
Aplicação de sistema de software para auxílio na análise de conteúdo

*Henrique M.R. de Freitas
Marcus V.M. da Cunha Júnior
Jean Moscarola*

As análises de documentos seguindo métodos clássicos, como o jurídico e o sociológico, apresentam um caráter racional em comum, mais ou menos intuitivo, pessoal e subjetivo. O método histórico tem problemas de validade, como o da autenticidade do texto, o da validade da interpretação e o da dependência da veracidade dos fatos. Todas essas análises, entre outras, apresentam os **senões** de um trabalho não-sistematizado, dependendo unicamente do valor e da competência do operador (pesquisador).

Com a multiplicação dos meios de comunicação (televisão, rádio etc.) e das propagandas tornou-se mais perceptível a necessidade de buscar significados, relações ou mesmo leis, de medir fatores de influência e, ainda, de prever certos eventos. Dessa necessidade de realizar pesquisas em novas direções nasceu um método moderno de análise de documentos: a **análise de conteúdo**.

As noções gerais expostas neste texto são baseadas principalmente nos trabalhos de Grawitz (1976) e de Gavard-Perret & Moscarola (1995). Primeiramente são apresentadas noções gerais sobre a análise de conteúdo, dando-se ênfase à sua aplicação e ao seu uso. Na seqüência são abordadas as etapas técnicas da análise de conteúdo, destacando-se o seu valor como instrumento de pesquisa. Sete níveis de análise lexical são apresentados de forma criativa e evolutiva, pressupondo-se o uso de ferramentas ou sistemas em microcomputador. No tópico seguinte, inspirando-se em dados qualitativos coletados em pesquisa, busca-se ilustrar algumas das técnicas de análise aqui enfatizadas. Por último, aborda-se a questão da utilidade e do interesse da análise de conteúdo.

NOÇÕES GERAIS SOBRE A ANÁLISE DE CONTEÚDO

Neste tópico é apresentada a definição de análise de conteúdo, bem como as suas qualidades, as suas fontes, os tipos de análise e as tendências.

Definição de análise de conteúdo

Os recursos que as ciências sociais oferecem para reflexão são essencialmente compostos por comunicações, sejam elas orais (textos de

Este estudo baseia-se em discussões e debates entre os autores, no âmbito de parceria de pesquisa do Programa de Pós-Graduação em Administração (PPGA) da Universidade Federal do Rio Grande do Sul (UFRGS) com a *Université de Savoie*, com o apoio do CNPq e da Fapergs. O sistema de pesquisa e análise de dados Sphinx Plus, de autoria do Professor Jean Moscarola e que vem sendo aperfeiçoado com o apoio da equipe de pesquisa do Grupo de Estudos em Sistemas de Informações e de Apoio à Decisão (GESID) do PPGA/UFRGS, serviu de suporte à realização da análise de conteúdo neste estudo. Os autores agradecem à colaboração da equipe de pesquisa do GESID e de Ângela Werner, bolsista IC CNPq do Grupo.

Recebida em fevereiro/97

Henrique M.R. de Freitas, Doutor em Gestão pela *Université Pierre Mendès France* (Grenoble, França), é Professor-Adjunto do Programa de Pós-Graduação em Administração (PPGA) da Universidade Federal do Rio Grande do Sul (UFRGS), membro do Grupo de Estudos em Sistemas de Informações e de Apoio à Decisão (GESID) do PPGA/UFRGS e Pesquisador do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).
E-mail: hfreitas@vortex.ufrgs.br

Marcus V.M. da Cunha Júnior é Mestrando em Administração do Programa de Pós-Graduação em Administração (PPGA) da Universidade Federal do Rio Grande do Sul (UFRGS).
E-mail: mvmcunha@portoweb.com.br

Jean Moscarola, Doutor em Gestão pela *Université de Paris*, é *Directeur* de l'IUP, *Université de Savoie*, Annecy (França) e autor do Sistema Sphinx.

discursos, anotações e relatórios de entrevistas, conversações) ou escritas (textos oficiais, artigos de jornais, cartas — tanto na área literária como na histórica ou na política). Qualquer que seja o nível que se deseja atingir e o objeto das pesquisas, os dados a reunir para compreender e explicar opiniões, condutas, ações, enfim, são quase sempre de origem **verbal**. A ação, quando apreendida, apresenta-se em um contexto de **palavras**: são sempre encontradas falas ou alocações e escritos ou textos, ou seja, documentações em palavras.

É importante que o pesquisador em ciências sociais possa analisar, de forma científica, esses materiais ou fontes e não se contente em ter uma simples impressão. A novidade do processo moderno da análise de conteúdo consiste em substituir o **impressionismo** (jogo de opinião e muito pessoal) por procedimentos mais padronizados, tendendo à quantificação, convertendo materiais **brutos** em dados passíveis de tratamento científico. Para tal, o texto será decomposto, ou seja, será estudado em função das idéias ou das palavras que contém; estas últimas escolhidas em razão de sua relação com o objetivo da pesquisa ou a questão investigada (por exemplo: Quantas vezes tal partido utilizou o termo **justiça social** durante a campanha eleitoral?).

Berelson (*apud* Grawitz, 1976:588) define a **análise de conteúdo** como “uma técnica de pesquisa para a descrição objetiva, sistemática e quantitativa do conteúdo manifesto das comunicações, tendo por objetivo interpretá-las”. De uma forma geral, Grawitz complementa que não se deve mais considerar indispensáveis os termos manifesto e quantitativa.

Qualidades: objetiva, sistemática e quantitativa

- **Objetiva** — deve-se proceder à análise segundo as regras preestabelecidas, obedecendo a diretrizes suficientemente claras e precisas de forma a propiciar que diferentes analistas, trabalhando sobre o mesmo conteúdo, obtenham os mesmos resultados (o que pressupõe terem eles concordado sobre os aspectos a analisar, as categorias a usar e a definição operacional de cada categoria).
- **Sistemática** — todo o conteúdo deve ser ordenado e integrado nas categorias escolhidas, em função do objetivo perseguido. Elementos de informação associados ou relativos ao objetivo não devem ser deixados de lado.
- **Quantitativa** — trata-se de poder evidenciar os elementos significativos, calcular a sua freqüência etc. Esta condição não é indispensável, visto que certas análises de cunho qualitativo buscam mais os temas do que a

sua exata medida ou importância. Não são muitas as ferramentas que colocam à disposição recursos de análise de conteúdo, associando o qualitativo ao quantitativo. No caso específico deste trabalho, foi utilizado o sistema Sphinx® que oferece diversas maneiras de navegar-se o texto tentando identificar temas, definir novas variáveis oriundas de critérios subjetivos, as quais podem ser tratadas quantitativamente e cruzadas com outras mais objetivas, possibilitando maiores riqueza e precisão nas inferências e conclusões. Esse sistema também coloca à disposição ferramenta para a codificação de textos, por meio da qual podem ser identificadas dimensões de análise e mesmo tipologias a partir da base de texto original.

Utilização de fontes

A análise de conteúdo pode servir para tratar:

- todo material de comunicação verbal evidenciado na sociedade em geral, seja nos seus textos escritos (documentos oficiais, livros, jornais, documentos pessoais, seja nos registros de voz ou imagem (rádio, televisão etc.) ou, ainda, em outras atividades que podem ser decompostas, como a análise de uma reunião, a da gestão e a do uso do tempo de um profissional qualquer;
- todo material especialmente gerado para a pesquisa psicossocial: discussão em grupos, entrevistas em profundidade, relatos de reuniões etc.

Tipos de análise de conteúdo

Análise de exploração e de verificação (análise não-dirigida e análise dirigida)

Deve-se fazer clara distinção, própria às pesquisas em ciências sociais, entre a análise de documentos que tem por objetivo a verificação de uma hipótese (o objetivo é preciso, sabendo-se o que se busca e podendo-se quantificar os resultados; bem mais facilmente rigorosa e sistemática, com bom número de regras aplicáveis e com técnicas acessíveis) e aquela cujo objetivo é explorar, buscando a definição de hipóteses, e que não pode ser padronizada, visto fazer ela apelo à intuição e à experiência.

