

# Estudo sobre como autores de artigos de revistas de Ciências da Comunicação verbalizam seus objetos de estudos em termos de palavras-chave

Sueli Mara Soares Pinto Ferreira  
Maria Helena Morais  
Marcos Mucheroni  
Josefina Perez

## RESUMO

Este estudo faz a análise das palavras-chave encontradas nos artigos publicados nas revistas brasileiras e portuguesas que participam do Portal Univerciência.org - Revistas de Acesso Aberto em Ciências da Comunicação. Partimos da ocorrência e frequência das palavras-chave empregadas pelos respectivos autores e editores na perspectiva de mapear características e formas de expressão referentes aos objetos de seus estudos. Os resultados possibilitaram uma melhor compreensão da área e das dificuldades na construção de parâmetros que orientem à inserção de palavras-chave, por apresentar um contexto interdisciplinar de vocabulários híbridos e distantes entre si. .

**PALAVRAS-CHAVE:** Atribuição de palavras-chave. Ciências da Comunicação. Representação do conhecimento.

## 1 Introdução

O uso de palavras-chave para a representação do conhecimento produzido e divulgado, em distintas fontes de informação, tem sido um dos principais recursos para o refinamento de busca em sistemas de recuperação de informação. Presume-se que tais palavras devam expressar de maneira precisa o conteúdo relevante do documento a que ela corresponde. Porém, vale refletir que tal presunção tem relação direta com o termo escolhido e o propósito de tal escolha. Por exemplo, no âmbito da web, recorrendo a recursos de *folksonomia*<sup>1</sup> os usuários/leitores podem agregar distintas palavras-chave aos conteúdos com os quais estejam interagindo, de acordo com suas preferências, interesses e experiências. Já no âmbito científico, normas internacionais definem que palavras-chave devem fazer parte de seus produtos e artefatos (como artigos de revistas, teses e dissertações, dentre outros) e estas devem ser determinadas pelos próprios autores. Portanto, poder-se-ia inferir que deveriam efetivamente representar – da melhor maneira possível – o conteúdo tratado no documento, sob a perspectiva e foco de seu próprio criador.

Identificar termos que efetivamente expressem e sintetizem o conteúdo e, ainda, utilizar palavras compreensíveis que também façam parte do universo linguístico dos potenciais leitores e que, ao mesmo tempo, se concatenem de maneira clara com a estrutura lógica definida e partilhada nas distintas disciplinas, são alguns dos percalços que os autores enfrentam na atribuição de palavras-chave.

Instrumentos como tesouros e vocabulários controlados têm sido recorrentemente utilizados pela comunidade científica para a seleção de termos a serem utilizados como palavras-chave, especialmente em sistemas de informação automatizados. Por exemplo, o DECS – Descritores em Saúde<sup>2</sup>, desenvolvido pela BIREME, é referência entre autores de artigos e teses na área. Seu uso constante e crescente leva à atualização, melhoria e refinamento do instrumento. Ao mesmo tempo, fornece à comunidade um maior controle do processo de representação do conhecimento e seleção de termos recorrentes e representativos.

A maturação e o entendimento que atingiram as chamadas ciências duras (Saúde e Tecnologia, por exemplo), quanto à necessidade de representar bem o conhecimento para recuperar eficazmente sua própria produção científica, não é equivalente em todas as áreas. Os campos das Ciências Sociais e Humanidades, dentre as quais as Ciências da Comunicação, destituídas tanto de instrumentos (vocabulários controlados e/ou tesouros) quanto de sistemas de informação específicos, vivenciam uma situação problemática frente à falta de fontes de referência terminológica para o uso de seus pesquisadores.

<sup>1</sup> O termo folksonomy foi proposto por Thomas Vander Wal, definido pela junção das palavras folks (povo) e taxonomy e se refere à categorização colaborativa de conteúdo por meio de palavras-chave livremente escolhidas pelos usuários. (LACERDA ; VALENTE, 2007).

<sup>2</sup> URL: <http://decs.bvs.br>

Tal constatação derivou uma série de perguntas: como os pesquisadores da área definem suas palavras-chave? A que instrumentos terminológicos recorrem quando necessitam de ajuda? Até que ponto estão cientes do papel que tais palavras ocupam quando inseridas em sistemas de recuperação da informação? Neste contexto, este estudo se propõe a analisar as palavras-chave publicadas nos artigos das revistas brasileiras e portuguesas que participam do Portal Univerciência.org - Revistas de Acesso Aberto em Ciências da Comunicação. Tal análise parte da verificação da ocorrência e frequência das palavras-chave selecionadas e empregadas pelos respectivos autores e editores, na perspectiva de se mapear características e formas de expressão referentes aos objetos de seus estudos.

■  
<sup>3</sup> Disponível em <http://www.univerciencia.org>

O Portal Univerciência.org<sup>3</sup> está sendo desenvolvido pela equipe do Centro de Estudos em Design de Sistemas Virtuais Centrado no Usuário, da Escola de Comunicações de Artes, Universidade de São Paulo (CEDUS/ECA/USP), e tem como foco principal contribuir para o desenvolvimento da pesquisa em Ciências da Comunicação, favorecendo o acesso aberto e irrestrito ao conhecimento gerado na área. Para tanto, por meio de uma parceria com representantes da comunidade científica, o Portal disponibiliza uma infra-estrutura eletrônica, apoio técnico e tecnológico para gerenciamento de revistas e, os editores parceiros, se responsabilizam pela inclusão e gestão dos conteúdos de suas revistas científicas. Cumprindo um de seus objetivos - analisar o comportamento da área a partir de distintas análises de sua produção - a equipe CEDUS desenvolveu esse estudo específico sobre o emprego de palavras-chave por distintos pesquisadores de artigos científicos.

