



## **Projeto SAPO: um barômetro de ciência e tecnologia na mídia<sup>1</sup>**

Sabine Righetti, Carlos Vogt, Yuriy Castelfranchi, Rafael Evangelista e Iuriatan Muniz.

Laboratório de Estudos Avançados em Jornalismo (Labjor), da Universidade Estadual de Campinas (UNICAMP). Fundação de Amparo à Pesquisa no Estado de São Paulo (FAPESP).

### **Resumo**

A proposta é apresentar o projeto SAPO (*Science Automatic Press Observer*), que tem o objetivo de identificar a presença de ciência na mídia impressa brasileira por meio de tecnologia. O SAPO consta de um *software*, desenvolvido pela equipe de pesquisadores, que captura e armazena, em um banco de dados, matérias jornalísticas de conteúdo científico. A coleta tem base em um filtro que faz a seleção por palavras-chave de ciência; cada palavra-chave recebe uma pontuação específica e a soma da pontuação determina se a matéria é ou não de conteúdo científico. Na sua fase inicial, o SAPO abrange os dois maiores jornais do país, *Folha de S.Paulo* e *O Estado de S.Paulo*, e já tem apresentado resultados significativos. O objetivo para as próximas fases é ampliar o número de jornais atingidos e expandir o SAPO para outros idiomas.

### **Palavras-chave**

Divulgação científica; Jornalismo Científico; Indicadores de ciência, tecnologia e inovação.

### **Corpo do trabalho**

#### **1.Introdução**

*O princípio basilar da ciência acadêmica é que os resultados da pesquisa devem ser públicos.*

[...] *A instituição fundamental da ciência, então, é o sistema de comunicação* (Ziman, 1987)

Não há ciência sem sua divulgação e comunicação ao público. A comunicação pública da ciência desempenha um papel central nas sociedades contemporâneas, não somente no sentido, enfatizado e bem estudado, da importância do conhecimento científico para a formação dos cidadãos e para a gestão das democracias, mas também por uma necessidade da própria ciência.

Se carreira, patrocínio, proteção aos cientistas sempre foram ligadas, pelo menos em parte, às suas capacidades de comunicar, é verdade que antigamente, pelo menos até as

---

<sup>1</sup> Trabalho apresentado na Sessão Jornalismo e Editoração, da Altercom – Jornada de Inovações Midiáticas e Alternativas Experimentais, evento componente do XXXI Congresso Brasileiro de Ciências da Comunicação.



primeiras décadas do século XX, grande parte das decisões relevantes para a ciência eram tomadas, de forma substancial, no interior das comunidades científicas.

As interações com o resto da sociedade, embora importantes para garantir recursos ou para interagir com o mundo da indústria, eram laterais, posteriores em certa medida à atividade científica: os cientistas, embora às vezes com pouco recursos, podiam viver na “torre de marfim” (Greco, 2002). Hoje, algumas decisões relevantes para a vida profissional e para o trabalho dos cientistas são tomadas com a participação de diferentes sujeitos, nem todos cientistas ou especialistas: são políticos, burocratas, empresários, militares, religiosos, movimentos sociais, consumidores e associações de pacientes que pedem, e frequentemente obtêm, o direito e a legitimidade para participar de decisões relevantes para o desenvolvimento da ciência.

Esse processo, bastante complexo e ainda pouco estudado, parece ter começado após as guerras mundiais, na Europa, e se aprofundou e se difundiu a partir da década de 1990. Os cientistas, às vezes com relutância, estão cada vez mais precisando aprender como sair de seus laboratórios e de suas universidades para se comunicarem com grupos sociais variados. Nesse contexto, toma forma e ganha corpo a discussão sobre a importância da comunicação científica.

