governo, parlamento, partidos políticos etc. e que são as responsáveis em última análise para a organização e funcionamento do corpo político na sociedade.

originárias é a construção de uma metodologia com vocação predominantemente empírica e o uso de instrumentos de medição dos fenômenos humanos menos centrais naquelas disciplinas. A ciência política pode ser definida como uma disciplina, a partir da segunda metade do século XX, que utiliza métodos concretos para interpretar comportamentos individuais e agregados em instituições especializadas na representação política dentro do Estado ou em instituições societais. As questões metodológicas são centrais para a área desde sua origem, como é possível perceber em Linz (1987), Ragin (1994) e Sartori (2002).

March e Olsen (1984) chamam atenção para cinco estilos de abordagem próprios da ciência política do século XX: (1) estilo contextual, interessado em integrar a política à sociedade; (2) reducionista, com uma visão dos comportamentos políticos individuais; (3) estilo utilitário, que trata a ação como o produto de interesse próprio; (4) estilo funcionalista, que se interessa com a história dos mecanismos eficazes para alcançar o equilíbrio; e (5) estilo instrumentalista, que define processos de tomada de decisão e apropriação de recursos como central para a vida política. As transformações pelas quais passou a sociedade, com evolução tecnológica, crises econômicas estruturais e ameaças à democracia, também resultaram em novos desafios para a ciência política (LIJPHART, 1989; HAN, 2012).

A nova maneira de produzir conhecimento sobre a disputa de poder político recebeu um grande estímulo no período pós-Segunda Guerra Mundial (AVILÉS, 2010). Foi quando as disputas pelo poder tornavam-se mais explícitas em escala mundial. Além disso, cresceu a escala de informação disponível sobre governos, o mundo experimentou a expansão da produção em massa de bens de consumo; houve inclusão em massa de faixas da sociedade no mercado urbano de consumo e de direitos políticos, com maior interação e

participação nas decisões que envolviam a política, em especial nas democracias ocidentais.

A aproximação entre discurso científico e dados da realidade para explicar as disputas de poder político cresce após a Segunda Guerra Mundial, quando o tema da sustentabilidade democrática ganha espaço. Cresce nos meios acadêmicos a ideia de estudar os agentes sociais como agentes políticos relevantes diante dos agentes estatais e dos agentes econômicos. Nesse período consolida-se a linha de pesquisas que considera a ação coletiva como uma das explicações necessárias para o fenômeno da disputa pelo poder, ou seja, a ação humana coletiva impulsionada por uma engenharia institucional que a valorize ganha espaço entre os cientistas políticos. O constructo “poliarquia”, de Robert Dahl (1961), é um dos exemplos resultantes das pesquisas desse período, já baseadas em fontes empíricas. Quando se pergunta “Quem Governa?”, Dahl (1961) vai buscar respostas na dinâmica política dos atores de uma pequena cidade dos Estados Unidos: New Haven. Ele usa dezenas de fontes de informações de mais de um século a respeito das disputas de poder, desigualdade e influências pessoais ou de grupos de interesse nas decisões políticas (DAHL, 1961). Mais ou menos no mesmo período, em 1975, na Europa, Susan Pharr e Robert Putnam (2000) recebem a encomenda de estudar a dinâmica política, em especial os desafios da democracia no século XX. Eles fazem isso a partir dos eixos *Estado, Economia e Sociedade*, usando dados disponibilizados por governos, aliados a informações de *surveys*. Pharr e Putnam (2000) percebem a relevância da cultura política para a explicação dos fenômenos de disputa de poder. Putnam deu continuidade a essa linha de pesquisa em estudo posterior apresentando uma tipologia para a cultura política em diferentes regiões da Itália (PUTNAM, 1996).

As ideias de participação ou isolamento em relação às práticas democráticas

institucionalizadas passam a ser estudadas, a partir de então, pela linha da cultura política (ALCÁNTARA, 2012). A informação sobre um crescimento na descrença à democracia coloca o indivíduo – não necessariamente o líder político – no centro da atenção dos cientistas políticos. É nesse contexto que Pharr e Putnam (2000) publicam trabalho sobre a crescente insatisfação com a democracia em países ricos, usando dados de *surveys* e outras fontes de países da Europa Central, Estados Unidos, Canadá e Japão (PHARR; PUTNAM, 2000). Outros estudos empíricos sobre política no mesmo período podem ser encontrados em Campbell et al. (1966), Fiorina (1981), Cain et al. (1987), Popkin (1991), Page e Shapiro (1992) e em Brennan e Lomasky (1993). No Brasil também foram produzidos estudos empíricos sobre a realidade política local já na metade do século XX (FIGUEIREDO, 1991; LAMOUNIER; CARDOSO, 1978; LAVAREDA, 1999; REIS, 1978; SIMÃO, 1956; SOARES, 1961, 2001). Em todos os casos parte-se de informações da realidade, que não estão sob o controle do pesquisador, para formular análises e, posteriormente, conclusões descritivas.

Além da disponibilidade de informações macrosociais de governos e países, nos anos 1960, a incorporação do behaviorismo e, principalmente, as técnicas de pesquisas comportamentalistas⁸ do campo da psicologia para o

das práticas políticas deram novo avanço à pesquisa empírica da ciência política (CHARNEY; ENGLISH, 2012). Desde então cientistas políticos vêm desenvolvendo instrumentos metodológicos tanto para pesquisas dos macrodados quanto com microdados⁹. Na década seguinte, nos anos 1970, a incorporação de técnicas quantitativas usadas principalmente em modelagens originais de áreas ligadas à economia, somadas aos métodos da psicologia social, permitiram o desenvolvimento de uma importante corrente de pesquisa da ciência política daquele período: a *rational choice*¹⁰. A teoria da escolha racional ainda é considerada o ápice da pesquisa empírica a partir de microdados e uma das experiências mais bem-sucedidas em produção de um arcabouço teórico a partir de análises do comportamento observado (CALVERT, 1995; MERRILL III; GROFMAN, 1999; OSTROM, 2005; SHEPSLE; BONCHEK, 1997; TSEBELIS, 2002). O avanço da ciência política se dá com a crescente disponibilização de dados de governos e de institutos privados e universitários de pesquisa individual, os *surveys*. Avilés (2010) lembra que, enquanto a filosofia política manteve-se, predominantemente e não de forma exclusiva, fazendo a narrativa dos fatos, o direito analisando a ação política em determinados marcos legais e a sociologia adaptando as grandes teorias às novas realidades sociais, a ciência política aprofundou-se na descrição e

8 O comportamentalismo é uma corrente teórica da psicologia social norte-americana com grande influência sobre as explicações de fenômenos sociais e políticos nas décadas de 1960 e 1970. Postula o estudo do comportamento social dos indivíduos como a melhor forma para explicar fenômenos sociais. Um dos principais representantes dessa escola para o uso das técnicas comportamentais nos estudos políticos foi Skinner (2003).

9 Adota-se, aqui, a definição de que microdados são aqueles que podem ser apreendidos diretamente dos respondentes às pesquisas, dizem respeito às características pessoais e, portanto, podem ser individualizadas até o nível microssoial. Já os macrodados são aqueles que dizem respeito a agregados de pessoas, portanto, não sendo possível, a partir deles, conhecer as preferências individuais. Eles dizem respeito ao comportamento agregado que, em última análise, não pode ser identificado em nenhum indivíduo isolado.

10 O texto não pretende fazer uma discussão exaustiva das características da *rational choice* como escola teórica aplicada à ciência política do século XX. A “novidade” dessa escola que nos interessa aqui é a consideração de que os agentes políticos atuam em favor de seus interesses, buscando a maximização de resultados em suas ações. Para uma crítica à aplicação dos conceitos de *rational choice* à ciência política, ver Green e Shapiro (1994).

verificação dos dados da realidade¹¹. Para tanto ela contou com a introdução de técnicas de pesquisa empírica interdisciplinares, que deram corpo à pesquisa com dados da realidade pelos cientistas políticos (MCEWEN, 1963).

Há quase três décadas o debate na ciência política se dá em termos epistemológicos e metodológicos. Trata-se de um tipo específico de conhecimento afetado por valores (ALMOND, 1990), contexto (NOHLEN, 2007) ou instituições. Para além desse debate epistemológico, se dá a sedimentação do campo de conhecimento em instituições formais, linhas de investigação, financiamento de pesquisas e redes colaborativas. A isso tudo se dá o nome de institucionalização da área, que acontece por etapas. No primeiro momento, o da instalação do campo de conhecimento autônomo, predomina a ortodoxia da abordagem. É o que Alcântara (2017) chama de a época de ouro da ciência política. Com um gap temporal de aproximadamente meio século, o processo de institucionalização da ciência política no Brasil ganha força na virada do século XX para o XXI.

No caso específico brasileiro¹², essa etapa dura praticamente todo o último terço do século XX, quando começa a se institucionalizar em programas acadêmicos organizados e desenvolvidos em ambientes universitários. Também houve um significativo crescimento no número de revistas acadêmicas especializadas na área no início do século XXI¹³. Entre

2010 e 2012 existiam 933 revistas avaliadas pela área de ciência política e relações internacionais na Capes. Dessas, 255 (24,11%) eram consideradas qualificadas, ou seja, com características adequadas a revistas acadêmicas. Já no período de 2013 a 2016 o número de títulos avaliados já era de 1.306, sendo 351 (26,8%) qualificadas¹⁴.

A pós-graduação, ao contrário das áreas tradicionais de conhecimento, é o espaço natural da institucionalização da ciência política no Brasil. O crescimento no número de programas de pós-graduação nessa área, no país, também é notável na última década. Entre 2008 e 2010 o Brasil tinha 2.380 programas de pós-graduação em todas as áreas e conhecimento. A ciência política apresentava 26 programas de pós-graduação no país (1,09% do total). No período seguinte, entre 2011 e 2013, a área da ciência política e relações internacionais já possuía 32 programas de pós-graduação. No entanto, a participação no total da pós-graduação brasileira caiu naquele período para 0,95%, pois o total de programas de pós-graduação no país tinha subido para 3.337. O mais recente período de avaliação (2013 a 2016) mostrou novo crescimento da área de ciência política acima da média nacional de pós-graduação. Eram 43 programas em ciência política e relações internacionais (1,02% do total de programas do país), embora maior participação que no período anterior, ainda abaixo da participação que tinha entre

11 Para não parecer que aqui se subestima a importância do behaviorismo na autonomização da área, sugiro a discussão realizada por John Gunnell (1988) a respeito do tema. Para eles, o behaviorismo teve papel fundamental para a consolidação do perfil empírico da ciência política, em especial nos Estados Unidos. O tema também é tratado por Feres (2000).

12 Para uma visão sobre o processo de institucionalização da ciência política em outros países da América Latina, ver Betancur e Mancebo (2017) e Relacip (2010).

13 Todas as informações quantitativas desta seção foram copiladas a partir do portal de avaliações quadriennais da Capes, disponível em: <<https://bit.ly/2vRDLmv>>.

14 O critério para avaliação de uma revista acadêmica por uma área da Capes é que pelo menos um professor de programa de pós-graduação credenciado à área tenha publicado um artigo no periódico durante o período de avaliação.

2008 e 2010. O total de programas de pós-graduação brasileiros apresentou um número de 4.175. Na última década a institucionalização da pós-graduação como campo específico de conhecimento acompanhou o avanço no número de periódicos acadêmicos e de programas de pós-graduação do país.

Com a institucionalização, o passo seguinte é a abertura para uma pluralidade maior de temas de pesquisa e abordagens mais heterodoxas. Por exemplo, o tema das identidades, antes mais próximo das pesquisas em antropologia, passou a integrar a agenda de politólogos quando discutem novas formas de identificação política em substituição à tradicional identidade partidária (TANAKA; DARGENT, 2015). A crise econômica iniciada em 2008 no hemisfério norte também contribuiu para a apresentação de novos temas à agenda de pesquisas, como globalização, desigualdades regionais, protecionismo, xenofobia no discurso político, extremismos no debate público e – talvez o mais presente – novas formas de populismos (ALCÁNTARA, 2017)¹⁵.

Os dados – a matéria prima da ciência que tem como vocação a empiria, descrevem fatos – ocorrências, características institucionais a partir de representações simbólicas tanto

quanto qualitativas. As fontes tradicionais desses dados no século passado eram os censos e as pesquisas amostrais que, em um estágio mais avançado, vinham sendo organizados em observatórios políticos, permitindo que as abordagens dos fenômenos passassem de diacrônicas para sincrônicas. Além da maior quantidade de dados disponíveis atualmente, há também novas dimensões analíticas, tais como as análises de sentimentos a partir de manifestações públicas em conversações de redes sociais on-line, além das explorações sobre as novas formas de ativismo político, inclusive aqueles que são declaradamente destinadas a confrontar instituições políticas tradicionais, como o WikiLeaks¹⁶.

A centralidade do subcampo metodologia, em especial as técnicas interdisciplinares de pesquisa, é considerada aqui o ponto de partida para a ciência política como campo de conhecimento próprio. Uma distinção detalhada entre ciência política moderna e os campos predecessores de análise política pode ser encontrada no verbete “Ciência Política”, de Bobbio (1998). Nesse sentido, o que caracteriza a adoção de técnicas de diferentes áreas é a fragmentação dos instrumentos metodológicos utilizados. Essa característica tem pontos positivos e negativos (KING,

15 Para além do campo das crises agudas, a evolução tecnológica gradual que alcança o ápice na virada do século XX para o XXI também trouxe novos desafios aos cientistas políticos. O aumento da capacidade de obtenção de informações, maior transparência de governos e disponibilidade de dados por instituições na rede mundial de computadores fez que a tradicional preocupação sobre o que fazer com a falta de informações fosse rapidamente substituída por o que fazer para selecionar os dados disponíveis (ALCÁNTARA, 2017). A internet e o Big Data são considerados a quarta revolução industrial e, como as anteriores (vapor, eletricidade e informática), também provocam mudanças estruturais nas relações políticas (GRAY et al., 2015; KITCHIN, 2014).

16 Também existem novos problemas a serem considerados. A captura de informações confiáveis é fundamental para qualquer disciplina com vocação empírica. *Data Mining* passa a ser ferramenta indispensável para os cientistas políticos do século XXI (QUAN-HAASE; MARTIN; MCCAY-PEET, 2015). Além disso, aquele primeiro objetivo da ciência política como disciplina autônoma, que era identificar, descrever e caracterizar o “demos”, empiricamente enfrenta um desafio adicional nos novos cenários de debate via internet. Não é tão simples distinguir o participante real do debate, o cidadão-usuário da internet, das ferramentas tecnológicas que replicam ou imitam debatedores. Os robôs, *trolls* e outros precisam ser separados do “demos” para uma adequada descrição do debate público digital. Mais que acumular informações, o importante é saber o que elas representam de fato e para que elas servem. Que fenômenos explicam, de fato? Assim, as estratégias metodológicas ganham ainda mais importância.

1991). A adaptação de técnicas de pesquisa de outras áreas para a ciência política oferece condições para a descrição de fenômenos até então pouco explorados. No entanto, é preciso cuidado com a excessiva especialização de técnicas, que podem resultar em uma inversão de valores: os instrumentos metodológicos passarem de meios para fins em si mesmos. Um exemplo desse risco é citado por King (1991), ao tratar dos modelos estatísticos aplicados à ciência política. Segundo o autor, esses modelos precisam ser usados para representar principalmente a essência da política, e não para dar centralidade aos mecanismos matemáticos de análise.

Os métodos quantitativos na ciência política

O objeto específico deste artigo, o subconjunto das estratégias de abordagem e técnicas de pesquisa empírica usadas na ciência política, chamado de métodos quantitativos, vem sendo discutido por metodólogos da área há mais de meio século (CARMINES; ZELLER, 1979; DAVIS, 1976; KENDALL; LAZARSELD, 1950; ROSENBERG, 1971), não sendo, portanto, um tema novo ou que possa ser esgotado aqui. Podemos definir a análise quantitativa como aquela que envolve uma coleção de dados ou informações representados por número e que indicam alguma característica do objeto de pesquisa. Esses números podem ser duração de períodos, quantidades, taxas ou escalas (GARWOOD, 2006). As quantidades podem ser medidas diretamente, por exemplo, quantidade de votos, ou de forma indireta, por exemplo, quantidade de adesão à democracia. Existem diferentes modos de

coleta dessas informações, e a maneira como isso acontece é importante, pois definirá a capacidade de extrapolação dos resultados em um estudo, sendo essa uma etapa fundamental para qualquer pesquisa empírica¹⁷.

Os dados usados em pesquisas empíricas podem ser coletados em seus ambientes naturais, sem nenhum tipo de controle, de toda população ou apenas a partir de uma amostra específica. As informações obtidas em ambientes naturais oferecem maior precisão e detalhamento. Também são os que requerem maior trabalho, pois o descontrole das fontes impede a validação de dados em muitos casos. Outra forma de coleta de informações é em ambientes semicontrolados, em que não há total domínio sobre todas as variáveis, mas o pesquisador é capaz de dar alguns estímulos específicos para avaliar respostas obtidas. A terceira forma de obter informações em pesquisas quantitativas é a partir de experiências realizadas em laboratório, totalmente controladas pelo pesquisador. Esse tipo de trabalho é raro em ciência política, pois a escala exigida, normalmente, em estudos sobre política é bem maior que a escala de trabalhos desenvolvidos em laboratórios (GARWOOD, 2006). A vantagem nesse tipo de coleta de dados é o total controle sobre todas as variáveis envolvidas. A desvantagem é a dificuldade em replicar os resultados nos ambientes sociais, fora dos laboratórios.

Nem sempre as reações e comportamentos registrados em ambientes totalmente controlados são replicados em sociedade. Conhecendo essa limitação, pesquisas recentes têm transferido as técnicas de experimentos de outras áreas para o campo da ciência política, em especial a partir de Druckman et al. (2011). No Brasil, os primeiros trabalhos publicados

17 Ainda que não seja objetivo do artigo apresentar a evolução do debate quanti-quali na ciência política, é preciso considerar sua relevância quando se trata de apresentar o subcampo da metodologia na ciência política. Sobre a discussão original a respeito do tema, sugiro King, Keohane, e Verba (1994) e Mahoney e Goertz (2006). Sobre como essa discussão vem sendo tratada no Brasil, ver Soares (2005) e Cano (2012).

a partir do conjunto de técnicas de experimentos controlados são os de Porto (2007), Turgeon e Rennó (2010) e Turgeon, Chaves e Wives (2014).

Mais importante que utilizar modelos matematicamente sofisticados é não perder de vista a necessidade de relacioná-los a uma teoria ou conceito da ciência política (KING, 1991). Sem isso, a descrição se distancia tanto da realidade que deixa de fazer sentido como explicação empírica e se aproxima de campos de conhecimento que não têm a finalidade de descrever as relações de poder político, tais como a estatística ou a matemática. Em pesquisas que usam métodos quantitativos, as características estudadas são chamadas de variáveis (dado que constantes não são capazes de explicar a ocorrência de fenômenos políticos, pois não variam) e representadas por números. As análises, técnicas e modelos são sempre a respeito dos números que representam as variáveis, e não sobre as características diretamente. Por isso pesquisadores precisam estar atentos para evitar que resultados de suas pesquisas fiquem restritos aos números/modelos estatísticos, que são meio de representação, e não o fenômeno político propriamente dito.

Uma das principais características da análise quantitativa é sua associação com uma epistemologia realista, partindo do princípio de que o conhecimento serve para aproximar as pessoas daquilo que existe no mundo empírico, conhecendo seus detalhes e, às vezes, as causas. Parte-se do princípio de que as coisas existem no mundo e que elas podem ser medidas e descritas cientificamente (GARWOOD, 2006). Dentre as técnicas quantitativas, as estatísticas

descritivas ocupam lugar de destaque na ciência política, pois são usadas para ilustrar e sumarizar “achados” da realidade. Também são capazes de identificar relações entre diferentes variáveis, como os diferentes coeficientes de correlação entre duas ou mais variáveis. Para além das descrições, ainda existem as técnicas inferenciais¹⁸, que permitem identificar efeitos distintos de diferentes intervenções. As mais comuns são a análise de variância, covariância e variância multivariada.

Seguindo o que propõe King (1991) há pelo menos dois grandes conjuntos de abordagens empíricas na ciência política. A primeira é o que chama de *inferência ecológica*, em que, a partir do método estocástico, se fazem inferências sobre dados individuais não observados. Aqui, espera-se que as explicações estejam no microcosmo político, em nível individual, que dificilmente pode ser apreendido diretamente em toda sua complexidade. Por isso a necessidade de inferir sobre dados não conhecidos. Por exemplo, a explicação da decisão de voto a partir do nível de escolaridade dos eleitores. Não é necessário conhecer a escolaridade de todos os eleitores para saber que existem padrões microsociais que relacionam o nível de escolaridade ao processo de decisão de voto. Evidentemente essa não é a única variável que define o voto, logo, é preciso considerar os efeitos de variações macro, porém, o que se busca aqui são os micro padrões gerais.