## 2 Ciências da comunicação: uma área recente do conhecimento

O campo da Comunicação possui um perfil multifacetado, compreendendo conhecimentos teóricos e práticos, englobando várias profissões legalmente reconhecidas e socialmente legitimadas que atendem às demandas das indústrias e organizações midiáticas, dos serviços públicos e privados e das organizações da sociedade civil.

Pode-se dizer que é uma ciência que tem raízes na práxis, pois as profissões ligadas à comunicação já existiam quando, em 1935, na Universidade do Distrito Federal, surgiram os primeiros cursos de graduação. Com o Estado Novo, a Universidade foi fechada e a iniciativa frustrada. Em maio de 1943, o Decreto-Lei 5.480 finalmente criou o ensino de comunicação social em nível superior, mas atendeu aos primeiros alunos apenas em 1948 (MENDES,

1999.). O Programa de Pós-graduação é iniciado em 1991 e, atualmente, conta com 37 programas de mestrado e doutorado, distribuídos em todas as regiões do país (COMISSÃO..., 2007).

No que se refere à sua área de atuação, de acordo com o CNPq e demais agências de financiamento do país, as Ciências da Comunicação estão inscritas na grande área classificada como “Ciências Sociais Aplicadas”, subdividindo-se em outras subáreas tais como: comunicação visual, jornalismo e editoração, rádio e televisão, relações públicas e propaganda. Porém, tal classificação de áreas e subáreas não é ponto pacífico entre seus especialistas.

Lopes (2000) mapeia as vertentes da produção científica nas Ciências da Comunicação, procurando mostrar como as áreas foram se constituindo a partir dos estudos e dos interesses dos especialistas. Analisando distintos autores, faz um apanhado de como se constituiu seu campo epistemológico, a saber:

- a) Claude Shannon destacou que o problema fundamental da comunicação é o de se reproduzir de um ponto dado, a mensagem selecionada de um outro ponto. Assim, propôs o logaritmo como unidade de medida da informação transmitida. Este sistema de transmissão começa por uma fonte de informação que produz só uma, ou uma seqüência de mensagens a serem transmitidas no terminal de recepção. De modo geral, o esquema explicativo desta teoria reduziu-se a três elementos básicos: os emissores, os canais e os receptores. (SHANNON, 1964 *apud* LOPES, 2000). Para Lopes (2000) o problema desta teoria é o seu formalismo matemático e a sua consequente circularidade, diminuindo a importância do significado e do conteúdo da mensagem. Ainda assim, essa perspectiva teve forte penetração, conseqüência talvez, do desenvolvimento exponencial das tecnologias. Segundo Lopes, a forte presença das máquinas e programas inibiu a discussão dos conteúdos, resumindo-se o problema à máxima de McLuhan, no qual o meio é a própria mensagem. A cibercultura é o exemplo maior do impacto das tecnologias da informação na formação das consciências contemporâneas para o bem e, talvez, como afirmam os mais pessimistas, para o mal (LOPES, 2000)<sup>4</sup>;
- b) Philippe Breton localizou precisamente a formulação do conceito de comunicação no contexto da Segunda Grande Guerra e identificou a importância da evolução da informática para a formação dos mitos contemporâneos, interferindo na significação econômica, política e cultural dos fenômenos comunicacionais e informacionais recentes. (BRETON, 1994, 1991, 1997 *apud*

■  
<sup>4</sup> Documento eletrônico.

LOPES, 2000). Os estudos de Mauro Wolf (1995) sobre mass media pavimentaram o caminho das pesquisas na área, muito embora não examine os fundamentos do processo informacional e comunicacional. O próprio Lopes (2000) menciona que durante as duas guerras mundiais, particularmente na segunda, a propaganda de massa transformou a sociedade. O uso da mídia impressa, alto-falantes, rádio e cinema generalizou-se, tanto no âmbito estatal quanto no privado e se tornaram meios para orientar e informar, mas também para doutrinar, esconder e manipular. Estes recursos rapidamente mostraram seu potencial, talvez muito maior que a força das armas, pois podem modelar consciências, instaurar ou derrubar expectativas, instigar ou desmotivar pessoas (WOLF, 1995 *apud* LOPES, 2000)<sup>5</sup>;

■  
<sup>5</sup> Documento eletrônico.

c) Armand e Michèle Mattelart deram o avanço à interdisciplinaridade ao eleger como temas principais a teoria da informação, a indústria cultural, o contexto atual diante das tecnologias da informação e as relações destes com a política, a economia, a sociedade e a cultura. (MATTELART, 1995 *apud* LOPES, 2000). Dominique Wolton também apontou que os estudos de comunicação são um “objeto interdisciplinar” e, por isto mesmo, a principal ameaça às Ciências da Comunicação é à instrumentalização operacionalizada pela especialização. É por meio de uma visão pluridisciplinar que é possível mantê-lo como um saber aberto aos vários conhecimentos das áreas das ciências humanas. Além da extensa interdisciplinaridade, percebe-se ainda que as Ciências da Comunicação constituem-se como uma área bastante permeável aos novos paradigmas tecnológicos, implicando em profundas alterações nas concepções científicas e, dentre esses novos paradigmas, foram agregados os estudos das recentes autodenominadas ciências cognitivas, alterando as concepções de produção, transferência e estocagem de informações entre emissores e receptores (WOLTON, 1997 *apud* LOPES, 2000)<sup>6</sup>;

