

(RE)APROXIMAÇÕES ENTRE BIBLIOMETRIA E ORGANIZAÇÃO DA INFORMAÇÃO: UM RELATO DE EXPERIÊNCIA

Raquel Juliana Prado Leite de Sousa¹

Resumo: Aborda o uso de processos, ferramentas e produtos usuais na Organização da Informação como procedimento de pesquisa na exploração textual, na normalização e no tratamento de grande quantidade de dados usualmente encontrados durante a realização de pesquisa bibliométrica para mensuração da produção científica. Descreve as experiências obtidas na elaboração de pesquisa bibliométrica e as soluções encontradas para a padronização de dados de citação, a fim de alertar sobre possíveis erros de resultados e prover sugestões e considerações que possam ser pertinentes a pesquisadores da área, em especial os iniciantes. O relato se baseia na experiência durante pesquisa de doutorado sobre levantamento bibliométrico da produção científica brasileira na área de Educação referente a 237 teses depositadas na Biblioteca Digital de Teses e Dissertações entre 1997 e 2016. Discorre sobre o uso da Análise Documental para a realização de leitura estratégica de *corpus* textual. Relata procedimentos adotados durante a coleta, tratamento e análise de dados para a diminuição de possíveis incongruências e dificuldades, os quais podem acarretar erros de resultados e atrasos na pesquisa. Sugere o uso de catálogos de biblioteca, bases de autoridades, modelos conceituais e regras de catalogação para a conferência e normalização de dados de autoria e título, a fim de facilitar a mensuração, evitar retrabalho e garantir o levantamento correto de indicadores de citações e frente de pesquisa e também como forma de ampliar a competência em informação do pesquisador bibliometrista.

Palavras-chave: Metodologia científica. Bibliometria. Cientometria. Organização da Informação. Produção científica. Relato de experiência.

1 INTRODUÇÃO

A análise bibliométrica, enquanto medida estatística da produção e circulação do conhecimento, impõe ao pesquisador enormes desafios procedimentais (SILVA; HAYASHI; HAYASHI, 2011). Para que os estudos bibliométricos obtenham relevância e possam gerar indicadores confiáveis, é necessário abranger um grande *corpus* de pesquisa e manipular diversas variáveis. Assim, é comum que o pesquisador encontre centenas e até milhares de dados que devem ser coletados, tabulados, processados em *softwares*, específicos ou não, e analisados, o que pode causar muitas dificuldades na organização e na prática da pesquisa, em especial aos bibliometristasⁱ iniciantes.

¹ Doutorado em Educação no Programa de Pós-Graduação em Educação da Universidade Federal de São Carlos. Mestrado em Ciência Tecnologia e Sociedade no Programa de Pós-Graduação em Ciência Tecnologia e Sociedade da Universidade Federal de São Carlos. Graduação em Biblioteconomia e Ciência da Informação na Universidade Federal de São Carlos, Comunicação Social – Jornalismo pela Universidade Estadual Paulista Júlio de Mesquita Filho e Licenciatura em Letras pelas Faculdades Integradas Maria Imaculada. Professora do curso de graduação em Biblioteconomia e coordenadora do curso de Especialização em Metodologias Ativas de Ensino do Centro Universitário Claretiano. E-mail: raquelsousa@claretiano.edu.br. ORCID <https://orcid.org/0000-0002-4126-0189>.

ⁱⁱ O termo bibliometrista se refere ao profissional que atua na análise bibliométrica, independentemente da formação em Biblioteconomia, e tem sido empregado no Brasil em diversos trabalhos sobre o tema. Segundo Vanz, Santin e Pavão (2018, p.



Ao desafio procedimental acarretado pela vastidão dos resultados avaliados, soma-se a baixa qualidade da normalização documentária das citações e referências presentes em publicações de cunho científico, como teses, dissertações, artigos etc. Diversos bibliometristas que realizam análise de citação apontam a falta de padronização como empecilho para o bom andamento da análise bibliométrica, o que pode acarretar erros na mensuração de citações e de frentes de pesquisa. Dito de outro modo, pode-se atribuir citação a autores que não o verdadeiro, deixar de contabilizar autorias, dispersar a contagem de citações de um mesmo autor e criar uma falsa classificação das frentes e dos *outliers*ⁱⁱ de pesquisadores.

Apenas em relação à citação da data de publicação de livros, por exemplo, Maroldi (2017) notou quatro ocorrências que podem influenciar a mensuração da vida média, do fator de envelhecimento e da perda anual da literatura citada, como: obras citadas duas vezes no mesmo trabalho, mas com anos diferentes, diferença entre o ano mencionado na citação e o registrado na referência, obras com ano de publicação diferente da edição referenciada, além de obras citadas diversas vezes com edições diferentes, o que não representa um erro, mas pode atrapalhar a contagem, tendo em vista que é necessário decidir se as várias versões serão contadas separadamente ou serão somadas, desconsiderando-se as edições.

Ao pesquisador inexperiente, que ainda não possui noção exata das dificuldades que poderá encontrar, essas questões podem passar totalmente despercebidas ou serem notadas apenas em fase já avançada da pesquisa ou até mesmo somente após seu término, quando já foram despendidas muitas horas de trabalho e esgotados os prazos de defesa ou de elaboração de relatórios finais, no caso de monografias de conclusão de curso ou pesquisas fomentadas por bolsa.

Silva, Hayashi e Hayashi (2011) recorrem ao conceito de competência em informação para identificar o conjunto de competências, habilidades e atitudes que se relacionam com as etapas e procedimentos realizados pelos bibliotecários e especialistas em indicadores científicos. Segundo os autores, há variáveis dependentes e independentes que influenciam a competência informacional dos bibliometristas: as variáveis dependentes dizem respeito a fatores que estão fora do controle do pesquisador, como: inconsistência das bases de dados; desatualizações, incoerências nos dados; restrições dos *softwares* e ausência de informações nos documentos.