A segunda abordagem é chamada de estratégia de análise de *variação e autocorrelação espacial*, para quando se aplicam conceitos que valem para tudo que está presente em uma unidade espacial qualquer. Nesse caso,

18 A definição de inferência usada aqui é a de que se trata da parte da estatística que busca identificar características de uma população não conhecida a partir de uma amostra conhecida e representativa dessa população. Para tanto, é preciso respeitar os pressupostos para a inferência a partir de uma amostra. O principal deles é a distribuição normal (em forma de curva de Gauss) dos casos na amostra, pois só é possível avaliar a probabilidade de ocorrência de determinado fenômeno ao acaso se a distribuição observada não apresentar nenhuma tendência. Para mais detalhes, ver Epstein e King (2002).

a unidade macro é usada para explicar determinado fenômeno (EPSTEIN; KING, 2002). Na autocorrelação espacial, por exemplo, espera-se encontrar alguma associação entre os votos de eleitores que vivem em periferias de grandes cidades. Aqui, não se busca explicar motivações individuais e é preciso considerar que nem todos os moradores de determinada região votarão da mesma forma. Não se quer entender a decisão individual, mas sim os padrões gerais – macrossociais – relacionados a determinado território geográfico com características próprias. Assim como o tempo percorrido pode ser uma variável explicativa importante em processos políticos, por exemplo (MORETIN; TOLOI, 2004). Mudanças em resultados de mesmos processos ocorridos em momentos distintos podem ser conseqüências do acúmulo de experiência ou informações pelos atores políticos.

Sempre que se usam modelos probabilísticos para explicar fenômenos políticos, por exemplo, é preciso considerar que eles dependem da teoria inferencial. Portanto, não faz sentido a aplicação desses modelos em pesquisas nas quais se pretende promover uma descrição ou quando não há uma amostra representativa da população. A metodologia empírica passa a ser um problema para a ciência política quando pesquisas quantitativas apresentam menos de política e mais de estatística (KING, 1986).

Um processo estocástico é definido em estatística como aquele que tem origem não determinística, ou seja, baseado em eventos aleatórios. Em outras palavras, é um conjunto de variáveis aleatórias com as mesmas chances de ocorrências em um mesmo espaço probabilístico (MORETTIN; TOLOI, 2004). King (1991) oferece algumas sugestões para evitar excessos em pesquisas empíricas. Em primeiro lugar, é preciso dar atenção à teoria inferencial e só usá-la quando for necessária. Segundo, reconhecer que as modelagens estocásticas

são mais sofisticadas, dado que os fenômenos políticos acontecem ao longo de períodos temporais. E, em complemento, a estatística descritiva e os elementos gráficos devem ter preferência, pois oferecem mais substância às análises quantitativas dos fenômenos reais a serem descritos.

Como todo conjunto de técnicas, a pesquisa quantitativa tem seus pontos fortes e fracos. Seu principal ponto positivo é produzir conhecimento cumulativo sobre o mundo e o comportamento humano. O ponto fraco das técnicas quantitativas deve-se ao fato de que muitos dados disponíveis não necessariamente representam aquilo que o pesquisador pretende analisar (GARWOOD, 2006). De qualquer maneira, os métodos quantitativos ainda são a melhor alternativa quando o pesquisador dispõe de dados da realidade e pretende testar uma hipótese teórica a respeito das relações empíricas entre diferentes variáveis. Com esse conjunto de técnicas é possível formular questões e testá-las estatisticamente, ou seja, testar a validade estatística das relações entre as variáveis. Uma vez utilizadas técnicas estatísticas para testar associações empíricas, é possível, em alguns casos, generalizar os resultados de estudos feitos em amostras representativas.

King e Powell (2008), em trabalho com o sugestivo título *How to not lie without statistics*, afirmam que a meta do cientista político é descrever, explicar e às vezes contribuir na formulação, implementação e avaliação de resultados de políticas públicas. Para eles, não há oposição entre teoria e empiria, dado que quanto mais “criatividade teórica”, maior a possibilidade de validação empírica (duas décadas antes, em 1985, Garry King havia apresentado em forma de conferência a primeira parte dessa discussão, intitulada *How not lie with statistics: avoiding common mistakes in quantitative political science*). Outros autores que também discutem o papel da metodologia para a ciência política no mesmo período são Ragin (1994);

Babbie (2005) e Gerring (2006). Para eles, a principal diferença entre a ciência política e os campos de conhecimento anteriores é o fato de se derivar teoria de um conjunto de dados existentes e acessados empiricamente.

Os campos de conhecimento que deram origem à ciência política não contavam com instrumentos metodológicos de pesquisas empíricas tão apurados e, muitas vezes, precisavam começar as explicações pela criatividade teórica. Os cientistas políticos, em função de sua origem tardia, invertem o processo: dada a disponibilidade de instrumentos metodológicos de acesso à realidade empírica, a ciência política conclui o processo de produção do conhecimento com teoria. Um ponto que não pode ser desconsiderado nesse tipo de definição é que, uma vez utilizado determinado conjunto de dados empíricos para sustentar uma teoria, esses mesmos dados não podem validar a teoria, como acontece nos demais campos de conhecimento ligados à ciência política. Em vez de validar teorias com dados da realidade, cabe ao empirista ajustar as teorias existentes com novos dados ou com informações de outras fontes.

A pesquisa empírica relevante consegue identificar interações suficientemente fortes entre duas variáveis, normalmente chamadas de X e Y (KING; POWELL, 2008). Quando se busca explicação de dependência de uma sobre a outra, a variável Y é a dependente e a X é a independente ou explicativa. No entanto, nem sempre buscam-se as relações de dependência. Também é possível pesquisar outro tipo de associações entre as variáveis, portanto, nem sempre haverá uma variável dependente a ser explicada por outra. O pesquisador coleta informações suficientes da realidade para usá-las em testes estatísticos a fim de explorar as possíveis relações entre variáveis e associar essas relações com teorias já existentes. Com isso, consegue reduzir a complexidade da teoria a nível suficiente

para explicar os “achados” empíricos ou, de outra forma, ele pode encontrar novos fatos observáveis da realidade, coletar e analisar as informações para ajustar a teoria a eles. Afinal de contas, se concordamos que o que distingue a ciência política dos demais campos de conhecimento anteriores a ela é a afinidade ao mundo empírico como ponto de partida e se considerarmos que os atores, as instituições e as relações políticas estão em constante transformação, há um evidente antagonismo com as grandes teorias generalizantes que visam explicar todos os fenômenos.

Os argumentos apresentados aqui apontam que a forma como se faz pesquisa empírica distingue a ciência política das disciplinas que deram origem a ela. Nesse sentido, King e Powell (2008) apresentam duas metas principais para os estudos feitos a partir de dados da realidade sobre disputa de poder político. A primeira é a “sumarização de dados”. Examinar o que a realidade apresenta, organizar e descrever determinado contexto é um dos mais importantes objetivos de um cientista político. A segunda meta é “fazer inferências”. Além de todas as limitações impostas pela teoria inferencial, essa meta só é indicada para quando já se produziu a sumarização de informações sobre o objeto da pesquisa. A inferência é a parte da pesquisa científica que usa informações que podem ser acessadas pelo pesquisador para fazer afirmações sobre o que ele não conhece. A pesquisa inferencial vai além da simples descrição daquilo que existe e pode ser acessado pelo pesquisador. Ela também pode ser o tipo de pesquisa contrafactual, muito importante para identificar falhas nas explicações teóricas ou até mesmo a superação de teorias em função das transformações no mundo empírico. É o tipo de pesquisa que estuda fatos que podem existir em algum momento, dada uma situação hipotética. A inferência contrafactual é importante para a explicação dos fenômenos empíricos e precisa

ser mais usada nas pesquisas sobre política, pois permite predições a partir da realidade e não na forma ensaísta ou puramente especulativa, como ocorre em outros campos de conhecimento (KING; KEOHANE; VERBA, 2000).

Todo campo de conhecimento científico é dotado de método, alguns métodos são específicos e outros compartilhados por vários campos de conhecimento. O método é, por natureza, um subcampo de qualquer área científica. Não defendemos aqui que a centralidade do método é uma exclusividade da ciência política. Advoga-se que, dado o fato de autonomização da ciência política como campo específico de conhecimento sobre as disputas de poder ter ocorrido simultaneamente à adoção de técnicas interdisciplinares empíricas nos estudos sobre esse objeto, os cientistas políticos não podem abrir mão de pensar sobre as ferramentas metodológicas que usam (RAINER et al., 2012; SARTORI, 2004; SLIPAK, 2012). Elas não são um fim em si mesmo e não podem afastar o trabalho de pesquisa do mundo empírico, tornando resultados excessivamente abstratos ao dar relevância a modelos que representam a realidade ao invés da própria realidade representada pelos modelos.

Conclusões

A decisão de fornecer subsídios para “alimentar” uma ciência empírica que estuda as relações de poder foi eminentemente política. Analisar a democracia, suas características, pontos fortes e fracos é uma forma de empoderar os próprios sistemas democráticos, evitando possíveis refluxos em favor de modelos autoritários que, segundo a visão predominante dos anos 1940, foram os responsáveis pelos conflitos mundiais entre 1914 e 1945. A ciência política empiricamente dirigida ganha força nesse contexto. No mesmo período, em 1936, a Fundação Gallup realiza o primeiro

grande *survey* amostral em disputa eleitoral norte-americana antecipando os resultados da eleição presidencial daquele ano. Essa e outras experiências chamaram atenção para a força preditiva da pesquisa científica empiricamente dirigida.

No Brasil, em 1942, é fundado o Instituto Brasileiro de Opinião Pública e Estatística (Ibope) de pesquisas para fornecer informações de mercado à economia e aos governos, principalmente a partir de *surveys* amostrais. Em 1950 o Ibope começa a produzir periodicamente o chamado *Boletim das Classes Dirigentes*, com amostra de 600 entrevistas na capital do Brasil (na época Rio de Janeiro), sobre avaliação de governo, nível de satisfação da população e opiniões sobre políticas públicas. O financiamento de pesquisas sobre cultura política para comparar a “saúde da democracia” entre países começa poucas décadas depois, ampliando constantemente o alcance desse tipo de ferramenta a cada década. Em 1973 é realizada a primeira onda de *survey* do Eurobarometer; em 1983 estreou o Word Values Survey; em 1995 foi a vez do Latinobarómetro e, em 1999, o Afrobarómetro.

Ainda que a ciência política brasileira tenha começado sua institucionalização em meados da década de 1960, com a abertura dos Programas de Pós-Graduação em Ciência Política da Universidade Federal de Minas Gerais, em 1966, e no Instituto de Pesquisas Universitárias do Rio de Janeiro, em 1969, a multiplicação de programas ocorrerá com mais força apenas nas primeiras décadas do século XXI. Praticamente dobrarmos o número de programas de pós-graduação e de revistas especializadas nas duas primeiras décadas do século XXI.

Se por um lado a disponibilização de dados empíricos e a incorporação de técnicas de pesquisa interdisciplinares caracterizaram a ciência política tal como a conhecemos hoje, não se pode esquecer que ela só existe

como ciência em função de sua capacidade de analisar e interpretar dados da realidade. Manter equilíbrio entre o avanço das técnicas de modelagem estatística e conexão com a realidade empírica tem sido o desafio dos cientistas políticos na transição do século XX para o XXI. O aprofundamento no uso de técnicas estatísticas muito especializadas, em especial nos métodos quantitativos, pode afastar o cientista político de seu objeto de análise, esterilizando os resultados. O quadro piora quando se desconsidera a teoria estatística que sustenta os testes, o que pode resultar em análises que não representam a realidade e que, do ponto de vista técnico, desrespeitam princípios teóricos básicos. O exemplo mais comum é o uso de técnicas inferenciais em análises descritivas ou que utilizam dados de toda população – um equívoco banal que faz toda diferença para a interpretação dos resultados. Outro exemplo do descompasso entre princípios técnicos e objetivos científicos está no uso de técnicas de regressão próprias para séries temporais. A função original de uma análise de série temporal é a predição. Para isso, a modelagem técnica exige que alterações sazonais ou outros tipos de mudanças sejam normalizados. Só assim será possível fazer predição com alguma segurança sobre o que ocorrerá no futuro – ainda não conhecidos –, mas que dependem em alguma medida do que já aconteceu no passado. Pois bem, na maioria das vezes o que o cientista político quer ao fazer uma análise de série temporal não é um modelo adequado para prever o futuro, mas informações organizadas sobre o que aconteceu do passado até o presente. Assim, o cientista político tem mais interesse

nos fatores que alteraram a direção da curva ao longo do tempo do que nos que garantiriam a sua manutenção no futuro. Usar modelagens para neutralizar efeitos externos à passagem do tempo em uma análise de série temporal de opinião política, por exemplo, é desconsiderar o principal objetivo dessa ciência: descrever a realidade empírica tal como ela se deu.

Os cientistas políticos, do início do século XXI, estão prontos para evitar os excessos que enfraquecem a própria ciência política? E, para evitar o excesso da teoria prescritiva deslocada da realidade, como se faz em outros campos do conhecimento? Ou, por outro lado, evitar o excesso de modelagem estatística em busca de “números mágicos” que de distantes da realidade empírica não representam nada além de informações sobre si mesmos? O século XXI começa com uma cultura de produção e disponibilização de dados sobre política nunca vista antes. Os cientistas têm fontes empíricas para descrever e analisar a realidade das disputas de poder político. Agora é preciso dar um passo atrás para evitar os efeitos danosos dos ensaios teóricos e das modelagens altamente especializadas para consolidar a ciência política. As recomendações básicas para esse cenário são: (1) dar mais atenção à sumarização dos dados da realidade, em especial no uso de representações gráficas, quando o objetivo for descrever o fenômeno; e (2) se for o caso de fazer predição, lembrar da necessidade de respeitar os pressupostos teóricos da teoria inferencial. Descrição e inferência são dois objetivos distintos para quem faz pesquisa científica empírica. Não são necessariamente excludentes, mas também não precisam estar presentes em um mesmo estudo para torná-lo relevante.

Referências

ALCÁNTARA, M. Cuando hablamos de Ciencia Política, ¿de qué hablamos? *Revista Mexicana de Sociología*, Coyoacán, v. 93, n. 4, p. 147-178, 1993.

- _____. *El oficio de político*. Madrid: Tecnos, 2012.
- _____. La Ciencia Política en el primer cuarto del siglo XX. *Revista Boliviana de Ciencia Política*, La Paz, v. 1. n. 1, p. 7-23, 2017.
- ALMOND, G. *A discipline civeded, schools and sects in political science*. Newbury Park: Sage, 1990.
- AVILÉS, L. M. V. La ciencia política en las ciencias sociales. *Revista Reflexiones*, San Pedro de Montes de Oca, v. 1, n. 89, p. 179-183, 2010.
- BABBIE, E. *Métodos de pesquisas de survey*. Belo Horizonte: Editora UFMG, 2005.
- BARREDA, M.; RODRÍGUEZ, L. M. *Análisis de la política: enfoques y herramientas de la Ciencia Política*. Madrid: Huygens Editores, 2016.
- BETANCUR, N.; MANCEBO, M. E. ¿Cómo se forma a los científicos políticos en Iberoamérica? Análisis de las titulaciones de grado en Ciencia Política en doce países. *Revista Española de Ciencia Política*, Madrid, v. 43, p. 161-185, 2017.
- BOBBIO, N. Ciência Política. In: BOBBIO, N.; MATEUCCI, N.; PASQUINO, G. *Dicionário de política*. Brasília: Editora UnB, 1998.
- BRENNAN, G.; LOMASKY, L. *Democracy & decision: the pure theory of electoral preference*. Cambridge: Cambridge University Press, 1993.
- CAIN, B. et al. *The personal vote: constituency service and electoral independence*. Cambridge: Harvard University Press, 1987.
- CALVERT, R. Rational actors, equilibrium and social institutions. In: KNIGHT, J.; SENED, I. *Explaining social institutions*. Ann Arbor: Michigan University Press, 1995.
- CAMINAL, M. *Manual de ciencia política*. Madrid: Tecnos, 1996.
- CAMPBELL, A. et al. *Elections and the political order*. New York: John Wiley and Sons, 1966.
- CARMINES, E.; ZELLER, R. *Reliability and validity assessment*. California: Sage, 1979.
- CANO, I. Nas trincheiras do método: o ensino da metodologia das ciências sociais no Brasil. *Sociologias*, Porto Alegre, n. 31, p. 94-119, 2012.
- CASTELLS, M. *Comunicación y poder*. Madrid: Alianza, 2009.
- CHARNEY, E.; ENGLISH, W. Candidate gene and political behavior. *American Political Science Review*, Cambridge, v. 106, n. 1, p. 1-34, 2012.
- COLOMER, J. M. *Ciencia de la política*. Barcelona: Ariel, 2009.

- DAHL, R. A. *Who governs? Democracy and power in an American city*. New Haven: Yale University Press, 1961.
- DAVIS, J. A. *Levantamento de dados em sociologia: uma análise estatística elementar*. Rio de Janeiro: Zahar, 1976.
- DRUCKMAN, J. N.; BOLSEN, T. Framing, motivated reasoning and opinions about emergent technologies. *Journal of Communication*, Oxford, n. 61, p. 659-688, 2011.
- EASTON, D. *The political system: an inquiry into the state of political science*. New York: Knopf, 1953.
- EPSTEIN, L. KING, G. The rules of inference. *The University of Chicago Law Review*, Chicago, v. 69. n. 1, p. 1-209, 2002.
- FERES, J. J. Aprendendo com o erro dos outros: o que a história da ciência política americana tem para nos ensinar. *Revista de Sociologia & Política*, Curitiba, v. 15, p. 97-110, 2000.
- FIGUEIREDO, M. *A decisão do voto*. São Paulo: Sumaré, 1991.
- FIORINA, M. P. *Retrospective voting in American national elections*. New Haven: Yale University Press, 1981.
- GARWOOD, J. Quantitative Research. In: JUPP, V. *The Sage Dictionary of social research methods*. London: Sage, 2006.
- GERRING, J. *Social science methodology: a criterial framework*. Cambridge: Cambridge University press, 2006.
- GOODIN, R.; KLINGEMANN, H. K. Political science: the discipline. In: GOODIN, R.; KLINGEMANN, H. K. *A new handbook of political science*. Don Mills: Oxford University Press, 1996.
- GRAY, E. et al. Small Big Data: using multiple data-sets to explore unfolding social and economic change. *Big Data & Society*, Thousand Oaks, v. 2, n. 1, p. 1-6, 2015.
- GREEN, D. P.; SHAPIRO, I. *Pathologies of rational choice theory: a critique of applications in political science*. New Haven: Yale University Press, 1994.
- GUNNEL, J. American political science, liberalism, and the invention of political theory. *American Political Science Review*, Cambridge, v. 82, n. 1, p. 71-87, 1988.
- GÜNTHER, H. Pesquisa qualitativa versus pesquisa quantitativa: esta é a questão? *Psicologia: Teoria e Pesquisa*, Brasília, v. 22, n. 2, p. 201-210, 2006.
- HATEMI, P. K.; BYRNE, E.; MCDERMOTT, R. "What is a gene and why does it matter for political science? *Journal of Theoretical Politics*, Thousand Oaks, v. 24, n. (3), 2012 (p. 305- a 327, 2012).
- HAN, B. C. *La sociedad del cansancio*. Herder: Madrid, 2012.
- HUNTINGTON, S. *El orden político en las sociedades en cambio*. Barcelona: Planeta, 1996.
- KENDALL, P.; LAZARSFELD, P. F. Problem of survey analysis. In: MERTON, R.; LAZARSFELD, P. F. (Orgs.). *Continuities in social research*. New York: Free press, 1950.

- KING, G. How not to lie with statistic: avoiding common mistakes in quantitative political science. *American Journal of Political Science*, Washington, DC, v. 30, p. 666-687, 1986.
- _____. On political methodology. *Political Analysis*, Cambridge, UK, v. 2, p. 1-30, 1991.
- KING, G.; KEOHANE, R. O.; VERBA, S. *El diseño de la investigación social. La inferencia científica de los estudios cualitativos*. Madrid: Alianza Editorial, 2000.
- KING, G.; POWELL, E. N. *How to not lie without statistics*. Massachusetts: Harvard University Cambridge, 2008.
- KITCHIN, R. Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, Thousand Oaks, v. 1, n. 1, p. 1-12, 2014.
- LAMOUNIER, B.; CARDOSO, F. H. (Orgs.). *Os partidos e as eleições no Brasil*. Rio de Janeiro: Paz e Terra, 1978.
- LAVAREDA, A. *A democracia nas urnas: o processo partidário-eleitoral brasileiro 1945-1964*. Rio de Janeiro: Editora Revan, 1999.
- LIJPHART, A. *Democracia en las sociedades plurales: una investigación comparativa*. Buenos Aires: Grupo Editor Latinoamericano, 1989.
- LINZ, J. J. La quiebra de las democracias. Madrid: Alianza Universidad, 1987.
- MAHONEY, J. A tale of two cultures: contrasting quantitative and qualitative research. *Political Analysis*, Cambridge, UK, n. 14, p. 227-249, 2006.
- MARCH, J. G.; OLSEN, J. P. The new institutionalism: organizational factors in political life. *American Political Science Review*, Cambridge, v. 78, n. 3, p. 734-749, 1984.
- MCEWEN, William P. The problem of social-scientific knowledge. New York: Bedminster Press, 1963.
- MERRILL III, S.; GROFMAN, B. A unified theory of voting: directional and proximity spacial models. Cambridge: Cambridge University press, 1999.
- MONROE, K. R. (Ed.). *Contemporary empirical political theory*. Berkeley: University of California Press, 1997.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de séries temporais*. São Paulo: Edgard Blücher, 2004.
- NOHLEN, D. Instituciones políticas en su contexto: las virtudes del método comparado. Buenos Aires: Rubinzal-Culzoni Editores: Buenos Aires, 2007.
- OSTROM, E. *Understanding institutional diversity*. Princeton: Princeton University Press, 2005.
- PAGE, B.; SHAPIRO, R. The rational public. Chicago: University of Chicago Press, 1992.