■  
<sup>6</sup> Documento eletrônico.

d) Régis Debray (1993) propôs a criação da “midiologia”, ou seja, tratar dos “meios simbólicos de transmissão e circulação” da informação, instituindo-a como disciplina (DEBRAY, 1993 *apud* LOPES, 2000). Retomou de McLuhan a perspectiva de que “o meio é a mensagem” (McLUHAN, 1993 *apud* LOPES, 2000);

e) Lucien Sfez, responsável pela crítica das teorias da comunicação existentes até a década de 1980, insistiu na importância da comunicação como área fundamental

do saber contemporâneo (SFEZ, 1990 *apud* LOPES, 2000)<sup>7</sup>.

■  
<sup>7</sup> Documento eletrônico.

Com base neste estudo apresentado por Lopes, verifica-se que a questão terminológica, usada para referenciar os estudos comunicacionais, é ampla e apresenta abrangente leque de significados semânticos e teóricos. Da mesma forma, Leydesdorff e Probst (s.d.), citando diversos autores, como Berger, 1991; Boure, 2006; Reeves e Borgman, 1983; Rice et al., 1988; So, 1988, evidenciam que

[...] importantes contribuições para o campo de estudos comunicacionais tem sido, historicamente, feito por pesquisadores de variadas disciplinas tais como ciência política, sociologia, psicologia e até mesmo a matemática. [...] Tais relacionamentos limítrofes, entretanto, permanecem difusos. Pesquisadores de estudos comunicacionais tendem a importar idéias de outros campos, muito mais do que exportar novas teorias e métodos para os demais campos" (LEYDESDORFF ; PROBST, 2009, p.3).

### 3 Procedimentos metodológicos

O *corpus* deste trabalho constitui-se das palavras-chave encontradas nas revistas brasileiras e portuguesas parceiras do Portal Univerciencia.org. Para seu desenvolvimento, no entanto, foi necessária a complementação dos metadados incluídos pelos editores nas versões eletrônicas de suas revistas científicas espelhadas no Portal Univerciencia.org.

Deste modo, procedeu-se à atualização manual e individual, fascículo a fascículo, dos metadados relativos às palavras-chave e aos resumos dos documentos, extraindo-os diretamente dos textos completos em PDF/HTML disponíveis na versão eletrônica das respectivas revistas. Durante tal processo, foi feita também a revisão dos metadados existentes no sistema, correção de erros de digitação, implementação de navegação nos distintos idiomas utilizados pelas revistas participantes (recurso disponível a partir da versão 2.2 do OJS)<sup>8</sup>, além da atualização da biografia e e-mail dos autores com link para o Currículo Lattes, quando existente.

Foram incluídos neste estudo 19 títulos de revistas, com coleções referentes aos anos de 1990 a 2008, contendo 162 fascículos, totalizando 2.284 documentos produzidos por 1.894 autores. Dentre tais documentos estão incluídos não somente artigos científicos, mas também entrevistas, resenhas e demais comunicações como mesas redondas em eventos, expediente e editorial. Seguramente todos os artigos científicos possuem palavras-chave, porém, em alguns casos, um ou outro tipo de documento também apresentava uma ou mais delas. Para este estudo, foram selecionadas todas as palavras-chave presentes no Portal, sem considerar se estavam em artigos ou qualquer outro tipo de documento.

■  
<sup>8</sup> O *Open Journal Systems* (OJS), traduzido no IBICT para Sistema Eletrônico de Editoração de Revistas (SEER), é um software de código aberto para publicação e gerenciamento de revistas, desenvolvido pelo *Public Knowledge Project* (<http://pkp.sfu.ca/ojs>), num trabalho conjunto entre a *Simon Fraser University*, *University of British Columbia* e a *Stanford University*. Disponível em <http://pkp.sfu.ca/ojs>, acessado em 19 abr. 2009.

Num primeiro levantamento, obteve-se o número inicial de 6.951 palavras-chave. Procedendo-se uma varredura para eliminar erros de sintaxe e espaçamentos extras de digitação, foram corrigidas 65 palavras. Isto representou uma redução de 0,9%, baixando a amostra deste estudo para o total de 6.886 palavras-chave.

A tabela 1 apresenta a distribuição do total de palavras-chave da amostra, segundo o total de documentos disponíveis por revista.

**Tabela 1** – Revistas com respectivos números de documentos, número de documentos que contém palavras-chave e o número de palavras-chave por revista.