Quando se deseja classificar certos elementos significativos em um texto ou um documento, buscando quantificação, deve-se saber o que se está tentando elaborar ou obedecer à intuição em detrimento de toda e qualquer sistematização. Quando se trata de grande massa de dados, e não se dispõe de hipóteses ou curiosidades claras para a exploração do texto, deve-se admitir a análise dirigida, tendo-se ciência do risco de deixar de lado ele-

mentos essenciais (e imprevistos) do escopo do estudo ou da análise.

Análise quantitativa e análise qualitativa

Uma diferença é essencial, segundo George (*apud* Grawitz, 1976:592): a análise qualitativa baseia-se na presença ou na ausência de determinada característica, enquanto a quantitativa busca identificar a frequência de temas, palavras ou símbolos considerados. Outras oposições qualitativo-quantitativo são: impressionismo *versus* sistematização, hipótese em vez de verificação, e flexibilidade e liberdade contra rigidez.

A noção de importância deve ser clara em cada uma dessas análises: para a quantitativa importa o que aparece seguidamente, o **número de vezes**, enquanto a qualitativa valoriza a novidade, o interesse, o tema, aspectos que permanecem na esfera do subjetivo. Tem-se, assim, sempre um dilema: adotar categorias específicas, retratando fielmente a realidade, mas com uma lista de temas cuja frequência será fraca, ou então reagrupar deliberadamente os dados em pequeno número de categorias, sacrificando informação talvez essencial que ficará perdida no resultado final.

Análise direta ou indireta, comunicação representativa e instrumental

A **análise quantitativa direta**, medida mais comumente utilizada, consiste em contabilizar as respostas tal qual elas aparecem. A **análise quantitativa indireta** pode por vezes, além do que se tem como resultado claro e manifesto, obter por **inferência** até mesmo aquilo que o autor deixou subentendido. Esse tipo de informação (escolha das palavras, ritmo do discurso etc.) pode ser muito revelador. Pode-se, então, a partir de uma análise quantitativa, buscar uma interpretação mais sutil, ou o que é **latente** sob a linguagem explícita utilizada no texto. A interpretação indireta, que vai além do que é dito, não se ampara apenas no qualitativo, ela pode perfeitamente apoiar-se em conteúdo quantificado.

Esse tipo de informação pode ser ainda mais preciosa quando se trata de **comunicação instrumental** (destinada a produzir certo efeito sobre o receptor), em oposição à comunicação puramente **representativa** (que informa sobre o estado ou a situação do emissor).

Na prática da análise de conteúdo é muito importante saber se a mensagem é representativa (caso em que se pode contentar em evidenciar o senso aparente, sem buscar **algo mais**) ou instrumental (caso em que se deve variar a estratégia de análise, sempre buscando seu objetivo).

Tendências

A técnica de análise de conteúdo ganhou flexibilidade e adaptou-se a **objetivos** variados e **etapas** diferentes de pesquisa.

- **Objetivos variados** — a análise de um texto de entrevista psicoterapêutica visará atingir uma interpretação **instrumental, indireta**, de natureza psicológica e **qualitativa** dos elementos inconscientes do paciente. Uma análise de conteúdo de um jornal será normalmente **direta, representativa, quantificando** as diversas rubricas e sendo ela mesma reveladora dos objetivos que o jornal persegue e da clientela à qual ele se destina.
- **Etapas diferentes** — por falta de hipótese *a priori* e de objetivo preciso de medida de determinado fator, o analista começará por se familiarizar com a sua documentação de base, relendo-a diversas vezes. Essa etapa, forçosamente qualitativa e intuitiva, possibilitará imaginar categorias que servirão para ordenar diversos elementos do texto e, na continuidade, quantificá-los.
- **Qualitativo e quantitativo não se excluem, completam-se** — a análise qualitativa fornece, inicialmente, as categorias necessárias a uma análise de conteúdo quantitativa e rigorosa.

A APLICAÇÃO E O USO DA ANÁLISE DE CONTEÚDO

Estudo da comunicação no contexto da relação emissor-receptor

Qualquer que seja a comunicação, estar-se-á na presença de um emissor que lança uma mensagem possuindo conteúdo e forma, visando atingir um objetivo, e endereçada a um ou vários receptores. Quem fala? Para dizer o quê? A quem? Como? Com que resultado? Esse é um esquema simples proposto por Lasswell (*apud* Grawitz, 1976:598) para os estudos da comunicação, o qual parece útil para a análise de conteúdo, com as questões propostas recobrando o conjunto dos problemas.

- **Quem fala?** ou estudo do emissor. Podem ser duas as situações: primeira, o emissor reage a um estímulo mais ou menos controlado pelo observador, como em uma entrevista dirigida; ou, caso mais freqüente, o analista não tem relação alguma com o emissor e busca com base no conteúdo as informações emitidas, podendo mesmo, por vezes, buscar identificar o autor ou as suas características a partir do texto-fonte.

- **Para dizer o quê?** Trata-se de estudar as características do conteúdo da mensagem, reconhecer as nuances e as diferentes **direções** do conteúdo, comparar temas, discursos, *slogans* ou mesmo a evolução dos textos de um autor. Seguidamente, trata-se de comparar materiais de textos diferentes. A análise de conteúdo contribui fortemente para a elaboração adequada de uma entrevista, facilitando a codificação das questões de resposta aberta ou dos relatos de uma reunião ou uma discussão em grupo, possibilitando uma sistematização.
- **A quem?** ou estudo do receptor. Determina-se quem é o receptor, a quem se dirige a mensagem. Pode ser interessante identificar também o que é latente e não só o explícito claramente (por exemplo, o estudo de uma mensagem que um dirigente endereça à sua nação, mas que de fato deseja passar à opinião internacional).
- **Como?** Comporta o estudo da forma ou dos meios pelos quais o texto busca produzir uma impressão. É algo essencialmente qualitativo e subjetivo, mas visto sob um aspecto quantitativo. Estuda-se os elementos que contribuem para produzir essa impressão: escolha das palavras, repetições, composição da frase etc. As categorias serão sempre qualitativas, mas o analista vai quantificar os dados associados.
- **Com que resultado?** A proposta é conhecer o efeito da mensagem sobre o receptor ou, então, como deseja-se, com os elementos certos e da melhor forma, influenciar determinado público em certo sentido ou, ainda, caso as estratégias do emissor sejam claras, prever o que ele busca e seus objetivos implícitos. Em outras palavras, por vezes é explicativo ou elucidativo, por vezes é mais preditivo (ou indicativo).

Ampliação da aplicação da análise de conteúdo

A multiplicação dos estudos sobre as comunicações, com a busca da identificação e da compreensão dos processos (além da análise do explícito ou do **trivial**), levou a considerar na análise, além das questões apresentadas, a influência do ambiente, tanto sobre o emissor como sobre o receptor, para medir o efeito de uma comunicação particular. A análise de conteúdo pode levantar ou mesmo resolver grandes problemas nas áreas de Sociologia e Psicologia Social (Grawitz, 1976:603).

ETAPAS TÉCNICAS DA ANÁLISE DE CONTEÚDO

Como em uma pesquisa, a primeira parte da análise de conteúdo está associada à idéia daquilo que se deseja

buscar ou investigar e ao exercício de precisar seu objetivo. O meio para obter-se uma resposta é a colocação de uma questão. Em uma entrevista, é o questionário que permite entrevistar; na análise de documentos, é por meio das categorias que o analista vai recolher os dados presentes nos documentos.

Primeira etapa A escolha das categorias (de análise)

Formulação

As categorias são as rubricas significativas em função das quais o conteúdo será classificado e eventualmente quantificado. No caso de uma enquete, pode-se prever as categorias já em forma de código. A análise de conteúdo deve, normalmente, permitir o aparecimento de variáveis e fatores de influência ignorados no início dos trabalhos. As categorias, quando não se tem idéia precisa delas, devem surgir com base no conteúdo. Na enquete de exploração as categorias constituem, elas próprias, o quadro, o objeto ou o contexto de análise (permitindo selecionar os dados no texto), enquanto na busca de verificação de uma hipótese elas podem ser predefinidas. Grawitz (1976:605) aconselha que se deve "prolongar o período de testes e estabelecer diversos sistemas de categorização ou codificação antes de adotar um esquema (estrutura ou sistema) de categorias".