- PHARR, S. J.; PUTNAM, R. D. *Disaffected democracies: what's troubling the trilateral countries?* Princeton: Princeton University Press, 2000.
- PINTO, J. *Introducción a la ciencia política*. Buenos Aires: Eudeba, 1995.
- POPKIN, S. L. *The reasoning voter: communication and persuasion in presidential campaigns*. Chicago: Chicago University Press, 1991.
- PUTNAM, R. D. *Comunidade e democracia: a experiência da Itália moderna*. São Paulo: FGV Editora, 1996.
- QUAN-HAASE, A.; MARTIN, K.; MCCAY-PEET, L. Networks of digital humanities scholars: the informational and social uses and gratifications of Twitter. *Big Data & Society*, Thousand Oaks, v. 2, n. 1, p. 1-12, 2015.
- RAINER B. et al. ¿De las “guerras” metodológicas al pluralismo metodológico? *Revista Española de Ciencia Política*, Madrid, v. 29, p. 11-38, 2012.
- RAGIN, C. *Constructing social research: the unit and diversity of method*. Thousand Oaks: Pine Forge Press, 1994.
- RELACIP. *La enseñanza de la Ciencia Política en las universidades de América Latina*. Rosario: Editorial Ciudad Gótica, 2010.
- REIS, F. W. (Org.). *Os partidos e o regime: a lógica do processo eleitoral brasileiro*. São Paulo: Símbolo, 1978.
- _____. O tabelão e a lupa. *Revista Brasileira de Ciências Sociais*, São Paulo, ano 6, n. 16. p. 27-41, 1991.
- ROSANVALLON, P. *La legitimidad democrática: imparcialidad, reflexividad y proximidad*. Barcelona: Paidós, 2010.
- ROSENBERG, M. *A lógica da análise do levantamento de dados*. São Paulo: Cultrix, 1971.
- SARTORI, G. La política como ciencia. In: SARTORI, G. *La política: lógica y método en las ciencias sociales*. 3. ed. Ciudad de México: Fondo de Cultura Económica, 2002. p. 225-260.
- _____. ¿Hacia dónde va la ciencia política? *Revista Política y Gobierno*, Ciudad de México, v. 11, n. 2, p. 349-354, 2004.
- SCHLICK, M. *El viraje de la filosofía*. Ciudad de México: Fondo de Cultura Económica, 1978.
- SIMÃO, A. O voto operário em São Paulo. *Revista Brasileira de Estudos Políticos*, Belo Horizonte, n. 1, p. 130-141, 1956.
- SHEPSLE, K.; BONCHEK, M. *Analyzing politics: rationality, behavior and institutions*. New York: W.W. Norton, 1997.
- SKINNER, B. F. *A ciência e o comportamento humano*. São Paulo: Martins Fontes, 2013.
- SLIPAK, D. Entre aporías y prescripciones. Una reflexión sobre la historia conceptual de lo político propuesta por Pierre Rosanvallon. *Foro Interno*, Madrid, v. 12, p. 61-80, 2012.

- PORTO, M. *Televisão e política no Brasil: a Rede Globo e as interpretações da audiência*. Rio de Janeiro: E-Papers, 2007.
- SOARES, G. A. D. Classes sociais, strata sociais e as eleições presidenciais de 1960. *Revista Sociologia*, São Paulo. v. 232, n. 3, p. 217-238, 1961.
- _____. *A democracia interrompida*. São Paulo: FGV Editora, 2001.
- _____. O calcanhar metodológico da ciência política no Brasil. *Sociologia, Problemas e Práticas*, Lisboa, v. 48, 27-52, 2005.
- TANAKA, M.; DARGENT E. (Eds.). *¿Qué implica hacer ciencia política desde el sur y desde el norte?* Lima: Pontificia Universidad Católica de Perú, 2015.
- TSEBELIS, G. *Veto players: how political institutions works*. Princeton: Princeton University Press, 2002.
- TOCQUEVILLE, A. *A democracia na América: leis e costumes*. São Paulo: Martins Fontes, 1998.
- TURGEON, M.; CHAVES, B. S.; WIVES, W. W. Políticas de ação afirmativa e o experimento de listas: o caso das cotas raciais na universidade brasileira. *Opinião Pública*, Campinas, v. 20, n. 3, p. 363-76, 2014.
- TURGEON, M.; RENNÓ, L. Informação política e atitudes sobre gastos governamentais e impostos no Brasil: evidências a partir de um experimento de opinião pública. *Opinião Pública*, Campinas, v. 16, n. 1, p. 143-59, 2010.
- VALLÉS, J. M.; MARTÍ, S. *Ciencia política: un manual*. Barcelona: Ariel, 2015.

Resumo

O papel da ciência política no discurso científico brasileiro: institucionalização e avanços de uma área em construção

O artigo discute a inserção do campo específico da ciência política como área própria dentro das ciências sociais. O processo de institucionalização internacional tem como ponto de partida, aqui considerado, o início do século XX. São apresentadas as principais etapas da institucionalização e características específicas da área. Na segunda parte, o texto trata da formalização da ciência política como campo de pesquisa no Brasil. O conjunto de técnicas chamado de métodos quantitativos aplicado a objetos empíricos é tratado como um dos indicadores da autonomização da politologia no Brasil. Outro deles é a formatação institucional em programas de pós-graduação e publicações científicas próprias dessa esfera.

Palavras-chave: Ciência Política; Pesquisa; Brasil.

Abstract

The role of Political Science in the Brazilian scientific discourse: institutionalization and advances of a field under construction

This article discusses the insertion of the specific field of Political Science as an area within Social Sciences. The starting point of the international institutionalization process is here considered in the beginning of the 20th century. The stages of institutionalization and specific characteristics of the area are presented. In the second part, the study discusses the formalization of Political Science as a field of research in Brazil. The set of techniques called quantitative

methods applied to empirical objects is one of the indicators of Political Science autonomy in Brazil. Another one is the institutionalization in postgraduate programs and scientific publications in the area.

Keywords: Political Science; Research; Brazil.

Résumé

Le rôle de la science politique dans le discours scientifique brésilien: institutionnalisation et avancées d'un domaine en construction

L'article traite de l'insertion du domaine spécifique de la science politique en tant que domaine des sciences sociales. Le processus d'institutionnalisation internationale prend comme point de départ le début du 20^{ème} siècle. Les principales étapes d'institutionnalisation et les caractéristiques spécifiques du domaine sont présentées. Dans la deuxième partie, l'article traite de la formalisation de la science politique en tant que domaine de recherche au Brésil. L'ensemble des techniques appelées méthodes quantitatives appliquées aux objets empiriques est considéré comme l'un des indicateurs de l'autonomie de la science politique au Brésil. Un autre indicateur est l'institutionnalisation des programmes de troisième cycle et des publications scientifiques.

Mots-clés: Science Politique; Recherche; Brésil.

O texto como dado: desafios e oportunidades para as ciências sociais¹Maurício Izumi²Davi Moreira³**Introdução**

A despeito da importância que a análise de conteúdo possui para as ciências sociais, a escassez de tempo, de recursos e a vulnerabilidade à qual o rigor do pesquisador está sujeito ao analisar de forma manual o conteúdo de grandes bases de documentos (*corpus*)⁴ fez que boa parte das pesquisas com esse propósito se limitasse à análise de pequenos acervos. No caso da ciência política, por exemplo, são raros os trabalhos como os do *Manifesto Research Group* que, desde os anos 1970, analisa a ênfase temática de manifestos partidários⁵ ou o *Comparative Agendas Project*, que coleta e analisa dados sobre agendas de políticas públicas em diferentes países⁶.

Mais recentemente, porém, o avanço tecnológico e científico permitiu que técnicas automatizadas de análise do conteúdo fossem desenvolvidas e aplicadas de forma simples a grandes acervos. Longe de se distanciar desse processo, as ciências sociais como um todo

têm se mantido presente nessa fronteira do conhecimento, e o campo de pesquisa *text as data* ganha cada vez mais força (MONROE; SCHRODT, 2008; ROBERTS, 2016).

Em linha com esta fronteira do conhecimento, a ciência social brasileira já apresenta de forma satisfatória resultados científicos obtidos pelo uso intensivo da análise automatizada de conteúdo. A tese de doutorado de Davi Moreira (2016), por exemplo, analisou mais de 127 mil discursos proferidos por deputados federais no Pequeno Expediente ao longo dos anos de 1999 a 2016, concluindo que a atividade parlamentar no âmbito da Câmara dos Deputados não é integralmente governada pela relação governo-oposição, como acontece no caso das votações nominais. Por sua vez, Izumi (2017) desenvolveu um modelo bayesiano da teoria da resposta ao item (TRI) para estimar posições políticas utilizando textos como dados. O autor estimou as posições políticas dos partidos brasileiros entre 1995 e 2014 utilizando os discursos dos senadores.

1 Ambos os autores contribuíram igualmente para este manuscrito, sendo sua ordem aleatoriamente definida.

2 Maurício Izumi é doutor em Ciência Política pela Universidade de São Paulo (DCP/USP) e pesquisador do Centro de Política e Economia do Setor Público da Fundação Getúlio Vargas (Cepesp/FGV). O autor contou com apoio da FAPESP, processo número 2018/08118-4. E-mail: mauricioizumi@hotmail.com

3 Davi Moreira é doutor em Ciência Política pela USP e pós-doutorando pela UFPE. Vencedor do Prêmio Capes de Tese 2017 na área de Ciência Política e Relações Internacionais. Especialista em análise automatizada de conteúdo, discursos políticos e métodos quantitativos para ciências sociais. Idealizador do projeto Retórica Parlamentar, implementado pelo Laboratório Hacker da Câmara dos Deputados do Brasil. E-mail: davi.moreira@gmail.com

4 Neste artigo, os termos coleção de documentos, base de documentos, acervo e *corpus* são usados como sinônimos.

5 Para mais detalhes, ver: <<https://bit.ly/1NthwLD>>. Acesso em: 27 jun. 2018.

6 Para mais detalhes, ver: <<https://bit.ly/2o0MEpP>>. Acesso em: 27 jun. 2018.

O resultado apresenta que, em vez de uma clivagem ideológica, os discursos dos senadores organizam os partidos em uma dimensão que representa o conflito entre governo e oposição. Esse resultado indica que presidentes exercem uma influência não apenas em como os senadores votam, mas também em como eles falam.

Nesse escopo, o principal objetivo do artigo é manter as ciências sociais brasileiras na fronteira desse processo, conforme se verifica em trabalhos como o de Campos, Feres e Guarnieri (2017), e apresentar ao leitor um leque atualizado das principais metodologias de análise automatizada de conteúdo⁷. Convém ressaltar que o espaço dedicado a cada um dos métodos apresentados não reflete sua importância diante dos demais, o campo de estudos que trata o texto como dado (*text as data*) é rico e diversificado, sendo a escolha da metodologia ou de um conjunto de métodos totalmente dependente da questão de pesquisa que se deseja responder. Sem esgotar o leque de métodos, técnicas e modelos, este artigo é um guia para essa intensa e instigante área de pesquisa, apresenta conceitos essenciais para a área e provê um arcabouço relevante para o uso do texto como dado (*text as data*) nas pesquisas em ciências sociais.

Este artigo está dividido em sete seções, além desta introdução. Na primeira, “O texto como dado”, apresentamos os principais conceitos, princípios e desafios da área. Na segunda, “Obtenção e pré-processamento dos dados”, apresentamos procedimentos básicos que são comumente necessários antes do uso de modelos automatizados de análise de conteúdo. Na terceira, “Semelhança entre textos”, apontamos técnicas para obter medidas que indiquem quão parecidos são os conteúdos de dois documentos.

Na quarta, “Métodos de classificação”, mostram-se técnicas para a classificação automatizada de documentos em categorias, sejam elas conhecidas ou não. Na quinta, “Métodos de escalonamento”, apresentamos os dois principais algoritmos para a extração de posições políticas a partir de textos. Na sexta, “Desenvolvimentos e aplicações mais recentes”, destacamos as novas tendências da área. Por fim, a última seção traz breves considerações.

Com foco sobre a língua portuguesa, ao menos um exemplo de aplicação ao caso brasileiro será apresentado nas partes de 2 a 5⁸. O acervo completo que será utilizado para aplicação das diferentes metodologias, com exceção da parte 3, refere-se aos 35.398 discursos proferidos pelos deputados federais no Pequeno Expediente durante a 54^a legislatura – 2011-2014 (MOREIRA, 2016, p. 54).

O texto como dado

Dada a complexidade da linguagem, o processo de geração, produção e seleção de dados que resultam na comunicação humana é ainda um mistério para a ciência. Tal complexidade faz com que os modelos estatísticos desenvolvidos falhem na tarefa de prover um relato preciso do processo de geração de dados utilizados na produção de conteúdo e, principalmente, em seu significado.

Os modelos de análise de conteúdo, portanto, não devem ser avaliados pelo quanto explicam do processo de geração dos dados, mas sim por sua performance ao auxiliar o pesquisador em suas atividades acadêmicas. Transformar palavras em números não substitui a leitura cuidadosa e atenta de documentos,

7 Outros trabalhos publicados já enfrentaram esse desafio. Recomendamos especialmente que o leitor tenha contato com os seguintes artigos: Grimmer e Stewart (2013), Lucas et al. (2015), Wilkerson e Casas (2017) e Welbers, Van Attenveldt e Benoit (2017).

8 Para conhecer técnicas e métodos de análise automatizada de conteúdo para múltiplas línguas disponíveis para uso nas pesquisas em perspectiva comparada, recomendamos a leitura do artigo de Lucas et al. (2015).

mas permite a análise sistemática de grandes bases de texto sem a necessidade de mão de obra em larga escala e de um enorme montante de recursos financeiros para financiamento de pesquisas, amplificando o potencial científico dos trabalhos acadêmicos.

Logo, o uso do texto como dado para análise automatizada de conteúdo permite: (1) a utilização de diferentes técnicas independentemente do idioma sob análise; (2) o cálculo de medidas de incerteza, sendo possível julgar se as diferenças entre os textos são substantivas ou apenas fruto de erros de mensuração e variação amostral; (3) reduzir a necessidade de intervenção humana, facilitando a replicabilidade dos resultados; (4) a análise de um volume de informações manualmente inviável.

Contudo, reconhecendo que “métodos de análise automatizada de conteúdo são modelos incorretos de linguagem” (GRIMMER; STEWART, 2013, p. 2), a performance de qualquer método automatizado não é garantida sem a consideração de ao menos quatro princípios (Ibidem):

1. Todos os modelos quantitativos de análise de conteúdo estão errados, mas alguns são úteis;
2. Métodos quantitativos de análise de conteúdo amplificam a capacidade humana, mas não a substitui;
3. Não há um método global para a análise automatizada de conteúdo;
4. Alidar, validar, validar.

Justamente por conhecer as vulnerabilidades dos modelos estatísticos de análise de conteúdo que esses jamais substituirão a leitura e análise humana. Como será apresentado, é por essa razão que a grande maioria dos modelos disponíveis atrelam sua capacidade inferencial a medidas e informações fornecidas pelo pesquisador antes e depois de produzidos os resultados. Para ser profícua, essa interação

deve resultar da leitura cautelosa de amostras do acervo a ser analisado e da análise cuidadosa dos resultados obtidos.

A escolha do modelo, da família de modelos ou de eventuais combinações a serem utilizadas é resultado dos objetivos de pesquisa almejados. Como será visto, há uma variedade de modelos disponíveis e nenhum deles se sobrepõe aos demais. Pelo contrário, são indicados para questões e aplicações diante do auxílio analítico que o pesquisador deseja receber e das perguntas que deseja responder. Mais do que sobrepor uma abordagem a outra, convém ao pesquisador identificar a abordagem que melhor lhe atende. Neste artigo, por exemplo, o maior ou menor espaço dedicado a cada um deles não deve ser visto como algo que reduz ou aumenta sua importância para propósitos científicos.

Como sugerem Grimmer e Stewart (2013), o que precisa ser evitado é o uso cego de qualquer método. Por essa razão, em acordo com esses pesquisadores, desencorajamos o uso de *softwares* comerciais para análise quantitativa de textos. A despeito de resultados positivos, por vezes é impossível conhecer o método aplicado e o pesquisador se torna refém dos resultados apresentados sem a possibilidade de interagir e interferir no procedimento de análise do modelo.

Na próxima seção apresentamos o processo de obtenção/coleta de dados e o pré-processamento dos documentos brutos.

Obtenção e pré-processamento dos dados

Obtenção dos dados

A disponibilidade de uma infinidade de informações publicadas na internet ampliou de forma exponencial a possibilidade de acesso a um volume expressivo de conteúdo. Atualmente, projetos de lei, documentos históricos, discursos, material de campanha eleitoral, entre outras

fontes, podem ser acessados de qualquer parte do planeta sobre qualquer parte do planeta.

Se de um lado o acesso foi ampliado, de outro multiplicaram-se as formas de obtenção de conteúdo para análise automatizada nas pesquisas em ciências sociais. A um custo baixíssimo, por exemplo, o conteúdo virtualmente publicado pode ser obtido com o uso de métodos de raspagem de dados (*web scraping*), pelo qual o computador é programado para acessar páginas na internet, copiar seu conteúdo e organizar os dados para o pesquisador. De forma adicional, como destacam Berinsky, Huber e Lenz (2012), mesmo quando dados não estão facilmente disponíveis na internet, plataformas on-line, como a *Mechanical Turk* da Amazon, podem ser alternativas eficientes para sua obtenção e compilação.

A maior dificuldade na obtenção de acervo para análise automatizada de conteúdo reside no interesse de análise de documentos históricos. Para tanto, é necessário o escaneamento em alta qualidade de documentos e o uso de um bom *software* para reconhecimento óptico de caracteres (OCR). Para teste fim, já existem pacotes⁹ na linguagem R¹⁰ que utilizam uma Interface de Programação de Aplicativos (API) do Google¹¹ para execução dessa tarefa. Além disso, existem diversos pacotes que facilitam a extração de dados da internet, como o “xml2” (WICKHAN; HESTER; OOMS, 2018), o “htr” (WICKHAM, 2014) e o “rvest” (Idem, 2016).

Para os exemplos de aplicação que serão apresentados ao longo deste artigo, os mais de 35 mil discursos proferidos pelos deputados federais no Pequeno Expediente ao longo dos anos de 2011

e 2014 foram obtidos através do webservice da Câmara dos Deputados com o uso de um programa desenvolvido em linguagem R para raspagem dos dados publicados (MOREIRA, 2016). Com o fantástico trabalho de registro e divulgação dos pronunciamentos parlamentares, praticamente tudo que é dito nos momentos institucionais previstos pelo Regimento Interno da Câmara dos Deputados (RICD) é gravado, arquivado e disponibilizado ao cidadão pela internet em linguagem de máquina sob escopo do projeto Dados Abertos da Câmara dos Deputados¹².

*Pré-processamento dos dados*¹³

Como já apontamos, dada a complexidade da linguagem, a análise automatizada de conteúdo pressupõe a perda de informação para ganho de eficiência dos modelos. Nesse sentido, o pré-processamento dos dados exige a adoção de procedimentos que reduzam a dimensionalidade com a qual os modelos de análise de conteúdo vão lidar. A título de exemplificação, apresentamos um roteiro básico comum para o pré-processamento dos dados.

O roteiro básico de pré-processamento consiste em cinco passos: (1) codificação de caracteres (*encoding*); (2) remoção de palavras e conteúdo desnecessário; (3) construção de uma sacola de palavras (*bag of words*); (4) *stemming*; (5) construção da matriz de documentos e termos (*Document Term Matrix – DTM*).

1. **Codificação de caracteres:** Como apontam Lucas et al. (2015), apesar dos métodos estatísticos para análise de conteúdo serem

9 Atualmente, o principal pacote para esta tarefa é o “tesseract” (OOMS, 2018).

10 O R é uma linguagem e um ambiente de programação estatística. Para mais detalhes, acesse: <<https://bit.ly/1gm1uk2>>. Acesso em: 8 jul. 2018.

11 Para mais detalhes, acesse: <<https://cloud.google.com/vision/>>. Acesso em: 29 jun. 2018.

12 Para mais detalhes, acesse: <<https://bit.ly/2w7kY6Z>>. Acesso em: 29 jun. 2018.

13 Para referências sobre o pré-processamento de acervos com múltiplas línguas, consulte Lucas et al. (2015).

agnósticos em relação à ligação, as ferramentas para pré-processar os textos não são. Isso significa que o pré-processamento precisa ter em conta o modo como o computador interpreta a informação a ele repassada. A codificação de texto é a maneira pela qual o computador traduz caracteres individuais e únicos em *bytes* para seu armazenamento na memória. Desse modo, dados de várias fontes diferentes e em línguas diferentes têm grande chance de estar em codificações distintas, sendo necessário ao analista converter todos os documentos do acervo para a mesma codificação. Em seguida, o analista deve garantir que o computador interprete corretamente a codificação dos caracteres¹⁴.