Já as variáveis independentes relacionam-se aos conhecimentos e experiências do pesquisador ou profissional sobre os fundamentos teóricos da Bibliometria e do campo de estudo em que esta será aplicada e precisam ser controladas, isto é, este **controle é possível por meio de procedimentos bem delineados. As variáveis independentes dependerão**

6), “o cargo “bibliometrista” (bibliometrician) é realidade em muitas bibliotecas europeias em países como Suécia, Alemanha e Holanda”.

ⁱⁱ *Outlier* é um valor discrepante que apresenta afastamento dos demais dados de uma série. Nos estudos bibliométricos, diz respeito aos autores de destaque que tiveram mais citações do que os demais autores que compõem a frente de pesquisa.

exclusivamente do próprio pesquisador, o que exige interesse e conhecimento prévio para o desenvolvimento de estudos bibliométricos. (SILVA; HAYASHI; HAYASHI, 2011, p. 124, grifo nosso)

Tendo em vista tais questões, o objetivo deste relato é descrever as experiências obtidas na elaboração de estudo bibliométrico e as soluções encontradas para a coleta, padronização e organização dos dados, a fim de prover sugestões e considerações que possam ser pertinentes à formação das competências informacionais dos pesquisadores da área, no que tange às variáveis independentes supracitadas. O relato aqui apresentado se refere à pesquisa de doutorado realizada entre 2015-2019 sobre análise bibliométrica e cientométrica da produção científica na área de Educação, especificamente centrada na temática do letramento, a qual visou ao levantamento de indicadores para mapeamento de perfil de teses depositadas na Biblioteca Digital de Teses e Dissertações (BDTD), análise de citações e arrolamento de frentes de pesquisa.

Face aos desafios impostos pela pesquisa, foram buscados procedimentos, produtos e ferramentas próprios da Ciência da Informação, mais especificamente da Organização da Informação, como uso de técnicas de Análise Documental, catálogos de bibliotecas, catálogo de autoridades e códigos/fundamentos de catalogação e a utilização do *Functional Requirements for Bibliographic Records* (FRBR), os quais serão descritos a seguir.

2 ANÁLISE DOCUMENTAL COMO APORTE METODOLÓGICO

Como fonte de dados, foi utilizada a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), repositório digital coordenado pelo IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia) em parceria com instituições científicas brasileiras. Para a pesquisa e a recuperação de dados, procedeu-se à busca avançada, utilizando-se os termos letramento, literacia e multiletramentos no campo assunto e selecionando a correspondência de busca ‘qualquer termo’ (equivale ao operador booleano OU), o que resultou na expressão:

(Assunto:letramento OU Assunto:literacia OU Assunto:multiletramentos)

Também foram utilizados os filtros “teses”, “língua portuguesa” e “data” de 1997 a 2016, o que resultou na recuperação de 250 registros.

As teses foram baixadas, salvas e organizadas entre os dias 03 e 07 de julho de 2017. Decidiu-se excluir as dissertações e utilizar apenas as teses, pois estas dizem respeito a resultados de pesquisas de doutorado, ou seja, trabalhos de cunho inédito e de fundamentação teórico-metodológica mais robusta.



Em decorrência da impossibilidade da leitura completa dos textos, muitos pesquisadores se baseiam no resumo para a triagem do *corpus* de pesquisa e até mesmo para o levantamento de dados. Entretanto, é comum que esse elemento pré-textual não traga informações muito precisas sobre referencial teórico, procedimentos, materiais e variáveis utilizados e encontrados na pesquisa. Além disso, em se tratando de monografias de conclusão de curso, a baixa qualidade dos resumos é comum, pois muitas vezes esse elemento é elaborado às pressas, no calor da pressão do prazo para a finalização do trabalho, pois as maiores preocupações dos pesquisadores são a manipulação dos dados e a redação do texto em si. Dessa forma, conclui-se que utilizar o resumo é uma decisão que pode comprometer a investigação científica quando o *corpus* é constituído de teses, dissertações e também de artigos.

Tendo em vista essa problemática, foi testado como alternativa o uso da Análise Documental (ou Análise Documentária) para a leitura das 250 teses levantadas, com emprego de técnicas próprias da Ciência da Informação para a extração dos principais conceitos dos documentos. A Análise Documental, atividade de base interdisciplinar fundamentada na Lógica e na Linguística, é um procedimento que faz parte do Tratamento Temático da Informação (TTI) e consiste na leitura estratégica e não linear de trechos representativos dos documentos, com vistas à identificação, seleção e descrição de assuntos, métodos e variáveis importantes dos textos. Essa atividade se trata de:

Técnica documental que permite, mediante uma operação intelectual objetiva, a identificação e transformação dos documentos em produtos que facilitem a consulta aos originais, em áreas de controle documental, e com o objetivo último de servir à comunidade científica. (GARCIA GUTIERREZ, 1984, p. 83)

Com base na superestrutura dos textos científicos, é possível fazer uma exploração extremamente rápida e eficaz do conteúdo. Partindo-se de elementos predeterminados (agentes, ações, ambiência etc.) são abordados alguns trechos, até que tais elementos sejam encontrados: título, subtítulos, gráficos, tabelas, palavras em destaque, referências e primeiros parágrafos da introdução, procedimentos, resultados e conclusões são trechos indicados para leitura (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992).