## **2. A ciência na mídia**

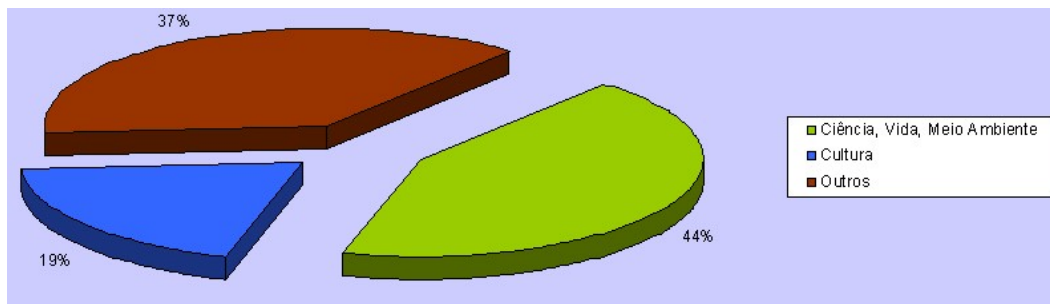
Quando se adentra na área de estudos sobre a ciência na mídia é preciso, primeiramente, delimitar o campo de estudos. Entende-se, nesse trabalho, ciência como um conceito mais amplo, uma “cultura científica”, ou seja, um grande ecossistema de símbolos, idéias, histórias, fatos, noções, que circulam e agitam a sociedade e tem, portanto, um fortíssimo reflexo midiático (Vogt & Polino, 2003).

Isso significa que a ciência está presente não somente nas matérias jornalísticas das editoriais de ciência, mas, por exemplo, nas cartas de leitores que, não divulgando nem explicando nada sobre transgênicos, expressam suas opiniões, valores, crenças a respeito de um tema importante em que a ciência entra na vida política e social do país.

A ciência está também nas páginas opinativas do jornal, em colunas econômicas ou nos artigos políticos, sobre decisões estratégicas no setor de inovação tecnológica ou política energética, por exemplo, o que pode ser notado na Figura 1 e na Figura 2, respectivamente, sobre a distribuição da Ciência, Tecnologia e Inovação (C,T&I) nos

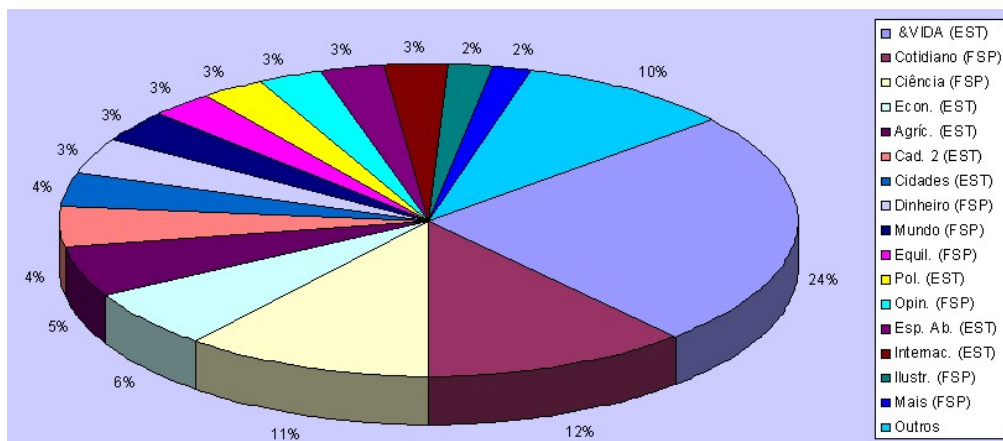
dois principais jornais diários do país, em termos de circulação, (*Folha de S.Paulo* e *Estado de S.Paulo*), em 2006.

**Figura 1 - Distribuição da C,T&I na mídia – grandes áreas**



Fonte: Projeto SAPO (Scientific Automatic Press Observser) – Labjor/UNICAMP

**Figura 2 - Distribuição da C,T&I na mídia– áreas específicas**



Fonte: Projeto SAPO (Scientific Automatic Press Observser) – Labjor/UNICAMP

Nota-se que os meios de comunicação de massa já não têm medo de tratar da atualidade das ciências e das tecnologias e recorrem a essas para esclarecer a atualidade em geral. A investigação e o desenvolvimento das ciências e das tecnologias nunca teve tão grande influência no nosso modo de vida e de trabalho, nas nossas concepções de espaço e tempo, nas nossas capacidades de intercâmbio e de comunicação em todo o planeta.