2. **Remoção de palavras e conteúdo desnecessário:** Harmonizada a codificação dos caracteres para a correta interpretação por parte do computador, o primeiro procedimento de redução de dimensionalidade a ser considerado é a remoção de palavras e elementos do texto que não representem substantivamente o alvo de interesse do pesquisador. Pontuação, números, artigos, pronomes e preposições (*stop words*) costumam ser retirados dos documentos¹⁵. Em seguida, é conveniente a retirada de uma lista predeterminada de palavras irrelevantes para os propósitos da pesquisa, bem como palavras pouco ou muito frequentes que apareçam em menos de 1% ou mais de 99%

dos documentos (HOPKINS; KING, 2010; QUINN et al. 2010)¹⁶. A título de exemplo, se considerarmos a seguinte sentença: “O Partido dos Trabalhadores é contra a reforma trabalhista”, nesta etapa do pré-processamento, teríamos: “partido trabalhadores contra reforma trabalhista”.

3. **Sacola de palavras (*bag of words*):** De forma geral, os métodos automatizados de análise de texto tratam documentos como um vetor composto por palavras, desconsiderando a ordem em que aparecem. Com essa abordagem, cada documento é representado como um único vetor, sendo seu comprimento igual ao número de palavras únicas que possui.
4. **Stemming¹⁷:** Até aqui, mesmo que os procedimentos adotados tenham diminuído enormemente a dimensionalidade do acervo de documentos ao transformá-lo em uma sacola de palavras, ainda é necessária a adoção de procedimentos que possam reduzir a complexidade do conteúdo a ser analisado. Nesse sentido, podemos pensar que determinado documento tenha em sua composição as seguintes palavras únicas: trabalho, trabalhador, trabalhista. Apesar de seus diferentes significados, cada uma dessas palavras pode ser reduzida ao seu radical, *trabalh*, dando ao pesquisador informações suficientes para sua análise e, assim reduzindo, o *n* de três palavras únicas para uma palavra que tem a soma das frequências anteriores. Para garantir que

14 No caso da linguagem R de programação, a função *encoding* é de grande importância. Para mais detalhes: <<https://bit.ly/2w81Cyw>>. Acesso em: 30 jun. 2018.

15 O pacote “tm” (FEINERER; HORNICK, 2018) da linguagem R pode ser utilizado para a remoção de *stop words*. Ele utiliza o acervo disponível no projeto *Snowball* para realizar as operações. Para mais informações acessar: <<https://bit.ly/2OYva8V>>. Acesso em: 30 jun. 2018.

16 Como apontam Lucas et al. (2015), para cada idioma, a escolha de quais palavras devem ser removidas é uma decisão substantiva que, em alguns casos, pode ter efeitos importantes nos resultados da análise (CAMPBELL; PENNEBAKER, 2003; FOKKENS et al., 2013).

17 O processo de *stemming* é uma aproximação do processo de lematização, que reduz palavras às suas formas básicas (JURAFSKY; MARTIN, 2009; MANNING; RAGHAVAN; SCHÜTZE, 2008).

palavras que variam apenas na flexão, número ou conjugação sejam consideradas iguais, reduzindo o número de dimensões contido no acervo, por meio da adaptação do algoritmo de Porter (1980) para o português já desenvolvida por diferentes projetos (*Snowball* e NILC-USP), podem ser obtidos os *stems* das palavras restantes¹⁸. Retomando a sentença de exemplo, nesta etapa do pré-processamento teríamos: “part trabalh contr reform trabalh”.

5. **Matriz de Documentos e Termos – DTM:** Com os procedimentos até aqui adotados, cada documento i ($i=1, \dots, N$) é representado por um vetor que contabiliza o número de vezes que cada uma das M palavras únicas ocorrem, $P_i = (P_{i1}, P_{i2}, \dots, P_{im})$. Cada P_{im} representa o número de vezes a m -ésima palavra ocorre no i -ésimo documento. Estruturando lado a lado os vetores de contagem forma-se a Matriz de Documentos e Termos (*Document Term Matrix* – DTM). Essa matriz será esparsa (com grande número de zeros) e conterá a frequência de cada termo (palavra ou *stem*) para cada documento do acervo. É essa matriz a matéria-prima para os modelos de análise automatizada de conteúdo.

Aplicação para o caso brasileiro – pré-processamento dos dados

Como destacado, a aplicação dos procedimentos apresentados deve levar em

consideração o modelo que será utilizado para a análise automatizada do conteúdo do acervo e os propósitos da análise. Desse modo, em posse do conteúdo dos 35.398 discursos proferidos pelos deputados federais no Pequeno Expediente durante a 54ª legislatura (2011-2014), os procedimentos de tratamento dos dados foram aplicados para o uso do *expressed agenda model* (GRIMMER, 2010) com o objetivo de se identificar os temas proferidos pelos parlamentares em seus discursos¹⁹.

Aplicados para todos os discursos cujos oradores proferiram no mínimo mais de uma fala²⁰, obteve-se uma *Document Term Matrix* (DTM), $d \times P$, com 33.941 linhas e 3.906 colunas, cujas linhas representam cada discurso e as colunas representam cada *stem* presente no *corpus*. É essa matriz que será utilizada na estimação dos tópicos²¹ das falas dos deputados na seção 4.

Como enfatizado ao longo de toda a seção, é possível ver, pela última coluna da Tabela 1, que a aplicação do pré-processamento dos dados reduz de forma drástica o volume informacional do conteúdo presente nos documentos. Para o caso em que foi aplicado, de mais de 150 mil palavras únicas, o modelo terá como matéria-prima de análise pouco menos de quatro mil *stems*²². Apesar da sensação de que não será possível obter resultados robustos após tamanha perda de informação, além de como será visto mais adiante neste artigo, pesquisadores têm mostrado que a representação do acervo por meio

18 Ver: <<http://snowball.tartarus.org/>> e <<https://bit.ly/2OXkVS6>>. Acesso em: 30 jun. 2018.

19 Mais detalhes sobre o *expressed agenda model* estão apresentados na quarta parte deste artigo.

20 Grimmer (2010) retira do *corpus* todas as palavras que estão presentes em 0,5% dos documentos e também retira dos documentos de cada autor as palavras que aparecem em mais de 90% dos documentos de cada um. Além desse procedimento, neste trabalho, foram retiradas palavras selecionadas após análise de uma amostra aleatória de discursos.

21 Neste artigo as palavras *tópico* e *tema* são usadas como sinônimos.

22 Além da DTM, o *expressed agenda model* recebe como argumento uma matriz de autores, $n \times 2$, cujas linhas representam cada autor, a primeira coluna representa a linha do primeiro discurso do autor na DTM e a segunda coluna representa a linha do último discurso deste autor na DTM.

da DTM é suficiente para inferências substantivas a respeito de seu conteúdo (HOPKINS; KING, 2010).

Na próxima seção apresentaremos um conjunto de técnicas utilizadas para mensurar quão similares são dois documentos.

Tabela 1
Resultado do tratamento aplicado à coleção de discursos

| | Oradores | Discursos (<i>D</i>) | Palavras únicas (<i>P</i>) |
|-----------|----------|------------------------|------------------------------|
| # Inicial | 591 | 35.398 | 153.111* |
| # Final | 552 | 33.941 | 3.906** |

* Número de palavras únicas antes do processo de *stemming*.

** Número de *stems* únicos.

Semelhança entre textos

Medidas de similaridade são ferramentas muito interessantes para aqueles que analisam o texto como dado e buscam identificar semelhanças entre textos. Essas medidas podem ter diversas aplicações nas ciências sociais. A primeira é o uso como uma técnica exploratória. Textos são entidades complexas e em grandes quantidades se torna difícil saber se existe algum padrão entre os documentos. A partir de medidas de similaridade podemos ver como os dados se estruturam e se relacionam. Além disso, elas podem ser utilizadas em tarefas mais complexas, como traçar a origem de projetos de lei (LI; LAROCHELLE; LO, 2014; WILKERSON; SMITH; STRAMP, 2015) e verificar como grupos de interesse e da sociedade civil influenciam o conteúdo das leis (GARRETT; JANSÁ, 2015; KROEGER, 2015). A seguir veremos dois tipos de métodos para análise de similaridade.

Similaridade de cosseno

Esse primeiro método busca avaliar quanto um documento é similar a outro como um todo. Assume-se que quanto maior a similaridade na frequência relativa das palavras utilizadas, maior será a similaridade do conteúdo entre os textos. Ou seja, busca-se obter uma medida que indique quão parecido é o uso das palavras em dois documentos.

Partimos da ideia de que um texto pode ser representado como um vetor (*bag of words*). Esse vetor é representado em um espaço de dimensão igual ao comprimento do universo de palavras utilizadas (vocabulário). Vamos supor que queremos comparar dois projetos de lei, PL1 e PL2. Vamos supor também que os deputados que redigiram esses projetos possuem um vocabulário reduzido. Eles conhecem apenas cinco palavras: “política”, “orçamento”, “presidente”, “futebol” e “copa”. Assim, o PL1 e o PL2 podem ser definidos como vetores (u e v , respectivamente) de dimensão igual a 5, R^5 . Na Tabela 2 apresentamos a frequência de palavras em cada projeto.

Tabela 2
Frequência de palavras no PL1 e no PL2

| PL | política | orçamento | presidente | futebol | copa |
|-----|----------|-----------|------------|---------|------|
| PL1 | 1 | 10 | 15 | 30 | 0 |
| PL2 | 0 | 15 | 20 | 0 | 3 |

A partir dessa tabela podemos construir os vetores u e v como $u = (1, 10, 15, 30, 0)$ e $v = (0, 15, 20, 0, 3)$. Uma forma de avaliarmos quão similares são esses vetores é calculando o produto interno entre eles, isto é, $u \cdot v = \sum_{i=1}^5 u_i v_i$. Quando os dois vetores possuem frequências altas para as mesmas palavras, maior o produto interno entre eles. Em outras palavras, se observamos um produto interno alto, temos evidências de que a distribuição do conjunto de palavras nos dois documentos é similar. No exemplo dado, o produto interno é igual $u \cdot v = (1 \times 0) + (10 \times 15) + (15 \times 20) + (30 \times 0) + (0 \times 3) = 450$.

Como é possível observar, essa medida ainda é problemática. Quanto maior o

comprimento do vetor ($|v| = \sqrt{\sum_{i=1}^5 v_i^2}$), maior será o produto interno. Assim, documentos

que possuem um vocabulário mais rico, bem como aqueles que utilizam palavras muito frequentes, provavelmente terão um produto interno maior.

A solução para esse problema é dividir o produto interno pelo produto dos comprimentos dos vetores, $\frac{u \cdot v}{|u| \cdot |v|}$. Mas, se lem-

brarmos da geometria analítica, em que $u \cdot v = |u| \cdot |v| \cdot \cos \theta$, temos que $\cos \theta = \frac{u \cdot v}{|u| \cdot |v|}$.

Logo, $\cos \theta$, que representa o cosseno do ângulo formado entre os vetores u e v , indica quão similares são os vetores. Essa medida é conhecida como similaridade de cosseno (JURAFSKY; MARTIN, 2009) e varia de zero a um. Quanto mais próxima de um, maior a similaridade dos vetores.

Quadro 1
Comparação entre o PL1375/2011 e o PL7480/2014

| PL1375/2011 | PL7480/2014 |
|---|--|
| <p>Altera a redação do art. 11 da Lei no 11.180, de 23 de setembro de 2005, no que se refere ao valor da bolsa-permanência do Programa Universidade para Todos – Prouni.</p> <p>O Congresso Nacional decreta:</p> <p>Art. 1º O art. 11 da Lei no 11.180, de 23 de setembro de 2005, passa a vigorar com a seguinte redação:</p> <p>Art. 11. Fica autorizada a concessão de bolsa-permanência, no valor de até um salário mínimo mensal, exclusivamente para custeio das despesas educacionais, a estudante beneficiário de bolsa integral do Programa Universidade para Todos – Prouni, instituído pela Lei nº 11.096, de 13 de janeiro de 2005, matriculado em curso de turno integral, conforme critérios de concessão, distribuição, manutenção e cancelamento de bolsas a serem estabelecidos em regulamento, inclusive quanto ao aproveitamento e à frequência mínima a ser exigida do estudante. (NR)</p> <p>Art. 2º Esta lei entra em vigor na data de sua publicação.</p> | <p>Altera a redação do art. 11 da Lei no 11.180, de 23 de setembro de 2005, com relação à concessão de bolsa-permanência para estudantes beneficiários do Programa Universidade para Todos (Prouni).</p> <p>O Congresso Nacional decreta:</p> <p>Art. 1º O art. 11 da Lei no 11.180, de 23 de setembro de 2005, passa a vigorar com a seguinte redação:</p> <p>Art. 11. Fica autorizada a concessão de bolsa-permanência, até o valor equivalente ao praticado na política federal de concessão de bolsas de iniciação científica, exclusivamente para custeio das despesas educacionais, a estudantes beneficiários de bolsa integral do Programa Universidade para Todos (Prouni), instituído pela Lei nº 11.096, de 13 de janeiro de 2005, conforme critérios de concessão, distribuição, manutenção e cancelamento de bolsas a serem estabelecidos em regulamento, inclusive quanto ao aproveitamento e à frequência mínima a ser exigida do estudante.</p> <p>Parágrafo único. Os critérios de concessão referidos no caput considerarão especialmente a situação de impossibilidade de compatibilidade entre a frequência ao curso, em turno parcial ou integral, e o exercício de atividade remunerada, no caso de o estudante não contar com renda própria ou familiar suficiente para prover sua subsistência. (NR)</p> |

Fonte: Dados abertos da Câmara dos Deputados.

No Quadro 1 apresentamos um exemplo concreto no qual comparamos dois projetos de lei. O primeiro é o PL1375/2011, de autoria da deputada professora Dorinha Seabra Rezende, do partido Democratas do Tocantins (DEM-TO), e o segundo é o PL7480/2014, de autoria do deputado Gustavo Petta, do Partido Comunista do Brasil de São Paulo (PCdoB-SP). A semelhança entre os projetos é clara. O cosseno do ângulo formado pelos dois vetores é igual a 0,85²³.

Uma limitação desse método é o fato de ele desconsiderar completamente o ordenamento das palavras. Vejamos um exemplo que ressalta essa limitação. Se observamos as seguintes frases: (1) “faça amor, não faça guerra”; e (2) “faça guerra, não faça amor”, a similaridade de cosseno será máxima pois ambas as frases utilizam o mesmo conjunto de palavras. No entanto, elas possuem sentidos completamente opostos. Ao desconsiderar o ordenamento das palavras, essa medida não é capaz de diferenciar esse tipo de detalhe. Obviamente, esse é um caso extremo. No geral, documentos com sentidos opostos tendem a utilizar palavras diferentes. A seguir apresentamos um método que enfrenta esse desafio.

Algoritmo de Smith-Waterman

Na similaridade de cosseno, estamos interessados em quanto um documento é similar a outro como um todo. Mas outra tarefa que um cientista social pode estar interessado é aquela em que buscamos os trechos de um documento que são semelhantes a trechos de outros documentos. Nessa tarefa estamos interessados na similaridade de apenas um pedaço do texto. A família de algoritmos que cumpre esse tipo função é a de alinhamento local e sua origem está na biologia molecular.

Nesse campo de estudos é comum querer saber quão similares são as sequências genéticas de organismos diferentes. O algoritmo mais utilizado para essa tarefa é o de alinhamento local de Smith-Waterman (SMITH; WATERMAN, 1981). Ele compara segmentos das sequências de todos os comprimentos possíveis de modo a maximizar a similaridade entre eles. A similaridade é baseada em uma função que atribui um valor positivo quando há um *matching* ($matching = 2$) e penaliza com valores negativos quando há um *mismatching* ($mismatching = -1$) ou um *gap* ($gap = -1$) nas sequências. Por exemplo, se observarmos as sequências “AAACGTCA” e “CGTA”, podemos tentar alinhar os trechos “AAAC” e “CGTA” (índice: $-1-1-1-1 = -4$). Outra possibilidade é alinhar “CGTCA” e “C#GTA” (índice: $2-1-1-1+2 = -1$). Aqui incluímos um *gap* (#) na segunda posição da segunda sequência. O melhor alinhamento possível das duas sequências (isto é, aquele que gera o maior valor no índice) é dado pelos trechos “CGTCA” e “CGT#A” (índice: $2+2+2-1+2 = 7$). Esse seria o alinhamento encontrado pelo algoritmo.

A diferença é que, em vez de comparar sequências de nucleotídeos em cadeias de DNA, vamos comparar sequências de palavras em textos. Esse algoritmo parece ser apropriado para diversos problemas, como a comparação de projetos de lei, por dois motivos. O primeiro é o fato de ele fazer comparações locais e não globais. Isso é relevante porque dois documentos semelhantes podem compartilhar apenas alguns trechos e serem bem diferentes em todos os outros. O segundo ponto está relacionado ao fato de ele permitir pequenas diferenças nos trechos ao adicionar os *gaps*. Isso é relevante, já que mudanças na linguagem são esperadas quando ideias migram de um contexto a outro²⁴.

23 Para mais detalhes, ver Izumi (2017).

24 Para uma aplicação ao caso norte-americano, ver Wilkerson, Smith e Stramp (2015).

Quadro 2

Resultado do algoritmo de Smith-Waterman para a comparação entre o PL1375/2011 e o PL7480/2014

| PL1375/2011 | PL7480/2014 |
|---|--|
| <p>Altera a redação do art 11 da Lei nº 11.180 de 23 de setembro de 2005 no ### que ##### se # refere ##### ao ## valor da bolsa permanência ##### ##### do Programa Universidade para Todos PROUNI</p> <p>O Congresso Nacional decreta</p> <p>Art 1º O art 11 da Lei no 11.180 de 23 de setembro de 2005 passa a vigorar com a seguinte redação</p> <p>Art 11 Fica autorizada a concessão de bolsa permanência no ### # valor ##### ## ##### ## ##### de até ##### um ## salário ##### mínimo ##### mensal ##### exclusivamente para custeio das despesas educacionais a estudante ##### beneficiário ##### de bolsa integral do Programa Universidade para Todos Prouni instituído pela Lei nº # # 11.096 de 13 de janeiro de 2005 matriculado em curso de turno integral conforme critérios de concessão distribuição manutenção e cancelamento de bolsas a serem estabelecidos em regulamento inclusive quanto ao aproveitamento e à frequência ##### mínima a ser exigida do estudante</p> | <p>Altera a redação do art 11 da Lei nº 11.180 de 23 de setembro de 2005 ## com ## relação ## à ##### concessão ## de ##### ## bolsa permanência para estudantes beneficiários do Programa Universidade para Todos Prouni</p> <p>O Congresso Nacional decreta</p> <p>Art 1º O art 11 da Lei no 11.180 de 23 de setembro de 2005 passa a vigorar com a seguinte redação</p> <p>Art 11 Fica autorizada a concessão de bolsa permanência ## até o valor equivalente ao praticado na política federal de ## concessão ## de ##### bolsas ##### de ##### iniciação científica exclusivamente para custeio das despesas educacionais a ##### estudantes ##### beneficiários de bolsa integral do Programa Universidade para Todos Prouni instituído pela Lei ## nº 11.096 de 13 de janeiro de 2005 ##### ## ##### ## ##### conforme critérios de concessão distribuição manutenção e cancelamento de bolsas a serem estabelecidos em regulamento inclusive quanto ao aproveitamento e à ##### frequência mínima a ser exigida do estudante</p> |

Fonte: Dados Abertos da Câmara dos Deputados.

No Quadro 2 apresentamos o resultado da aplicação do algoritmo de Smith-Waterman para o PL1375/2011 e o PL7480/2014. Como podemos observar, esse algoritmo conseguiu captar de forma precisa a semelhança entre os projetos. A ementa foi mantida com a inclusão de alguns *gaps*, assim como o artigo 11. O 1º artigo se manteve inalterado. Embora o artigo 2º de ambos projetos sejam idênticos, ele não foi alinhado, pois sua inclusão levaria a uma penalidade muito grande, já que o parágrafo único aparece em apenas um dos projetos.

Métodos de classificação

A classificação automatizada organiza o acervo de documentos em categorias, sejam elas conhecidas ou não. Utilizaremos esse critério básico, conhecimento ou desconhecimento

das categorias nas quais deve ser classificado o acervo, para apresentar métodos de cassificação automatizada.

Categorias conhecidas

Ancorados na teoria, pesquisadores em ciências sociais almejam classificar documentos em categorias já conhecidas. Inúmeras são as aplicações, mas entre elas descata-se duas: (1) o uso de métodos de dicionário; e (2) métodos de aprendizagem supervisionada (*supervised learning methods*).

