

Classificação temática para o mapeamento de campos científicos

estudo de caso na área de comunicação social

Carlos Alberto Ávila Araújo

Como citar: ARAÚJO, Carlos Alberto Ávila. Classificação temática para o mapeamento de campos científicos: estudo de caso na área de comunicação social. *In:* FUJITA, Mariângela Spotti Lopes; MARTELETO, Regina Maria; LARA, Marilda Lopes Ginez de (org.). **A dimensão epistemológica da ciência da informação e suas interfaces técnicas, políticas e institucionais nos processos de produção, acesso e disseminação da informação.** Marília: Fundepe; São Paulo: Cultura Acadêmica, 2008. p. 253-268. DOI: <https://doi.org/10.36311/2008.978-85-98176-17-8.p253-268>



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição-NãoComercial-SemDerivações 4.0 (CC BY-NC-ND 4.0).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-No comercial-Sin derivados 4.0 (CC BY-NC-ND 4.0).

CLASSIFICAÇÃO TEMÁTICA PARA O MAPEAMENTO DE CAMPOS CIENTÍFICOS: ESTUDO DE CASO NA ÁREA DE COMUNICAÇÃO SOCIAL

Carlos Alberto Ávila Araújo¹

RESUMO: apresenta-se uma proposta de mapeamento temático de campos científicos a partir da criação de sistemas facetados. Inicialmente, discutem-se alguns dos princípios da teoria da classificação facetada. A seguir, são analisados dois trabalhos de mapeamento temático da área de comunicação social que tiveram como objeto empírico 754 teses e dissertações. Por fim, apresenta-se o resultado de um mapeamento temático destas mesmas 754 teses e dissertações, utilizando-se um sistema facetado. São comparados os resultados e evidenciadas as possibilidades abertas pelo uso de um sistema facetado.

PALAVRAS-CHAVE: Mapeamento temático. Produção científica. Classificação facetada.

ABSTRACT: In this study, a proposal of a thematic mapping of scientific disciplines using faceted systems will be presented. Initially, some aspects of the theory of classification will be analyzed. Next, two thematic mappings of the area of social communication that use as empirical material 754 thesis and dissertations will be analyzed. Finally, the results of a thematic mapping of this same 754 thesis and dissertations using a faceted system will be presented. The set of results will be compared, indicating the possibilities created by using a faceted system.

INTRODUÇÃO

A ciência é uma forma de conhecimento que busca conhecer, em princípio, toda a realidade. Assim é que diferentes ramos, áreas ou disciplinas científicas se debruçam sobre fenômenos físicos, químicos, biológicos, chegando aos fenômenos humanos e sociais. Dos processos microscópicos aos gigantes fatos astronômicos, dos mitos e religiões às burocracias contemporâneas, tudo aquilo passível de ser conhecido pelo ser humano converte-se em objeto de pesquisa para a ciência.

Nesse processo, também a própria ciência tornou-se um objeto de problematizações e pesquisas. Dentro mesmo do conhecimento científico foram-se construindo tradições de reflexão sobre a ciência: a validade de seus fundamentos, seus processos e produtos, a relação que ela estabelece com o contexto mais amplo em que está inserida.

¹ Professor Adjunto da Escola de Ciência da Informação da UFMG

Entre as diversas áreas da ciência que a tomam como o próprio objeto de estudo estão aquelas de natureza mais filosófica, tais como a teoria crítica elaborada por Adorno e Horkheimer ainda na década de 1940, denunciando a vinculação da ciência a processos de dominação e o fracasso do projeto iluminista, transformado em um projeto de “mistificação das massas”; a filosofia da ciência, que nas décadas de 1960 e 1970 foi palco de intensos debates envolvendo pesquisadores como Popper, Feyerabend, Lakatos e Habermas a respeito do caráter político e da dimensão ideológica da ciência; a perspectiva de Kuhn que, numa abordagem histórica, analisa a ciência a partir do conceito de “paradigma”.

Outro ramo que se desenvolveu ao longo do século XX foi a sociologia da ciência. Uma primeira perspectiva, de natureza funcionalista, foi inaugurada por Merton e se desenvolveu nas décadas seguintes com os estudos sobre as comunidades científicas (Crane, Crawford, Ben Storer). Uma outra perspectiva, de cunho marxista, cujo autor mais proeminente é Bourdieu, enfatizou as disputas e conflitos existentes no interior do campo científico. Recentemente, abordagens mais etnográficas, originadas com as pesquisas de Latour e Woolgar, e construtivistas, como as de Knorr-Cetina, têm ganhado projeção.

Também parte significativa da ciência da informação tem a ciência como objeto de estudo. Os trabalhos sobre as comunidades científicas e os “colégios invisíveis” realizados por Price e outros constituem um exemplo. Ou, então, os estudos sobre a comunicação científica, desde as tentativas de representação do fluxo de conhecimento, com as pioneiras pesquisas de Garvey e Griffith, aos estudos sobre as características dos canais formais e informais por autores como Meadows, ou mesmo a teoria do crescimento epidêmico da literatura de Goffman.