Assim, como nunca aconteceu em toda a história, fala-se em comunicação científica e tecnológica e governos nacionais ou regionais têm apoiado a criação e as atividades no campo da cultura científica e tecnológica.

As pessoas, cada vez mais, têm se mostrado interessadas por temas de ciência e tecnologia. E no Brasil não é diferente: em recente pesquisa de Percepção Pública da



Ciência e Tecnologia<sup>2</sup>, divulgada em 2007 e realizada em 2006, com mais de 2 mil pessoas de todo país, pelo Ministério da Ciência e Tecnologia (MCT), 41% dos brasileiros demonstraram “muito interesse” por temas ligados à ciência e à tecnologia têm da preferência. O número chegou perto do tema “esportes” (47% dos entrevistados) e ganhou de assuntos como “política” (20%).

Na medida em que a ciência ganha espaço nos meios de comunicação, emerge a necessidade da construção de indicadores que estejam voltados para a produção científica e tecnológica e que meçam, de alguma maneira, os impactos dessa produção nas dimensões sociais, como, por exemplo, os indicadores de inovação tecnológica e os indicadores de percepção pública da ciência.

A inclusão de indicadores de percepção pública da ciência e da tecnologia entre outros indicadores de C,T&I (como os indicadores de produção científica, de recursos humanos trabalhando com pesquisa, de dispêndios com pesquisa, entre outros) faz-se necessária em vista de oferecer aos tomadores de decisão para o desenvolvimento científico e tecnológico, dentro da aceitação pública, informações e indícios de como as ações devem ser encaminhadas.

Dentro do escopo da construção de indicadores de percepção pública de ciência e tecnologia, novas metodologias têm surgido e entrado na discussão acadêmica. Por exemplo, as pesquisas de opinião, dirigidas ao levantamento de tendências de comportamento político e social, e têm se tornado um importante veículo para a obtenção de informações sobre atitudes relacionadas ao meio ambiente, consumo de informações científicas, conhecimentos de descobertas científicas e tecnológicas e opiniões sobre seu impacto na vida cotidiana (Figueiredo *et al.*, 2005).

Uma recente pesquisa de opinião sobre os alimentos transgênicos, realizada no Brasil pelo *Greenpeace*, em parceria com o Instituto Brasileiro de Opinião Pública e Estatística (IBOPE), em 2003, constatou, em uma amostra nacional, um alto grau de conhecimento da população sobre o tema, e opiniões claras sobre medidas de controle<sup>3</sup>. Os resultados confirmam a idéia de que a presença das questões públicas e também as científicas nos *meios de comunicação* ampliam o acesso às informações relativas a essas

---

<sup>2</sup> As pesquisas de percepção pública da ciência e da tecnologia têm tradição de décadas nos países desenvolvidos e começaram a ser desenvolvidas na região da Ibero-América nas proximidades da década de 1990. O objetivo dessas pesquisas é avaliar o conhecimento, interesse e a apropriação da sociedade em relação ao conhecimento científico (ou, como alguns autores definem, a “alfabetização científica”) para guiar tomadas de decisão ou políticas públicas na área.

<sup>3</sup> Nesta pesquisa, 63% dos entrevistados já haviam tido alguma informação sobre produtos transgênicos, 74% escolheriam produtos não transgênicos se tivessem que escolher, 92% consideraram que as informações sobre componentes transgênicos deveriam vir no rótulo dos alimentos, e 73% responderam considerar que esses produtos deveriam ser proibidos. Relatório Ibope, *Pesquisa sobre transgênicos*, dezembro de 2003.



questões, evidenciando-se o seu impacto na vida cotidiana da população (Figueiredo *et al.*, 2005).