A Análise Documental é utilizada como base para os processos de classificação de assunto, indexação, resumo, recensão e elaboração de índices, entretanto, sua utilização como procedimento metodológico para realização de pesquisa científica se mostrou eficaz, pois permitiu a leitura estratégica das 250 teses em apenas cinco dias. Ressalta-se que a experiência do analista, o conhecimento do assunto e as condições do ambiente de trabalho (ruído, iluminação, etc.) interferem na dinâmica da leitura.

Como suporte ao procedimento, sugere-se a utilização da Norma Brasileira de Redação 12.676 'Métodos para análise de documentos – determinação de seus assuntos e seleção de termos de indexação'

da Associação Brasileira de Normas Técnicas (1992) e a utilização do modelo de leitura documental proposto por Fujita e Rubi (2006), reunidos no Quadro 1.

Quadro 1 – modelo de leitura documental

Conceito	Questionamento (norma 12.676)	Parte da estrutura textual
Objeto	O documento possui um objeto sob efeito de uma atividade/observação?	Introdução (objetivos)
Ação	O assunto contém um conceito ativo (ação, operação, processo etc.)?	Introdução (objetivos)
Agente	Possui um agente que praticou essa ação?	Introdução (objetivos)
Métodos do agente	Refere-se a modos específicos de realizar a ação (instrumentos, técnicas, métodos)?	Procedimentos metodológicos
Local ou ambiência	Todos esses fatos estão condicionados a um lugar ou ambiente específico?	Procedimentos metodológicos
Causa e efeito	São identificadas variáveis?	Resultados Discussão
Ponto de vista do autor, perspectiva	Há um ponto de vista não comumente associado ao assunto?	Conclusões

Fonte: adaptado de Fujita e Rubi (2006).

Após a Análise Documental, fez-se necessário excluir 13 trabalhos que, embora estivessem descritos com as palavra-chave letramento, ou literacia ou multiletramentos, não abordavam os aspectos sociais do ensino da leitura e da escrita, fugindo do escopo da pesquisa. Dessa forma, chegou-se a um *corpus* de 237 teses. Assim, percebeu-se que os procedimentos de Análise Documental empregados no dia a dia das bibliotecas durante o TTI podem e devem ser utilizados como método científico para leitura textual como forma de substituir a análise do resumo.

Ao preencher o quadro do modelo de Fujita e Rubi (2006) com as informações de cada texto componente do *corpus* analisado, será possível, também, obter uma fonte de dados de pesquisa a ser utilizada como base de análise. Assim, a leitura documental pode servir não apenas para a exploração prévia dos *corpora*, mas também para levantamento de fonte prospectiva de dados, inclusive em trabalhos que utilizem outros procedimentos metodológicos além da bibliometria.

3 COLETA E TRATAMENTO DE DADOS BIBLIOMÉTRICOS

Após a Análise Documental, as teses foram salvas e identificadas com códigos numéricos. Esse procedimento foi muito importante para a organização dos *corpora*, pois permitiu que teses específicas fossem acessadas com facilidade sempre que necessário. Sugere-se não salvar os arquivos com seus nomes originais, o que pode causar confusão caso se precise consultar algum texto para conferência de dados, por exemplo.

Para a codificação, utilizou-se o ano da tese seguido de um numeral cardinal para cada trabalho. Exemplo: 2001(1), 2001(2), 2001(3), 2002(1), 2002(2) etc. Foi criada uma pasta para a separação dos trabalhos por ano de defesa, o que é um recurso dispensável, uma vez que a codificação citada já permite a ordenação e a identificação dos arquivos. A codificação pode ser criada conforme a preferência e a necessidade, mas aconselha-se que não seja dispensada. Novos documentos criados durante os procedimentos podem ser salvos tendo como base a mesma codificação utilizada para os arquivos dos *corpora*, o que gerará imediata integração de todos os dados gerados durante a pesquisa.

Com o auxílio do *software Excel*[®], foi criada uma planilha de coleta e registro dos dados contendo os campos: código do trabalho (citado acima), autor, gênero do autor, orientador, gênero do orientador, título do trabalho, ano de defesa, instituição de ensino, programa de pós-graduação, dependência administrativa, região do país, agência de fomento e palavras-chave. Em seguida, os dados foram transportados para o *VantagePoint*[®], *software* de análise bibliométrica que possibilita a mensuração de grande conjunto de dados e a geração de indicadores.

Ressalta-se que a digitação ou cópia de dados para o *Excel*[®] ou outro *software* congênere deve ser feita com muita atenção, pois um único caractere errado ou inconsistente, como troca de letras, falta ou excesso de espaçamento, acentuação etc., poderá acarretar erro de contagem dos dados. Por exemplo, uma autora digitada como 'Ecléa Bosi' e 'Eclea Bosi' pode ser contada como duas autoras diferentes por causa da falta de acentuação no segundo caso. Assim, sugere-se que os dados sejam digitados e revisados minuciosamente antes de serem transpostos para o *VantagePoint*[®], sejam eles nomes pessoais e institucionais, títulos de obras, palavras-chave, localidades geográficas etc., pois isso evitará retrabalho para a correção e garantirá a exatidão dos resultados. Uma boa estratégia para evitar o erro citado acima é excluir quaisquer sinais diacríticos (acentos) e possíveis sinais gráficos.