Há ainda todo um ramo da ciência da informação dedicado à análise da ciência. Trata-se da bibliometria, que, de técnica quantitativa e estatística de medição de bibliografias, passou a concentrar-se nos índices de produção e disseminação do conhecimento científico, a ponto mesmo de ter surgido o termo “cienciometria” para designar exatamente o exercício da bibliometria no campo científico. Trata-se de um ramo de estudos que se desenvolveu a partir da elaboração de leis empíricas sobre o comportamento da literatura. Entre seus principais marcos de desenvolvimento estão o método de medição de produtividade de cientistas (lei de Lotka, de 1926), a lei de dispersão do conhecimento científico de Bradford (elaborada em 1934) e o modelo de distribuição e frequência de palavras num texto (lei de Zipf, publicada em 1949).

A área mais importante da bibliometria é a análise de citações, que pode ser definida como a técnica que investiga as relações entre os documentos citantes e os documentos citados considerados como unidades de análise, no todo ou em suas diversas partes (autor, título, origem

geográfica, ano, idioma de publicação). Na década de 1960, com a utilização do computador, a técnica, que já havia sido realizada desde séculos anteriores, ganha novo fôlego, sendo o ano de 1963 considerado um grande marco, com o surgimento do primeiro índice de citações, o Science Citation Index (SCI) por Eugene Garfield, fundador do ISI – Institute of Scientific Information. Esse tipo de análise permite a identificação e descrição de uma série de padrões na produção do conhecimento científico, tais como autores mais citados, autores mais produtivos, tipo de documento mais utilizado, idade média da literatura utilizada, obsolescência da literatura, entre muitos outros.

Atualmente diversos indicadores de ciência e tecnologia são produzidos e utilizam contribuições bibliométricas, tais como o Science and Technology Indicators (EUA), o Science & Technology Indicators (Ásia), o S&T Indicators for the European Research Area (Europa) e o Main Science and Technology Indicators (dos países da OECD).

Por fim, um último ramo de estudos sobre a ciência é aquele que busca realizar mapeamentos temáticos de uma determinada disciplina ou área do conhecimento científico. Estudos dessa natureza buscam definir os principais assuntos ou temáticas de uma ciência particular e determinar quanto da atividade científica é dedicada a cada um deles, tomando como objeto empírico conjuntos de teses e dissertações, artigos de periódicos ou pesquisas financiadas por agências de fomento, ou, ainda, tomando toda a produção de um conjunto de pesquisadores, de docentes de universidades, de institutos de pesquisa, entre outras possibilidades.

Estes estudos podem constituir-se em simples elaborações de listas de assuntos e contagem de frequências desses assuntos em produtos científicos ou envolver um trabalho mais elaborado na construção da listagem de assuntos. Essa segunda opção está sendo desenvolvida nas últimas décadas com o apoio das teorias da classificação. Assim, a abordagem dos *boundary objects* (ALBRECHTSEN; JACOB, 1998), as pesquisas em *domain analysis* (HJORLAND, 2002; BEGHTOL, 1995) e os estudos em visualização de literaturas (WHITE; MCCAIN, 1997) constituem importantes exemplos de aplicação de princípios da teoria da classificação para o mapeamento temático de campos científicos.

O presente trabalho busca apresentar uma nova proposta que vem somar-se a esses esforços que estão sendo realizados para construir melhores sistemas classificatórios com o objetivo de quantificação temática das pesquisas realizadas em campos científicos. Trata-se da proposta de construção de um sistema facetado com o objetivo específico de mapeamento temático científico. O trabalho desenvolvido na pesquisa aqui relatada consistiu na construção de um sistema facetado para a área de comunicação social. Como forma de avaliação das potencialidades abertas pela utilização de um sistema facetado para tal fim, foi realizada

ainda uma aplicação desse sistema a um conjunto de 754 dissertações e teses defendidas entre 1992 e 1996 em todo o país, para então se proceder a uma comparação com duas outras pesquisas que realizaram mapeamentos temáticos destas mesmas 754 dissertações e teses, porém sem utilizar uma estrutura facetada. Essa comparação, apresentada no final deste texto, pretende apontar as contribuições que um sistema facetado pode trazer para o mapeamento temático de disciplinas científicas.

CLASSIFICAÇÃO FACETADA

Não cabe, nos limites deste trabalho, uma apresentação detalhada e profunda do que é a teoria da classificação facetada. Serão vistos a seguir, portanto, apenas alguns aspectos, indispensáveis para a compreensão do papel que um sistema facetado pode ter no mapeamento de campos científicos, dado que este é apenas um objetivo possível para a classificação facetada, mas não o único nem mesmo o principal – a classificação facetada surgiu com o objetivo de organizar coleções de documentos em sistemas dinâmicos e multidimensionais. Assim, antes de mais nada, deve-se ter em mente que a classificação facetada constitui uma das teorias da classificação bibliográfica.

As teorias da classificação bibliográfica, de uma forma geral, buscam promover uma classificação sistemática, lógica, que reflete crítica e filosoficamente sobre os elementos de ligação que servem para a reunião de conceitos. É na lógica aristotélica (LANGRIDGE, 1977, p. 24) que são encontradas as primeiras contribuições para a formulação de uma teoria da classificação. A primeira contribuição de Aristóteles diz respeito à divisão dicotômica dos objetos em gênero e espécie. Trata-se de uma hierarquização conceitual que divide um tema geral em espécies a partir da aplicação de uma característica classificatória.