Revista	Documentos	Documentos com palavras-chave	Número de palavras-chave
Comunicação & Educação	579	489	2.202
Revista Famecos	353	270	824
Galáxia	245	179	675
Comunicação, Mídia e Consumo	129	126	516
Estudos em Jornalismo e Mídia	158	120	467
Revista Em Questão	135	113	419
Libero	144	103	414
Revista Fronteiras – Estudos midiáticos	107	92	313
Media & Jornalismo	96	58	275
Comunicação & Sociedade	75	41	198
MATRIZES	34	30	126
Rastros	66	30	116
Verso e Reverso	28	24	93
Contracampo	25	18	73
Acervo de Mídia Regional	23	18	69
E-Com	17	12	37
Rebej – Revista Brasileira do Ensino do Jornalismo	12	8	37
Organicom	13	7	32
Observatório (OBS*)	43	0	0
Totais	2.282	1.738	6.886

Pode-se verificar a expressividade da revista **Comunicação & Educação** com relação ao número de palavras. Das 6.886 palavras extraídas, 2.202 (32%) pertencem a esta publicação. Por outro lado, a revista **Matrizes**, apesar de oferecer menos documentos, usou a quantidade máxima de cinco palavras-chave por artigo, o que impactou percentualmente a amostra, tanto quanto a revista **Comunicação e Educação**. Apesar do expressivo número de documentos apresentados pela revista **Comunicação**

e **Educação**, estes não impactam sobremaneira a amostra, pois sua média é de 3,5 palavras-chave. Ou seja, seu impacto dá-se, não pelo número de palavras-chave, mas, pelo número de artigos que a revista possui.

Observa-se que há uma média de 4 palavras-chave por revista, o que apenas comprova que as revistas, de maneira geral, seguem a norma da Associação Brasileira de Normas Técnicas – ABNT (NBR-6021) – que prescreve o emprego de três a cinco palavras-chave por documento. Apenas 166 artigos (9,6% do total de 1.713 documentos que continham palavras-chave) apresentaram mais do que cinco palavras-chave por artigo. Apesar de não ser o foco deste estudo, vale observar que o número de palavras-chave é um fator a ser considerado quando pensamos nas limitações materiais de uma revista. No mundo eletrônico, talvez esta questão não seja relevante e, deste modo, o número de palavras-chave poderia ser definido com base nas necessidades de cada artigo ou área do conhecimento.

A partir da referida preparação do *corpus* a compor a amostra deste estudo, e por meio do SQL (*Structured Query Language*), extraíram-se da base de dados do OJS as palavras-chave de todos os documentos publicados nas edições das revistas ativas do Portal Univerciencia.org. Foram selecionadas aquelas produzidas no idioma português, o que permitiu verificar a enorme adoção de termos estrangeiros para o vocabulário da área.

Tais palavras-chave foram exportadas para uma planilha eletrônica, passando novamente por um trabalho de normatização e conversão dos textos em minúsculas para efeitos de comparação.<sup>9</sup>

A partir disto, deu-se início ao processo de avaliação tendo como variáveis de estudo, a ocorrência e frequência no uso dos termos. Por um tempo pensou-se também em analisar se cada palavra atribuída como palavra-chave de fato cumpria a “[...] função essencial de representar e transmitir os conhecimentos especializados” (KRIEGER, 2001, p. 23). Para tanto, procurou-se relacionar as palavras-chave a uma das áreas e subáreas das Ciências da Comunicação, utilizando como referência a classificação da CAPES<sup>10</sup>. Esse percurso mostrou-se bastante árduo já que a própria característica interdisciplinar da área dificulta sobremaneira essa classificação, situação presente também nas

<sup>9</sup> As palavras encontradas não foram alteradas de acordo com a reforma ortográfica, pois esta entrou em vigor em 1 de janeiro de 2009 e os documentos foram produzidos antes desta data.

<sup>10</sup> Disponível em <http://www.capes.gov.br/avaliacao/tabela-de-areas-de-conhecimento>. Acesso em 18 abr. 2009.

**Tabela 2** – Distribuição das revistas do Portal Univerciencia.org por área de atuação

Áreas	Monotemática	Interdisciplinares	Total
Comunicação	9	5	14
Jornalismo	3	2	05
Totais			19



revistas analisadas, conforme a tabela abaixo:

Esse procedimento de mapeamento das palavras-chave frente às distintas áreas e subáreas das Ciências da Comunicação, embora não inócuo, foi abandonado dado o volume de trabalho, a necessidade de equipe interdisciplinar e o investimento maior de tempo. Portanto, para este estudo, optou-se por focar o reconhecimento das palavras-chave empregadas na representação dos documentos, com vistas a delinear o que seria o vocabulário utilizado nessa área do conhecimento.

Assim, como primeiro passo, definiu-se o mapeamento da ocorrência e da frequência no uso dos termos para, num segundo momento, identificar o perfil terminológico das distintas palavras atribuídas como palavras-chave. Procurou-se assim verificar a vigência dos conceitos mais utilizados na área das Ciências da Comunicação.

## 4 Análise dos dados

### 4.1 Análise quanto à ocorrência e frequência das palavras-chave

Para fins de análise adotamos o termo **ocorrência** para nos referirmos ao aparecimento das palavras (ou seja, consideramos sua unicidade) e **frequência** para contabilizar a quantidade de aparecimentos de um único termo.

Analisando-se a amostra composta por 6.886 palavras, identificou-se um total de 3.143 palavras distintas: algumas com frequência alta, como por exemplo, **comunicação** que aparece 174 vezes (frequência 174) entre as palavras-chave definidas para o total de 2.282 documentos, e outras, que aparecem uma única vez (frequência igual a um). Entre as 3.143 palavras distintas, há 2.370 que aparecem uma única vez, ou seja, 34,4% do conjunto de palavras distintas. Considerando o total de palavras - o universo da amostra estudada - aquelas com frequência igual a 1 correspondem a 75,4%.

Tal resultado é bastante expressivo, podendo evidenciar: pulverização de pesquisas, demasiada criatividade e liberdade na determinação de palavras-chave, dificuldade ou desconhecimento no processo de definição de palavras-chaves. De qualquer forma, seja qual for o motivo, tal fato tende a complicar e a dificultar uma melhor e mais acurada recuperação da informação gerando, portanto, menor visibilidade e acessibilidade à produção da área. Na análise das palavras atribuídas como palavras-chave apresentadas a seguir, no item 4.2, poder-se-á identificar algumas destas evidências mencionadas.