As categorias devem ter origem no documento objeto da análise ou em certo conhecimento geral da área do conhecimento ou da atividade no qual ele se insere: das respostas, caso seja uma entrevista; dos objetivos, intenções, crenças do emissor, no caso de um texto. Considerando sempre elementos **ausentes** que podem ser significativos.

Características

A escolha das categorias é o procedimento essencial da análise de conteúdo, visto que elas fazem a ligação entre os objetivos de pesquisa e os seus resultados. O valor da análise fica sujeito ao valor ou à legitimidade das categorias de análise. O objetivo perseguido deverá pautar a escolha ou a definição do que deve ser quantificado. As categorias devem ser exaustivas (percorrer todo o conjunto do texto), exclusivas (os mesmos elementos não podem pertencer a diversas categorias), objetivas (características claras, de modo a permitir seu uso por diferentes analistas em um mesmo texto) e pertinentes (em relação aos objetivos perseguidos e ao conteúdo tratado).

A pertinência pressupõe a adaptação das categorias ao quadro de referência que permeia o texto ou a situação.

Excessos a evitar

- Impor esquema de análise muito rígido desde o início, o que impedirá a apreensão da complexidade do conteúdo;
- elaborar o esquema de forma superficial, centrando-se no explícito e esquecendo-se do conteúdo latente;
- escolher categorias muito detalhadas e muito numerosas, reproduzindo quase que todo o texto pensando em **não perder nada**;
- estabelecer categorias muito **grosseiras**, as quais não distinguirão suficientemente os elementos que reagrupam.

Deve-se estabelecer um nível de medida adequado.

Padronização

Seria interessante dispor de certo número de categorias aplicáveis a diversos casos (idade, sexo, profissão, religião, nacionalidade e nível de instrução são alguns exemplos). No entanto, de resto, não é fácil adotar um padrão. Certas áreas organizam-se para tal, como os pesquisadores de/ em *marketing* (organizados na *American Marketing Association* (AMA) ou na ESOMAR, associação européia com a mesma finalidade, por exemplo).

Os esquemas de análise, para serem satisfatórios, devem respeitar ou convir a um sistema conceitual geral aceito por todos, mas também a um conteúdo específico que muda a cada nova pesquisa.

Segunda etapa

Os problemas oriundos da quantificação do conteúdo

A dificuldade de análise associa-se à ambigüidade, à variedade e à complexidade das fontes.

O problema da amostra

Este problema não existe nas ocasiões em que se estuda um texto em sua totalidade ou uma série de discursos. No entanto, quando se estuda algo vasto, como vários anos de um jornal, ou muito variados, como as propagandas, deve-se forçosamente delimitar a pesquisa, escolhendo certo tipo, certa rubrica, certo setor (por exemplo, somente os editoriais de um jornal). Pode-se também, em vez de limitar o tema, limitar uma amostra que seja representativa do todo.

As regras para constituir uma amostra são:

- delimitar o universo ao qual se aplica a generalização;
- assegurar que cada unidade desse universo tenha uma probabilidade conhecida de estar contida na amostra;

- proceder a uma amostragem independente de toda a correlação entre as unidades do universo;
- escolher amostra relativamente ampla, para que comporte somente reduzido provável erro de amostragem.

Quais são as características que podem assegurar a representatividade da amostra de um documento?

Talvez elas pudessem ser definidas em função das hipóteses formuladas, mas seria pouco científico e não compreenderia todos os fatores em jogo. Segundo Berelson (*apud* Grawitz, 1976:613), o problema de amostragem implica **três decisões** em matéria de análise de comunicações:

- **escolha da fonte** — depende do objetivo. Se o analista deseja ter condições de justificar suas conclusões gerais a partir do material analisado, deverá ser capaz de formular as razões que o levaram a definir determinado universo de conteúdo;
- **número e data das mensagens** — Qual a quantidade de mensagens a reter para que a amostra seja representativa? Qual o período a considerar? O tema pode determinar a época, mas não o período. Assim, deve-se estar atento para o fato de a unidade escolhida do universo a estudar ter probabilidade conhecida de estar contida na amostra e de ser independente de toda a correlação entre as unidades desse universo;
- **conteúdo observado** — trata-se da amostra da amostra, difícil de definir, dada a variedade de objetivos, o que conduz a usar seguidamente métodos bem-flexibilizados, que aceitam em um mesmo estudo tipos de amostra levantados a partir de fontes, períodos e conteúdos diferentes.

Diversas unidades de quantificação

Se houver o interesse em quantificar, deve-se escolher índices a reter para categorizar e decidir o tamanho (ou o comprimento) dos elementos, segundo os quais se vai decompor o conteúdo, o que envolve alguns conceitos:

- **Unidade de registro** — é a primeira unidade de análise (e a mais estreita), de tamanho variável: **palavra**, menor unidade — pode-se contar quantas vezes determinada palavra foi utilizada e em que contexto foi usada; **tema**, fragmento significativo correspondente à idéia que recobre uma das categorias — é elemento menos preciso do que a palavra e pode ter três linhas ou quatro páginas, sendo difícil obter-se resultados de grande fidelidade devido à grande diferença de percepção entre o analista e outro quanto a um mesmo conteúdo; **item**, corresponde a um conteúdo total, um livro, um filme, um discurso.

- **Unidade de contexto** — é mais sutil e flexível do que a unidade de registro e não se ampara em quantificação rigorosa. Trata-se de segmento mais amplo de conteúdo, ao qual se refere quando se deseja compreender a unidade de registro. Por exemplo, caso a unidade de registro seja a palavra, a unidade de contexto poderá ser a frase.
- **Unidade de numeração** — trata-se da forma de contagem, da forma de medir-se as partes, o número de linhas, por exemplo, ou a quantidade de centímetros quadrados em um artigo de jornal etc. Enquanto a unidade de registro indica a forma de decompor o conteúdo (elemento significativo), a unidade de numeração implica na forma de medir esses diversos **pedaços**. Desejando uma medida precisa, deve-se recorrer a uma unidade de numeração. Para marcar a diferença entre dois temas, um tratado 15 vezes em três linhas e outro tratado duas vezes em 100 linhas, é necessário ter uma unidade de numeração (número de linhas) distinta da unidade de registro (número de vezes em que aparece o tema). As unidades de numeração referem-se ao espaço ou ao tempo: parágrafos, linhas, centímetros, minutos de registro. Elas devem ser passíveis de medida objetiva, fiel e verificável, e se prestar a manipulações matemáticas (desde a adição).

É essencial a **distinção** entre a **unidade de registro** e a **unidade de numeração**.

VALOR DA ANÁLISE DE CONTEÚDO COMO INSTRUMENTO DE PESQUISA

A análise de conteúdo é uma técnica de refino, portanto delicada, e que exige, para satisfação da curiosidade do investigador, muita dedicação, paciência e tempo, além de intuição, de imaginação para perceber o que é importante e de criatividade para escolher as categorias. Ao mesmo tempo, o investigador deve ter disciplina e perseverança, rigor ao decompor um conteúdo ou ao contabilizar resultados ou análises.

- **Fidelidade** — Por definição, a análise de conteúdo deve ser objetiva e os resultados devem ser independentes do instrumento de medida, sendo conveniente reduzir ao máximo as diferenças entre os pontos de vista dos analistas. No entanto, esse é um problema antigo nas ciências sociais: O que vale mais à pena, uma medida exata ou dados ricos? A fidelidade não pode ser, então, encarada da mesma forma quando se trata de uma análise quantitativa de um conteúdo claro e manifesto e quando se trata de uma análise de tipo mais qualitativo em que se busca identificar intenções latentes, em que vale mais a presença ou a ausência de

determinado elemento e não a sua frequência. Neste caso, o analista é conduzido a minimizar a questão da fidelidade. Já na análise quantitativa, a princípio o analista vai se deter na frequência dos elementos, não se importando tanto com a validade de certos dados isolados.