Para que existam gêneros e espécies é imprescindível a existência de um princípio classificatório ou “característica de uma classificação”, que é o “elemento que serve para reunir os grupos segundo as semelhanças que apresentam” (BARBOSA, 1969, p. 14). Essa é a segunda contribuição de Aristóteles para a teoria da classificação: a elaboração dos cinco predicados, isto é, os cinco tipos de relações existentes num arranjo lógico:

- Gênero: classe ou grupo de seres ou objetos que possuem um determinado número de características em comum;
- Espécie: ser ou coisa que possui uma diferença específica que a distingue de seu gênero próximo; a espécie é obtida do gênero pelo acréscimo de uma diferença;
- Diferença: é a característica que serve para gerar uma espécie; cada acréscimo de diferença gera uma nova espécie;

- Propriedade: algo próprio de cada elemento de uma classe mas que não é imprescindível à definição da classe;
- Acidente: qualidade não-obrigatória a todos os elementos de uma classe, isto é, que pode, ou não, estar presente em um conceito.

Um exemplo de como os cinco predicados se relacionam num processo classificatório pode ser o seguinte: um conjunto de cadeiras é dividido, conforme a cor, em subconjuntos contendo, cada um, cadeiras azuis, verdes, brancas e amarelas. Neste exemplo, “cadeiras” representa o gênero; “cor” representa a diferença; cadeiras azuis, cadeiras verdes, cadeiras brancas e cadeiras amarelas são as espécies; azul, verde, branca e amarela são propriedades das cadeiras; grande, média e pequena são acidentes. Ou seja, uma cadeira azul pode ser grande ou pequena, isso não afeta seu pertencimento na espécie “cadeiras azuis”. Nos sistemas de classificação construídos a partir dessa lógica,

[...] as espécies são, portanto, obtidas pela diferença específica, ou seja, as qualidades ou atributos que, somados ao próprio gênero, as distinguem. Diremos, então, que a cada derivação conceitual as espécies daí decorrentes adquirem pelo menos *um* atributo a mais que seu gênero próximo, tornando-se mais intensas ou compreensíveis na medida em que aumenta o número de diferenças (DODEBEI, 2002, p. 80).

Os predicados estão envolvidos na construção de uma estrutura de classificação. Alguns princípios lógicos são requeridos, ainda, para uma adequada estrutura conceitual. Conforme Dodebei, três são esses princípios lógicos. O primeiro é o princípio da completude, que determina que: “a divisão do conceito deve ser completa, adequada e ordenada por complexidade crescente, isto é, enumerar todas as espécies de que o gênero se compõe, do simples ao complexo ou do abstrato ao concreto” (DODEBEI, 2002, p. 82). A autora apresenta um exemplo: caso se tenha, como classe, “educação”, e se queira dividi-la pelos níveis de instrução, é preciso que todas as espécies possíveis sejam enumeradas – no caso, “pré-escolar”, “1º grau”, “2º grau” e “3º grau”. O segundo é o princípio da irredutibilidade, segundo o qual “a divisão deve garantir que a cada dedução conceitual os conteúdos sejam irredutíveis entre si, isto é, não se deve enumerar mais do que os elementos verdadeiramente distintos entre si, de maneira que nenhum esteja compreendido no outro.” (DODEBEI, 2002, p. 82).

A autora fornece outro exemplo: caso se tenha como gênero “homem” e se queira dividir esse gênero pelo local de nascimento, podem ser geradas, como espécies, “americanos”, “brasileiros”, “cariocas”. Neste caso há um erro, pois o conceito de “cariocas” está subordinado ao conceito de “brasileiros” e não pode ocupar o mesmo lugar na derivação conceitual.

O terceiro princípio, e o mais importante, é o da mútua exclusividade, que postula que “para cada derivação conceitual deve-se

usar apenas uma característica do conceito” (DODEBEI, 2002, p. 83). O exemplo da autora é o seguinte: do gênero “livros” obtem-se as espécies “filosofia”, “50 cm” e “direito”. O erro lógico, aqui, é que foram consideradas duas características de derivação ao mesmo tempo: o assunto do livro e seu tamanho. Para serem respeitados os princípios lógicos, em primeiro lugar, os livros devem ser divididos pelo assunto e, a seguir, pelo tamanho – ou vice-versa. BARBOSA (1969, p. 24) também se refere a essa propriedade, afirmando que a característica escolhida deve “ser consistente ou exclusiva”, isto é, deve-se dividir um assunto apenas por um critério de divisão, para a seguir se usar-se outro. De acordo com a autora, quando isso não acontece, tem-se uma “classificação cruzada”, em que um assunto tanto pode estar numa classe como em outra ou outras. Um outro autor ressalta, ainda, a importância de respeitar o princípio da exclusividade mútua em sistemas classificatórios:

Apenas um princípio de divisão deve ser usado de cada vez para produzir classes mutuamente exclusivas. Se elas se sobrepõem então é impossível se ter certeza a que classe um determinado objeto pertence. Esse erro é conhecido como classificação cruzada (LANGRIDGE, 1977, p. 24).