É extremamente comum que as monografias de conclusão de curso não tragam uma área específica para a indicação da agência de fomento, caso o pesquisador tenha sido contemplado com bolsa, pois esse tipo de informação não é exigido pelas Normas Brasileiras de Redação ou pelos modelos de formatação

elaborados pelas instituições de ensino. Esse dado costuma ser registrado apenas na área de agradecimentos, quando o autor expressa gratidão às pessoas que contribuíram para a elaboração do trabalho. Por isso, durante a etapa de coleta desta pesquisa, para a confirmação das agências de fomento financiadoras das teses, foi consultado o currículo Lattes dos autores. Sugere-se que, para o levantamento de dados sobre financiamento de pesquisas, sejam utilizados não apenas as informações constantes nos agradecimentos das teses, dissertações e artigos, mas também uma fonte oficial como o Lattes, tendo em vista que foi verificado, neste trabalho, que alguns se esqueceram de citar a agência de fomento nas teses.

As palavras-chave também são elementos que podem acarretar erro de dados e resultados, em especial quando o *corpus* possui trabalhos antigos, os quais muitas vezes não trazem as palavras-chave. Deve-se atentar, também, que essas palavras podem ser retiradas de três locais diferentes: do resumo do próprio texto, da ficha catalográfica da obra e da base de dados onde está depositada. Entretanto, as palavras-chave de um mesmo texto podem ser diferentes nos três locais. Assim, para uniformizar os resultados, sugere-se que se retirem esses dados de uma única fonte dentre as três citadas.

As palavras-chave que aparecem após os resumos ajudam a revelar como os autores dos trabalhos identificam as próprias pesquisas; as retiradas da base de dados podem ser mais uniformes, em especial quando se utiliza alguma lista ou linguagem de indexação, o que deve ser confirmado no momento da coleta; já as retiradas das fichas catalográficas podem provocar falta de padronização, sabendo-se que parte das fichas do *corpus* pode ter sido elaborada pelos autores e parte redigida pelo bibliotecário da instituição em que a pesquisa foi feita e depositada (o autor do trabalho utiliza a linguagem natural para descrever os assuntos de sua pesquisa; já os bibliotecários que fazem a indexação no momento do depósito da obra nas bibliotecas costumam utilizar linguagens de indexação). Assim, para manter a consistência dos dados tabulados, sugere-se fazer uma verificação prévia nos trabalhos e na base de pesquisa, para se decidir de antemão de onde serão retiradas as palavras-chave, atentando-se para a uniformidade dos dados levantados e para possíveis diferenças semânticas na identificação dos assuntos.

Outra etapa extremamente importante é a normalização das citações, o que evitará o máximo de ruídos na tabulação e análise. Para a pesquisa em questão, foram despendidos quatro meses de trabalho para a normalização das 33.850 referências devido à grande quantidade de erros e inconsistências. Outra dificuldade comumente encontrada diz respeito à formatação das referências que são copiadas de arquivos do tipo pdf: após a cópia, ao transpô-las para o processador de texto ou planilha, quase sempre aparecem quebras de parágrafo ao final de cada linha, o que faz com as referências fiquem quebradas, ou seja: se uma referência tem 3 linhas, por exemplo, cada uma delas ficará como um parágrafo independente e ficarão afastadas caso se necessite classificar as referências em ordem alfabética ou numérica. A exclusão das

quebras é essencial e pode dificultar e até impedir a utilização de *softwares* de normalização, pois a referência poderá ficar fragmentada em duas ou mais partes, fazendo com que o programa a veja como referências distintas. Sabendo-se que o tempo despendido para a exclusão de quebras e espaçamentos desnecessários é enorme, sugere-se que o bibliometrista avalie se compensa fazer a normalização ‘à mão’ ao mesmo tempo em que corrige a formatação dos parágrafos e codifica os tipos documentais para separação ou se será mais produtivo arrumar a formatação do arquivo para somente depois utilizar um *software* de normalização e, em seguida, codificar as referências.

Sugere-se que não sejam contabilizadas as referências não citadas (as constantes apenas na bibliografia consultada) e as citações de citações (*apud*), mesmo que estas apareçam nas referências. A propósito, apesar de comumente ocorrer, a inclusão da citação de citação nas referências é um erro, uma vez que o texto original não foi consultado, e sim apenas o texto que o citou. Como alternativa, as citações de citações podem ser contadas separadamente, caso se verifique essa necessidade, o que permitirá a mensuração dos textos que foram citados mesmo sem acesso ao seu conteúdo original, tendo em vista que isso pode ser usual em áreas do conhecimento em que obras importantes não estejam mais disponíveis para acesso.

Também se aconselha excluir os textos indicados como ‘mimeo’, ‘versão preliminar’, ‘no prelo’, ‘inédito’ e ‘a sair’, pois pode ser impossível confirmar seus dados em caso de alterações. A justificativa é a seguinte: entre a versão preliminar e a publicação da obra, pode haver mudanças, em decorrência do processo de editoração de periódicos e livros. Uma modificação muito comum é a alteração do título das obras por parte das editoras, o que impossibilita a conferência dos dados e seu agrupamento. Por exemplo, uma obra no prelo intitulada ‘Análise bibliométrica na Educação’ foi citada pelo autor de uma tese; entretanto, após a editoração final, decidiu-se alterar seu título para ‘Educação e bibliometria’. Esse segundo título, quando citado, será contabilizado como obra diferente da primeira. Como um texto ainda não publicado não possui validade formal, sugere-se não contabilizá-lo, a fim de evitar incongruências de contagem.

Em se tratando de referências, a normalização é fundamental para a consistência da mensuração de autoria, obras e atualidade das fontes (datas). Para isso, faz-se necessária, antes da tabulação e mensuração, a correção de detalhes como:

- a) Erros de digitação (nomes de autor e título): BORDIEU, P. em vez de BOURDIEU, P.
- b) Inclusão ou exclusão de artigos iniciais em títulos: (A) Pedagogia da autonomia.
- c) Troca da ordem de autoria: PASSERON, J.; BOURDIEU, P. ao invés de BOURDIEU, P.; PASSERON, J.