Dentre as palavras-chave empregadas com maior frequência, a **tabela 3** ilustra as que apareceram mais do que dez vezes.

**Tabela 3** – Palavras-chave empregadas com maior frequência.

Frq	Palavras	Frq	Palavras	Frq	Palavras	Frq	Palavras
174	comunicação	32	novas tecnologias	20	pesquisa	14	educação à distância
139	televisão	30	imaginário	20	pós-modernidade	14	epistemologia
132	jornalismo	30	leitura crítica	20	professor	14	imprensa
84	mídia	30	política	19	democracia	14	mídias
76	educação	28	cinema e educação	18	Brasil	14	modernidade
74	cultura	28	dramaturgia	18	cidade	14	propaganda
56	internet	27	comunicação/ educ	18	corpo	14	sentido
51	cinema	27	imagem	18	jornal	13	enunciação
46	história	27	jornal e educação	18	subjetividade	13	espaço público
45	telenovela	26	globalização	18	telejornalismo	12	documentário
43	recepção	26	memória	17	arte	12	estética
42	publicidade	25	ética	17	cotidiano	12	media
40	escola	24	informação	15	ciência	12	notícia
39	identidade	23	rádio	15	ideologia	12	rede globo
37	cibercultura	22	conhecimento	15	leitura	12	sociabilidade
36	discurso	22	linguagem	15	literatura	12	teatro
35	tecnologia	22	semiótica	15	narrativa	11	ensino fundamental
34	meios de comunicação	21	ensino	15	relações públicas	11	interação
32	cidadania	21	violência	15	representação	11	interatividade
32	consumo	20	fotografia	14	ciberespaço	11	jornalistas

Como mostra tal tabela, os termos **comunicação**, **televisão**, **jornalismo** e **mídia** (tanto no singular como no plural) encabeçam a lista de palavras mais frequentes. A palavra-chave **comunicação** isolada, embora caracterize o conteúdo do artigo, é tão ampla que se torna ineficaz para uma busca. É bom que se diga que um dos metadados do padrão OAI-PMH, formato Dublin Core, adotado pelo OJS, prevê a classificação dos documentos por área de conhecimento, desobrigando o autor a inseri-la como palavra-chave.

Também se observou a ocorrência de nomes de instituições, locais geográficos e palavras destituídas de sentido quando analisadas isoladamente. Em alguns casos, foi possível observar que o nome da instituição inserida como palavra-chave referia-se à afiliação do autor, não sendo objeto de estudo do referido documento. Portanto, evidencia-se um descompasso entre os metadados disponíveis no OJS (que prevê um campo específico para inserção do vínculo institucional dos autores) e a visão geral sobre o processo de recuperação da informação por parte de quem atribui as palavras-chave.

Enfim, considerando que as palavras-chave são parte do conjunto de metadados e, portanto, campos de busca e recuperação de informação em um ambiente controlado, é de suma

importância que seu preenchimento seja o mais correto possível. Muito embora este não seja o foco deste estudo – o qual analisa as palavras-chaves em seu contexto e não o preenchimento correto dos distintos campos e metadados dos documentos – vale reforçar a importância de tal procedimento, não somente para melhoria da recuperação do conteúdo nas respectivas revistas, como principalmente nas futuras bibliotecas digitais federadas (também conhecidas por provedores de serviço ou harvesters<sup>11</sup>), onde certamente tais revistas deverão ter interesse em participar com a melhor presença possível.

<sup>11</sup> Na área de Ciências da Comunicação, ver o Portal da Produção Científica em Ciências da Comunicação implementado pelo CEDUS – [www.univerciencia.org](http://www.univerciencia.org) – que, em dezembro de 2009, continha 20.191 documentos (artigos, trabalhos em eventos, teses, dissertações, livros etc.) publicados por 68 instituições de 17 países.

#### 4.2 Análise das palavras atribuídas como palavras-chave

Para uma análise das palavras-chave utilizadas, passou-se a buscar correlações com as áreas/subáreas das Ciências da Comunicação. Assim, obteve-se 337 palavras (quase 10%) que pertencem **não somente** ao escopo da comunicação, podendo apresentar-se destituídas de sentido quando isoladas, por exemplo: **pesquisa, produção, processos, projeto, teoria, rádio**. Outras palavras utilizadas foram mais complexas, tendo em vista a impossibilidade de se identificar a qual área/subárea pertenciam, como: **contato, tédio, cor, escuta, escolha, sensível**.

Tais resultados suscitaram a necessidade de uma leitura mais aprofundada do resumo ou do próprio documento. Sem este aprofundamento, por inúmeras vezes foi, efetivamente, impossível identificar a relação das palavras empregadas com o conteúdo do documento em si. Novamente, as questões de pesquisa ficaram evidentes: os autores desconhecem como atribuir palavras-chave? Este fato é decorrência da característica híbrida da produção científica na área das Ciências da Comunicação? Faltam instrumentos de referência como vocabulários ou tesouros?

Outra situação também ficou evidente: certas palavras-chave apresentaram uma frequência expressiva, no entanto, algumas delas não são adequadas para uma boa recuperação da informação, pois, são amplas demais, comprometendo a busca. Foram os casos de comunicação e mídia. Estes termos deveriam ser qualificados, como em: comunicação governamental, comunicação de massa, comunicação digital ou mídia impressa, mídia eletrônica, mídia televisiva.