- **Validade lógica** — Será que o instrumento mede o que ele se propõe a medir? Uma análise é válida quando a descrição quantificada que oferece do conteúdo (objeto do estudo) é significativa para o problema originalmente colocado e quando reproduz fielmente a realidade dos fatos que traduz. Naturalmente, é condição essencial de partida a representatividade da amostra, assim como a observação satisfatória das condições técnicas inerentes a cada etapa. Quanto mais limitado ou delimitado o objetivo, mais fácil será a comprovação da validade, bem como uma análise quantitativa terá sua validade mais facilmente demonstrada do que uma qualitativa.
- **Inferência** — É necessária atenção especial, pois por vezes uma mesma expressão tem dupla interpretação ou, então, interpretação negativa ou positiva de acordo com o seu contexto no conteúdo analisado.
- **Validade empírica e medida da validade** — A predição inerente é justa ou precisa? Trata-se de questão de difícil resposta. Em vez de convicções, indicações, recomenda-se prudência e humildade nas conclusões. Entretanto, acredita-se que a experiência e a capacitação do analista referendarão o valor de suas análises.

COLOCANDO EM EVIDÊNCIA CERTAS POTENCIALIDADES, APLICAÇÕES E LIMITES DAS FERRAMENTAS DE ANÁLISE DE DADOS TEXTUAIS

Neste texto, o interesse é simplesmente valorizar a utilidade do uso da análise lexical. Ela permite, por exemplo, analisar em profundidade o modo de expressão específico de cada pessoa (ou grupo) envolvida em um debate. O trabalho sobre um texto organizado (**lematizado**) permite utilizar categorias gramaticais, como substantivo, verbo, adjetivo, de forma a organizar a primeira apreensão ou impressão do conteúdo do corpo do texto.

O termo **lematizador** é aqui definido, com base no francês **lemmatiseur** (utilizado pela equipe de pesquisa do Professor Jean Moscarola da *Université de Savoie*, França), como uma ferramenta informatizada que auxilia o analista marcando no texto as diferentes categorias gramaticais, como substantivos, verbos, adjetivos, nomes etc. Usar-se-á, então, o termo **texto lematizado** para aquele que já foi trabalhado com o **lematizador**.

Pode-se evidenciar indicadores lexicais, como a banalidade (ou a trivialidade), especificidade que orienta a pesquisa seletiva das frases mais triviais ou das palavras mais significativas ou originais de determinado grupo ou determinada categoria. Pode-se, igualmente, com esses recursos, construir indicadores *ad hoc*, por exemplo para avaliar o grau de engajamento dos atores ou integrantes de um grupo ou uma discussão.

Esses métodos são de fácil emprego graças ao progresso dos sistemas (*software*) e das técnicas de análise, de navegação e de estatísticas lexicais. A grande variedade de vias de investigação e de novas leituras que esses sistemas aplicativos colocam à disposição dos pesquisadores não deve, entretanto, mascarar ou esconder a dificuldade da interpretação final inerente a essa atividade. A dedicação e a paciência, aliadas à curiosidade, devem ser características marcantes do pesquisador nesse campo.

A rápida evolução das ferramentas de análise de dados textuais (ou análise de conteúdo) coloca à disposição do pesquisador novos métodos de trabalho combinando:

- a estatística lexical e os métodos de análise de dados;
- a análise sintática;
- a navegação lexical e a leitura assistida por computador.

O sistema Sphinx®, um dos *softwares* disponíveis no **mundo dos sistemas**, integra essas abordagens em um mesmo ambiente (ver funções principais no quadro da página seguinte) e autoriza a qualquer instante, o retorno ao texto e sua leitura. Seu uso permite distinguir sete níveis de análise, a seguir descritos e ilustrados sumariamente com base em Moscarola (1995). De qualquer forma, **independentemente do sistema a ser utilizado**, do princípio e da técnica de exploração dos dados qualitativos disponíveis, os esquemas a seguir apresentados poderão contribuir para melhor estratégia de análise.

Aproximação lexical sumária

- **Nível 1** — reduzir o texto ao alto do léxico: o exame das palavras mais freqüentes dá idéia de seu conteúdo (figura 1).

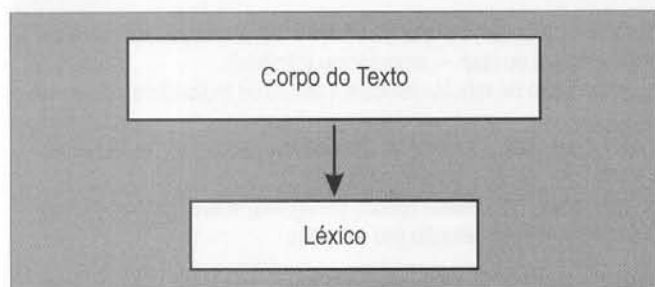


Figura 1: Representação da Aproximação Lexical Sumária

Aproximação lexical controlada

- **Nível 2** — reduzir o texto ao seu léxico e controlar, via navegação lexical, a validade e o fundamento das interpretações elaboradas a partir do léxico (figura 2).

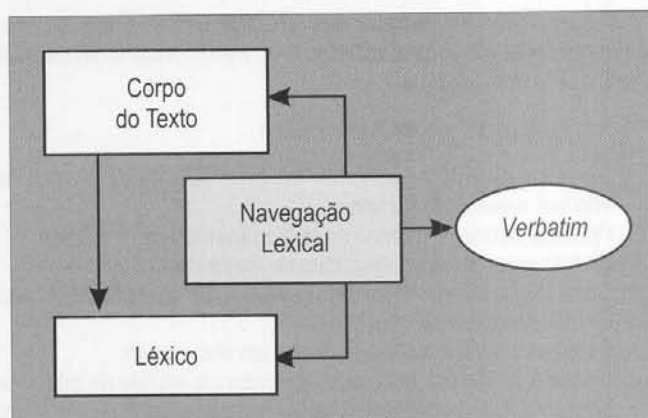


Figura 2: Representação da Aproximação Lexical Controlada

Aproximação lexical seletiva

- **Nível 3** — trabalhar sobre um léxico reduzido após ter eliminado as palavras-ferramentas (definição daquelas palavras que se pode **deixar de lado** ou desprezar, todas armazenadas em um dicionário específico), concentrando a atenção no exame de nomes, verbos, adjetivos, tudo via utilização de dicionários e do lematizador (figura 3).

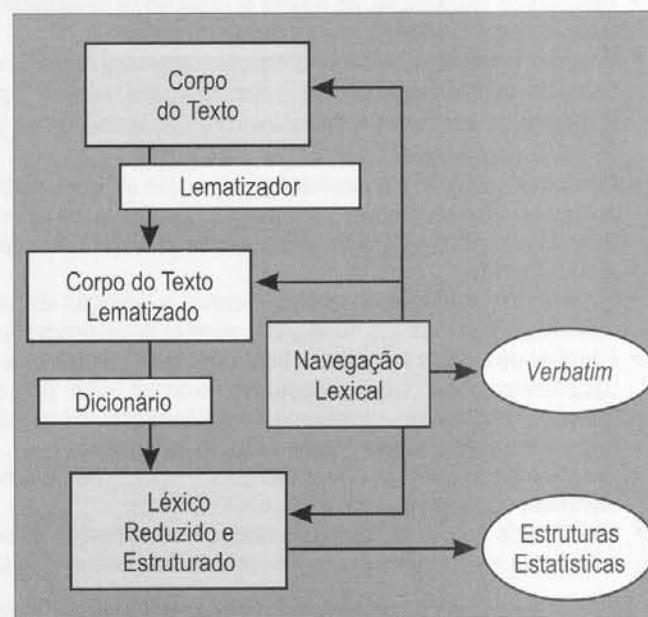


Figura 3: Representação da Aproximação Lexical Seletiva

ANÁLISE DE DADOS TEXTUAIS

Ferramentas de Uso Geral para Múltiplas Aplicações

Devem permitir acesso, leitura e análise de textos de toda e qualquer natureza

Um sistema em microcomputador (Sphinx® ou outro) destina-se àqueles que desejam analisar textos, qualquer que seja a origem dos mesmos: entrevistas não-diretivas, documentos técnicos ou literários, documentos de autor, bases de dados textuais (bases bibliográficas, patentes, cadastros etc).