- d) Inversão dos elementos do título: TFOUNI, L. V. Alfabetização e letramento ao invés de TFOUNI, L. V. Letramento e alfabetização e SOARES, M. Letramento e alfabetização no lugar de SOARES, M. Alfabetização e letramento.

Para o tratamento e a confirmação dos dados de citação, sugere-se a utilização de ferramentas e procedimentos da área de Organização da Informação, como bases de autoridades, catálogos de bibliotecas e o Código de Catalogação Anglo Americano (AACR2).

A normalização *a priori* também é essencial para a definição da tipologia documental das referências, ou seja, para identificar se se trata de artigo, tese, livro, capítulo etc.

Para a contagem das referências, após tabulação, deve-se dar um comando para que o *software*, seja ele editor de texto ou de planilhas, organize alfabeticamente esses dados. Se há inversão de ordem de autores e elementos do título, erro de digitação ou acréscimo/exclusão de artigos, a alfabetização ficará errada, comprometendo o agrupamento, a contagem e a análise, acarretando erros nos resultados.

Para a separação das referências por tipologia documental, sugere-se sua identificação com códigos numéricos ainda durante o processo de normalização, o qual foi discutido anteriormente. Por exemplo, na pesquisa em questão, os livros foram codificados com o numeral 1 antes da referência, 2 para capítulos de livros e 3 para artigos de periódicos científicos, sendo que os tipos documentais que não fizeram parte do *corpus* de citações não foram identificados por nenhum número. Sugere-se que essa codificação seja feita durante o processo de normalização e de formatação dos documentos (para exclusão de quebras de parágrafo e espaçamentos), porque isso vai poupar tempo do bibliometrista, evitando que se necessite avaliar todas as referências duas ou até três vezes, ou seja, uma para formatar as referências, outra para normalizar e outra para codificar e separar os tipos documentais.

Após a normalização e a classificação numérica, basta selecionar toda a listagem de referências e dar o comando para que o *software* as classifique por ordem numérica crescente. Assim, obter-se-á um agrupamento por tipo documental, permitindo que os dados sejam separados em arquivos diferentes e contabilizados com segurança.

3.1 CONTROLE DE AUTORIDADE PARA UNIFORMIZAÇÃO DE REFERÊNCIAS

Outro grande problema que pode influenciar os resultados das pesquisas, incorrendo em erros e complicações na contagem, é a falta de congruência nas citações dos nomes dos autores das obras referenciadas. Sejam por falhas de digitação, dificuldades de identificar a forma correta, a escolha entre

grafias diferentes ou a alteração do nome ao longo do tempo, a discrepância na identificação de prenomes e sobrenomes acarretará erros de contabilização das citações e frente de pesquisa.

Assim, para a padronização da forma dos nomes de autores pessoais e entidades, foram utilizadas as regras do Código de Catalogação Anglo-americano (2004), bem como a confirmação no catálogo de autoridades da Biblioteca Nacional (Brasil)ⁱⁱⁱ e do Library of Congress Authorities^{iv}, da Biblioteca do Congresso dos Estados Unidos.

Um catálogo de autoridade é uma relação das formas corretas de identificação de nomes de autores pessoais e institucionais (entidades) ou assuntos (SOUSA, 2019), ou seja, uma ferramenta que auxilia a reconhecer, por exemplo, se a autoria deve ser identificada como IBGE (forma autorizada) ou como Instituto Brasileiro de Geografia e Estatística (forma não autorizada). Essa listagem toma como base as regras já consolidadas dos códigos de catalogação, as quais não serão aqui discutidas devido a sua quantidade e extensão.

O Quadro 2 apresenta inconsistências comuns nas variações de alguns nomes, com exemplos que foram encontrados nas citações analisadas na referida pesquisa.

Quadro 2 – variações de nome de autores do *corpus* de citações.

Inconsistência	Exemplo
Autores com nome completo <i>versus</i> nome abreviado	BOHN, Hilário I. BOHN, Hilário Inácio
Exclusão de nomes e sobrenomes do meio	SOARES, Magda Becker SOARES, Magda
Autores com mudança de sobrenome (em decorrência de alteração de estado civil, p. e.)	LONGO, Maria do Rosário MORTATTI, Maria do Rosário Longo
Erros de digitação	BAHKTIN, Mikail BAKHTIN, Mikail
Sobrenome duplo	VÁZQUEZ, Adolfo Sanchez SÁNCHEZ VÁZQUEZ, Adolfo
Inclusão / exclusão de pronome	VAN DIJK, Teun Adrianus DIJK, Teun Adrianus van
Alteração histórica	FREYRE, Gilberto (forma antiga) FREIRE, Gilberto (forma atual)
Sobrenomes transliterados de formas diferentes (representação dos caracteres de uma escrita pelos de outra)	Выготский, Лев Семёнович (original em russo) VIGOTSKI, Lev Semyonovich VYGOSTKI, Lev Semyonovich VYGOSTKII, Lev Semyonovich VYGOSTKY, Lev Semyonovich VYGOSTKIJ, Lev Semyonovich VIGOSTKY, L. S.

Fonte: autoria própria. Os nomes marcados em negrito representam as formas autorizadas.

ⁱⁱⁱ http://acervo.bn.br/sophia_web/index.html

^{iv} <http://authorities.loc.gov/>

Outro problema que se verificou na identificação da autoria foi referente à substituição dos nomes dos autores pelo traço sublinear equivalente a seis espaços e ponto (_____.), conforme a regra 9.1.1 da antiga NBR 6023 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2002). Alguns pesquisadores, ao elaborar as referências de seus trabalhos, utilizam o traço sublinear considerando apenas o primeiro autor da obra, excluindo segundos e terceiros autores e a expressão *et al.* de outras obras de autoria compartilhada. No exemplo abaixo, encontrado nas referências do trabalho em questão, as duas obras seriam contabilizadas como de autoria de Pierre Bourdieu. Entretanto, o segundo livro é de autoria compartilhada com J. C. Passeron.