Os termos **televisão e jornalismo**, embora frequentes, são convenientes, uma vez que se constituem como modalidades da comunicação de maneira geral, ou seja, não é a frequência dos termos que os torna mais adequados, mas, sua especificidade dentro da área.

Uma tentativa de mapeamento do universo de palavras aqui

utilizadas como palavras-chave, teve como resultado a identificação de distintas categorias possíveis de serem encontradas. Ou seja, na amostra estudada, observou-se a existência de palavras-chave representando as seguintes categorias:

- a) **nomes próprios:** 122 ocorrências envolvendo diferentes personagens e de diferentes áreas (autores, atores, pesquisadores, filósofos etc.), mas com baixa frequência (entre 1 e 3). Em alguns casos, tem-se o uso de abreviaturas como: **A. Pereira**. A menos que esta seja realmente a maneira como o autor é conhecido (supondo-se que o nome referenciado seja de alguém conhecido da área), para se garantir melhor recuperação, o ideal é a inclusão do nome sem abreviatura (no caso, Antonio Pereira);
- b) **nomes geográficos:** 39 ocorrências e no geral, baixa frequência (entre 1 e 3), **destacando-se apenas Brasil (18), América Latina (10) e Portugal (4);**<sup>12</sup>
- c) **jargões:** 9 ocorrências (fortemente ligados aos estrangeirismos, migrando em maior quantidade da língua inglesa, sendo que alguns já se encontram aportuguesados). Muitos deles ligados ao jornalismo, como por exemplo: **agenda setting, briefing, brand equity, clichê, lead, tablóide, fait divers, cauda longa** (do inglês, long tail) e **design**;
- d) **sinônimas:** 20 ocorrências, envolvendo 40 palavras. Alguns exemplos são: **auto-aprendizagem e auto-didatismo, auto-referência e auto-referencialidade, cultura midiática e cultura das mídias, ética do jornalismo e ética jornalística**, dentre outras. Há o caso das palavras **recepção e audiência**, freqüentemente usadas como sinônimos, mas que provocam certa ambiguidade, pois “**recepção**” também foi usada para mencionar aspectos cognitivos com relação à veiculação de certos conteúdos, resvalando para o campo da semiótica. Além deste caso, **dramaturgia e teledramaturgia**, embora não sendo sinônimos, foram assim usados;
- e) **singular e plural:** 70 ocorrências. Neste caso, evidencia-se a necessidade de verificá-las individualmente, pois algumas apresentam significados diferentes quando usadas no plural. É o exemplo de **linguagem e linguagens, audiência e audiências**;
- f) **neologismos:** 31 ocorrências, sempre associados aos **estrangeirismos**. É de se notar o número considerável de novos termos criados ou absorvidos da língua inglesa para o português. Destacamos ainda os neologismos em português **Dromocracia e docudrama**. Outro destaque

<sup>12</sup> É de se destacar que no formato eletrônico os termos de localização geográfica podem ser incluídos em outro campo e não necessariamente nas palavras-chave. O Dublin Core (padrão de metadados) prevê um campo especificamente criado para este fim: é o Coverage que, dentro do OJS, se subdivide em cobertura geográfica, histórica e cronológica, desobrigando o autor de atribuí-las nas palavras-chave.

refere-se ao uso do termo **web** (significa rede, em inglês), **wiki** (associado à rapidez, refere-se a softwares que permitem conhecimento colaborativo), **ciber** (também grafado cyber), **folk** (em inglês, povo), **info** e **bunker** (do alemão, significa abrigo) são freqüentemente usados como prefixos/sufixos e deram origem a vários conceitos (Quadro 1):

O mesmo acontece com o termo **log** que significa o registro constante dos eventos em um computador. Esse fato

Prefixo	Conceitos
ciber/cyber	cibercidade, cibercultura (s), ciberespaço, cibericonocracia, ciberjornal, cibernética, cibersubjetividade, ciborgue, cyberspace, cybersubcultura
folk	Folkcomunicação, folkmídia, folksonomia
web	web 2.0, weblog, web sites, webcomunicação, webjornalismo, webradio, websemântica
wiki	wikifilosofia
info	infoentretenimento, infoera, infográficos e infotenimento
bunker	bunkerização

Quadro 1: Prefixos estrangeiros que deram origem a novos conceitos

deu origem aos conceitos de **Blogs**, **Fotolog** e **Weblog**, recuperando a noção de algo que se faz diariamente;

- g) **estrangeirismos**: 30 ocorrências. Palavras que nomeiam **estilos**, **processos** ou **comportamentos** como é o caso de *Graphic Novels* (termo já consagrado para se referir a um tipo de Histórias em Quadrinhos); *data minning* (que se refere à mineração de dados/processo de busca de dados); *newsmaking* e *cultura remix* (referindo-se a mixagem de sons, usado para conceituar um tipo de cultura contemporânea); *funk*, *flanêur* (do francês, flunar), *otaku* (do japonês, refere-se à opção por viver uma vida virtual). E ainda *Storyspace*, *cabri-géomètre* e *ohmynews* (nomes de programas/software) e *Crash*, *Minority Report* e *Monty Python* (nomes de filmes), *Newsware* (nome de site). Há ainda a expressão *made in china* (definindo produtos e ao mesmo tempo, um qualificativo). Outras a serem citadas: *games*, *gatekeeper*, *on-line*, *readymade*, *reality shows*, *release e remake*, *short list e smart mobs*, *link*, *ombudsman*, *off*, *self e face/off*;
- h) **associações ou instituições**: 45 ocorrências, com frequência entre 1 e 2, destacando-se **Universidade de São Paulo (6)**. Nesta categorização, um dos problemas detectados foi o uso de **siglas (19 ocorrências)** sem que houvesse o nome por extenso. Como exemplo,