Métodos de dicionário

É comum que pesquisadores queiram saber se um documento tem conotação positiva ou negativa sobre um tópico qualquer. Podemos

usar o método de análise de sentimentos para esse tipo de tarefa. A análise de sentimentos está associada ao uso da análise quantitativa de textos para extração de estados afetivos contidos em documentos. Esses estados afetivos são conhecidos como sentimentos ou opiniões. Em geral, na análise de sentimentos, o foco está sobre opiniões que expressam sentimentos positivos ou negativos. Queremos saber se um consumidor possui opinião positiva ou negativa sobre o produto ou serviço que ele adquiriu; se as propagandas de um candidato em campanha política possuem conotação positiva ou negativa; se a cobertura jornalística sobre determinado candidato é positiva ou negativa (LIU, 2012; PANG; LEE, 2008).

De modo mais preciso, um sentimento é definido como uma quádrupla (g, s, h, t) , em que g é o alvo do sentimento, s é o sentimento a respeito do alvo, h é o detentor da opinião e t é o momento em que a opinião foi expressada. Um bom exemplo está na frase “eu tenho ódio e nojo à ditadura”, proferida por Ulysses Guimarães no momento da promulgação da Constituição de 1988. Nessa frase, Ulysses Guimarães é o detentor da opinião (h). O alvo de seu sentimento negativo (s) é a ditadura (g) e seu sentimento foi expressado no dia 5 de outubro de 1988 (t).

Uma das abordagens mais simples na análise de sentimentos, ou de modo mais geral, para a classificação de documentos em categorias preestabelecidas, é a abordagem por meio de dicionários anotados (TABOADA et al., 2011). Como o nome sugere, nessa abordagem, o sentimento de um documento é determinado com o auxílio de um dicionário no qual as palavras são anotadas com sua orientação semântica

(positiva ou negativa)²⁵. Assim, palavras como “ódio” e “nojo” possuem conotação negativa, já palavras como “amor” e “gostar” têm conotação positiva. Desse modo, documentos que utilizam mais palavras positivas do que negativas são classificados como documentos que expressam um sentimento positivo.

Formalmente, sejam W_{dp} o número de vezes que a palavra p aparece no documento d e s_p o sentimento associado à palavra p . Para simplificar, $s_p = -1$, quando o sentimento associado for negativo e $s_p = 1$, quando o sentimento associado for positivo. O sentimento s_d de um documento d é:

$$s_d = \sum_{p=1}^M \frac{s_p W_{dp}}{\sum_{p=1}^M W_{dp}}$$

Classificamos um documento como positivo, se $s_d > 0$ e o classificamos como negativo, caso contrário.

Métodos de aprendizagem supervisionada (*Supervised learning methods*)

Métodos de aprendizado supervisionado replicam a familiar tarefa de codificação manual, porém com enorme redução de custos e grande ganho de escala. Sua implementação pressupõe a classificação manual de uma amostra do acervo em um conjunto predeterminado de categorias. Essa amostra classificada, conhecida como conjunto de treinamento ou *training set*, é usada para treinar modelos estatísticos, cuja principal aplicação é a classificação do restante do acervo, conjunto de teste ou *test set*, nas categorias predeterminadas. Ao final da classificação, procedimentos de

25 O Grupo de Processamento da Linguagem Natural da Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS) possui o *OpLexicon*, um léxico de sentimento para a língua portuguesa (SOUZA; VIEIRA, 2012; SOUZA et al., 2011). Na atual versão, esse dicionário apresenta 31.719 termos dos quais 14.254 carregam sentimentos negativos, 8.469 sentimentos positivos e 8.996 sentimentos neutros. Mais informações podem ser encontradas em: <<https://bit.ly/2OZGxNP>> Acesso em: 7 jul. 2018.

validação devem ser adotados para se averiguar a performance do modelo utilizado.

Vejam que o método de classificação a partir de dicionários anotados não é supervisionado. Essa abordagem tem como vantagem não depender de regras ou procedimentos criados fora do domínio específico do conteúdo que está sob análise. Tal característica evita problemas de uso de referências externas, como pode ocorrer no caso de modelos de dicionário. Métodos de aprendizado supervisionado requerem que os pesquisadores desenvolvam regras de codificação manual para as categorias de interesse. Tal necessidade força os analistas a desenvolverem definições coerentes de conceitos para aplicações particulares, o que leva à clareza sobre a classificação pretendida. Outra vantagem dos métodos de aprendizado supervisionado é a facilidade de validação e verificação da performance do modelo utilizado.

Como apontam Grimmer e Stewart (2013), todos os métodos de aprendizagem supervisionada pressupõem três etapas básicas após os procedimentos de pré-processamento já apresentados: (1) construir um conjunto de treinamento (*training set*); (2) aplicar o método de aprendizado supervisionado; e (3) validar a saída do modelo.

1. **Construção de um conjunto de treinamento – *training set***

Esquema de codificação: Para a construção de um conjunto de treinamento, deve ser criado um esquema de codificação manual que supere dificuldades relacionadas a ambiguidades na linguagem, a atenção limitada dos codificadores e o entendimento sobre conceitos presentes

no acervo. Com um livro de códigos elaborado, devem-se realizar exercícios de testes para que sejam identificadas ambiguidades no esquema de codificação ou nas categorias negligenciadas. Esse procedimento leva, subsequentemente, a uma revisão do livro de códigos, que então precisa ser aplicado a um novo conjunto de documentos para assegurar que as ambiguidades tenham sido suficientemente resolvidas. Logo, somente após os codificadores aplicarem o esquema de codificação aos documentos sem perceber ambiguidades, o esquema estará pronto para ser aplicado ao restante do conjunto de dados²⁶.

Seleção do conjunto de treinamento:

Idealmente, os documentos presentes no conjunto de treinamento devem ser representativos do acervo. Logo, para um bom desempenho do modelo, é aconselhável que o conjunto de treinamento seja construído a partir de uma amostra aleatória da coletânea à qual pertencem (HAND, 2006). Isto posto, resta saber qual a quantidade ideal de documentos para o conjunto de treinamento. Hopkins e King (2010) indicam quinhentos como regra geral, sendo cem documentos já suficientes para alguns casos. No entanto, o número ideal dependerá da aplicação específica de interesse, pois, conforme o número de categorias aumenta, mais documentos são necessários em cada categoria do conjunto de treinamento para uma boa performance do modelo. Como será exposto adiante, uma vantagem do uso da aprendizagem supervisionada para

26 Como apontam Grimmer e Stewart (2013), a criação de esquemas de codificação é uma literatura rica em ciências sociais. Para mais detalhes, ver Krippendorff (2004), Neuendorf (2002), Weber (1990) e a documentação disponível no pacote do R “ReadMe” (HOPKINS; KING, 2017).

classificação é a possibilidade de usar o processo de validação para verificar se não é necessário aumentar o N de documentos do conjunto de treinamento.

2. O uso do método de aprendizagem supervisionada

Os métodos de aprendizagem supervisionada são diversos, mas compartilham de uma estrutura comum (HASTIE; TIBSHIRANI; FRIEDMAN, 2001). Como apresentam Grimmer e Stewart (2013), suponha que existam N_{train} documentos ($i=1, \dots, N_{\text{train}}$), sendo esse o conjunto de treinamento manualmente classificado em uma das K categorias, ($k=1, \dots, K$). A categoria de cada documento i é representada por $Y_i \in \{C_1, C_2, \dots, C_k\}$ e todo o conjunto de treinamento é representado como $Y_{\text{train}} = (Y_1, \dots, Y_{N_{\text{train}}})$. Cada modelo de aprendizagem supervisionada assume que existe uma função, f , não observada que associa a configuração de palavras dos documentos às categorias preestabelecidas, $Y_{\text{train}} = f(P_{\text{train}})$. Assim, com base no conjunto de treinamento, o algoritmo busca compreender essa associação e replicá-la aos demais documentos, $\hat{Y}_{\text{train}} = \hat{f}(P_{\text{train}})$. Com base nessa estrutura comum, apresentamos a seguir um método de inferência da relação entre palavras e categorias para classificação individual de documentos.

Naive Bayes (MARON; KUHNS, 1960)²⁷: Esse é um dos mais simples e poderosos métodos de classificação individual. Nele, o conjunto de treinamento é usado para aprender sobre a distribuição de palavras para documentos de cada categoria k . Essa distribuição é usada para classificar cada um dos documentos no conjunto restante do acervo. Com base na regra de Bayes, o modelo opera essa classificação ao inferir a probabilidade de que o documento i pertença à categoria k , dado o perfil de palavras P_i . Aplicando a regra de Bayes, $p(C_k | P_i) \propto p(C_k) p(P_i | C_k)$, sabemos que é necessário estimar $p(C_k)$ e $p(P_i | C_k)$. Sendo o conjunto de treinamento uma boa representação do acervo, temos que o estimador de máxima verossimilhança de $p(C_k)$ é dado pela proporção de documentos do conjunto de treinamento em cada categoria k . Por sua vez, a estimação de $p(P_i | C_k)$ é mais complexa e necessita do pressuposto (*naive assumption*) de que, dada a categoria de um documento, as palavras são geradas de forma independente, $p(P_i | C_k) = \prod_{m=1}^M p(P_{im} | C_k)$. Obviamente, tal pressuposto é equivocado, visto que o uso de palavras é altamente correlacionado. Contudo, mesmo com esse pressuposto, o modelo ainda é capaz de capturar informações úteis para classificação dos documentos. Usando essa suposição, a estimativa de $p(P_{im} | C_k)$ é dada por:

$$\hat{p}(P_{im} = j | C_k) = \frac{\text{\#documentos do conjunto de treinamento na categoria } k \text{ e com a palavra } m \text{ usada } j \text{ vezes}}{\text{\#documentos na categoria } k}$$

²⁷ Em vez de classificar cada documento do acervo, pode ser de interesse do pesquisador somente ter informação sobre a proporção de documentos presentes em cada categoria. Entre os modelos disponíveis para mensuração de proporções está o ReadME, desenvolvido por Hopkins e King (2010).

Dado o número de zeros presentes na DTM, algumas contagens específicas de palavras nunca ocorrem no conjunto de dados. Grimmer e Stewart (2013) apontam que a solução comum é adicionar uma pequena quantidade a cada probabilidade. Desse modo, o classificador estimado para Naive Bayes fica:

$$\hat{f}(P_v) = \arg \max_k \left[\hat{p}(C_k) = \prod_{i=1}^M \hat{p}(P_{im} | C_k) \right]$$

Não sendo as palavras condicionalmente independentes, o classificador de Naive Bayes se ajusta perfeitamente ao primeiro princípio apresentado por Grimmer e Stewart (2013), de que todos os modelos quantitativos de análise de conteúdo estão errados, mas alguns são úteis²⁸.

3. Validação

Se o método de aprendizagem supervisionada aplicado tiver bom desempenho, ele será capaz de se assemelhar à classificação manual nas tarefas de codificação de documentos em categorias ou medir a proporção de documentos em categorias. Esse objetivo claro implica um padrão preciso para sua avaliação, ou seja: a comparação da saída da codificação automatizada com a saída da codificação manual. Logo, o procedimento de validação ideal se divide em três: (1) que o ajuste inicial do modelo seja realizado no conjunto de treinamento; (2) depois que um modelo final é escolhido, que um segundo conjunto de documentos

codificados manualmente – o conjunto de validação – seja usado para avaliar o desempenho do modelo; (3) que o modelo final seja então aplicado ao restante do acervo para completar a classificação. Para validação, pode ser aplicado o procedimento de validação cruzada – *cross-validation* (EFRON; GONG, 1983; HASTIE; TIBSHIRANI; FRIEDMAN, 2001). Nele, o conjunto de treinamento é particionado aleatoriamente em V ($v = 1, \dots, V$) grupos, procedimento conhecido como *V-fold cross-validation*. Para cada grupo v , o modelo é treinado nos outros $V-1$ grupos, depois aplicado ao grupo V -th para avaliação de seu desempenho.

*Categorias desconhecidas*²⁹

Vimos que com um conjunto definido de categorias os métodos de aprendizagem supervisionada auxiliam o pesquisador na tarefa de classificação do acervo. No entanto, não é difícil encontrar situações nas quais o conjunto de categorias não seja conhecido pelo pesquisador. Com base nos dados utilizados para aplicação deste artigo e analisados por Moreira (2016), pode, por exemplo, ser do interesse do pesquisador identificar quais tópicos são enfatizados pelos deputados federais nos discursos proferidos ao longo da 54ª legislatura (2011-2014)³⁰. Uma vez que a atividade do representante político se debruça sobre inúmeras esferas da sociedade e da vida, predeterminar categorias temáticas de fala dos deputados federais pode limitar o conhecimento a ser obtido sobre o acervo. Para

28 Naive Bayes é apenas um exemplo de uma rica literatura que inclui outros modelos como: Random Forests (BREIMAN, 2001), Support Vector Machines (VENABLES; RIPLEY, 2002) e redes neurais (BISHOP, 1995).

29 Dedicamos esta seção aos algoritmos de *fully automated clustering* (FAC). Para detalhes sobre algoritmos de *computer assisted clustering* (CAC), ver Grimmer e King (2011).

30 Esse exemplo prático será apresentado mais adiante nesta seção.

enfrentar esse desafio, a seguir apresentamos métodos de aprendizado não supervisionado (*unsupervised learning methods*).

Métodos de aprendizado não supervisionado (*unsupervised learning methods*)

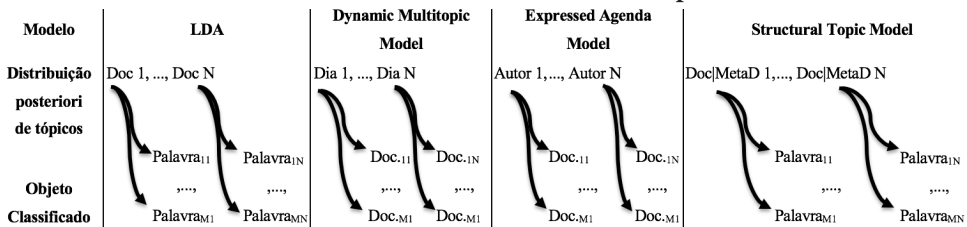
A classe de métodos de aprendizado não supervisionado revela características subjacentes ao texto sem a necessidade de imposição direta de categorias de interesse. Logo, ao invés de exigir que pesquisadores determinem previamente as categorias conhecidas, os métodos de aprendizado não supervisionado usam premissas e propriedades de modelagem dos textos para estimar um conjunto de categorias e, simultaneamente, atribuir a elas os documentos (ou partes de seu conteúdo). Ao informar para o algoritmo o número k de categorias nas quais os documentos devem ser alocados, há a oportunidade de se descobrir qual a composição de categorias que tenha melhor aderência ao conteúdo que está sendo analisado. Dada a incerteza do pesquisador sobre a performance do modelo em relação ao acervo que está sendo analisado, procedimentos de validação são essenciais.

Modelos de tópico (*topic models*)³¹. Para fins deste artigo, apresentaremos métodos de

aprendizagem não supervisionada também conhecidos como modelos de tópico (*topic models*). Os modelos de tópicos possuem duas principais características. Em primeiro lugar, definem estatisticamente um tópico como função densidade de probabilidade sobre palavras. Para um tópico k ($k=1, \dots, K$), essa função de probabilidade é representada com um vetor M^k , q_k , em que q_{mk} descreve a probabilidade de o k -ésimo tópico usar a m -ésima palavra. Logo, para estimar um tópico, os modelos usam a ocorrência de palavras entre documentos e pressupõem, em sua grande maioria, o uso da DTM obtida através do pré-processamento apresentado anteriormente.

Em segundo lugar, os modelos de tópicos compartilham uma estrutura hierárquica básica. Como apresenta a Figura 1, para cada modelo temos qual o elemento sobre o qual os tópicos estão distribuídos. Em outras palavras, o elemento que terá uma distribuição de tópicos que somada resulta em um. Em seguida, na parte inferior da hierarquia, palavras ou documentos que são atribuídos a um único tópico. O *Expressed Agenda Model*, por exemplo, pressupõe que cada documento seja classificado em apenas um tópico.

Figura 1
Estrutura comum entre modelos de tópico



Fonte: Adaptado de Grimmer e Stewart (2013, p. 18)

31 Estes modelos diferem-se de modelos de clusterização como o *K-means*. O objetivo do algoritmo K-means é identificar uma partição de documentos que minimiza o quadrado da distância euclidiana entre *clusters*. Para mais detalhes, ver MacQueen (1967). Para conhecer outros modelos, recomendamos consultar Grimmer e Stewart (2013).

Latent Dirichlet Allocation – LDA (BLEI; NG; JORDAN, 2003)³². Como primeiro e mais difundido modelo de tópico, o LDA (BLEI; NG; JORDAN, 2003) assume que cada documento é constituído por diferentes (uma mistura de) tópicos. Para cada documento i , p_{ik} representa a proporção do documento dedicada ao tópico k , sendo $p_i = (p_{i1}, p_{i2}, \dots, p_{ik})$ as proporções pelos tópicos. *A priori*, o LDA assume que a proporção de cada documento advém de uma distribuição Dirichlet, $p_i \sim \text{Dirichlet}(a)$, em que a representa o parâmetro da distribuição.

Em cada documento, as palavras são extraídas de acordo com a distribuição dos tópicos. Suponha que um documento contenha um total de palavras N_j , ($j=1, \dots, N_j$). Como Grimmer e Stewart (2013) apresentam, o LDA assume que um processo de duas etapas gera cada palavra. Para obter a j -ésima palavra no i -ésimo documento, o primeiro passo é extrair seu tópico, $t_{ij} \sim \text{Multinomial}(1, p_i)$. Condicional ao tópico atribuído, a palavra é extraída se a j -ésima palavra no i -ésimo documento está atribuída ao k -ésimo tópico, então extrai-se do tópico correspondente, $P_{ij} \sim \text{Multinomial}(1, q_k)$.

Foi com base na inovação promovida pelo LDA que a ciência política se apresentou mais uma vez na fronteira do conhecimento metodológico ao desenvolver outros modelos de tópico. Entre eles, o *Dynamic Multitopic Model* (QUINN et al., 2010), o *Expressed Agenda Model* (GRIMMER, 2010) e o *Structural Topic Model* (ROBERTS et al., 2013, 2014).

Resguardadas as inovações promovidas por cada um dos modelos, conforme a Figura 1, todos apresentam uma estrutura semelhante ao LDA.

***Dynamic Multitopic Model* (QUINN et al., 2010)**. Com base no caso dos discursos em plenário do Senado americano, o modelo supõe que a cada dia exista uma distribuição diferente de atenção a uma diversidade de tópicos. De forma análoga à atribuição de palavras a tópicos de determinado documento no LDA, cada discurso proferido no Senado americano é atribuído a um tópico. Por fim, uma *priori* dinâmica é usada para fazer inferências sobre a proporção de discursos de cada dia alocada a cada tópico.

***Expressed Agenda Model* (GRIMMER, 2010)**³³. Projetado para medir como os autores dividem sua atenção sobre temas, o modelo apresenta outra maneira de explorar a mesma estrutura inaugurada pelo LDA. Sua principal suposição é que cada autor divide sua atenção a um conjunto de tópicos. Assim, condicionado a tal distribuição de atenção dos autores, o tópico de cada documento é extraído. Mais adiante nesta seção será apresentado um exemplo de aplicação do *expressed agenda model* para o caso brasileiro.

***Structural Topic Model – STM* (MONROE et al., 2015; ROBERTS et al., 2013, 2014)**³⁴. Com base na mesma estrutura do LDA, o STM inova com duas características. Em primeiro lugar, permite que, ao nível dos documentos, seus metadados³⁵, entendidos

32 Para aplicação do LDA, recomendamos o pacote “topicmodels” da linguagem R (GRÜN; HORNIK, 2011). Para mais detalhes, ver: <<https://bit.ly/2o0VjZb>>. Acesso em: 3 jul. 2018.

33 Até o momento da elaboração deste artigo, o autor do modelo não disponibilizou o pacote em linguagem R prometido em Grimmer (2010).

34 Para estimação de tópicos com o STM, é possível utilizar gratuitamente o pacote “stm” na linguagem R (ROBERTS; STEWART; TINGLEY, 2018).

35 Metadados podem ser entendidos como dados sobre outros dados. Ou seja, informação que ajude a compreender características de um dado específico. No caso desta seção, nos referimos a dados que podem ajudar a compreender características dos documentos analisados como: data de publicação, autoria, georreferenciamento de onde foi criado, número de pessoas diferentes responsáveis pela produção do documento etc.

enquanto covariáveis, sejam incorporados ao modelo. Logo, informações como autoria e data de publicação podem ser incluídas para contribuir com a estimação dos tópicos. Em segundo lugar, ele permite estimar a correlação entre os tópicos de modo a identificar, por exemplo, quando dois tópicos podem ocorrer simultaneamente num documento. Tal informação pode auxiliar o pesquisador na identificação de temas que transcendam os tópicos.

Assim, com o objetivo de estimar a relação entre metadados e tópicos, no STM, estes são definidos como uma mistura sobre palavras em que cada palavra tem uma probabilidade de pertencer a um tópico. Logo, um documento é uma mistura sobre tópicos, o que significa que um único documento pode ser composto de vários tópicos do mesmo modo, como no caso do LDA.