Nesse contexto, o papel da mídia nas pesquisas de percepção pública da ciência e da tecnologia torna-se explícito e as metodologias que relacionam exposição à mídia, participação midiática de temas de C,T&I e a percepção pública da C,T&I se fortalecem. Aqui se define o campo de estudos no qual o projeto SAPO está inserido, o que será abordado na próxima seção.

### 3. O projeto SAPO

O projeto SAPO (*Science Automatic Press Observer*) tem o objetivo de identificar a presença de ciência na mídia impressa brasileira por meio de tecnologia. O SAPO consta de um *software*<sup>4</sup> que funciona como sistema de coleta, seleção, organização e mensuração da presença e do impacto da ciência, tecnologia e inovação (C,T&I) na mídia. Trata-se de um banco de dados integrado com indicadores quantitativos, medidos automaticamente, da presença e do impacto das questões ligadas a CT&I em jornais diários impressos<sup>5</sup>.

Pelo sistema, há um grupo de palavras-chave relacionadas a assuntos científicos, cada uma com uma determinada pontuação. O conjunto dessas palavras-chave, no corpo do texto jornalístico, resulta em uma pontuação que define se o texto *é* de conteúdo científico, *pode ser* de conteúdo científico (e, nesse caso, o texto vai para uma base de verificação humana) ou *não é* de conteúdo científico.

O sistema, basicamente, se constitui de:

1. Um conjunto de algoritmos especificamente preparados para a extração e cópia do texto integral (disponibilizado em sua versão eletrônica) de alguns dos jornais brasileiros de maior tiragem e difusão.
2. Um sistema de classificação e filtragem de matérias, capaz de selecionar de forma automática aqueles textos que tratem de temas relacionados às ciências; políticas científico-tecnológicas e para inovação; biomedicina e meio ambiente. Veremos a seguir qual é a eficiência desse sistema de seleção e quais critérios guiam a escolha das matérias a selecionar.

---

<sup>4</sup> O software foi desenvolvido pela empresa SOLIS, com apoio da FAPESP, sob a demanda da equipe de pesquisadores do LABJOR/UNICAMP que, inclusive, criaram, a partir de estudos, o grupo de palavras-chave de ciência.

<sup>5</sup> A opção pela prioridade aos textos publicados na mídia diária impressa (disponibilizados *online*) é comum em estudos clássicos da imprensa. Atualmente, o SAPO trabalha com os dois principais jornais diários do Brasil em termos de circulação, *Folha de S.Paulo* e *O Estado de S.Paulo*, de acordo com o MídiaDados (2007).



3. Um banco de dados estruturado e um buscador inteligente que, organizam e gerenciam, a partir de vários meta-dados (veículo, caderno, data, fonte, autor, gênero jornalístico etc.), matérias selecionadas como sendo “de ciência”.
4. Um conjunto de equações capazes de gerar indicadores sobre a frequência, relevância e outras características do material coletado e armazenado.

Para verificar a correspondência entre o material coletado pelo sistema e o que é publicado nos jornais, comparamos matérias contidas no banco de dados em três “semanas construídas” (os sete dias da semana, porém escolhidos casualmente num intervalo de seis meses de disponibilidade do jornal) para os jornais coletados, com o material disponível também na versão impressa dos jornais. Após ajustamentos iniciais nos veículos de coleta (alguns jornais mudaram sua produção, introduziram ou eliminaram cadernos etc), conseguimos coletar todos os cadernos dos jornais com um excelente nível de correspondência entre o material no banco de dados e o que sai nas bancas.

O material coletado, de conteúdo científico (selecionado) ou de conteúdo não científico (descartado), pode ser consultado de maneira fácil e ágil. O banco de dados reúne o texto jornalístico na íntegra, data de publicação da matéria, nome do autor e ainda destaca as palavras-chaves de conteúdo científico. Assim, o usuário do SAPO pode realizar pesquisas no banco de dados do SAPO utilizando um termo específico para a busca (por exemplo, a busca pela expressão “mal de chagas” ou “células-tronco”) e em recortes temporais específicos.