Devem atender a objetivos variados

- Revisar, examinar ou visualizar um texto ou uma base de dados para ter-se idéia rápida e aproximada de seu conteúdo.
- Analisar a estrutura do texto.
- Procurar certos elementos do texto para produzir extratos.
- Efetuar uma análise sistemática de conteúdo.
- Caracterizar ou identificar certos elementos de estilo ou de enunciação.
- Elaborar dicionários temáticos.
- Construir o índice ou o sumário de um documento.
- Aplicar a determinado texto os métodos de análise de dados textuais.
- Controlar certas características do texto baseadas na construção de indicadores estatísticos.

Devem contemplar diversas áreas de aplicação

- Análise de dados de enquetes e de todas as formas de entrevistas.
- Pesquisa lingüística, lexicologia e terminologia.
- Estudos e criação literária, ensino literário e lingüística.
- Pesquisa documentária.
- Vigília tecnológica e concorrencial.
- Criação publicitária.
- Jornalismo.
- Análise de interações de um grupo de estudos ou de executivos.
- Aprendizagem a partir dos discursos de dirigentes ou políticos.
- Compreensão de um processo de decisão coletivo.

Por uma ferramenta com espectro amplo de aplicação

- **Funções de importação, de edição (e recortes) e de estruturação** do corpo de um texto existente e disponível na forma de um ou mais arquivos tipo texto.
- **Funções de análise lexical** — produção automática e gestão do léxico do texto.
- **Funções de navegação lexical** — com navegação no texto a partir de elementos do léxico ou de variáveis de contexto, extração de elementos do texto em função de seu conteúdo lexical, busca (e identificação) de contexto, produção de *verbatim* (extratos do texto segundo certo critério).
- **Funções de revisão e de anotação** — correção do texto, marcação no texto, agrupamento de palavras ou expressões, produção de textos derivados ou expurgados a partir do texto base em análise (exportados).
- **Funções de gestão de dicionários** — com produção, enriquecimento e utilização de dicionários de formas gráficas ou de expressões compostas.
- **Funções de produção de índice** — busca e definição de palavras para compor o índice ou o sumário de um documento, com atribuição automática dos números de páginas correspondentes.
- **Funções de análise sintática** — busca da classe gramatical das palavras, lematização (mudança das formas derivadas ou infletidas das palavras, como plurais ou conjugações de um verbo, para sua forma original ou raiz — singular ou infinitivo).
- **Funções estatísticas** — contagem e desdobramento das ocorrências, construção de tabelas lexicais, cálculo de indicadores lexicais, cálculos de especificidades, caracterização de contextos.
- **Funções de análise de dados textuais** — produção de variáveis de origem textual, análise fatorial, classificação, **tratamento integrado** de dados textuais e de outras variáveis.
- **Funções de gestão de bases de dados** — modificação automática dos níveis de análise (texto, parágrafo, frase, proposição ou qualquer outra unidade de senso) e restrição das variáveis de contexto em função da seleção (ou recorte).

Todas essas funções devem estar disponíveis em um sistema em microcomputador, *for windows®*, de forma amigável, flexível, para que o usuário ou o analista tenha autonomia e tranqüilidade para exercer sua curiosidade e elaborar análises e inferências em profundidade.

Estatísticas lexicais e quantificação do texto

- **Nível 4 — pesquisa das especificidades lexicais:** estabelecer a estatística das palavras do texto segundo uma variável externa não-textual (figura 4).

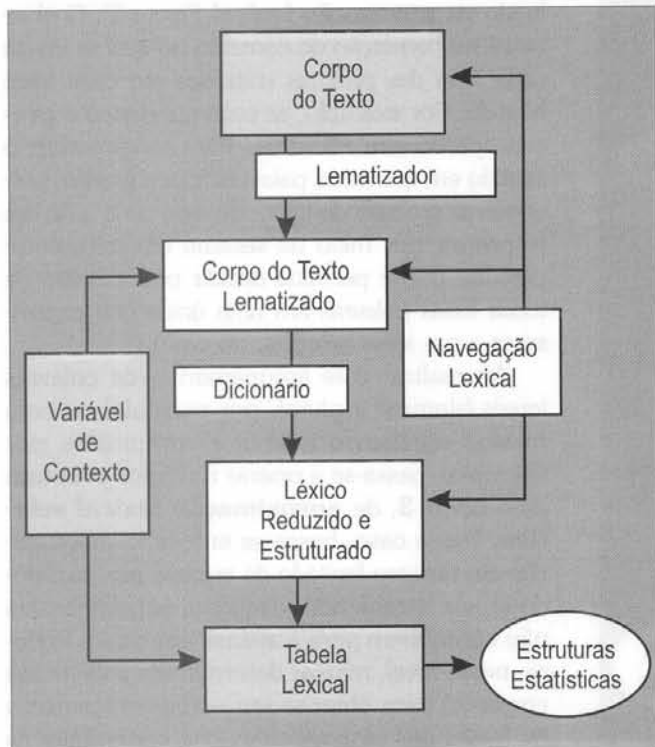


Figura 4: Representação da Pesquisa das Especificidades Lexicais

O valor de uma análise de conteúdo depende da qualidade da elaboração conceitual feita a priori pelo pesquisador.

- **Nível 5 — codificação lexical e medidas:** descrever o texto por uma variável nominal, codificando a presença nas respostas de conjunto de palavras selecionadas no léxico (figura 5).
- **Nível 6 — quantificação do texto:** calcular as intensidades lexicais de certos campos lexicais, a partir de um texto **lematizado e aliviado** das palavras-ferramentas (retirando-se do texto as palavras que não inte-

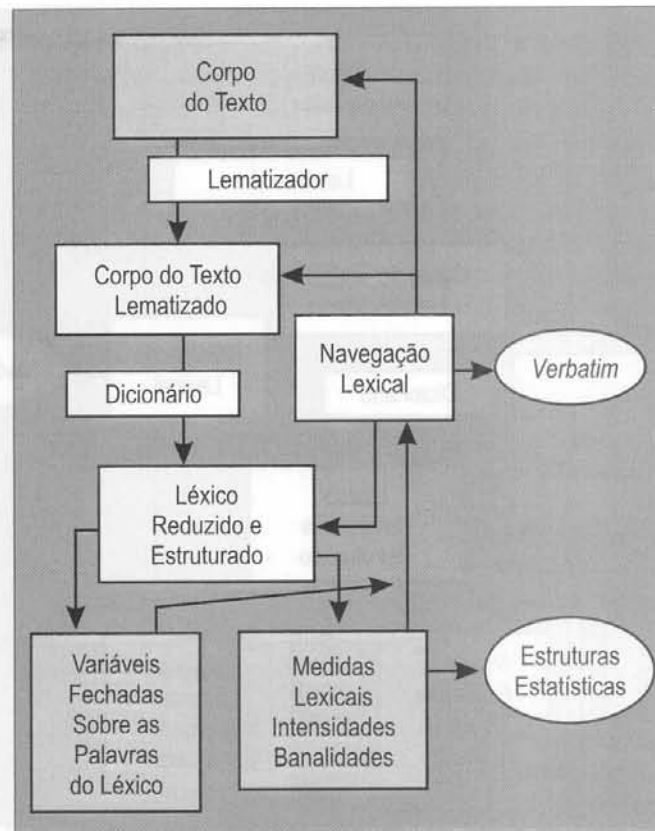


Figura 5: Representação de Codificação Lexical e Medidas e da Quantificação do Texto

ressam, por procedimento ou **cliques**). Cada resposta é descrita pela intensidade com que se manifestam os diferentes campos estudados (figura 5).

Análise multivariada dos dados textuais

- **Nível 7 — análise de dados efetuada sobre as variáveis de origem lexical:** aplicar os métodos de análise multivariada sobre tabelas lexicais, variáveis de origem lexical consideradas entre elas (análise fatorial, classificação) ou em relação a outras variáveis de contexto. Integra-se, assim, a análise de dados clássica com a análise de dados textual (figura 6).