O autor apresenta um exemplo. Sua grande classe é “pessoas” que, dividida, resultou nas classes “jovens”, “de meia idade”, “idosas”, “homens” e “mulheres”. Homens jovens, por exemplo, neste caso, podem tanto ser colocados em “jovens” como em “homens”. As características (idade e gênero) “devem ser aplicadas uma de cada vez, em qualquer ordem que satisfaça ao propósito em questão” (LANGRIDGE, 1977, p. 25).

Estas noções estão na base da definição de classificação como hierarquia (SHERA; EGAN, 1969, p. 55) e foram fundamentais na formulação dos primeiros sistemas de classificação bibliográfica, normalmente conhecidos como sistemas de classificação hierárquicos por organizarem os conceitos em estruturas de gênero/espécie, identificando características essenciais e acidentais. Nesses sistemas, a estrutura conceitual é obtida pela aplicação sucessiva de características de divisão. Entre os primeiros sistemas de classificação bibliográfica, os mais importantes são os de Cutter, a Classificação Decimal de Dewey (CDD), a Classificação Decimal Universal (CDU) e a classificação da *Library of Congress*.

Na primeira metade do século XX surgiu um outro tipo de classificação bibliográfica. Alguns autores estabelecem que se trata do momento em que a teoria tradicional é confrontada com a teoria moderna, ou que a teoria descritiva é confrontada com a teoria dinâmica (CAMPOS, 2001, p. 31). A grande novidade é o desenvolvimento da teoria da classificação facetada:

A Teoria da Classificação Facetada é desenvolvida por Shiyali Ramamrita Ranganathan na década de 30, a partir da Colon Classification, tabela

de classificação elaborada para a organização do acervo da Biblioteca da Universidade de Madras, na Índia [...] Ranganathan foi aquele que conseguiu estabelecer princípios para uma nova teoria da classificação bibliográfica e o fez tendo como base o próprio conhecimento (CAMPOS, 2001, p. 26, 28).

O diferencial do sistema classificatório proposto por Ranganathan é a utilização de uma estrutura dinâmica, multidimensional, com a introdução do termo faceta “que ficou sendo, nos modernos estudos sobre teoria da classificação, o substituto de característica” (BARBOSA, 1969, p. 16). Seu ponto de partida foi uma outra idéia de Aristóteles, a definição das dez categorias do ser, isto é, as formas sob as quais os seres e objetos se apresentam.

Refletindo sobre essas categorias em suas obras (principalmente em *Five Laws of Library Science*, de 1931, e *Prolegomena to Library Classification*, de 1937), Ranganathan percebeu que cada uma delas poderia servir de princípio classificatório mas, nos sistemas utilizados até então, hierárquicos e monolíticos, apenas uma categoria podia ser utilizada por vez como característica de classificação.

Assim, se for tomado como exemplo um conjunto de cadeiras, pode-se pensar em agrupá-las de diferentes formas. Usando-se a categoria quantidade, seriam separadas as grandes das médias e das pequenas. Usando-se a categoria qualidade, pensada aqui como a cor, separar-se-iam as brancas das azuis e verdes. Ou, ainda, usando-se a categoria sofrimento de ação, separar-se-iam as produzidas manualmente das produzidas industrialmente. Num sistema hierárquico, contudo, não podem ser utilizados esses três princípios ao mesmo tempo, mas apenas um. Se for escolhida a categoria quantidade, então se terão, para o gênero cadeira, espécies grandes, médias e pequenas. Em cada espécie haverá cadeiras de diferentes cores. Se for a intenção, agora, separar por cores, terá de ser aplicado um novo princípio de classificação *dentro* das espécies já formadas, que agora se convertem em gêneros. Haverá, assim, cadeiras pequenas brancas, pequenas azuis e pequenas verdes; médias brancas, médias azuis e médias verdes; grandes brancas, grandes azuis e grandes verdes. Ou seja, a categoria qualidade, definida neste exemplo como relativa à cor, fica aqui dispersa em relação ao gênero geral “cadeira”. Caso se queira ter todas as brancas juntas e separadas das verdes e das azuis, deve-se abandonar o primeiro princípio classificatório e substituí-lo por outro. Não se pode, dessa forma, dividir as cadeiras entre as brancas, as azuis e as pequenas. Aqui há um erro lógico de classificação, a aplicação de diferentes critérios classificatórios, gerando conceitos de natureza diferente que convivem na mesma operação classificatória. A categoria “pequena” não compartilha com as demais a característica que define a classe.

Ao mesmo tempo que refletia sobre essa questão, Ranganathan percebeu, na época em que estudou na Inglaterra, que a CDD era largamente utilizada, mas freqüentemente adaptada e modificada. Quando, a partir de 1931, começou a compor um novo sistema, idealizou uma estrutura bem mais elástica do que as já existentes, isto é, que permitisse a síntese, o agrupamento de vários componentes para a especificação do assunto. Para isso, adotou o uso de dois pontos como símbolo para correlacionar idéias diferentes. Sua classificação, a Colon Classification (“Classificação de dois pontos”), foi publicada pela primeira vez em 1933.

O que importa, aqui, não é exatamente a Colon, mas o princípio que anima sua construção: o conceito de “faceta”, definida como “a totalidade das subdivisões resultantes da aplicação de uma única característica” (PIEDADE, 1977, p. 22). A utilização do termo “faceta” não implica apenas uma mudança terminológica, mas uma mudança na concepção do processo classificatório:

Segundo Ranganathan, analisar um assunto por *facetas* significa que cada aspecto desse assunto pode ser visto como as manifestações de certas *características* ou *facetas* que obedecem a postulados pré-determinados. O sistema torna-se, assim, multidimensional e ilimitado (BARBOSA, 1969, p. 166).