O uso de modelos de aprendizagem não supervisionada

Como apontam Grimmer e Stewart (2013), todos os métodos de aprendizagem não supervisionada pressupõem duas etapas básicas após os procedimentos de pré-processamento apresentados: (1) a definição do número de categorias; e (2) validação.

1. **Definindo o número de categorias:** Determinar o número de categorias é uma das questões mais complexas no aprendizado não supervisionado³⁶. Ao mensurarem quão bem modelos se ajustam aos dados, medidas de ajuste estatístico tornam-se inúteis diante da brusca redução de informação que o uso de métodos não supervisionados para análise de conteúdo pressupõem após o pré-processamento dos

dados. Os textos pré-processados representam uma simplificação substancial dos documentos, sendo o objetivo do uso de métodos não supervisionados a revelação de informações substantivamente relevantes. Logo, em vez de ajuste estatístico, a seleção de modelos deve ser tratada como um problema de mensuração substantiva. Em linha com Grimmer e Stewart (Ibidem), acreditamos que a abordagem *mixed-method* fornecida por Quinn et al. (2010) é adequada. Nela, os modelos candidatos são ajustados variando-se o número de categorias para, em seguida, ser realizada uma avaliação manual e qualitativa para seleção do modelo final com base na qualidade das categorias obtidas.

2. **Validação:** Se, de um lado, o uso da aprendizagem não supervisionada reduz os custos de análise manual do acervo por parte do pesquisador antes do uso do modelo, de outro, a carga de trabalho para a validação de seus resultados é imensa. É a validação extensiva das categorias estimadas e dos documentos classificados que permite a realização de inferências concretas sobre o acervo. Veremos, no exemplo prático a seguir, abordagens para validação dos resultados do *expressed agenda model* aplicado ao caso brasileiro³⁷.

Aplicação para o caso brasileiro – *expressed agenda model*

O *expressed agenda model* foi utilizado por Moreira (2016) com o objetivo de identificar as ênfases temáticas proferidas pelos deputados

36 Alguns métodos tentam eliminar essa decisão e estimar o número de categorias (FREY; DUECK, 2007), mas estudos mostram que o número estimado é fortemente dependente do modelo (WALLACH et al., 2010). Também não é conveniente verificar medidas de ajuste do modelo (CHANG et al., 2009).

37 É importante apontar que as validações realizadas aqui são apenas um subconjunto do que pode ser empregado.

Federais em seus discursos no Pequeno Expediente ao longo da 54ª legislatura (2011-2014).

1. **Definição do número de tópicos:**

Uma vez realizado o pré-processamento dos dados, de acordo com os procedimentos apresentados anteriormente³⁸, o primeiro desafio imposto pelo modelo é a definição do número k de tópicos presente no *corpus*, ou seja, a quantidade de temas abordados em cada uma das legislaturas analisadas.

Para a definição do número k de tópicos, utilizaram-se duas estratégias: (1) o uso de um modelo não paramétrico para clusterização de texto baseado no *Dirichlet process prior* (BLEI; LAFFERTY, 2006; GRIMMER, 2010); e (2) a estimação de diferentes modelos.

O modelo não paramétrico resultou em 36 tópicos contidos no acervo. No entanto, dadas as ponderações já apresentadas, esse resultado não foi considerado de forma definitiva e a estipulação da quantidade k de tópicos contou com uma avaliação qualitativa do resultado de diferentes modelos para cada legislatura. A avaliação qualitativa permite que o valor k seja definido pela coesão substantiva identificada pelo analista através da análise dos *stems* mais associados a cada tópico em diferentes modelos e da leitura de amostras aleatórias de documentos presentes nas categorias estimadas por cada modelo.

Foram estimados e analisados os resultados de modelos que variaram de 5 a 80 tópicos. Por um lado, comparados entre si, quanto menor o número de tópicos de um modelo, maior é a diversidade de discursos

classificados em cada um, resultando em categorias muito genéricas. Por outro, quanto maior o número de tópicos do modelo, maior é a quantidade de tópicos tratando sobre o mesmo tema. Por essa razão, com o auxílio da evidência estatística do modelo não paramétrico, foi possível analisar de forma qualitativa os resultados dos 75 modelos estimados para a definição de um resultado de 39 tópicos para a 54ª legislatura. O resultado do modelo com as 39 categorias podem ser encontrados na Tabela 3. Na primeira coluna apresenta-se o rótulo dado a cada tópico após a leitura de uma amostra de ao menos dez discursos aleatoriamente selecionados de cada um deles. Na segunda coluna, é possível verificar até o quinto *stem* com maior informação mútua³⁹ em cada tópico. Na terceira, é apresentado o percentual de documentos do *corpus* classificado em cada um dos tópicos.

2. **Validação:**

Para que o resultado apresentado fosse considerado consistente, distintas formas de validação foram adotadas para averiguar se os resultados são substantivamente relevantes. Os tópicos foram validados por meio de quatro procedimentos: (1) dado que a matéria-prima para a análise dos tópicos são os *stems* das palavras contidas nos discursos, verificou-se quais os dez *stems* mais associados a cada um pelo do cálculo de sua informação mútua (GRIMMER, 2010, 2013); (2) foram lidos ao menos dez discursos aleatoriamente selecionados para rotulação de cada tópico; (3) sendo cada discurso pertencente a um tópico, é

38 Além da DTM desenvolvida e já apresentada, o *expressed agenda model* também solicita uma matriz adicional que informe o índice do primeiro e do último discurso de cada um dos deputados na DTM, pressupondo que as linhas desta estejam organizadas por orador com mais de um discurso.

39 A informação mútua entre um tópico e um *stem* mede a quantidade de informação que este provê sobre a possibilidade de um tópico gerar um documento aleatoriamente selecionado no *corpus*.

analisada sua pertinência temporal conforme a frequência dos tópicos ao longo da legislatura; (4) é qualitativamente analisada a dedicação de parlamentares selecionados a tópicos específicos, de modo que seja

possível identificar se há coerência entre a classificação temática dos discursos e perfis parlamentares amplamente conhecidos e difundidos na sociedade e na ciência política brasileira.

Tabela 3
Temas dos discursos proferidos na legislatura 54

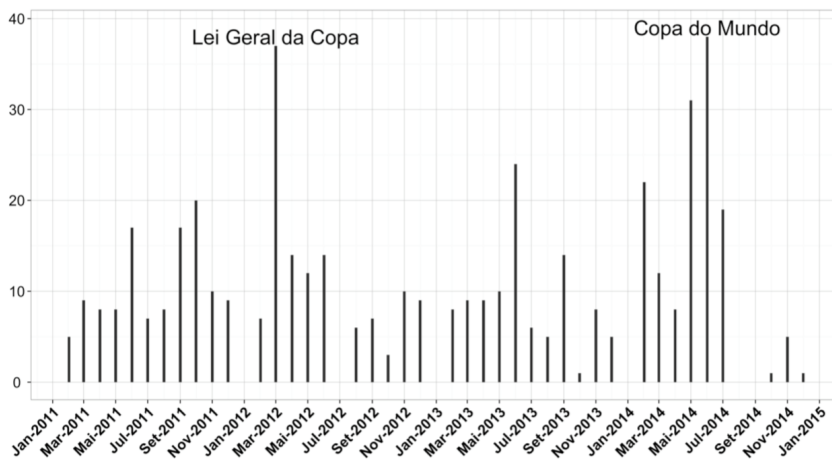
| Rótulo | Stems | % |
|---------------------------------|---|----------|
| <i>Votação</i> | vot, mat, votaca, sim, projet | 6,1 |
| <i>Datas comemorativas</i> | dia, registr, jos, vid, jornal | 6,0 |
| <i>Trabalho</i> | trabalh, direit, empreg, dia, projet | 4,9 |
| <i>Questões municipais</i> | municipi, prefeit, nov, cidad, recurs | 4,9 |
| <i>Protestos e corrupção</i> | pov, polit, trabalh, precis, dilm | 4,8 |
| <i>Projetos de lei</i> | projet, lei, aprov, trabalh, comissa | 4,5 |
| <i>Questões regionais</i> | ciudad, prefeit, municipi, regia, trabalh | 4,0 |
| <i>Educação</i> | educaca, escol, professor, ensin, alun | 3,9 |
| <i>Economia</i> | econom, setor, polit, ano, invest | 3,8 |
| <i>Sistema político</i> | polit, reform, vot, eleitoral, campanh | 3,4 |
| <i>Agentes de saúde</i> | saud, agent, atend, recurs, trabalh | 3,4 |
| <i>Medida Provisória</i> | med, provisor, vot, emend, lei | 3,1 |
| <i>Questões regimentais</i> | comissa, lid, titul, art, suplent | 3,1 |
| <i>Segurança pública</i> | polic, seguranc, milit, trabalh, crim | 2,9 |
| <i>Empresas</i> | empres, petrobr, milho, trabalh, servic | 2,8 |
| <i>Transporte</i> | transport, rodov, port, sant, sul | 2,6 |
| <i>Agropecuária</i> | agricultur, produtor, produca, produt, famili | 2,6 |
| <i>Sistema de Justiça</i> | justic, tribunal, suprem, defensor, process | 2,5 |
| <i>Recursos e investimentos</i> | milho, mil, invest, municipi, recurs | 2,3 |
| <i>Direitos Humanos</i> | direit, human, comissa, pesso, trabalh | 2,3 |
| <i>Gênero</i> | mulh, violenc, polit, trabalh, dia | 2,2 |
| <i>Pessoa com deficiência</i> | pesso, deficienc, trabalh, direit, vid | 2,1 |
| <i>Crédito financeiro</i> | banc, nord, econom, jur, trabalh | 1,9 |
| <i>Meio ambiente</i> | ambient, mei, ambiental, are, codig | 1,8 |
| <i>Programas federais</i> | program, famil, bols, social, rend | 1,7 |
| <i>Educação superior</i> | univers, educaca, ensin, curs, estud | 1,7 |
| <i>Questão hídrica</i> | agu, sec, nord, regia, municipi | 1,6 |
| <i>Saúde</i> | saud, hospital, atend, canc, doenc | 1,5 |
| <i>Esporte</i> | cop, mund, esport, futebol, estadi | 1,4 |
| <i>Medicina</i> | medic, saud, trabalh, program, atend | 1,4 |
| <i>Criança e adolescente</i> | crianc, adolescent, violenc, direit, sexual | 1,3 |
| <i>Questão indígena</i> | indigen, terr, indi, pov, direit | 1,2 |
| <i>Servidor público</i> | servidor, trabalh, servic, pec, direit | 1,2 |
| <i>Energia</i> | energ, eletr, consumidor, tarif, cont | 1,1 |
| <i>Amazônia</i> | amazon, manaus, zon, franc, regia | 1,1 |
| <i>Drogas e violência</i> | drog, crack, pesso, usuari, saud | 0,8 |
| <i>Idoso e Previdência</i> | idos, aposent, pesso, trabalh, projet | 0,8 |
| <i>Questão racial</i> | negr, dia, polit, populaca, racial | 0,7 |
| <i>Estatuto da Juventude</i> | juventud, jovens, polit, direit, trabalh | 0,5 |

Os resultados de dois dos quatro procedimentos de validação: a leitura atenta de uma amostra aleatória dos discursos presentes em cada tópico para rotulagem adequada e a análise de raízes com a maior informação mútua em cada um dos tópicos podem ser encontrados na Tabela 3. Avançamos a seguir, portanto, no sentido de avaliar a pertinência temporal dos tópicos e a ênfase temática esperada de alguns parlamentares de perfil amplamente conhecido pela sociedade brasileira e a ciência política nacional.

Pertinência temporal de tópicos selecionados. É possível averiguar a validade de um tópico por meio de sua pertinência temporal. Verifica-se se os discursos relacionados a cada tópico estão em acordo com debates desenvolvidos ao longo da legislatura e, em especial, se condizem com a ocorrência de eventos exógenos à instância de produção do discurso.

A paixão brasileira pelo futebol somada à escolha do país para sediar a Copa do Mundo FIFA de 2014 e a Copa as Confederações FIFA de 2013 produziu efeitos sobre toda atividade política nacional. Além da previsão de investimentos em transporte e infraestrutura, a atividade legislativa federal contou com a necessidade de produzir ordenamento jurídico específico para a realização do evento. Ainda assim, a atividade parlamentar a seu respeito não se restringiu à aprovação de dispositivos legais para sua execução. Conforme a Figura 2, apresenta, em 2012, para aprovação da Lei Geral da Copa (Lei Ordinária 12.663/12), mas, sobretudo, em 2014 – ano do evento – as falas proferidas no Pequeno Expediente trataram de destacar a temática do esporte, a relevância da Copa do Mundo em território nacional e seus efeitos políticos.

Figura 2
Pronunciamentos classificados na categoria Esporte ao longo da 54ª Legislatura



Ênfase temática de deputados federais selecionados. A principal contribuição do *expressed agenda model*, em comparação

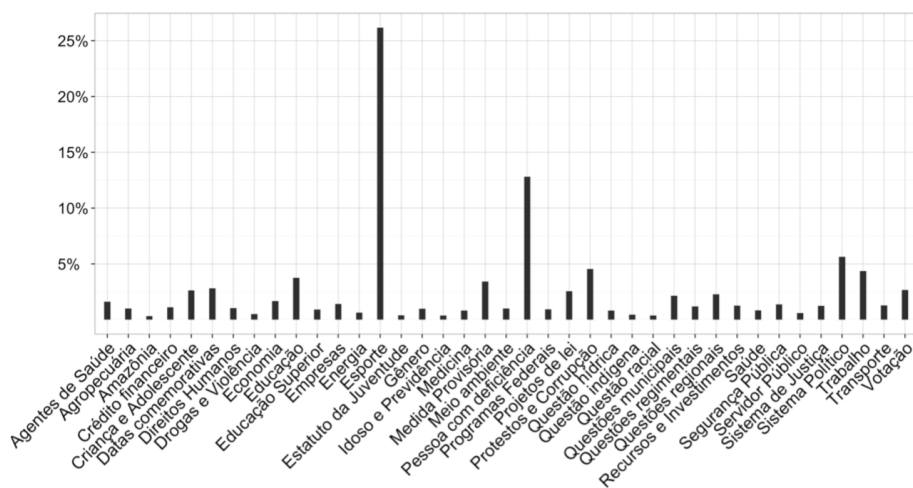
com as demais metodologias utilizadas na classificação de conteúdo de forma não supervisionada, é sua estrutura hierárquica

que permite identificar a ênfase de temática de autores⁴⁰. Por tal razão, como última estratégia de validação dos resultados obtidos pelo modelo para as legislaturas analisadas, foram averiguadas as ênfases temáticas de deputados federais cujo perfil é amplamente difundido e conhecido na ciência política nacional.

O então deputado federal Romário, do Partido Socialista Brasileiro (PSB-RJ), era conhecido em todo o país em função de sua atuação como jogador de futebol. Durante seu mandato como deputado federal, o ex-jogador elegeu dois grandes temas para sua atuação:

o esporte e as pessoas com deficiência. Foi autor de 21 projetos de lei, entre eles o PL4129/12, que institui a Semana Olímpica nas escolas públicas, e o PL 7916/14, que dispõe sobre a contratação de Apaes e Pestalozzis – entidades sem fins lucrativos – como prestadoras de serviços do Poder Público, com especialização em educação especial. Ao longo da 54ª legislatura, o deputado federal Romário realizou 26 pronunciamentos no Pequeno Expediente, e a Figura 3 apresenta a ênfase temática de seus discursos estimada em conjunto com a dos demais oradores dessa legislatura.

Figura 3
Ênfase temática dos pronunciamentos realizados pelo deputado federal Romário (PSB-RJ) na 54ª legislatura



40 No primeiro *Hackathon* da Câmara dos Deputados, realizado no mês de outubro de 2013, foi desenvolvida uma aplicação *web* que permite ao usuário saber qual tema cada deputado federal mais enfatizou em seus discursos e, ainda, qual a importância daquele tema para determinado parlamentar quando comparado aos demais. O projeto Retórica Parlamentar, assim denominado, foi desenvolvido por Davi Moreira, pelo doutor em Ciência Política, Manoel Galdino, e por Luis Carli, então doutorando em Visualização de Dados pela Faculdade de Arquitetura e Urbanismo, ambos pela Universidade de São Paulo. A equipe do Laboratório Hacker e do Departamento de Taquigrafia, Revisão e Redação da Câmara dos Deputados deu continuidade ao projeto e ele pode ser visto em: <<https://bit.ly/2i9kOEu>> e <<https://bit.ly/2OYDOnJ>>. Acesso em: 7 jun. 2018.

Como se pode constatar, as falas proferidas pelo deputado federal Romário possuem nítida relação com sua atuação parlamentar.

Em conjunto com os outros dois procedimentos de validação, a ênfase temática dos deputados federais identificada pelo *expressed agenda model* e ilustrada com a apresentação dos resultados para o tema do esporte ao longo da legislatura e os pronunciamentos do deputado federal Romário indica que o modelo estimado foi satisfatório.

Atribuir textos a categorias é o uso mais comum de métodos de análise de conteúdo na ciências sociais. Vimos nesta seção que os métodos automatizados podem mitigar o custo da atribuição de documentos às categorias e amplificar a quantidade de classificação que os humanos executam de forma manual. Graças ao desenvolvimento tecnológico e o baixo custo dos equipamentos informáticos domésticos, tarefas antes impossíveis e questões de pesquisas antes restritas a um grande montante de recursos humanos e financeiros podem agora ser acessadas com sucesso.

Métodos de escalonamento

A utilização da análise quantitativa de textos para a extração de posições políticas/ideológicas de partidos, políticos e eleitores é uma área extremamente promissora na ciência política. O teste de modelos de competição partidária, por exemplo, depende do conhecimento das posições dos principais atores envolvidos no jogo político.

Se, por um lado, existem diversos modelos que estimam os “pontos ideais” desses atores a partir de votações nominais (CLINTON; JACKMAN; RIVERS, 2004; POOLE; ROSENTHAL, 2007) e pesquisas de opinião (ALDRICH; MCKELVEY, 1973; POWER; ZUCCO, 2009), por outro, a utilização e disponibilidade desses dados são limitadas. Votações nominais representam apenas uma

pequena amostra daquilo que é discutido e decidido no interior de um parlamento (CARRUBBA et al., 2006). Nesse sentido, as posições políticas extraídas a partir delas são mais um produto do que a causa do processo político sob investigação. Além disso, apenas membros do corpo legislativo participam de votações nominais. Importantes atores como presidentes, ministros e eleitores não podem ter suas posições políticas estimadas. Pesquisas de opinião também são problemáticas. Elas possuem problemas de comparabilidade interpessoal (BRADY, 1985; KING et al., 2004) e são limitadas temporalmente. Não podemos voltar ao passado e conduzir pesquisas com atores políticos de outras épocas.

O emprego de textos como fonte primária para a estimação de posições políticas não é apenas uma alternativa devido à limitação de outras fontes. Em realidade, a utilização de palavras – sejam elas escritas ou faladas – é a forma mais básica de como o conflito político é expressado (GRIMMER; STEWART, 2013; MONROE; SCHRODT, 2009). Antes de qualquer votação nominal, parlamentares discutem a matéria em plenário. Partidos apresentam suas ideias em programas. Cidadãos discutem política nas redes sociais. Em resumo, é por meio de palavras que os indivíduos podem expressar suas preferências políticas e é através de seu uso estratégico que a política se materializa.

Uma das iniciativas pioneiras no uso de textos para inferir posições políticas foi o *Comparative Manifesto Project* (BUDGE et al., 2001). Esse projeto utiliza técnicas de análise de conteúdo para codificar os programas de mais de mil partidos políticos desde 1945 até hoje em mais de cinquenta países. Essa codificação é empreendida manualmente por uma equipe formada por um grande número de indivíduos treinados. Como é possível imaginar, esse projeto envolve grande quantidade de recursos financeiros que dificilmente estão disponíveis para a maior parte dos pesquisadores.

No entanto, o emprego de técnicas que dependem quase exclusivamente de recursos computacionais tem tornado a tarefa de estimar posições políticas a partir de textos acessíveis a qualquer pesquisador⁴¹. Atualmente, as duas técnicas mais populares são o *Wordscores* (LAVER; BENOIT; GARRY, 2003) e o *Wordfish* (SLAPIN; PROKSCH, 2008).

Wordscores

O *Wordscores* é um algoritmo supervisionado para estimar posições políticas (LAVER; BENOIT; GARRY, 2003). Nessa família de algoritmos, são apresentados ao computador alguns dados de entrada e as saídas esperadas. Chamamos esse conjunto de entradas e saídas de *training set*. A partir desse conjunto de informações o algoritmo “aprende” a classificar novos documentos. Esses formam o conjunto do *test set*. Assim, no *Wordscores* temos dois conjuntos de textos. O primeiro é formado pelos textos de referência (*training set*). Nele temos documentos cujas posições políticas são definidas, *a priori*, em uma dimensão conhecida pelo analista. Na grande maioria dos casos, essa dimensão está associada à escala ideológica esquerda-direita. O segundo conjunto é formado pelos textos cujas posições políticas são desconhecidas (*test set*), mas gostaríamos de conhecer. Observamos apenas o número de vezes que cada palavra aparece em cada texto. De modo intuitivo, o algoritmo classifica os documentos do *test set* em um contínuo entre os documentos de referência a partir da similaridade da frequência relativa de palavras.