BOURDIEU, P. *O poder simbólico...*

_____. *A reprodução: elementos para uma teoria do sistema de ensino...*

A inadequação acima acarretaria dois erros: contagem de frente de pesquisa e de tipo de autoria (autor único, dois autores, três autores e mais de três autores), podendo interferir, também, na mensuração de *outliers*. Assim, aconselha-se que as referências que trazem o traço sublinear para substituir autores sejam confirmadas, com substituição do traço pela indicação do(s) autor(es) antes mesmo da tabulação dos dados.

Verificou-se, nesta pesquisa, que o traço sublinear foi utilizado para substituir autoria compartilhada entre dois ou três autores, mas sem que a ordem de autoria fosse levada em conta. No exemplo abaixo, apesar dos autores serem os mesmos, há diferença na ordem; caso fosse mantido o traço sublinear para a substituição de autoria, isso acarretaria erro de contagem de citação.

BARTON, D.; HAMILTON, M.; IVANIC, R. *Situated literacies: reading and writing in context...*

HAMILTON, M.; BARTON, D.; IVANIC, R. *Worlds of literacy...*

Outro problema que a inversão da ordem dos autores acarreta diz respeito à atribuição de frações erradas de autoria. A literatura aponta para quatro modos diferentes de contabilizar a autoria nos casos em que os textos são assinados por duas ou mais pessoas (POTTER, 1981; TRIFUNAC, 2005; TOL, 2011).

1. Na contagem direta, proposta por Lotka, atribui-se a autoria apenas ao primeiro autor, ignorando-se os colaboradores.
2. Na contagem completa (ou normal), credita-se uma autoria completa para cada pessoa.

3. Na contagem ajustada, atribui-se uma fração da contribuição total, ou seja, havendo três autores, por exemplo, cada um recebe 1/3 da contribuição.
4. Na contagem *rank weight*, compartilha-se o crédito com base na ordem dos autores, sendo que o primeiro autor recebe uma fração maior que o segundo, que recebe uma maior que o terceiro, e assim sucessivamente.

Dessa forma, percebe-se que a inversão da ordem de autoria não acarreta erro de mensuração da produtividade dos autores somente quando se utiliza as contagens completas ou ajustadas, o que reforça a necessidade de haver grande rigor ao normalizar as referências sob o risco de gerar falhas nos resultados da pesquisa.

A grande quantidade de variações nos nomes e de erros na substituição de autoria pelo traço sublinear também chama a atenção para a importância de sensibilizar os pesquisadores sobre o rigor necessário para a documentação de seus trabalhos. Erros podem ter como consequência a perda de citação ou até a atribuição de citação a autores, livros ou periódicos erroneamente.



3.2 USO DO FRBR PARA UNIFORMIZAÇÃO E CONTAGEM DE TÍTULOS

É comum que as obras sejam citadas em várias línguas, pois alguns pesquisadores têm acesso às traduções e outros preferem o texto original. Não é raro que algumas obras sejam traduzidas para a mesma língua por tradutores diferentes ou que a mesma obra seja traduzida diretamente do idioma original e também de outra tradução. Por exemplo, um livro de Bakhtin pode ter sido traduzido do russo diretamente para o português; o mesmo livro pode ter sido traduzido do russo para o francês e, depois, do francês para o português. Como a velha máxima diz que não existe tradução sem traição, cada versão em idioma distinto implica mudança de expressão. Convém esclarecer que a tradução é a conversão de um texto estrangeiro para sua língua nativa (no caso o português), já a versão é o contrário, a passagem de um texto em português para outro idioma.

Alguns textos podem ser publicados, por exemplo, como artigo e também como capítulo de livro, o que não faz dele dois textos diferentes.

Assim, para que não houvesse dispersão na contagem das referências por conta das questões supracitadas, foi utilizado o modelo conceitual do *Functional Requirements for Bibliographic Records* (FRBR), mais especificamente as entidades do grupo 1, para a identificação do *corpus* de análise.

O FRBR se baseia no modelo entidade-relacionamento e é usado na Ciência da Informação para o reconhecimento e a descrição das entidades envolvidas em um documento. No grupo 1 temos as entidades obra, expressão, manifestação e item. Obra é uma entidade abstrata que expressa uma criação intelectual, por exemplo, a obra *Marxismo e filosofia da linguagem*, de Bakhtin. A expressão é uma entidade que se refere à realização intelectual, que pode apresentar alteração no conteúdo. Para *Marxismo e filosofia da linguagem*, temos uma expressão original do autor em russo e outra expressão em português, por exemplo. A manifestação é a materialização desse conteúdo em um suporte físico, cada manifestação gera itens físicos. Portanto, a obra é realizada pela expressão, que está contida na manifestação, que é exemplificada pelo item.

Dessa forma, para a contagem das referências foi utilizado o conceito de obra, ou seja, em caso de traduções e versões, foi selecionado o texto original e todas as referências foram contabilizadas nela. Para a confirmação das traduções e versões, foram consultados catálogos em linha de bibliotecas universitárias. Nos casos de artigos posteriormente republicados como capítulo de livro, ou vice-versa, também sugere-se fixar a obra original como base para a contagem.