É importante ressaltar, porém, que o sistema possui algumas limitações:

- a) O sistema coleta somente *textos*: imagens, charges, tiras, publicidades etc não fazem parte do *corpus* coletado;
- b) Há algumas diferenças entre a versão impressa do jornal que um assinante recebe em sua casa e a versão que o mesmo assinante pode acessar no site do jornal como sendo a versão “impressa”. Por exemplo, a versão “impressa” disponível *online*, conterà, em alguns casos, mais cadernos que qualquer versão física comprada nas bancas, pois esta última só tem o caderno local correspondente à cidade de residência (do Rio de Janeiro, Niterói, Campinas, São Paulo etc.), enquanto a versão *online* pode ter todos;
- c) Há outras pequenas, porém interessantes, diferenças entre versão impressa e sua “cópia” *online*. Algumas matérias da versão impressa disponíveis no site são extremamente parecidas com o que saiu realmente no papel, porém não idênticas:



o título e o primeiro parágrafo podem ser ligeiramente diferentes ou, às vezes, o tamanho ligeiramente ajustado. Isso, provavelmente, deve-se a modificações de última hora ocorridas durante a diagramação e edição final do jornal, devido a exigências de publicidade ou de notícias de última hora, e que foram executadas de forma diferente em diferentes versões impressas do jornal.

Feitas tais ressalvas, podemos afirmar que o sistema de coleta é extremamente confiável e que, no SAPO, a tecnologia representa um instrumento importante, tanto para pesquisas acadêmicas quanto como ponto de partida para o desenvolvimento de produtos e aplicações tecnológicas destinadas a entidades públicas e empresas.

O SAPO pode oferecer à comunidade de pesquisadores, de variadas áreas, informações diárias sobre a cobertura da mídia em CT&I, permitindo enxergar não só quanto este tema freqüenta o leitor, mas também como o leitor o freqüenta nas páginas dos jornais (Vogt *et al.*, 2005).

#### **4. Próximas etapas e considerações**

Como abordado neste trabalho, a necessária discussão sobre os impactos da ciência, da tecnologia e da inovação nos diversos campos da atividade humana e a forma na qual a C,T&I passa a ser parte viva da nossa cultura deve ser baseada em instrumentos de análise adequados e para permitir uma análise mais profunda. Nesse sentido, a proposta do SAPO surge de uma perspectiva inovadora, em que a tecnologia é usada para realizar um trabalho quantitativo, até então realizado de maneira nada tecnológica.

O índice criado pelo SAPO não pretende ser um reflexo exato sobre o quanto de ciência é divulgado na mídia. Trata-se de um recorte específico, a partir de critérios discutíveis, porém fixos, que, a partir do momento em que são adotados e mensurados historicamente, adquirem sentido e passam a indicar flutuações e processos interessantes. Mas, mesmo assim, o SAPO já se mostra um instrumento bastante valioso para a reflexão e pesquisa sobre o quanto e como ciência, tecnologia e inovação aparecem na mídia impressa, e conta com um desenvolvimento bastante maduro principalmente no que se refere ao armazenamento e qualificação das matérias inseridas.

Nas próximas etapas do trabalho, dois caminhos principais podem ser apontados. O primeiro, e mais evidente, é a integração de novos veículos ao sistema, para que a amostra de pesquisa seja mais representativa e diversificada. O segundo é o uso de



inteligência artificial para a seleção das matérias consideradas “de ciência”, aprimorando, assim, o sistema já desenvolvido.

Além disso, estuda-se a possibilidade de se “abrir” o sistema, por meio de um site, para consulta de pesquisadores que tenham interesse em verificar a presença de ciência ou de parte específica dela na mídia.<sup>6</sup>

Assim, por meio da tecnologia, é possível gerar um conjunto muito interessante de informações estatísticas capazes de identificar “cronicidades” e “epidemias” relativas à frequência do tema CT&I na mídia, o que pode dar fôlego para novos estudos na comunicação, ou em outras áreas e, quem sabe, pode pautar políticas públicas na questão da promoção e do incentivo à divulgação científica.