Esse raciocínio permite a simultaneidade de critérios classificatórios sem que se incorra nos “sérios defeitos” da “classificação cruzada” (PIEDADE, 1977, p. 16), porque o que muda é a estrutura do sistema, que passa a aceitar a convivência de subdivisões de naturezas diferentes dentro de cada assunto.

SISTEMA FACETADO PARA MAPEAMENTO TEMÁTICO

Os sistemas facetados podem ser construídos para atender a diferentes finalidades. A mais comum, como já citado, é dirigida para a sua utilização em bibliotecas e outros sistemas de informação com o objetivo de organizar os documentos e propiciar formas dinâmicas e diferenciadas de acesso a eles. Contudo, esse não é o único objetivo possível. Entre essas várias aplicações, destaca-se a utilização dos princípios de classificação facetada no mapeamento temático de campos científicos.

No contexto brasileiro, vários são os estudos que buscam mapear e quantificar temáticas estudadas numa disciplina científica específica para a produção de um diagnóstico dessa disciplina. Entre esses estudos podem ser citados, como exemplos, na área de Ciência da Informação, trabalhos de Oliveira (1998), Witter e Pécora (1997), Mueller e Pecegueiro (2001), Frota (1998) e Bufrem (1996). Outros exemplos, em outras áreas de conhecimento, são as pesquisas que buscam mapear os assuntos estudados

em áreas como contabilidade (RICCIO; SAKATA; CARASTAN, [19—?]), gerência de operações (ARKADER, 2003), marketing (VIEIRA, 2003), finanças (LEAL et al., 2003) e recursos humanos (TONELLI et al., 2003).

Utilizar os princípios da classificação facetada para a quantificação de temas estudados numa área de conhecimento pressupõe um tipo de trabalho com especificidades próprias. A primeira dessas especificidades diz respeito ao impacto que os três princípios lógicos da derivação conceitual possuem na totalização das frequências. Para evidenciar como isso acontece, será utilizado um exemplo hipotético, no caso, mais uma vez, as cadeiras.

Supondo-se que, num determinado conjunto de cadeiras, fossem somadas as azuis, as brancas, as pequenas e as médias, obtendo-se, por exemplo, 20% de cadeiras azuis, 35% de brancas, 20% de pequenas e 25% de médias. Essa informação não diz o suficiente pois, se é sabido que 20% de cadeiras são azuis, esse dado *deveria* informar que 80% são de outras cores. Mas se 45% das cadeiras foram “indexadas” em categorias de tamanho, não se possui a informação sobre a cor delas, de forma que é possível existirem outras cadeiras azuis no meio daquelas classificadas como “pequenas” ou “médias”. Isso significa que dizer que 20% das cadeiras são azuis não é uma afirmação segura – é possível que esse número seja maior.

A recíproca é verdadeira: nada garante que, entre as 55% de cadeiras classificadas pela cor, existam apenas pequenas, ou médias, ou grandes. Pode-se ter uma cadeira pequena entre aquelas classificadas como “azuis” e, então, o percentual de cadeiras pequenas deveria ser maior do que 20.

Enfim: uma distribuição estatística entre assuntos pesquisados num conjunto de trabalhos precisa realizar a soma entre categorias obtidas por meio de um mesmo princípio classificatório, isto é, relativos a uma mesma *faceta* do assunto. Esse problema lógico, como visto, é conhecido como “classificação cruzada”. Ele gera vários tipos de complicação. Além das distorções no somatório das frequências de cada característica, ela também causa confusão na definição da forma de contabilização de cada cadeira: caso se tenha uma cadeira azul média, por exemplo, não é evidente em qual grupo ela deve ser contabilizada, podendo haver duas opções.

COMPARAÇÃO ENTRE OS DOIS SISTEMAS

É esse raciocínio que está na base da proposta da pesquisa relatada neste artigo, visto que se buscou construir um sistema facetado para a área de comunicação social e utilizar esse sistema para a comparação com os resultados obtidos em outros trabalhos de mapeamento temático da comunicação social que utilizam outras formas de classificação.

Muitos estudos que buscam mapear campos específicos do conhecimento científico, e aqui se incluem alguns relativos à área de comunicação social, optam por não realizar uma discussão conceitual e acabam utilizando, como lista de assuntos para a contagem de frequência, termos obtidos a partir de critérios classificatórios diferentes, promovendo classificações cruzadas. Normalmente resolve-se essa questão classificando-se os estudos a partir da característica que se evidencia com maior importância (quando um estudo pode ser contabilizado em mais de uma classe).