O acesso aos dados documentais ou aos materiais da pesquisa qualitativa é, assim, facilitado. Entretanto, qualquer que seja a potência da ferramenta, a análise de dados textuais nada faz além de **aflorar** a questão do senso. Apesar de poder permitir a redução do **ruído** de linguagem, a economia das longas e às vezes fastidiosas leituras e a produção bem mais rápida de conclusões e relatórios, conduz todos a uma leitura interpretativa de certa forma perigosa, visto ser ela rápida e correr o risco de ser falsamente objetiva.

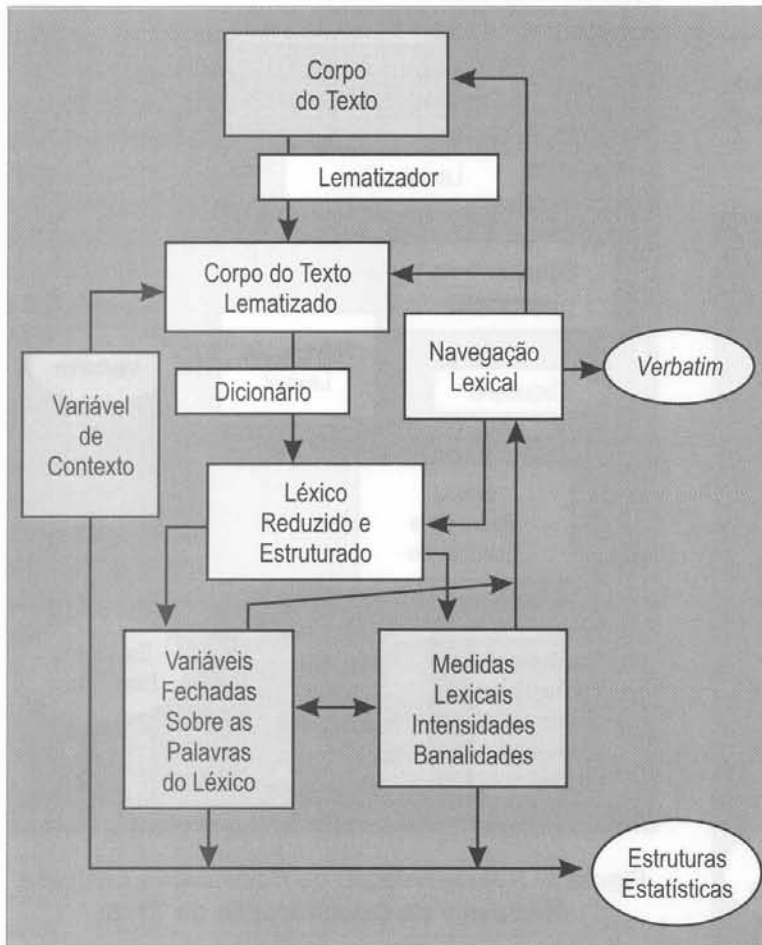


Figura 6: Representação da Análise de Dados Efetuada Sobre as Variáveis de Origem Lexical

APLICAÇÃO PRÁTICA DA ANÁLISE QUALITATIVA NA ÁREA DE MARKETING

Para ilustrar alguns dos níveis de análise de conteúdo, descritos no tópico anterior, serão utilizados dados qualitativos oriundos de pesquisa realizada no segundo semestre de 1994 para o SEBRAE/RS e a FIERGS (IEL), com coordenação do PPGA/UFRGS (Ruas *et alii*, 1994). A pesquisa, que tinha como objetivo a avaliação dos Programas de Qualidade na indústria gaúcha, durou quatro meses e abrangeu 120 empresas (30 pequenas, 30 médias e 60 grandes). Foram efetuadas entrevistas com o diretor industrial e o diretor da qualidade ou de recursos humanos de cada empresa.

A questão submetida à análise de conteúdo refere-se ao monitoramento do ambiente em que a organização se insere. O executivo industrial foi questionado sobre "os critérios utilizados para o acompanhamento do mercado", caso a empresa realizasse tal acompanhamento.

A primeira análise que se pode propor relaciona-se com o **nível 2**, de **aproximação lexical controlada**. Nesse nível, o corpo do texto (formado por todas as respostas emitidas pelos 120 entrevistados), composto por 374 palavras sendo 184 diferentes (constituindo o **léxico**), é analisado via **navegação lexical** (figura 7). O objetivo é ter convicção do contexto no qual se insere cada uma das palavras utilizadas em cada frase emitida. Por exemplo, as palavras **preço** e **preços** apareceram 19 vezes. Para compreender o sentido em que essas palavras foram usadas, percorreu-se o corpo do texto, ou seja, cada uma das respostas, por meio de sistema em microcomputador, o que permitiu decidir pela reunião de todas essas palavras em uma única que expressasse a sua idéia (**preços**, no caso).

Ao realizar esse agrupamento, de palavras iguais (singular e plural, por exemplo) ou com mesmo significado (cliente e consumidor, por exemplo), passa-se a operar na lógica proposta pelo **nível 3**, de **aproximação lexical seletiva**. Nesse caso, busca-se enfocar a investigação em número limitado de argumentos, excluindo-se, via ferramental adequado, as palavras que não contribuem para a análise em curso. Pode-se, nesse nível, marcar determinada palavra (ou conjunto) para obter-se um *verbatim* (conjunto de frases dos respondentes que contenham as palavras selecionadas para análise). Por exemplo, se se marcar no **léxico** a palavra **vendedores** (reagrupamento de **vendedores** e de **representantes**, no caso desta análise) e se solicitar o *verbatim*, o resultado será o que consta no quadro 1.

Quadro 1

Verbatim da Palavra Vendedores

# vendedores (4)	
Entrevistado n° 1:	Através de questionário aos clientes, e dos REPRESENTANTES.
Entrevistado n° 8:	Comentários, palavra dos clientes, VENDEDORES etc.
Entrevistado n° 67:	Reuniões com REPRESENTANTES de vendas: — Situação de mercado e relacionamento com clientes. — Fidelidade: renovação dos pedidos; Não existe questionário formal; 1995: Pesquisa com questionário junto aos clientes e fornecedores.
Entrevistado n° 73:	Aceitabilidade pelo cliente REPRESENTANTES.