temos: CECAE, CELACOM, CEERT, CS, CTI, IAMCR, INTERCOM, IBERCOM, INCOR, MPB, MTV, PSDB, PT entre outras. Embora algumas siglas sejam amplamente reconhecidas como **GLBT** (gays, lésbicas, bissexuais e transgêneros), outras vão ainda se consolidando. É o caso da sigla **TICS** (Tecnologias da informação e Comunicação);

- i) **fatos jornalísticos**: 12 ocorrências. Como exemplo, podem ser citados: **acidente com air bus da Tam, menina afegã e World Trade Center**. Todos eles carregam sentido temporal;
- j) **verbos no infinitivo**: 2 ocorrências. Usados como palavras-chave, não representam conceito e ao mesmo tempo, é pouco provável que apareçam numa consulta. Foram encontrados os verbos **ouvir** e **conseguir**, o primeiro referia-se à percepção da comunicação pela audição e, o segundo, à obtenção do fato jornalístico.
- k) **novos conceitos**: foram encontrados os casos de **hibridismo metodológico, heterotopia** (criado por Foucault, e significa coexistência num espaço impossível) e **holo-fractal** (estilo de composição), **pictorialismo, semiosfera** e muitos outros que necessitariam aqui de definições a serem buscadas numa outra etapa do trabalho.

Outro fato que chamou a atenção foi o uso de **sinais gráficos** nas palavras-chave e que comprometem a compreensão, por exemplo, os dois pontos, as aspas e a barra inclinada. No primeiro caso, efetiva-se uma correlação entre termos. Quanto às aspas, estas conferem uma conotação diferenciada para o termo. Um exemplo são as palavras-chave “**teorias**” e “**tabloidização**”. No primeiro caso, o termo é ambíguo e, entre aspas, dá margem a outras interpretações; quanto ao segundo termo, o uso das aspas parece ter a intenção de mostrar que a palavra é um neologismo. Quanto às palavras **recepção/audiência**, separadas por barra inclinada, tem o mesmo efeito que separadas por hífen, embora a relação que se estabeleça seja de sinonímia. No caso dos **dois pontos**, como em **Editoração: ética**, estabelece-se uma relação de qualificar a primeira palavra. O mesmo parece acontecer em **perfil profissional: informação e perfil profissional: comunicação**.

Outro aspecto observado foi a **ambiguidade** de certas palavras, o que exigiu o retorno sistemático ao contexto de uso, a fim de verificar, no resumo ou no texto completo do artigo, o sentido que assumiram para o autor no momento de sua atribuição. Exemplo disso foi a palavra **Ellis**, referindo-se ao **Modelo de Ellis**. A supressão da palavra **Modelo** implica num menor alcance de significação por incompletude do sentido. Outro exemplo foi a

palavra **funções** (que apareceu por duas vezes). Resgatando o contexto de seu uso, verificou-se que uma das ocorrências se referia às **funções da linguagem** e, na segunda ocorrência, referia-se a **funções do observatório da imprensa**. O mesmo ocorreu ainda com a palavra **fragmentação** (que se referia à **fragmentação das relações sociais**) e **escuta** (referindo-se à maneira como as novas tecnologias interferiram na distribuição e no consumo da música).

## 5 Considerações finais

Quis-se nesta ainda incipiente iniciativa, evidenciar o potencial de análise das palavras chave utilizadas e publicadas nas revistas brasileiras e portuguesas da área de ciências da informação, em especial daquelas espelhadas no Portal Univerciencia.org. Também foi objeto deste estudo ressaltar a importância da adoção de princípios terminológicos na produção e na circulação da produção científica em Ciências da Comunicação, via periódicos on-line.

Embora se tenham claros os limites do trabalho quantitativo, este se constituiu como um passo importante no exame proposto, oferecendo interessantes e oportunas evidências no que se refere ao comportamento da área em estudo.

Numa primeira análise, é possível concluir que nem todas as palavras-chave atribuídas podem receber o estatuto de termo, pois nem sempre representam eficazmente o conteúdo dos artigos, fato que implica negativamente na recuperação da informação, resultando em prejuízo tanto para o usuário quanto para o produtor do artigo.

Um dos motivos parece resultar do processo de atribuição das palavras-chave que, no âmbito da materialidade da revista, ao serem separadas por vírgulas, mantém um vínculo de significação entre elas. No caso de uma busca digital, este vínculo se perde e, as palavras isoladas, não comportam mais o significado do momento em que foram atribuídas.

Outra peculiaridade identificada, por este estudo, foi a grande quantidade de termos que não carregam uma “marca” temática que demonstre a qual área do conhecimento pertence. Talvez isto se deva à própria natureza das Ciências da Comunicação, ou seja, o corpus coletado é oriundo de um domínio bastante híbrido, pois, uma das principais características desta área é a interdisciplinaridade, fruto dos frequentes flertes mantidos, por exemplo, com a sociologia, a filosofia, a educação, dentre outras.