Entre esses trabalhos que têm por objetivo analisar os assuntos das pesquisas realizadas na área de comunicação, dois serão aqui destacados. O primeiro é o trabalho de Stumpf e Capparelli, que tiveram como objeto as 754 dissertações e teses defendidas nos programas de pós-graduação em comunicação do Brasil no período entre 1992 e 1996. Buscando realizar um mapeamento quantitativo da área, os autores trabalharam com uma listagem de assuntos obtida a partir da combinação de diversas bases de dados e publicaram, como resultado, um mapeamento que lista a frequência dos 20 assuntos principais. Esse resultado é apresentado a seguir, na Tabela 1:

Tabela 1 - Assuntos estudados nas teses e dissertações em comunicação no Brasil (1992-1996), segundo classificação adotada por STUMPF e CAPPARELLI

Assunto	N	%	Assunto	N	%
Semiótica	100	13,2	Imprensa	39	5,2
Literatura	77	10,2	Comunic. organizacional	36	4,8
Televisão	62	8,2	Imagem	36	4,8
Jornalismo	60	8,0	Psicanálise	36	4,8
Arte	56	7,4	Narrativa	24	4,5
Discurso jornalístico	55	7,3	Música	33	4,4
Educação	45	6,0	Recepção	31	4,1
Novas tecnologias	45	6,0	História	30	4,0
Cultura	41	5,4	Filosofia	29	3,8
Cinema	39	5,2	Linguagem	26	3,4

Fonte: Stumpf; Capparelli, 2000, p. 248.

Outro trabalho relevante, centrado no mesmo objetivo, é o de Cicilia Peruzzo, que se debruça sobre as mesmas 754 dissertações e teses para realizar o mapeamento temático e sua quantificação. A autora preferiu construir uma lista própria de termos, em vez de utilizar os termos de uma base de dados existente. Como resultado, apresentou um mapeamento composto de 46 assuntos, como se pode ver na Tabela 2:

Tabela 2 - Assuntos estudados nas teses e dissertações em comunicação no Brasil (1992-1996), segundo classificação adotada por Peruzzo

Assunto	N	%	Assunto	N	%
Jornalismo	107	14,2	Teatro	11	1,5
Literatura	71	9,4	Desenv. tecnologias da com.	10	1,3
Televisão	37	4,9	Pintura/Xilogravura	10	1,3
Cultura	36	4,8	Marketing	10	1,3
Estética	36	4,8	Psicanálise/Inconsciente	10	1,3
Publicidade/Propaganda	33	4,4	Teorias e metod. da com.	10	1,3
Relações Públicas	33	4,4	Comunicação Científica	10	1,3
Cinema	32	4,2	Outros	10	1,3
Com. Popular/Alternativa	22	2,9	Comunicação e Saúde	9	1,2
Música	20	2,7	Corpo e Comunicação	8	1,0
Meios/Tec. com. na educ.	20	2,7	Comunicação Rural	5	0,7
Sociologia/Ciência Política	19	2,5	Dança	5	0,7
Linguagem e Significação	18	2,4	Empresas/Instit. de com.	5	0,7
Filosofia	15	2,0	Religiões e com. religiosa	4	0,5
Pensad./Jornalistas/Artistas	14	1,8	Humor	4	0,5
Com., tecnologias e arte	14	1,8	Futebol/Esporte	4	0,5
Educação	14	1,8	Grafito	3	0,4
Fotografia	13	1,7	Tradução	3	0,4
Inf. no capit./Ind. Cultural	12	1,6	Comunicação Grupal	2	0,3
Rádio	12	1,6	Política Cultural	2	0,3
Arquitetura	12	1,6	Transcodificação	2	0,3
Editoração	12	1,6	Cerâmica/Escultura	2	0,3
Vídeo	11	1,5	Romantismo/Modernidade	2	0,3

Fonte: Peruzzo, 2002, p. 56-60.

Como se tratam de listas hierárquicas, cujos termos foram obtidos a partir de critérios classificatórios diferentes, o que se pode notar nos dois casos é a possibilidade de um documento ser classificado sob termos diferentes, por exemplo, um trabalho sobre telejornalismo pode ser contabilizado tanto sob o termo jornalismo como sob televisão. As contagens dos autores citados acima, portanto, precisam constantemente fazer esse tipo de opção ao realizar a contagem dos assuntos.

Há ainda outra questão. A informação de que os estudos sobre jornalismo somam 8,0% (no primeiro caso) ou 14,2% (no segundo caso) não diz o suficiente, pois, se é sabido que 8,0% das dissertações e teses são sobre jornalismo (ou 14,2%), esse dado deveria informar que 92%

(ou, no segundo caso, 85,8%) não são sobre jornalismo. Contudo, não se pode afirmar isso, pois é possível que, entre as dissertações e teses classificadas sob outros termos (por exemplo, televisão, semiótica, arte) é possível que existam algumas sobre jornalismo, que apenas não foram classificadas sob esse termo. Isso significa que afirmar que 8,0% (ou 14,2%) dos documentos analisados são sobre jornalismo não é uma informação segura – é possível que esse número seja maior.

Isso acontece porque as listas de assuntos geradas nos dois mapeamentos misturam termos obtidos a partir de critérios classificatórios diferentes: alguns se referem, por exemplo, a habilitações profissionais do campo; outros, a meios de comunicação; outros, a assuntos veiculados nos processos comunicativos.

Enfim, uma distribuição estatística entre assuntos pesquisados num conjunto de trabalhos precisa realizar a soma entre categorias obtidas por meio de um mesmo princípio classificatório, isto é, relativos a uma mesma *faceta* do assunto. Esse problema lógico, como visto, é conhecido como “classificação cruzada”. Ele gera vários tipos de complicação. Além das distorções no somatório das frequências de cada característica, ela também causa confusão na definição da forma de contabilização de cada assunto: caso se tenha por assunto telejornalismo, por exemplo, não é evidente em qual grupo ele deve ser contabilizada, podendo haver pelo menos duas opções.