Toma-se como exemplo o livro *Social literacies*, de Brian V. Street, traduzido para o português como *Letramentos sociais*. As referências das duas expressões foram contabilizadas como uma obra, representada

pelo título original em inglês. Sem fazer essa distinção, a contagem seria dividida em duas, o que incorreria em dispersão do número de citações do livro, ou seja, do seu conteúdo intelectual. Ressalta-se que o conhecimento do pesquisador sobre o tema da pesquisa e sobre a bibliografia da área ajuda no reconhecimento de questões como essa. Sugere-se que os títulos originais sejam conferidos em catálogos de bibliotecas, verificando-se a indicação de títulos uniformes, equivalentes ou notas de tradução de títulos.

A utilização da entidade obra do FRBR também foi uma solução encontrada para a dispersão de datas e edições, mencionada anteriormente: utilizando os dados da referência da primeira edição da obra ou da mais antiga citada, encontrou-se uma justificativa plausível para desconsiderar datas e edições na primeira etapa da pesquisa. Dessa forma, sugere-se que a contagem inicial das referências desconsidere as expressões diferentes geradas por novas edições/datas de publicação.

Para o estudo da vida média da literatura citada, aconselha-se que sejam analisadas as datas e edições somente das obras mais citadas. Por exemplo: se em um *corpus* de 3.000 livros citados foram encontradas 56 obras mais citadas, apenas as datas e edições dessas 56 obras serão identificadas, em uma nova etapa do trabalho. Ao considerar a vida média somente das obras mais citadas, poupa-se tempo de separação de datas e edições de obras que não estão entre as mais citadas.

4 CONTAGEM DE AUTORIA DE CAPÍTULOS DE LIVROS

Os capítulos de livros correspondem a um grande volume entre os textos citados pelos autores de teses e dissertações. *Stricto sensu*, os capítulos não são tipos documentais diferentes dos livros, mas é importante diferenciar um capítulo que é a seção de uma obra e um capítulo que é uma obra em si, retomando a entidade obra do FRBR.

No caso de livro subdividido em seções, temos um todo orgânico de conteúdo segmentado em partes; por exemplo, Estética da criação verbal é uma obra de Bakhtin dividida em sete capítulos. Esse tipo de livro é diferente de uma coletânea de obras diversas, escritas por autores diferentes, na maioria das vezes, e reunidas com um título coletivo: neste caso, cada um dos capítulos é uma obra única, a qual se aproxima mais dos gêneros ensaio ou artigo científico, pois cada um tem seus próprios objetivos, referenciais, procedimentos e resultados. Por exemplo, em Bakhtin: conceitos-chave, temos oito obras de autores diferentes reunidas por Beth Brait.

Nos estudos sobre análise de citação, os pesquisadores não diferenciam os capítulos que são seção de uma obra dos capítulos que são uma obra independente, contando-os conjuntamente. Para a pesquisa em questão, não foi feita essa separação, pois essas diferenças foram percebidas em fase já avançada de

trabalho, quando já se havia finalizado a mensuração dos 5.207 capítulos do *corpus*, mas notou-se que os capítulos mais citados fazem parte de coletâneas. Tendo em vista esse resultado, sugere-se, para as futuras análises de citação, a separação dos dois tipos de capítulo, o que vai gerar dois resultados: um com os capítulos do tipo seção mais citados e outro com os capítulos do tipo ensaio/artigo mais citados. Essa diferenciação, com certeza, gerará ranques diferentes dos que são obtidos quando não se faz essa separação, o que pode impactar os resultados de capítulos mais citados e da produtividade e frente de autores de capítulos, revelando dados significativos.

Outro detalhe importante diz respeito à contagem da produtividade dos autores dos livros de onde foram retirados os capítulos, ou seja, das obras-fonte. Segundo o Código de Catalogação Anglo-americano (2004, p. 21-6), “autor é a pessoa fundamentalmente responsável pela criação do conteúdo intelectual ou artístico de uma obra”. Na autoria compartilhada, os vários autores são responsáveis pelo conteúdo total da obra. Nos livros com organizador, há várias obras, cada uma com uma autoria diferente. A pessoa que reúne os textos para publicação não pode ser considerada como responsável intelectual pelo todo do conteúdo, e sim autor de um dos capítulos, no máximo. Atribuir a autoria ao organizador do livro seria equivalente a atribuir a autoria de todos os artigos de um periódico científico ao editor-chefe. Entretanto, na contagem da autoria desse tipo de livro, os organizadores, editores, compiladores e coordenadores costumam ser contados como autores. Mas essas indicações de responsabilidade não podem ser confundidas com autoria, pois esse tipo de obra se trata de coletânea de textos de vários autores (CÓDIGO..., 2004).

Portanto, sugere-se que os livros com organizador sejam contabilizados separadamente dos demais, a fim de verificar se os livros são de autoria individual, dois autores, três autores, mais de quatro autores ou coletânea.

5 CONSIDERAÇÕES FINAIS

Os desafios e as dificuldades impostas pelas pesquisas bibliométricas podem implicar em erros que afetarão os resultados e colocarão em xeque a credibilidade do estudo e, também, a do pesquisador. Algumas variáveis independentes, isto é, as que estão sujeitas à experiência e perícia do bibliometrista, conforme citadas por Silva, Hayashi e Hayashi (2011), não podem ser confundidas com variáveis dependentes, ou seja, com fatores que não são passíveis de serem controlados por ele, pois certas inconsistências nos dados podem e devem ser corrigidas.

Dentre essas variáveis, podem ser citados os erros e/ou inconsistências nas referências bibliográficas, como confusão na indicação de títulos de obras de periódicos, erros no registro de datas e edições e, em especial, a indicação de autoria. Para isso, sugere-se que os dados recebam um tratamento prévio rigoroso,

valendo-se de padrões já estabelecidos na Ciência da Informação não somente no que tange aos procedimentos bibliométricos, mas também aos padrões estabelecidos para a representação da informação, dada a importância dessa área perante o desenvolvimento teórico e prático para a identificação e a uniformização de dados dos registros do conhecimento frente às necessidades de seus usuários.