A primeira etapa para implementar o *Wordscores* é escolher os textos de referência e definir quais são suas posições políticas. Essa etapa é fundamental e envolve

o conhecimento substantivo do contexto no qual os dados são gerados. Os autores fornecem algumas diretrizes para a escolha desses textos. Em primeiro lugar, é importante que os textos de referência utilizem o mesmo léxico que os textos virgens. Por exemplo, se queremos classificar discursos de parlamentares, é recomendável que os textos de referência também sejam discursos de parlamentares. Textos de diferente natureza, como programas de partidos, utilizam um conjunto de palavras muito diferente do empregado em discursos, portanto, trazem poucas informações. A segunda orientação é selecionar textos que cubram todo o espectro ideológico. Idealmente, é recomendável escolher textos que ocupem os extremos da escala, além da posição central. Por fim, a última recomendação é que os textos de referência possuam um conjunto diversificado de palavras. Assim, devemos evitar o uso de documentos curtos como textos de referência, porque os do *test set* serão analisados no contexto do universo de palavras dos textos de referência (*training set*).

Após a escolha dos textos devemos atribuir valores às suas posições políticas. Por exemplo, se nossos textos de referência são discursos de um partido de esquerda, um de direita e um de centro, podemos atribuir os valores de -1, 1 e 0 para cada partido, respectivamente. Outra possibilidade é utilizar as posições estimadas a partir de outros dados, como pesquisas de opinião. Com isso, completamos nosso *training set*.

A segunda etapa é gerar *scores* para as palavras dos textos de referência. Esse *score* é a média da posição política atribuída *a priori* (-1, 1 ou 0, no exemplo acima) ponderada pela probabilidade de observar um documento,

41 É possível aplicar o *Wordscores* e o *Wordfish* utilizando o pacote “austin” do R (LOWE, 2015). Para mais detalhes, ver: <<https://bit.ly/2MsqgnN>>. Acesso em: 7 jul. 2018.

dado que estamos analisando uma palavra em particular. Isto é, $S_{pd} = \sum_r (P(r|p) \times A_{pd})$, em

que $P(r|p) = \frac{P(r|p)P(p)}{P(p)}$ é a probabilidade

de observar um documento de referência r , dado que estamos observando a palavra p ; e A_{pd} é a posição política da palavra p na dimensão política d .

Por exemplo, suponha que os textos de esquerda, centro e direita tenham mil palavras cada. Suponha também que a palavra “copa” apareça dez vezes no texto de esquerda, vinte no de direita e trinta no de esquerda. A probabilidade de observarmos o documento de esquerda, dado que observamos a palavra “copa”, é de 17% (0,01 / 0,06). Para o documento de direita temos 33% (0,02 / 0,06) e para o documento de centro temos 50% (0,03 / 0,06).

Assim, o *score* da palavra “copa” será: $0,17(-1) + 0,5(0) + 0,33(1) = 0,16$. Isto é, dada a frequência relativa de palavras nos textos de referência, se soubéssemos apenas que a palavra “copa” aparecia em um documento qualquer, esperaríamos que sua posição política seria de 0,16.

Com o *score* para todas as palavras no universo dos textos de referência podemos estimar a posição política dos textos virgens. Essa terceira etapa nada mais é do que calcular o *score* médio das palavras ponderando pela frequência relativa de palavras em cada documento virgem. Ou seja, $S_{vd} = \sum_p (P(p|v) \times S_{pd})$, em que $P(p|v)$ é a probabilidade de observar a palavra p no documento virgem v .

Para colocar os textos do *test set* na mesma escala que os textos de referência, podemos aplicar a seguinte transformação: $S_{vd}^* = (S_{vd} - S_{vd}) \left(\frac{SD_{rd}}{SD_{vd}} \right) + S_{vd}$, em que S_{vd}^* é o

score médio dos textos do *test set* e SD_{rd} e SD_{vd} são os desvios-padrão amostrais dos textos de referência e virgens, respectivamente.

Podemos também calcular medidas de incerteza para os *scores*, como a variância: $V_{vd} = \sum_p P(r|p) (S_{pd} - S_{vd})^2$. Com isso é possível implementar testes como diferenças de médias e avaliar se as diferenças entre as posições políticas estimadas para dois documentos são estatisticamente significativas.

Embora o *Wordscores* constitua um grande avanço na análise quantitativa de textos, ele não é livre de problemas. O principal é o fato de ele depender fortemente da escolha dos textos de referência (*training set*). Em situações extremas é possível que, com a escolha diferente de textos de referência, um mesmo pesquisador encontre resultados diferentes para um mesmo conjunto de dados. A segunda limitação, como apontado por Lowe (2008), é a possibilidade de as diferenças entre os textos estarem mais relacionadas com o estilo linguístico do autor do que com as posições políticas. Como todas as palavras adicionam a mesma quantidade de informação sobre o documento, temos que palavras politicamente relevantes em um contexto sejam igualmente ponderadas a palavras pouco informativas.

Wordfish

A segunda técnica mais popular para estimar posições políticas a partir de textos é o *Wordfish* (SLAPIN; PROKSCH, 2008). Ao contrário do *Wordscores*, esse é um algoritmo não supervisionado pois não depende da escolha de textos de referência, ou seja, da construção de um *training set*. Dessa forma, não há a limitação de diferentes pesquisadores chegarem a resultados diferentes a partir do mesmo conjunto de dados. Outra vantagem em relação ao *Wordscores* é o fato de ele não atribuir o mesmo peso para todas as palavras. O *Wordfish* estima a importância das palavras para discriminar as posições políticas. Assim, palavras

politicamente relevantes em uma dimensão têm peso maior na tarefa de localizar os documentos no espectro político.

O *Wordfish* é baseado em modelos da TRI, tal como aqueles utilizados para estimação de pontos ideais a partir de votações nominais (CLINTON; JACKMAN; RIVERS, 2004). Mas, em vez de utilizar votos dados a projetos, ele opera com a frequência relativa de palavras. Nesse sentido, acredita-se que o uso relativo das palavras forneça informações relevantes sobre as posições políticas dos atores.

O modelo assume uma distribuição de Poisson para a contagem de palavras. Isto é, cada palavra j de um documento i , Y_{ij} , é gerada a partir de uma distribuição de Poisson com parâmetro $\lambda_{ij} > 0$, $Y_{ij} \sim \text{Poisson}(\lambda_{ij})$. A escolha dessa distribuição foi determinada por sua simplicidade. Ela possui apenas um parâmetro (λ) que, ao mesmo tempo, representa a média e a variância. Outra coisa importante de se notar é a existência um pressuposto de que a probabilidade de observamos uma palavra em um documento é independente da posição das outras palavras no mesmo documento. Embora esse pressuposto seja falso, ele é frequentemente utilizado na análise quantitativa de textos.

A forma funcional do modelo é dada por $\lambda_{ij} = \exp(\alpha_i + \psi_j + \beta_j \times \omega_i)$, em que α é um conjunto de efeitos fixos por documento. A inclusão desses parâmetros ocorre porque alguns documentos são mais longos do que outros; ψ é um conjunto de efeitos fixos por palavra. Sua inclusão é importante porque algumas palavras são mais frequentes do que outras. Já β é o parâmetro de discriminação, uma estimativa de quanto a palavra j é importante para distinguir as posições políticas. Por fim, ω é a estimativa da posição política. Esse é o principal parâmetro de interesse.

O principal interesse dos autores foi estimar as posições políticas dos partidos a partir de seus programas. Em geral, os programas

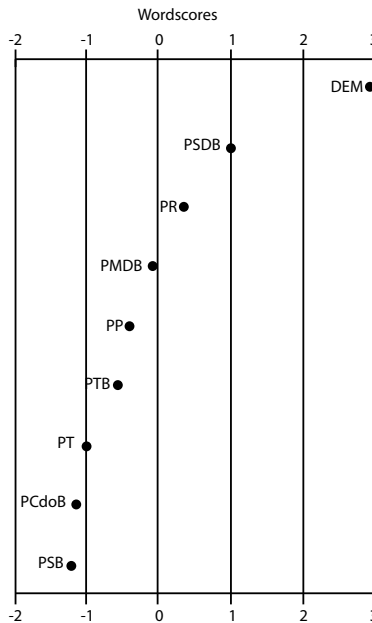
apresentam as posições dos partidos em um grande leque de assuntos. Portanto, a expectativa é a de que, ao aplicar o algoritmo, o resultado deve ser uma escala ideológica. Se o analista está interessado em extrair as posições dos partidos no que diz respeito a um tema específico, a primeira etapa deve ser selecionar os textos sobre esse tema.

A limitação desse modelo é o fato de ele necessitar que os documentos cubram uma grande quantidade de temas para extrair as posições ideológicas. Assim, se os parlamentares focarem seus discursos em determinadas áreas temáticas, provavelmente não teremos resultados consistentes ao aplicarmos o *Wordfish*. Como a variação no uso das palavras não será determinada pelas preferências políticas, mas pelos tópicos, a diferença nas posições estimadas também refletirá essa diferença (LAUDERDALE; HERZOG, 2016).

Aplicação para o caso brasileiro – Wordscores

Nesta seção aplicaremos o *Wordscores* aos discursos dos deputados durante o ano de 2011, utilizando a DTM obtida pelo pré-processamento apresentado anteriormente. Para essa aplicação, agrupamos os discursos por partido. Assim, todos os documentos de parlamentares de um mesmo partido compõem um único documento. O objetivo desse procedimento é tentar garantir que os textos tratem de uma grande quantidade de temas. Como individualmente cada parlamentar tende a focar seus esforços em determinadas áreas temáticas, ao agruparmos os discursos, temos uma variedade maior. Selecionamos também apenas os partidos com pelo menos 300 discursos. Ao todo temos nove partidos: Partido dos Trabalhadores (PT), Partido Trabalhista Brasileiro (PTB), Partido Progressista (PP), Movimento Democrático Brasileiro (PMDB), Partido da República (PR), Partido da Social Democracia Brasileira (PSDB), PSB, PCdoB e DEM.

Gráfico 1
Posições políticas dos partidos políticos brasileiros (2011)



Para essa aplicação, selecionamos como textos de referência os discursos do PT e do PSDB. Essa escolha parece ser adequada, já que no período atual os dois partidos têm polarizado as disputas políticas. Como é possível observar pelo Gráfico 1, o *Wordscores* classificou os principais partidos políticos em uma escala ideológica. Do lado esquerdo (negativo) temos PSB, PCdoB e PT, do lado direito (positivo) temos DEM, PSDB e PR, e no centro temos PTB, PP e PMDB.

Manter a metodologia das ciências sociais brasileiras na fronteira do conhecimento humano para que técnicas contemporâneas possam ser utilizadas para questões de pesquisa sobre o presente, o passado e o futuro é um dos objetivos desse artigo. Por isso, para além dos métodos apresentados até aqui, na próxima seção indicamos desenvolvimentos e avanços recentes que apontam para onde caminha o uso do texto como dado nas ciências sociais.

Desenvolvimentos e aplicações mais recentes

O volume de conteúdo publicado cresce rapidamente. Porém a forma de publicação não se restringe mais ao mundo físico como jornais, livros, panfletos etc., mas também – e especialmente – atinge o mundo virtual em mídias sociais e outros meios baseados na internet que reduzem custos de publicação e expressão social por meio do uso do texto como dado, ampliando as fontes de acervo para análises em ciências sociais (BARBERÁ, 2012). Não há sinal de que essa tendência mude e, como aponta Simon Jackman (MAGALHÃES et al., 2013), ao ser perguntado sobre as perspectivas em metodológica em ciência política:

“Eu acredito que tratar textos como dados é algo grande. Isso significa que tudo é dado. Há dados em todos os lugares. Textos vem juntos com análise dimensional. O que é uma lei? Como leis se

relacionam entre si? O que é um discurso político? Qual é a retórica de um político? Como você faria um trabalho quantitativo sobre isso? Eu acredito que isso vai ser grande.”

As ciências sociais se mantêm nessa fronteira do conhecimento e avanços não param de ocorrer. A seguir, apresentamos uma visão geral da tendência no campo de análise do texto como dado.

Olhar o passado com as lentes do presente

Sem dúvida alguma uma das áreas mais promissoras de uso e aplicação dos métodos de análise automatizada de conteúdo é a análise de processos históricos⁴². Como o título dessa subseção apresenta, há, hoje, a oportunidade de abordar questões de pesquisa sobre o passado com a possibilidade de aplicação do ferramental metodológico contemporâneo. A seguir, veremos dois exemplos dessa aplicação.

Medindo novidade, transitoriedade e ressonância

Em trabalho recente, Barron et al. (2018) analisaram mais de 40 mil discursos durante os debates no interior do parlamento da Revolução Francesa. Os autores traçaram a criação, destruição e propagação de ideias a partir dos padrões de uso das palavras. Os resultados corroboram evidências qualitativas de que parlamentares da esquerda traziam inovações aos debates, ao passo que parlamentares de direita agiam de modo a preservar os padrões anteriores. Esse processo foi dirigido em parte por algumas importantes figuras da época,

como Robespierre e Pétion de Villeneuve. Esses políticos radicais não apenas introduziram novas estratégias discursivas e padrões de uso de palavras em taxas maiores que os outros parlamentares, mas também o fizeram de modo que suas ideias se mantiveram ao longo do tempo.

Discursos do presente, que são muito diferentes de discursos proferidos no passado, indicam que novas ideias são trazidas ao debate político. Por outro lado, grandes desvios dos padrões discursivos de hoje em relação a padrões do futuro indicam a natureza transitória dessas ideias. Discursos que trazem novos padrões ao debate e levam a discussão para uma nova direção são aqueles que têm ressonância. Ou seja, o desequilíbrio entre alta novidade e baixa transitoriedade é o que caracteriza a ressonância de uma ideia.

A partir de uma definição bayesiana de surpresa (ITTI; BALDI, 2006), uma medida de divergência entre as distribuições *a priori* e *a posteriori* é utilizada para análise. Em outras palavras, dizemos que novos dados observados (D) trazem uma surpresa quando a distribuição, *a posteriori* resultante da observação desses dados, é significativamente diferente da distribuição *a priori*⁴³. Assim, desenvolvem medidas de novidade, transitoriedade e ressonância.

Os autores classificaram os discursos em tópicos (K = 100) utilizando o LDA e analisaram como a combinação desses tópicos se desvia de discursos anteriores (novidade) e posteriores (transitoriedade). Grandes desvios comparando padrões de hoje com os do passado indicam que os tópicos são novos (novidade). Grandes desvios comparando

42 Uma iniciativa interessante com este enfoque pode ser encontrada em: <http://www.history-lab.org/>. Acessado em 09 de julho de 2018.

43 Essa ideia é operacionalizada por meio de uma medida de entropia relativa ou divergência de Kullback-Leibler (BARRON et al., 2018).

padrões de hoje com os do futuro indicam que os padrões não foram retidos (transitoriedade). Por sua vez, discursos que trazem novos padrões ao debate e levam a discussão a uma nova direção são aqueles que possuem ressonância.

Medindo complexidade linguística

Spirling (2015) estimou o impacto da expansão do sufrágio sobre a complexidade linguística dos discursos dos membros do parlamento inglês na era vitoriana. Com a *Second Reform Act*, de 1867, houve a inclusão de grande contingente de novos eleitores. Além disso, houve redução dos requerimentos necessários, em termos de propriedade, para participar do processo eleitoral. Com isso, o resultado da Reforma foi a extensão do sufrágio para uma grande massa de trabalhadores urbanos pouco educados, quando não analfabetos. Em reação a esse novo cenário eleitoral, os parlamentares reduziram a complexidade de seus discursos para atingir esses eleitores.

Para medir essa mudança de graus na compreensão dos discursos o autor lançou mão de métricas comumente utilizadas em pesquisas na área de educação. Essas medidas levam em consideração a relação entre o número de sílabas e o número de palavras em um documento (FLESCH, 1948).

Os autores analisaram mais de 650 mil discursos proferidos por mais de 3.500 parlamentares entre 1832 e 1915. Os resultados apontam que, logo após a Reforma, os parlamentares alteraram seus discursos de modo a torná-los de mais fácil compreensão para o novo eleitor mediano, alguém mais pobre e menos educado do que o eleitor mediado anterior.

Big Data, o texto como dado e causalidade

Enquanto lê esta sentença, tente imaginar quantas buscas no Google foram realizadas e registradas ao redor do planeta. Na era do *Big Data* a todo instante volumes massivos de dados são produzidos (LAZER et al., 2009) e as ciências sociais, com o uso da análise automatizada de conteúdo, têm papel crucial na transformação desses dados em informação e conhecimento. O *Big Data* oferece a oportunidade de produzir conhecimento a partir de um volume de dados inviável há apenas alguns anos.

Diante desse desafio, para além de conhecimentos computacionais, analisar habilmente a massa de conteúdo que tem sido intensamente produzida também exige uma avaliação rigorosa (PATTY; PENN, 2015), um desenho de pesquisa cuidadoso e a implementação criativa de técnicas estatísticas (GRIMMER, 2015). Como Grimmer (Idem) aponta, para que a análise do *Big Data* realmente forneça respostas aos problemas da sociedade, deve-se reconhecer que ela é tanto sobre ciência social quanto sobre ciência da computação.

O papel da descrição. Vimos neste artigo que o ganho de escala obtido por meio do uso dos métodos apresentados se coloca como ferramenta promissora diante dos desafios do *Big Data*. Assim, as oportunidades para inferências descritivas são abundantes em *Big Data*, têm potencial enorme para reorientar teorias e questões de inferência causal estabelecidas nas ciências sociais⁴⁴. Conforme Grimmer (2015), a análise de coleções de conteúdo, de acervos de notícias, de postagens em mídias sociais, entre outros, podem contribuir para respostas a importantes questões das ciências

⁴⁴ No âmbito da ciência política, o projeto *VoteView* talvez seja o melhor exemplo de como projetos puramente descritivos afetam teorias e as questões de inferência causal da literatura (MCCARTY; POOLE; ROSENTHAL, 2006; POOLE; ROSENTHAL, 2007).

sociais como: o conhecimento sobre a agenda da grande mídia, o quanto a política está presente nas redes sociais, o posicionamento do público a respeito de um tema específico, ou reações diante de propostas de campanha, bem como as próprias campanhas eleitorais realizadas em âmbito virtual. A chance de inferência descritiva de conteúdo em tamanha escala cria para as ciências sociais a oportunidade de fazerem perguntas causais e criarem teorias anteriormente impossíveis (MONROE et al., 2015).

Causalidade. Diante dos desafios da *Big Data*, ser capaz de aplicar técnicas estatísticas a conjuntos de dados massivos para a obtenção de informações descritivas do conteúdo publicado é apenas o princípio da contribuição que as ciências sociais podem dar a essa revolução computacional. Somada a esse princípio, está a oportunidade do uso de metodologias de análise automatizada de conteúdo para a identificação de efeitos causais baseados no texto como dado. Combinados a experimentos, o uso do texto como dado pode ser útil para a descoberta de medidas que testem teorias de interesse das ciências sociais a partir de grandes coleções de texto. Trabalhos que conectam a literatura do texto como dado (LAVER; BENOIT; GARRY, 2003; PENNEBAKER; MEHL; NIEDERHOFFER, 2003; QUINN et al., 2010), com a crescente literatura sobre inferência causal nas ciências sociais (HERNAN; ROBINS, 2018; IMBENS; RUBIN, 2015; PEARL, 2009) nesse sentido, já têm sido publicados.

Egami et. al. (2018) apresentam uma estrutura conceitual para fazer inferências causais com tratamentos obtidos a partir da análise automatizada de conteúdo, fornecendo uma estrutura rigorosa para inferências causais baseadas em texto. Em linha semelhante, Fong e Grimmer (2016) apresentam um novo modelo experimental e um modelo estatístico para, simultaneamente, descobrir

tratamentos presentes num acervo e, ainda, estimar seus efeitos causais. De forma adicional, focados em surveys com questões de respostas abertas, Roberts et al. (2014) mostram como o fato de o STM admitir a inclusão de covariáveis pode ser útil para a análise de respostas a experimentos em survey, concluindo que tal abordagem possui grau razoável de sucesso quando comparada à codificação manual.

A combinação do potencial da análise automatizada de conteúdo com a era do *Big Data* e a oportunidade de produção de inferências causais é, sem dúvida alguma, um dos campos mais promissores para a área de metodologia em ciências sociais. Certamente, as ciências sociais brasileiras têm muito a contribuir com esse processo e deve se manter na fronteira do conhecimento.

Considerações finais

O principal objetivo deste artigo foi apresentar ao leitor um leque atualizado das principais metodologias de análise automatizada de conteúdo. Sem esgotar a atual variedade de métodos, técnicas e modelos, trata-se de um guia inicial para essa intensa e instigante área de pesquisa. No Quadro 3 apresentamos uma visão geral dos principais métodos para análise quantitativa de textos revisados neste trabalho.