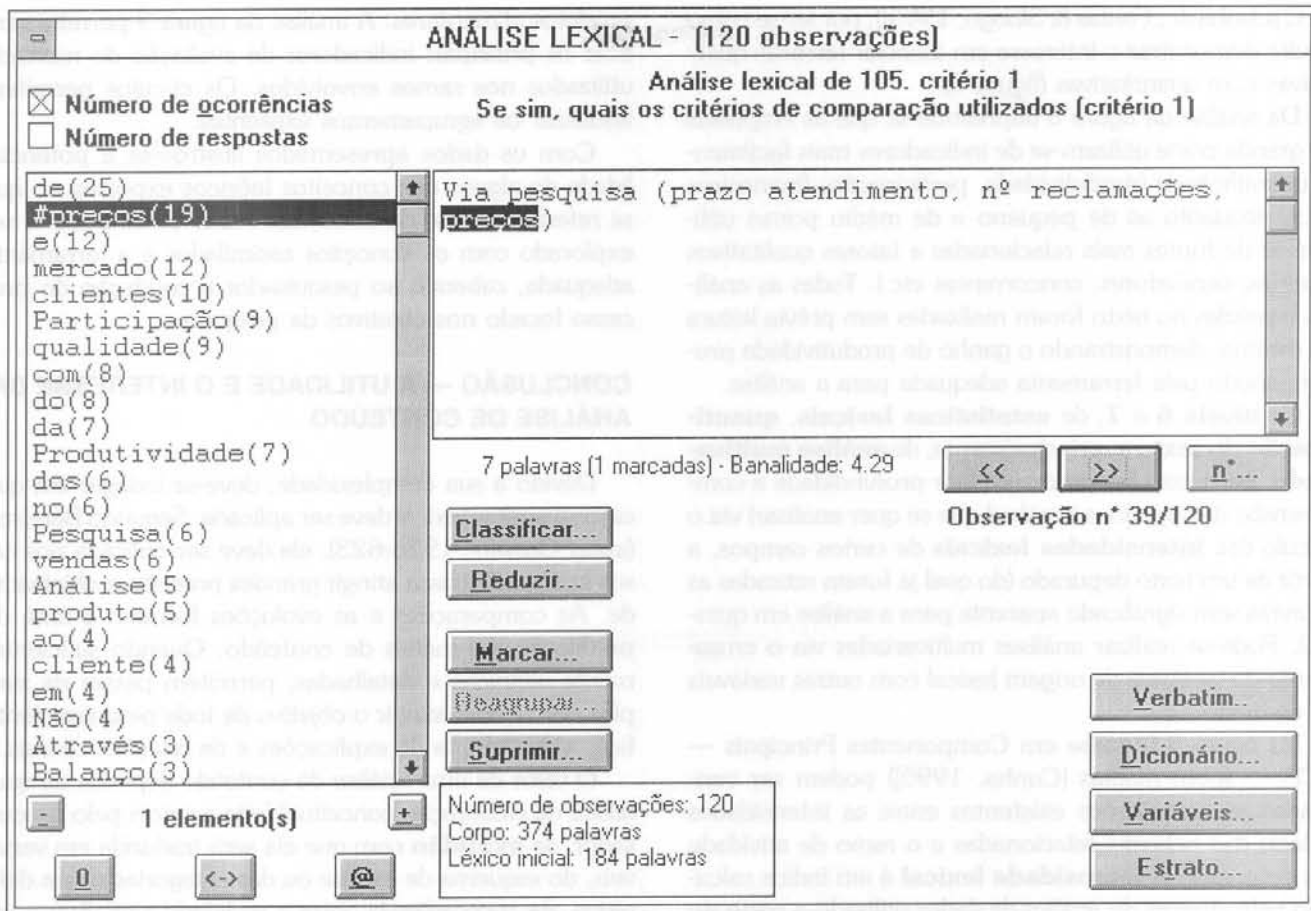


Figura 7: Tela de Navegação Lexical (Sphinx Plus®)

Considera-se que exercícios de reagrupamento de palavras foram realizados, resultando no elenco de palavras apresentadas no quadro 2 (cada uma representando certa quantidade de idéias ou respostas, todas potencialmente significativas para o objetivo da investigação em curso. O símbolo (#) significa que um reagrupamento de palavras foi realizado e o número entre parênteses após a palavra representa a quantidade de vezes que ela foi citada.

Diversos tratamentos podem derivar dessa lista de palavras-chave importantes para a análise. Ilustrar-se-á a

seguir o uso dessas técnicas de análise qualitativa nos níveis 4 e 5 e, finalmente, nos níveis 6 e 7.

Os **níveis 4 e 5**, de **estatísticas lexicais e quantificação do texto**, permitem buscar a compreensão das opiniões (variável que se deseja analisar) pelo simples cruzamento com uma variável externa. Considerou-se no exemplo o porte das organizações envolvidas: pequeno, médio e grande.

Primeiramente, foi necessário criar uma variável de múltipla resposta, neste caso denominada de **critério de avaliação da concorrência**, sendo as modalidades de

resposta as palavras da lista do quadro 2, cada qual com a identificação SIM ou NÃO, ou seja, se está ou não presente na resposta de cada um dos entrevistados. Feito isso, realizou-se o cruzamento bivariado entre porte e critério de avaliação da concorrência, optando-se de imediato pela representação em uma carta fatorial (Análise Fatorial das Correspondências —

Quadro 2

Elenco de Palavras-Chave para a Investigação

#Preços(19)	#Clientes(15)	Mercado(12)	#Vendas(10)	Participação(9)
Qualidade(9)	#Pesquisa(8)	#Concorrência(7)	#Financeiros(7)	
Produtividade(7)	Produto(5)	#Benchmarking(3)	#Vendedores(4)	

AFC [Cunha Jr., Freitas & Slongo, 1995]), por ter-se como intuito demonstrar o interesse em associar técnicas qualitativas com quantitativas (figura 8).

Da análise da figura 8 depreende-se que as empresas de grande porte utilizam-se de indicadores mais facilmente quantificáveis (produtividade, participação, financeiros etc.), enquanto as de pequeno e de médio portes utilizam-se de fontes mais relacionadas a fatores qualitativos (clientes, vendedores, concorrentes etc.). Todas as análises baseadas no texto foram realizadas sem prévia leitura do mesmo, demonstrando o ganho de produtividade proporcionado pela ferramenta adequada para a análise.

Os níveis 6 e 7, de **estatísticas lexicais**, **quantificação do texto** e, principalmente, de **análise multivariada**, permitem buscar com maior profundidade a compreensão das opiniões (variável que se quer analisar) via o cálculo das **intensidades lexicais** de certos campos, a partir de um texto depurado (do qual já foram retiradas as palavras sem significado aparente para a análise em questão). Pode-se realizar análises multivariadas via o cruzamento de variáveis de origem lexical com outras variáveis de contexto.

Na figura 9 (Análise em Componentes Principais — ACP — sobre médias [Cunha, 1995]) podem ser verificadas as associações existentes entre as intensidades lexicais das palavras selecionadas e o ramo de atividade das empresas. A **intensidade lexical** é um índice calculado pelo sistema de análise de dados utilizado a partir do número de palavras na resposta a uma questão e do número de repetições da palavra selecionada nessa mesma resposta, ou seja, a intensidade lexical é a relação **repetição da palavra/comprimento da resposta**. Considerou-se no exemplo os seguintes ramos de atividade das organizações envolvidas: metalurgia, química, bebidas, mecânica, material de transporte, papel/papelão, fumo, vestuário e calçados, produtos minerais não-metálicos,

produtos alimentares. A análise da figura 9 permite verificar os principais indicadores de avaliação do mercado utilizados nos ramos envolvidos. Os círculos permitem visualizar os agrupamentos existentes.

Com os dados apresentados ilustrou-se a potencialidade de alguns dos conceitos teóricos expostos no que se refere à análise de conteúdo. Muito ainda poderia ser explorado com os conceitos assimilados e a ferramenta adequada, cabendo ao pesquisador a condução do processo focado nos objetivos da pesquisa.

CONCLUSÃO — A UTILIDADE E O INTERESSE DA ANÁLISE DE CONTEÚDO

Devido à sua complexidade, deve-se indagar em que casos a análise lexical deve ser aplicada. Segundo Berelson (*apud* Grawitz, 1976:623), ela deve ser aplicada nos casos em que se busca atingir grandes precisão e objetividade. As comparações e as evoluções formam a área de predileção da análise de conteúdo. Quando suficientemente definidas e detalhadas, permitem passar da simples descrição e atingir o objetivo de toda pesquisa científica: a descoberta de explicações e de relações causais.

O valor de uma análise de conteúdo depende da qualidade da elaboração conceitual feita *a priori* pelo pesquisador, da **exatidão** com que ela será traduzida em variáveis, do esquema de análise ou das categorias e, em definitivo, da concordância entre a realidade a analisar e estas categorias. Para que tal análise mereça ser realizada, as questões orientadas pelas categorias devem explicitar uma hipótese interessante, assim como devem corresponder ao material selecionado.