Dessa forma, conclui-se que o mesmo rigor seguido na Organização da Informação para a representação documental deve ser adotado nos procedimentos metodológicos das pesquisas bibliométricas, com a utilização de processos, ferramentas e produtos tipicamente empregados nessa área. As regras de catalogação, os catálogos de bibliotecas e a análise documental mostraram-se importantes não apenas no que se refere à confirmação e uniformização de elementos e dados, mas também como aporte teórico para a resolução de desafios que surgiram durante a pesquisa. A Análise Documental, em especial, revelou-se um procedimento que pode ser empregado como método científico para a leitura textual em substituição à leitura do resumo ou do texto completo, pois permite agregar mais confiabilidade tanto à exploração inicial como ao levantamento de dados de pesquisa. A utilização do FRBR mostrou-se essencial como modelo para a separação e tabulação do *corpus* das referências analisadas e para a identificação de tipologias documentais, apontando para a necessidade de diferenciação nas futuras análises de referência entre os capítulos que são seções de obras dos capítulos que são uma obra em si.

Por fim, espera-se que os bibliometristas tenham plena consciência de que suas competências interferem em todo o processo de pesquisa, mas que é essencial não se pautar em decisões pessoais ao se deparar com desafios procedimentais, mas sim em escolhas embasadas em teorias e métodos já estabelecidos, sendo extremamente profícua a (re)aproximação entre a análise bibliométrica e a Organização da Informação.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676**: métodos para análise de documentos, determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023**: informação e documentação, referências, elaboração. Rio de Janeiro, 2002.

CÓDIGO de catalogação anglo-americano. 2. ed. Revisão de 2002. São Paulo: FEBAB, Imprensa Oficial, 2004.

FUJITA, Mariangela Spotti Lopes; RUBI, Milena Polsinelli. Um modelo de leitura documentária para a indexação de artigos científicos: princípios de elaboração e uso para a formação de indexadores. **DataGramZero**, Rio de Janeiro, v. 7, n. 3, p. 1-19. jun. 2006.

GARCIA GUTIERREZ, Antonio Luis. **Linguística documental**: aplicación a la documentación de la comunicación social. Barcelona: Mitre, 1984.



MAROLDI, Alexandre Masson. **Estudos bibliométricos sobre educação indígena: frente de pesquisa, vida média e obsolescência da literatura citada em teses e dissertações.** 2017. Tese (Doutorado em Educação) – Universidade Federal de São Carlos, 2017. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/9451>. Acesso em: 19 nov. 2018.

POTTER, William Grey. Lotka's law revisited. **Library Trends**, v. 31, p. 21-39, 1981.

SILVA, Márcia Regina da; HAYASHI, Carlos Roberto Massao; HAYASHI, Maria Cristina Piumbato Innocentini. Análise bibliométrica e cientométrica: desafios para especialistas que atuam no campo. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 2, n. 1, p. 110-129, jan./jun. 2011. Disponível em: <http://www.revistas.usp.br/incid/article/view/42337>. Acesso em: 16 jun. 2017.

SOUSA, Raquel Juliana Prado Leite de. **Linguagens documentárias.** Batatais: Claretiano, 2019.

TOL, Richard S. J. Credit where credit's due: accounting for co-authorship in citation counts. **Scientometrics**, v. 89, n. 1, p. 291-299, Oct. 2011. Disponível em: <https://link.springer.com/content/pdf/10.1007%2Fs11192-011-0451-5.pdf>. Acesso em: 31 jul. 2018.

TRIFUNAC, Mihailo D. On publication rates in earthquake engineering. **Soil Dynamics and Earthquake Engineering**, v. 25, p. 413-420, 2005. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0267726105000953>. Acesso em: 09 nov. 2020.

VANZ, Samile Andrea de Souza; SANTIN, Dirce Maria; PAVÃO, Caterina Marta Groposo. A bibliometria e as novas atribuições profissionais em bibliotecas universitárias. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 9, n. 1, p. 4-24, mar./ago. 2018. Disponível em: <https://www.revistas.usp.br/incid/article/view/137741/140235>. Acesso em: 09 nov. 2020.

(RE)APPROXIMATIONS BETWEEN BIBLIOMETRICS AND INFORMATION ORGANIZATION: AN EXPERIENCE REPORT

Abstract: It deals with the use of processes, tools and usual products in the Knowledge Organization as a research procedure in the textual exploration, in the normalization and in the treatment of a large amount of data usually found during bibliometric research to measure scientific production. It describes the experiences obtained in the elaboration of bibliometric research and the solutions found for the standardization of citation data, in order to alert about possible errors of results and to provide suggestions and considerations that may be pertinent to researchers in the field, especially the beginners. The report is based on experience during doctoral research on a bibliometric survey of the Brazilian scientific production in the area of Education related to 237 theses deposited in the Digital Library of Theses and Dissertations between 1997 and 2016. It discusses the use of the Subject Analysis for reading strategic of textual corpus. It reports procedures adopted during the collection, treatment and analysis of data for the reduction of possible incongruities and difficulties, which can lead to errors of results and delays in the research. Suggests the use of library catalogs, author bases and cataloging rules for the conference and normalization of authorship and title data, in order to facilitate measurement, avoid rework and ensure correct survey of citation indicators and research front and also as a way of increasing the researcher's competence in information.

Keywords: Scientific methodology. Bibliometrics. Scientometrics. Knowledge organization. Scientific production. Experience report.