Como buscamos destacar ao longo de todo o artigo, o uso de métodos automatizados para análise do texto como dado (*text as data*) é algo ainda muito recente na história da humanidade. Mesmo sob a contribuição de diferentes áreas do conhecimento, sendo a análise de conteúdo um campo tradicional de dedicação das ciências sociais, não há melhor área para guiar e contribuir com esse avanço. Muito ainda será feito e consideramos importante que as ciências sociais brasileiras acompanhem a fronteira desse processo.

Quadro 3
Visão geral dos métodos para análise quantitativa de textos

| Família | Técnica | Objetivo |
|--|---|---|
| Semelhança entre textos | Similaridade de cosseno | Medir quão similares são dois documentos. |
| | Algoritmo de Smith-Waterman | Encontrar quais os trechos mais similares entre dois documentos. |
| Métodos de classificação em categorias conhecidas | Dicionário (Análise de sentimentos) | Classificar documentos em categorias conhecidas com auxílio de um dicionário anotado. |
| | Supervisionado (classificador de Naive Bayes) | Classificar documentos em categorias conhecidas a partir de um conjunto de treinamento. |
| Métodos de classificação em categorias desconhecidas | Não supervisionado (LDA, Dynamic Multitopic Model, Expressed Agenda Model, STM) | Classificar documentos quando não se conhece as categorias previamente. |
| Métodos de escalonamento | <i>Wordscores</i> | Estimar posições políticas em uma dimensão predeterminada a partir de documentos de referência. |
| | <i>Wordfish</i> | Estimar posições políticas quando não se conhece previamente referências da dimensão. |

Os trabalhos de Moreira (2016) e Izumi (2017), apresentados no início deste artigo, são um exemplo notável de que muito ainda pode e deve ser feito pelas ciências sociais no país. Se, de um lado, no Pequeno Expediente na Câmara dos Deputados há evidências consistentes para concluir que os temas enfatizados não são governados pela relação governo-oposição (MOREIRA, 2016), de outro, no Senado Federal o posicionamento político resgata a importância dessa variável (IZUMI, 2017). Afinal de contas, há um padrão geral de fala dos parlamentares no

Congresso Nacional? Para além dos temas, qual é o posicionamento político dos deputados federais nos discursos proferidos no Pequeno Expediente? Ademais, qual seria a ênfase temática presente nos discursos dos senadores?

As diferentes conclusões apresentadas nesses trabalhos e as inúmeras lacunas entre elas demonstram quão rica é esta agenda de pesquisa e quão importante é a dedicação das ciências sociais brasileiras a essa temática. A paisagem no horizonte é deslumbrante e está pronta para ser explorada.

Referências

- ALDRICH, J.; MCKELVEY, R. A method of scaling with applications to the 1968 and 1972 presidential elections. *The American Political Science Review*, Washington, DC, v. 71, n. 1, p. 11-130, 1973.
- BARBERÁ, P. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, Cambridge, UK, v. 23, n. 1, p. 76-91, 2015.
- BARRON, A. et al. Individuals, institutions, and innovation in the debates of the French Revolution. *Proceedings of the National Academy of Sciences*, Washington, DC, v. 115, n. 18, p. 4607-4612, 2018.

- BERINSKY, A.; HUBER, G.; LENZ, G. Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political Analysis*, Cambridge, UK, v. 20, n. 3, p. 351-368, 2012.
- BISHOP, C. *Neural networks for pattern recognition*. Gloucestershire: Clarendon Press, 1995.
- BLEI, D. M.; LAFFERTY, J. D. Dynamic topic models. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 23., 2006, New York. *Proceedings...* New York: ACM, 2006. pp. 113-120.
- BLEI, D.; NG, A.; JORDAN, M. Latent dirichlet allocation. *Journal of Machine Learning Research*, Cambridge, MA, v. 3, n. 1, p. 993-1022, 2003.
- BRADY, H. The perils of survey research: inter-personally incomparable responses. *Political Methodology*, Oxford, UK, v. 11, n. 3-4, p. 269-291, 1985.
- BREIMAN, L. Random forests. *Journal of Machine Learning Research*, Cambridge, MA, v. 45, n. 1, p. 5-32, 2001.
- BUDGE, I. et al. *Mapping policy preferences: estimates for parties, electors, and governments, 1945-1998*. Oxford, UK: Oxford University Press, 2001.
- CAMPBELL, S. PENNEBAKER, J. The secret life of pronouns flexibility in writing style and physical health. *Psychological Science*, Washington, DC, v. 14, n. 1, p. 600-65, 2003.
- CAMPOS, L. A., FERES JR., J.; GUARNIERI, F. 50 Anos da Revista DADOS: uma análise bibliométrica do seu perfil disciplinar e temático. *Dados*, Rio de Janeiro, v. 60, n. 3, p. 623-661, 2017.
- CARRUBBA, C. et al. Off the record: unrecorded legislative votes, selection bias and roll-call vote analysis. *British Journal of Political Science*, Cambridge, UK, v. 36, n. 4, p. 691-704, 2006.
- CHANG, J. et al. Reading tea leaves: how humans interpret topic models. In: BENGIO, Y. et al. *Advances in neural information processing systems*. Cambridge, MA: MIT Press, 2009. p. 288-296.
- CLINTON, J.; JACKMAN, S.; RIVERS, D. The statistical analysis of roll call data. *American Political Science Review*, Washington, DC, v. 98, n. 2, p. 355-370, 2004.
- EFRON, B.; GONG, G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *American Statistician*, Abingdon, v. 37, n. 1, p. 36-48, 1983.
- EGAMI, N. et al. *How to make causal inferences with text*. Working paper. 2018. Disponível em: <<https://bit.ly/2M-tXMdq>>. Acesso em: 21 jul. 2018.
- FEINERER, I. HORNIK, K. tm: Text Mining Package. *R package*, [s.l.], 2018. Disponível em: <<https://bit.ly/2K-cAx2w>>. Acesso em: 21 jul. 2018.
- FLESCH, R. A new readability yardstick. *Journal of Applied Psychology*, Washington, DC, v. 32, n. 3, p. 221-233, 1948.

- FOKKENS, A. et al. Offspring from reproduction problems: what replication failure teaches us. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 51., 2013, Sofia. *Proceedings...* Sofia: Association for Computational Linguistics, 2013. (Volume 1: Long Papers). p. 1691-1701.
- FONG, C.; GRIMMER, J. Discovery of treatments from text corpora. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 54., 2016, Berlin. *Proceedings...* Berlin: Association for Computational Linguistics, 2016. p. 1-10. FREY, B.; DUECK, D. Clustering by passing messages between data points. *Science*, Washington, DC, v. 315, n. 5814, p. 972-976, 2007.
- GARRETT, K.; JANSA, J. Interest group influence in policy diffusion networks. *State Politics & Policy Quarterly*, Thousand Oaks, v. 15, n. 3, p. 387-417, 2015.
- GRIMMER, J. A Bayesian hierarchical topic model for political texts: measuring expressed agendas in Senate press releases. *Political Analysis*, Cambridge, UK, v. 18, n. 1, p. 1-35, 2010.
- _____. We are all social scientists now: how big data, machine learning, and causal inference work together. *PS: Political Science & Politics*, Cambridge, UK, v. 48, n. 1, p. 80-83, 2015.
- GRIMMER, J.; KING, G. General purpose computer-assisted clustering and conceptualization. *Proceedings of the National Academy of Sciences*, Washington, CD, v. 108, n. 7, p. 2643-2650, 2011.
- GRIMMER, J.; STEWART, B. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, v. 21, n. 3, p. 267-297, 2013.
- GRÜN, B.; HORNIK, K. Topicmodels: AN R Package for fitting topic models. *Journal of Statistical Software*, Innsbruck, v. 40, n. 13, p. 1-30, 2011.
- HAND, D. Classifier technology and the illusion of progress. *Statistical Science*, Bethesda, v. 21, n. 1, p. 1-14, 2006.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning*. New York: Springer, 2001.
- HERNAN, M.; ROBINS, J. *Causal inference*. Boca Raton: CRC Press, 2018.
- HOPKINS, D. KING, G. A method of automated nonparametric content analysis for social science. *American Journal of Political Science*, Washington, DC, v. 54, n. 1, p. 229-247, 2010.
- HOPKINS, D. et al. ReadMe: software for automated content analysis. *Gari King*, Cambridge, MA, 2017. Disponível em: <<https://bit.ly/2Mq7HRI>>. Acesso em> 21 jul. 2018.
- IMBENS, G.; RUBIN, D. *Causal inference in statistics, social, and biomedical sciences*. Cambridge, UK: Cambridge University Press, 2015.
- ITTI, L.; BALDI, P. Bayesian surprise attracts human attention. In: JORDAN, M. I.; LECUN, Y.; SOLLA, S. A. (Eds.). *Advances in neural information processing systems: proceedings of the first 12 conferences*. Cambridge, MA: The MIT Press, 2006.

- IZUMI, M. *Velhas questões, novos métodos: posições, agenda, ideologia e dinheiro na política brasileira*. 2017. 113 f. Tese (Doutorado em Ciência Política) – Universidade de São Paulo, São Paulo, 2017.
- JURAFSKY, D.; MARTIN, J. *Speech and natural language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River: Prentice Hall, 2009.
- KING, G. et al. Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, Cambridge, UK, v. 98, n. 1, p. 191-207, 2004.
- KRIPPENDORFF, K. *Content analysis: an introduction to its methodology*. New York: Sage, 2004.
- KROEGER, M. *Plagiarizing policy: model legislation in state legislatures*. Working paper. 2015. Disponível em: <<https://bit.ly/2o0lpf5>>. Acesso em: 21 jul. 2018.
- LAUDERDALE, B. HERZOG, A. Measuring political positions from legislative speech. *Political Analysis*, Cambridge, UK, v. 24, n. 3, p. 374-394, 2016.
- LAVER, M.; BENOIT, K.; GARRY, J. Extracting policy positions from political texts using words as data. *American Political Science Review*, Washington, DC, v. 97, n. 2, p. 311-331, 2003.
- LAZER, D. et al. Life in the network: the coming age of computational social science. *Science*, New York, v. 323, n. 5915, p. 721, 2009.
- LI, W.; LAROCHELLE, D.; LO, A. *Estimating policy trajectories during the financial crisis*. Working paper. 2014. Disponível em: <<https://bit.ly/2MtZfjN>>. Acesso em: 21 jul. 2018.
- LIU, B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, London, v. 5, n. 1, p. 1-167, 2012.
- LOWE, W. Understanding wordscores. *Political Analysis*, Cambridge, UK, v. 16, n. 4, p. 356-371, 2008.
- _____. Austin: do things with words. *Conjugateprior*, Princeton, 2015. Disponível em: <<https://bit.ly/2BCFGAY>>. Acesso em: 21 jul. 2018.
- LUCAS, C. et al. Computer-assisted text analysis for comparative politics. *Political Analysis*, Cambridge, UK, v. 23, n. 2, p. 254-277, 2015.
- MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: LE CAM, L. M.; NEYMAN, J. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 1967. (Volume 1: Statistics). p. 281-297.
- MAGALHÃES, R. et al. Perspectives on political methodology: interview with Simon Jackman. *Leviathan*, São Paulo, n. 7, p.158-175, 2013.

- MANNING, C.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to information retrieval*. Cambridge: Cambridge University Press, 2008.
- MARON, M.; KUHN, J. On relevance, probabilistic indexing and information retrieval. *Journal of the ACM (JACM)*, New York, v. 7, n. 3, p. 216-244, 1960.
- MCCARTY, N. POOLE, K. ROSENTHAL, H. *Polarized America: the dance of ideology and unequal riches*, Cambridge: MIT University Press, 2006.
- MONROE, B.; SCHRODT, P. Introduction to the special issue: the statistical analysis of political text. *Political Analysis*, Cambridge, UK, v. 16, n. 4, p. 351-355, 2008.
- MONROE, B. et al. No! Formal theory, causal inference, and big data are not contradictory trends in political science. *PS: Political Science & Politics*, Cambridge, UK, v. 48, n. 1, p. 71-74, 2015.
- MOREIRA, D. *Com a palavra os nobres deputados: frequência e ênfase temática dos discursos dos parlamentares brasileiros*. 2016. 204 f. Tese (Doutorado em Ciência Política) – Universidade de São Paulo, SP, 2016.
- NEUENDORF, K. *The content analysis guidebook*. Thousand Oaks: Sage, 2002.
- OOMS, J. Tesseract: Open Source OCR Engine. *R package*, [s.l.], 2018. Disponível em: <<https://bit.ly/2whiySw>>. Acesso em 21 jul. 2018.
- PANG, B.; LEE, L. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, Hanover, v. 2, n. 1-2, p. 1-135, 2008.
- PATTY, J.; PENN, E. Analyzing big data: social choice and measurement. *PS: Political Science and Politics*, Cambridge, UK, v. 48, n. 1, p. 95-101, 2015.
- PEARL, J. *Causality*. Cambridge, UK: Cambridge University Press, 2009.
- PENNEBAKER, J. W.; MEHL, M. R.; NIEDERHOFFER, K. G. Psychological aspects of natural language use: our words, ourselves. *Annual Review of Psychology*, Palo Alto, v. 54, n. 1, p. 547-577, 2003.
- POOLE, K.; ROSENTHAL, H. *Ideology and congress*. New Brunswick: Transaction Publishers, 2007.
- PORTER, M. F. An algorithm for suffix stripping. *Program: Electronic Library and Information Systems*, Belfast, v. 14, n. 3, p. 130-137, 1980.
- POWER, T.; ZUCCO, C. Estimating ideology of Brazilian legislative parties, 1990-2005: a research communication. *Latin American Research Review*, Pittsburgh, v. 44, n. 1, p. 218-246, 2009.
- QUINN, K. et al. How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, Washington, DC, v. 54, n. 1, p. 209-228, 2010.

- ROBERTS, M. E. Introduction to the Virtual Issue: recent innovations in text analysis for social science. *Political Analysis*, Cambridge, UK, v. 24, n. 10, p. 1-5, 2016.
- ROBERTS, M.; STEWART, B.; TINGLEY, D. stm: R Package for Structural Topic Models. *R package*, [s.l.], 2018. Disponível em: <<https://bit.ly/2wc0rOT>>. Acesso em: 3 jul. 2018.
- ROBERTS, M. E. et al. *The structural topic model and applied social science*. Advances in neural information processing systems workshop on topic models: computation, application, and evaluation. Cambridge, MA: Harvard University, 2013.
- _____. Topic models for open-ended survey responses with applications to experiments. *American Journal of Political Science*, Washington, DC, v. 58, n. 4, p. 1064-1082, 2014.
- SLAPIN, J.; PROKSCH, S.-O. A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, Washington, DC, v. 52, n. 3, p. 705-722, 2008.
- SMITH, T.; WATERMAN, M. Identification of common molecular subsequences. *Journal of Molecular Biology*, Amsterdam, v. 147, n. 1. p. 195-197, 1981.
- SOUZA, M.; VIEIRA, R. Sentiment analysis on Twitter data for Portuguese language. In: INTERNATIONAL CONFERENCE COMPUTATIONAL PROCESSING OF THE PORTUGUESE LANGUAGE, 10., 2012, Coimbra. *Proceedings...* Coimbra: University of Coimbra, 2012. p. 241-247.
- SOUZA, M. et al. Construction of a Portuguese opinion lexicon from multiple resources. In: BRAZILIAN SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY, 8., 2011, Uberlândia. *Proceedings...* Uberlândia: Federal University of Uberlândia, 2011. pp. 59-66.
- SPIRLING, A. Democratization and linguistic complexity: the effect of franchise extension on parliamentary discourse, 1832-1915. *The Journal of Politics*, Chicago, v. 78, n. 1, p. 120-136, 2015.
- TABOADA, M. et al. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, Cambridge, MA, v. 37, n. 2, p. 267-307, 2011.
- VENABLES, W. N.; RIPLEY, B. D. *Modern applied statistics with S*. 4. ed. New York: Springer, 2002.
- WALLACH, H. et al. An alternative prior for nonparametric Bayesian Clustering. In: International CONFERENCE ON ARTIFICIAL INTELLIGENCE AND STATISTICS, 13., 2010, Sardinia. *Proceedings...* Sardinia: Chia Laguna Resort, 2010. p. 892-999, 2010.
- WEBER, R. P. *Basic content analysis*. Newbury Park: Sage, 1990. (University Paper Series on Quantitative Applications in the Social Sciences).
- WELBERS, K.; VAN ATTEVELDT, W.; BENOIT, K. Text analysis in R. *Communication Methods and Measures*, Abingdon, v. 11, n. 4, p. 245-265, 2017.

WICKHAM, H. *httr: Tools for Working with URLs and HTTP. R package*, [s.l.], 2016. Disponível em: <<https://bit.ly/2PwgzT0>>. Acesso em: 20 jul. 2018.

_____. *rvest: Easily Harvest (Scrape) Web Pages. R package*, [s.l.], 2018. Disponível em: <<https://bit.ly/2wee0fl>>. Acesso em: 21 jul. 2018.

WICKHAM, H.; HESTER, J.; OOMS, J. *xml2: Parse XML. R package*, [s.l.], 2018. Disponível em: <<https://bit.ly/2MrMzdi>>. Acesso em: 20 jul. 2018.

WILKERSON, J.; CASAS, A. Large-scale computerized text analysis in political science: Opportunities and challenges. *Annual Review of Political Science*, Palo Alto, v. 20, p. 529-544, 2017.

WILKERSON, J.; SMITH, D.; STRAMP, N. Tracing the flow of policy ideas in legislatures: a text reuse approach. *American Journal of Political Science*, Washington, DC, v. 59, n. 4, p. 943-956, 2015.

Resumo

O texto como dado: desafios e oportunidades para as ciências sociais

A comunicação é instrumento fundamental para as relações humanas. É por meio dela, por exemplo, que valores são construídos, símbolos sociais são estabelecidos, tradições são repassadas, debates são concretizados, a política se materializa e o conflito político se expressa. Foco de análises dos cientistas sociais há séculos, a análise do conteúdo transmitido na comunicação sempre esteve restrita à necessidade de volumes relevantes de recursos para a avaliação manual de grandes acervos. Revertendo esse quadro limitado, recentes desenvolvimentos tecnológico, computacional e científico permitem que as ciências sociais potencializem sua investigação reduzindo drasticamente os custos envolvidos na análise de grandes acervos. Por intermédio de novos métodos desenvolvidos, atualmente, é possível verificar comportamentos que antes não eram observáveis, medir quantidades anteriormente imensuráveis e testar hipóteses até então impossíveis de serem testadas. Nesse escopo, o principal objetivo deste artigo é manter as ciências sociais brasileiras na fronteira desse processo e apresentar ao leitor um leque atualizado das principais metodologias de análise automatizada de conteúdo. Sem esgotar suas inúmeras possibilidades, este artigo é um guia para a inovadora e instigante área de pesquisa do texto como dado.

Palavras-chave: Análise Automatizada de Conteúdo; Semelhança entre Textos; Métodos de Classificação; Métodos de Escalonamento; *Big Data*.

Abstract

The text as data: challenges and opportunities for Social Sciences

Communication is a fundamental tool for human relations. It is through communication that values are constructed, social symbols are established, traditions are passed on, debates are realized, politics are materialized and political conflict is expressed. A focus in analyses of social scientists, the analysis of the content transmitted in communication has always been restricted to the need for a great amount of research funds for the manual assessment of large collections. Changing this limited scenario, recent technological, computational and scientific developments allowed social scientists to analyse larger collections of documents with low cost. Currently, through the development of new methods, it is now possible to identify behaviors that could not be observed, to measure quantities that could not be quantified, and to test hypothesis that could not be tested. In this sense, the main objective of this study is to maintain Brazilian Social Sciences at the frontier of this process and present to the reader the latest methodologies for automated content analysis. Without exhausting its several possibilities, this article is a guide to the innovative area of researching text as data.

Keywords: Automated Content Analysis; Similarity between Texts; Classification Methods; Scheduling Methods; Big Data.

Résumé

Le texte en tant que donné : défis et opportunités pour les Sciences Sociales

La communication est un outil fondamental pour les relations humaines. C'est par la communication que des valeurs sont construites, des symboles sociaux sont établis, des traditions sont transmises, des débats sont réalisés, des politiques sont matérialisées et des conflits politiques sont exprimés. Un accent dans les recherches des sociologues, l'analyse du contenu transmis dans la communication a toujours été limitée au besoin d'une grande quantité de fonds de recherche pour l'évaluation manuelle de grandes collections. En changeant cet scénario limité, les récents développements technologiques, informatiques et scientifiques ont permis aux sociologues d'analyser des plus grandes collections de documents à bas prix. Actuellement, grâce au développement de nouvelles méthodes, il est désormais possible de identifier comportements qui étaient inobservables, de mesurer des quantités auparavant incommensurables et tester des hypothèses jusqu'alors impossibles. Dans ce sens, l'objectif de cet article est de maintenir les Sciences Sociales brésiliennes à la frontière de ce processus et de présenter au lecteur les méthodologies les plus récentes pour l'analyse de contenu automatisée. Sans épuiser ses nombreuses possibilités, cet article est un guide sur le domaine innovant de la recherche des textes en tant que donnés.

Mots-clés: Analyse de Contenu Automatisée ; Similarité entre textes ; Méthodes de Classification ; Méthodes de Planification ; Big Data.