Como o uso de técnicas de análise **qualitativa** é tema latente na comunidade científica brasileira, espera-se ter contribuído, com este trabalho, para resgatar os seus conceitos e ilustrar algumas de suas potenciais aplicações. ◆

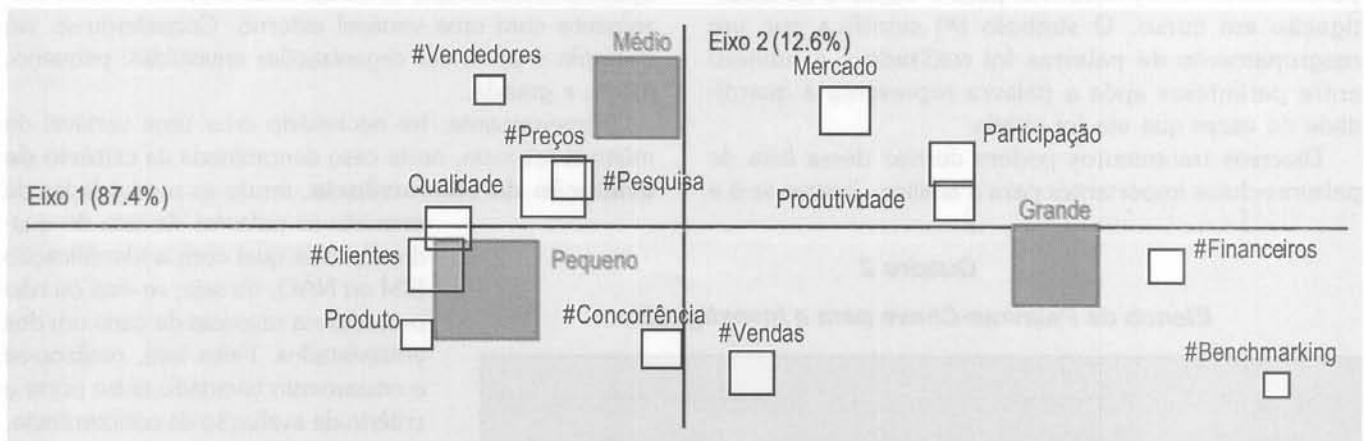


Figura 8: Análise de Correspondência do Cruzamento Envolvendo o Porte e os Critérios de Monitoração do Mercado

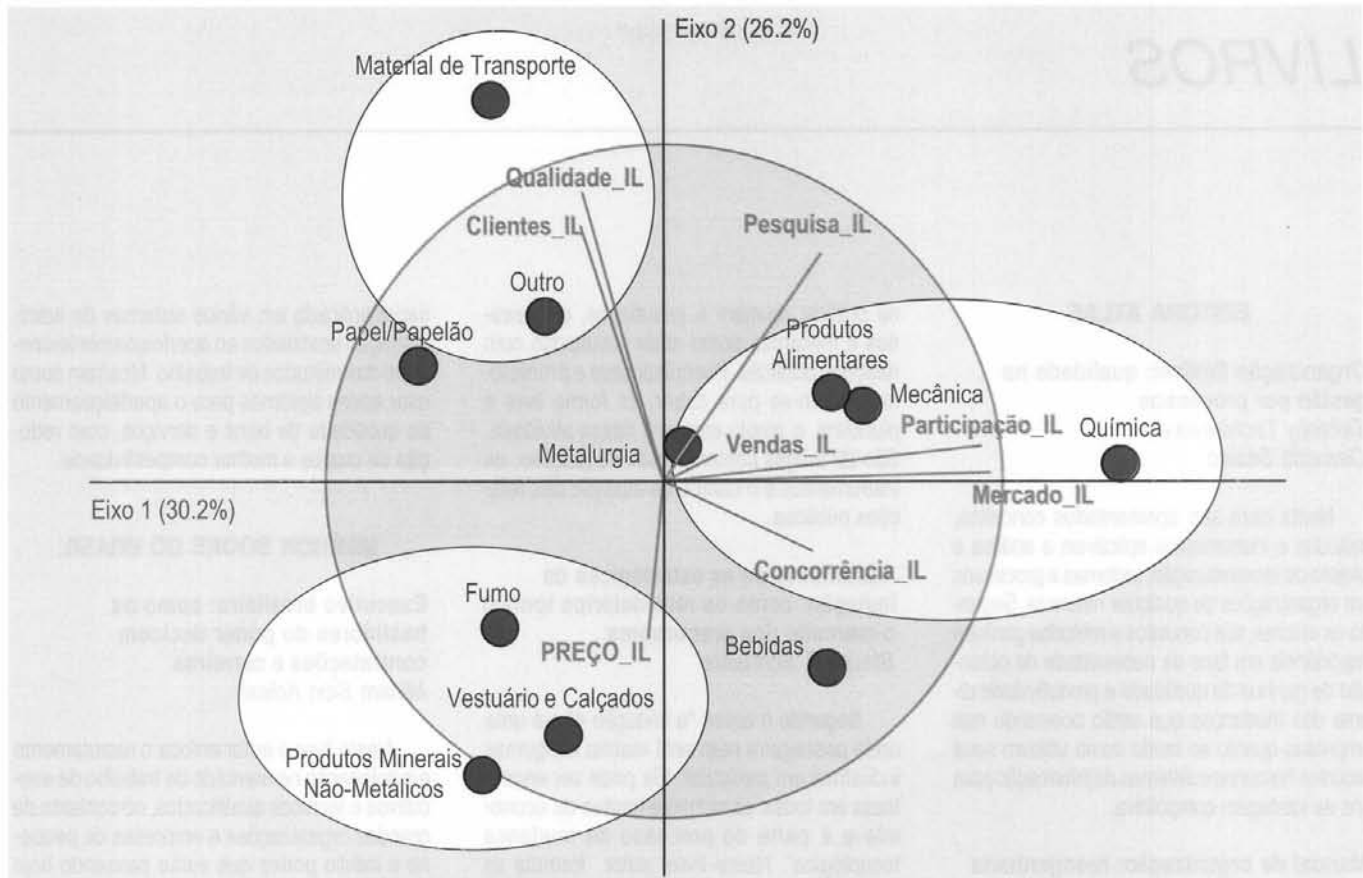


Figura 9: Análise em Componentes Principais Envolvendo o Ramo de Atividade da Empresa e a Intensidade Lexical das Palavras Selecionadas

REFERÊNCIAS BIBLIOGRÁFICAS

CUNHA Jr., M.V. *O processo de segmentação de mercado em função dos atributos do produto: o caso da Cooperativa Central Gaúcha de Leite (CCGL) no mercado de Porto Alegre.* Porto Alegre, DCA, Curso de Administração, FCE/UFRGS, 1995. 98 p. [Trabalho de conclusão de curso de graduação.]

CUNHA Jr., M.V.M.; FREITAS, H.; Slongo, L.A. A pesquisa de marketing como fator de interação universidade-empresa: estudo de caso aplicado utilizando uma ferramenta (Sphinx®) de estruturação e de análise quantitativa e qualitativa de dados. *Caderno de Pesquisa nº 06/95 do PPGA/UFRGS*, Porto Alegre - RS, 1995. ENANPAD, 19. João Pessoa - PB, *Anais*, v.1, n.5, p.109-129, Marketing, 25-27 set. 1995.

GAVARD-PERRET, M.L. & MOSCAROLA, J. De l'énoncé à l'énonciation: pour une relecture de l'analyse lexicale en marketing. *Cahier du GEREG*, Annecy-França, IUP/GEREG e CERIAM, Université de Savoie, 1995. 34 p.

GRAWITZ, Madeleine. *Méthodologie des sciences sociales.* Paris, Dalloz, 1976. [Cap. II, p.586-627]

MOSCAROLA, J. *Balladur, Chirac, Jospin, les mots d'une campagne: quelques exemples d'analyse lexicale avec le Sphinx.* Annecy-França, IUP/GEREG, Université de Savoie, 1995. 8 p. [Workpaper]

RUAS, R.L.; PIRES, M.; FREITAS, H.; ANTUNES, E.; CUNHA

Jr, M. e equipe técnica. *Pesquisa indicadores da qualidade e produtividade.* Participação como consultor técnico (metodologia e análise de dados). Pesquisa em parceria PPGA/UFRGS, IEL/FIERGS, SEBRAE, dez. 1994. 68 p. [Relatório Executivo]

BIBLIOGRAFIA COMPLEMENTAR

BACHELET, & MOSCAROLA, J. La communication interne: diagnostic par l'analyse du discours des acteurs. *Cahier du GEREG*, Annecy-França, IUP/GEREG, Université de Savoie, 1995. 10 p.

LE SPHINX LEXICA (v.2.0). Versão francesa em BETA teste. Annecy-França, Le Sphinx Développement, 1995. [Manual do Usuário]

MOSCAROLA J. *Analyse de contenu et analyse de données: solutions logiciels pour une intégration quali/quant.* França, Actes JADT, 1993.

MOSCAROLA, J. La communication politique vue par l'analyse lexicale. *Cahier du GEREG*, Annecy-França, IUP/GEREG, Université de Savoie, n.32, 1994.

MOSCAROLA, J. Les actes de langage: protocoles d'enquêtes et analyse des données textuelles. COLLOQUE CONSENSUS EX-MACHINA. Paris, La Sorbonne, Avr. 1994.