Como exemplo das potencialidades trazidas pela classificação facetada no mapeamento de campos científicos, evoca-se aqui o resultado da aplicação de um sistema facetada construído especificamente para este fim. Esse trabalho distingue seis facetas para a área: modelo de fenômeno comunicativo; âmbito relacional do processo; meios, veículos e suportes; assuntos das mensagens; disciplinas de interface; habilitações profissionais. Entre os resultados encontrados destacam-se:

- a) A primeira faceta apareceu em 90,98% das dissertações e teses. Destas, o foco “significação” atingiu 38,99%, seguido de “efeitos” com 18,29%, “cultura” com 11,37% e “técnica” com 9,99%;
- b) A segunda faceta apareceu em 79,97% das dissertações e teses, com o foco “comunicação de massa” em primeiro com 60,78%, a seguir “comunicação institucional” com 22,31% e “comunicação comunitária” com 12,77%;
- c) A terceira faceta apareceu em 84,88% das dissertações e teses, sendo o foco mais freqüente “jornal”, com 15,47%, seguido de “livro” com 14,66% e “televisão” com 13,05%;
- d) A quarta faceta foi encontrada em 79,58% dos documentos analisados, ficando o foco “sociedade” em primeiro com 13,70%, em segundo “educação” com 8,36% e em terceiro “política” com 7,40%;

- e) A quinta faceta foi verificada em 38,20% das dissertações e teses, sendo “semiótica”, com 28,46%, o foco com maior índice, seguido de “filosofia” com 11,26% e “psicanálise” com 10,67%;
- f) A última faceta apareceu em apenas 22,41% das dissertações e teses, sendo “jornalismo” o foco com maior frequência, atingindo 43,68%, vindo a seguir “relações públicas” com 15,06% e “publicidade” com 14,46%.

A utilização do modelo facetado apresenta uma série de vantagens. Por exemplo, o sistema facetado permite explicitar os critérios de divisão dos assuntos, dando clareza sobre em que sentido ou aspecto do assunto geral o assunto específico está sendo compreendido. No exemplo citado acima, “política” é um termo que pode tanto ser foco da faceta 4, relativa aos assuntos das mensagens, quando da faceta 5, referente a disciplinas de interface. Um sistema facetado permite identificar, sempre, cada um desses sentidos, ao contrário de um sistema hierárquico. Afinal, no sistema facetado, o princípio de classificação, que origina cada termo, é necessariamente explicitado.

Um segundo ganho diz respeito ao fato de o sistema, ao mesmo tempo em que considera e explicita as dimensões que compõem a área estudada - no nosso exemplo, a área de comunicação - e alerta para o fato de existirem outras dimensões além daquelas escolhidas para a formação das facetas - no caso, seis -, ele também permite a identificação do peso que cada uma das dimensões possui na totalidade do campo, através da indicação da quantidade de casos em que cada faceta se aplica - por exemplo, a faceta 3, referente aos meios e veículos de comunicação, ocorre em 84,88% das dissertações e teses, já a faceta 5, que identifica disciplinas de interface, ocorre em apenas 38,20%.

Essa informação revela, assim, o peso que a formação de interfaces (faceta 5) possui em relação à totalidade de teses e dissertações, garantindo uma informação mais confiável sobre o quanto a área recorre a outras disciplinas e campos científicos do que a simples constatação das frequências obtidas por termos que representam disciplinas científicas - na medida, mais uma vez, em que vários termos classificados por outras dimensões, como pelos meios de comunicação ou pelas profissões, poderiam deixar de ser contabilizados como movimento de interface. Com isso, o sistema facetado permite ainda mais um ganho metodológico na análise de campos científicos, que é a possibilidade de problematizar a importância relativa de cada uma das dimensões (ou facetas) que compõem este campo - permitindo problematizar, por exemplo, a relevância de linhas de pesquisa, de subáreas, etc.

Do ponto de vista dos dados quantitativos, o maior ganho diz respeito à exatidão dos valores encontrados a respeito de cada assunto. Assim, por exemplo, o mapeamento identifica que, entre as dissertações

e teses que têm por assunto algum meio ou veículo de comunicação, 13,05% são sobre televisão. Esse fato informa, com precisão, a recíproca, que 86,95% delas não são sobre televisão, são sobre outros meios de comunicação (livro, jornal, rádio, etc). Da mesma forma, quando se tem que, entre os documentos analisados que tratam de alguma habilitação profissional, 43,68% são sobre jornalismo, esse dado informa que os restantes 56,32% são sobre as demais habilitações profissionais (publicidade, relações públicas, etc). Obtidos desta forma, os dados encontrados permitem uma avaliação precisa sobre o “peso” de cada assunto dentro da área analisada, pois permite a comparação com os demais assuntos obtidos pelo mesmo princípio classificatório.

O que se pode concluir da comparação entre os resultados do mapeamento temático feito com o auxílio da classificação facetada e os outros dois é que a utilização de um sistema facetado abre uma nova possibilidade de visualização de assuntos pesquisados em uma disciplina científica, identificando outros resultados e buscando evidenciar aquilo que outros tipos de mapeamentos temáticos não podem detectar. Abrem-se, assim, ricas possibilidades de novas leituras dos dados para o mapeamento e a avaliação da atividade científica.