Fica evidenciada a necessidade da construção de um vocabulário controlado na área, tanto para apoio e uso da própria comunidade, como para garantir melhor, mais eficiente e apropriada recuperação da informação produzida, garantindo seu acesso e



visibilidade. Diante dessa realidade caberia um reexame aprofundado das palavras-chave em uso nas diversas fontes de informação existentes, no sentido de trilhar o caminho da identificação de uma terminologia na área das Ciências da Comunicação.

### **Study on how to authors of articles in journals sciences communication verbalize their objects of study by keywords**

#### **ABSTRACT**

This study is the analysis of keywords found in the articles published in the Brazilian and Portuguese participating Portal Univerciência.org - Open Access Journals in Communication Sciences. We start with the occurrence and frequency of keywords used by the authors and editors with a view to map features and forms of expression relating to the objects of their studies. The results allowed a better understanding of the area and the difficulties in building criteria to guide the insertion of keywords to present an interdisciplinary context of hybrid vocabularies and far between.

**KEYWORDS:** Assigning keywords. Communication sciences. Knowledge representation.

### **Estudio sobre como los autores de artículos en revistas de las Ciencias de la Comunicación verbalizan sus objetos de estudio al definir las palabras clave**

#### **RESUMEN**

Este estudio haz un análisis de las palabras clave encontradas en los artículos de las revistas brasileñas y portuguesas que hacen parte del Portal Univerciencia – Revistas de Acceso Abierto en Ciencias de la Comunicación. Empieza con la ocurrencia y la frecuencia de las palabras clave empleadas por los respectivos autores y editores con el intento de mapear características y maneras de expresar sus objetos de estudio. Los resultados han permitido una mejor comprensión del área y de las dificultades en la construcción de criterios para orientar a la inserción de palabras clave, por tratarse de un contexto interdisciplinario entre vocabularios híbridos y distantes entre sí.

**PALABRAS CLAVE:** Asignación de palabras clave. Ciencias de la Comunicación. Representación del conocimiento.

#### **Referências**

COMISSÃO DE APERFEIÇOAMENTO DE PESSOAL DE ENSINO SUPERIO. **Relação de cursos recomendados e reconhecidos em Ciências Sociais Aplicadas, área de Comunicação.** 2007. Disponível em: <http://conteudoweb.capes.gov.br/conteudoweb/ProjetoRelacaoCursosServlet?acao=pesquisarLes&codigoArea=60900008&descricaoArea=CI%CANCIAS+SOCIAIS+APLICADAS+&descricaoAreaConhecimento=COMUNICA%C7%C3O&descricaoAreaAvaliacao=CI%CANCIAS+SOCIAIS+APLICADAS+I>. Acesso em: 18 abr. 2009.

KRIEGER, Maria da Graça; Maciel, Anna Maria B. (Org).



**Temas de terminologia.** Porto Alegre: Ed. Universidade / UFRGS, 2001.

LACERDA, J.; VALENTE, P. A Emergência em sistemas baseados em folksonomias. **Estudos em Jornalismo e Mídia**, Florianópolis, v. 4, n. 2, 2007. Disponível em <<http://www.revistas.univerciencia.org/estudos/article/view/5826/5297>>. Acesso em: 18 abr. 2009.

LEYDESDORFF, Loet ; PROBST, Carole. The Delineation of an interdisciplinary specialty in terms of a Journal Set: the case of Communication Studies. **Journal of the American Society for Information Science & Technology** (forthcoming), New York, v.60, n.8, p. 1709-1718, 2009. Disponível em: <<http://arxiv.org/ftp/arxiv/papers/0911/0911.1053.pdf>>. Acesso em: 19 dez. 2009.

LOPES, Luis Carlos. Ciências da Comunicação, possibilidades e problemas. **Revista Ciberlegenda**, Rio de Janeiro, v.3, 2000 Disponível em: <<http://www.uff.br/mestcii/lclop3.htm>>. Acesso em: 13 abr. 2009.

MENDES, Ricardo Fontes. **A Profissionalização do jornalismo no Brasil**. 1999. Disponível em: <http://www.saladeprensa.org/art40.htm>. Acesso em: 19 abr. 2009.

PORTAL UNIVERCIENCIA. **Sobre o Portal**. Disponível em: <<http://www.univerciencia.org>>. Acesso em: 13 abr. 2009.

#### **Maria Helena Morais**

*Bolsista de iniciação científica no Centro de Estudos em Design de Sistemas Virtuais Centrado no Usuário na Escola de Comunicações e Artes da Universidade de São Paulo (CEDUS/ ECA/USP).  
E-mail: mhjmorais@gmail.com*

#### **Sueli Mara S.P. Ferreira**

*Professora Livre Docente do Departamento de Biblioteconomia e CEDUS/ECA/USP.  
E-mail: smferrei@usp.br*

#### **Marcos Mucheroni**

*Professor Doutor do Departamento de Biblioteconomia e Documentação e Pesquisador no Centro de Estudos em Design de Sistemas Virtuais Centrado no Usuário na Escola de Comunicações e Artes da Universidade de São Paulo (CEDUS/ ECA/USP).  
E-mail: mmucheroni@hotmail.com*

#### **Josefina Perez**

*Analista de sistemas do Centro de Estudos em Design de Sistemas Virtuais Centrado no Usuário na Escola de Comunicações e Artes da Universidade de São Paulo (CEDUS/ ECA/USP).  
E-mail: josipkp@gmail.com*