### Referências bibliográficas

BAUER, M.W.; GASKELL, G. **Pesquisa qualitativa com texto, imagem e som. Um manual prático.** Petrópolis, RJ: Vozes, 2002.

CASTELFRANCHI, Y. Scientists to the streets: Science, politics and the public moving towards new osmoses. *Jcom*, Vol. 1, n. 2, Trieste: Junho 2002.

FIGUEIREDO, S. P.; VOGT, Carlos Alberto; KNOBEL, Marcelo. **Percepção pública da C&T: um importante instrumento de apoio a políticas públicas.** In: XI Seminario Ibero-Americano de Gestión Tecnológica - Innovación Tecnológica, cooperación y desarrollo, 2005, Salvador - BA. Anais. São Paulo: PGT-USP. v. 1.

GRECO P. Communicating in the post-academic era of science. *Jcom*, Vol. 1, n. 1, Trieste, Março 2002. Disponível em: <<http://jcom.sissa.it/editorial/edit0101.pdf>>.

GRECO, P. “Il modello Venezia. La comunicazione nell’era post-accademica della scienza”, em: **La comunicazione della scienza, Atti del I and II Convegno Nazionale**, Roma: ZadigRoma, 2004.

MídiaDados 2007. Relatório anual publicado pelo Grupo de Mídia.

MCT e CNPq/Ibope. *O que o brasileiro pensa da ecologia?*, Relatório de pesquisa, Brasília, 1992.

VOGT, C.; MARQUES DE MELO, J.; et al. “S&T na mídia impressa brasileira: tendências evidenciadas na cobertura nacional dos jornais diários sobre ciência & tecnologia (biênio

---

<sup>6</sup> Atualmente, alguns pesquisadores já fazem uso do *software* do SAPO para suas respectivas pesquisas. Um estudante de Sistemas de informação e biblioteconomia, da Universidade Federal de São Carlos (Ufscar), tem participado do desenvolvimento do SAPO na sua iniciação científica. Outra pesquisadora, uma doutoranda de medicina na Universidade Federal de São Paulo (UNIVESP), utiliza o SAPO para dar embasamento à sua discussão sobre a presença de algumas doenças específicas na mídia. O SAPO também está começando a ser usado por um grupo de pesquisadores do CinAPce (Cooperação Interinstitucional de Apoio a Pesquisa sobre o Cérebro) que visam verificar a presença de doenças neurológicas, em especial a epilepsia, na mídia.



2000-2001)”. Em: Guimarães, AND. (Org). **Produção and Circulação do Conhecimento: Política, Ciência, Divulgação**. Campinas: Pontes Editores, 2001.

VOGT, C.A.; POLINO, C. (Orgs.) **Percepção pública da ciência: Resultados da pesquisa na Argentina, Brasil, Espanha and Uruguai**. Campinas, SP: Editora da UNICAMP; São Paulo: FAPESP, 2003.

VOGT, C.; KNOBEL, M.; EVANGELISTA, R.; PALLONE, S.; CASTELFRANCHI, Y. “Percepção Pública da Ciência and Tecnologia: uma Abordagem Metodológica para São Paulo”. Em: Landi, F. (Org.). **Indicadores de Ciência, Tecnologia and Inovação do Estado de São Paulo - 2004**. São Paulo, FAPESP: 2005, cap. 12.

ZIMAN, J. **Il lavoro dello scienziato**, Roma-Bari: Laterza, 1987 (Ed. original: An introduction to science studies; The philosophical and social aspects of science and technology. Cambridge: Cambr. Univ. Press, 1984).

\_\_\_\_\_. **Real Science: What it is, and what it means**. Cambridge: Cambr. Univ. Press, 2000.