REFERÊNCIAS

ALBRECHTSEN, Hanne; JACOB, Elin. Classification systems as boundary objects in diverse information ecologies. *Advances in Classification Research*, New Jersey, v. 8, nov. 1998.

ARISTÓTELES. Categorías. In: AZCÁRATE, Patricio (Org). *Obras completas de Aristóteles*. Buenos Aires: Anaconda. 1947, Tomo II, p. 335-394.

ARKADER, Rebecca. A pesquisa científica em gerência de operações no Brasil. *RAE: Revista de Administração de Empresas*, São Paulo, v. 43, n. 1, p. 70-80, jan./mar. 2003.

BARBOSA, Alice Príncipe. *Teoria e prática dos sistemas de classificação bibliográfica*. Rio de Janeiro: Instituto Brasileiro de Bibliografia e Documentação, 1969.

BEGHTOL, Claire. Domain analysis, literary warrant and consensus: the case of fiction studies. *Journal of the American Society for Information Science*, Maryland, v. 46, n. 1, 1995, p. 30-44.

BUFREM, Leilah Santiago. *Linhas e tendências metodológicas na produção acadêmica discente do mestrado em Ciência da Informação do Instituto Brasileiro de Informação em Ciência e Tecnologia – Universidade Federal do Rio de Janeiro (1972-1995)*. Curitiba: UFPR, 1996. Tese para concurso de professor titular.

CAMPOS, Maria Luiza de Almeida. *Linguagem documentária: teorias que fundamentam sua elaboração*. Niterói: Ed. da UFF, 2001.

- DODEBEI, Vera Lúcia. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intexto; Rio de Janeiro: Interciência, 2002.
- FROTA, Maria Guiomar. A delimitação das unidades de análise em ciência da informação. *Ciência da Informação*, Brasília, v. 27, n.31, p. 262-267, set./dez. 1998.
- HJORLAND, Birger. Domain analysis in information science. *Journal of Documentation*, v. 58, n. 4, 2002, p. 122-162.
- LANGRIDGE, Derek. *Classificação: abordagem para estudantes de biblioteconomia*. Rio de Janeiro: Interciência, 1977.
- LEAL, Ricardo et al. Perfil da pesquisa em finanças no Brasil. *RAE: Revista de Administração de Empresas*, São Paulo, v. 43, n. 1, p. 91-104, jan./mar. 2003.
- MEADOWS, Arthur Jack. *A comunicação científica*. Brasília: Briquet de Lemos, 1999.
- MUELLER, Suzana; PECEGUEIRO, Cláudia. O periódico *Ciência da Informação* na década de 90: um retrato da área refletido em seus artigos. *Ciência da Informação*. Brasília, v. 30, n. 2, p. 47-63, maio/ago. 2001.
- OLIVEIRA, Marlene. *A investigação científica na Ciência da Informação: análise da pesquisa financiada pelo CNPq*. 1998. Tese (Doutorado) – Universidade de Brasília, Brasília, 1998.
- PERUZZO, Cicília. Em busca dos objetos de pesquisa em comunicação no Brasil. In: WEBER, Maria Helena; BENTZ, Ione; HOHLFELDT, Antonio. *Tensões e objetos da pesquisa em comunicação*. Porto Alegre: Sulina, 2002.
- PIEPADE, Maria Antonietta. *Introdução à teoria da classificação*. Rio de Janeiro: Interciência, 1977.
- RICCIO, Edson; SAKATA, Marici; CARASTAN, Jacira. A pesquisa contábil nas universidades brasileiras – 1962/1994. Disponível em: <<http://www.tecsi.fea.usp.br/riccio/artigos>>. Acesso em: 6 out. 2004.
- SHERA, Jesse; EGAN, Margaret. *Catálogo sistemático: princípios básicos e utilização*. Brasília: Ed. UnB, 1969.
- STUMPF, Ida; CAPPARELLI, Sérgio. Produção discente dos programas de pós-graduação em Comunicação (1992-1996). *Revista de Biblioteconomia & Comunicação*, Porto Alegre, v. 8, p. 241-250, jan./dez. 2000.
- TONELLI, Maria José et al. Produção acadêmica em recursos humanos no Brasil: 1991-2000. *RAE: Revista de Administração de Empresas*, São Paulo, v. 43, n. 1, p. 105-122, jan./mar. 2003.
- VIEIRA, Francisco. Narciso sem espelho: a publicação brasileira em marketing. *RAE: Revista de Administração de Empresas*, São Paulo, v. 43, n. 1, p. 81-90, jan./mar. 2003.
- WHITE, Howard; McCAIN, Katherine. Visualization of literatures. In: WILLIAMS, Martha (Ed). *Annual Review of Information Science and Technology*, New Jersey, ASIS, 1997. v. 32.

WITTER, Geraldina Porto; PÉCORA, Gláucia Maria. Temática das dissertações e teses em Biblioteconomia e Ciência da Informação no Brasil (1970-1992). In: WITTER, Geraldina Porto (Org). Produção científica. Campinas: Átomo, 1997.