

**Referência completa para citação:**

FREITAS (H.), CUNHA Jr. (M. V. M.) e MOSCAROLA (J.). Pelo resgate de alguns princípios da análise de conteúdo: aplicação prática qualitativa em marketing. Angra dos Reis - RJ: Anais do 20ºENANPAD, ANPAD, Marketing, 23 - 25 de Setembro 1996, p. 467 - 487.

---

**PELO RESGATE DE ALGUNS PRINCÍPIOS DA ANÁLISE DE CONTEÚDO:<sup>1</sup>  
APLICAÇÃO PRÁTICA QUALITATIVA EM MARKETING**

Henrique M. R. de Freitas<sup>2</sup>  
Marcus V. M. da Cunha Jr.<sup>3</sup>  
Jean Moscarola<sup>4</sup>

**Resumo**

Este artigo apresenta um resgate de alguns princípios e fundamentos da análise de conteúdo: partindo-se de noções gerais, enfatiza-se a sua aplicação e uso. Trata-se de contribuição ao embrião de literatura disponível em português sobre o assunto, seu mérito sendo o da organização e seleção dos tópicos abordados. São descritas as etapas técnicas da análise de conteúdo, destacando-se o seu valor como instrumento de pesquisa. Sete (7) níveis de análise lexical são apresentados de forma criativa e evolutiva, pressupondo-se o uso de ferramentas ou sistemas em microcomputador. Conclui-se abordando a questão da utilidade e do interesse da análise de conteúdo.

**Palavras-Chave**

Análise lexical - Análise de conteúdo - Pesquisa - Estatística - Análise qualitativa e quantitativa de dados - Suporte - Autonomia - Usuário Final - Sistema

**Abstract**

This paper presents a literature review of content analysis. At first, we have the concept and some general principles. After, we describe the use and application, the stages of this method to analyze a text, and the importance of this tool or technique to the research on social sciences.

**Key-Words**

Lexical analysis - Content analysis - Research - Statistics - Qualitative and quantitative data analysis - Support - Autonomy - End-user - System.

---

<sup>1</sup> Este estudo baseia-se em discussões e debates realizados quando da vinda ao Brasil do Professor Jean Moscarola (Annecy, França), em Dezembro de 1994, em parceria de pesquisa do PPGA/UFRGS com a Université de Savoie, com o apoio do CNPq e da Fapergs; e aprofundado na visita do Professor Henrique Freitas à França, em Janeiro de 1995, com apoio do CNPq; além de diversos contatos via Internet e de um intenso trabalho de cooperação realizado em Fev/Março de 1996, em nova visita do Prof. Moscarola à equipe do GESID no Brasil. O sistema de pesquisa e análise de dados Sphinx Plus®, de autoria do professor francês e sendo aperfeiçoado com o apoio da equipe gaúcha, serviu de suporte à realização da análise de conteúdo neste estudo. Este texto foi produzido com o apoio da equipe de pesquisa do GESID (Grupo de Estudos em Sistemas de Informações e de Apoio à Decisão) do PPGA/UFRGS (Porto Alegre/RS). A bolsista IC CNPq do GESID, Ângela Werner, colaborou na edição final deste texto.

<sup>2</sup> Professor-adjunto do PPGA/UFRGS; Doutor em Gestão (Grenoble, França), Coordenador do GESID-PPGA/UFRGS, Pesquisador CNPq

<sup>3</sup> Mestrando em Administração, PPGA/UFRGS.

<sup>4</sup> Directeur de l'IUP, Université de Savoie, Annecy (França), Doutor em gestão pela Université de Paris, Autor do Sphinx

## 1. Introdução

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

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

As noções gerais a seguir apresentadas são baseadas principalmente nos textos de GRAWITZ (1976) e de GAVARD-PERRET & MOSCAROLA (1995).

Na seqüência, são apresentadas noções gerais sobre a análise de conteúdo (seção 2), dando-se ênfase na seção 3 para a sua aplicação e uso. Na seção 4 são apresentadas as etapas técnicas da análise de conteúdo, a seção 5 destacando o seu valor como instrumento de pesquisa. Sete (7) níveis de análise lexical são apresentados de forma criativa e evolutiva na seção 6, pressupondo-se o uso de ferramentas ou sistemas em microcomputador. Inspirando-se em dados qualitativos coletados em pesquisa, busca-se - seção 7 - ilustrar algumas das técnicas de análise aqui enfatizadas. O artigo é concluído com a seção 8, abordando a questão da utilidade e do interesse da análise de conteúdo.

## 2. Noções gerais sobre a ANÁLISE DE CONTEÚDO

Nesta seção, descreve-se a definição de análise de conteúdo, bem como as suas qualidades, as suas fontes, os tipos de análise, finalizando com as tendências.

### 2.1 - Definição de Análise de Conteúdo

Os recursos que as ciências sociais oferecem à nossa reflexão são essencialmente compostos de comunicações; que se trate de comunicações orais (textos de discursos, anotações e relatórios de entrevistas, conversações) ou escritas (textos oficiais, artigos de jornais, cartas, tudo isto tanto na área literária, quanto histórica ou política). Qualquer que seja o nível que se deseje atingir e o objeto de nossas pesquisas, os dados a reunir para compreender, explicar opiniões, condutas, ações, enfim, estes dados são quase sempre de origem *verbal*. A ação, quando nós a apreendemos, apresenta-se em um contexto de *palavras*... sempre encontramos falas ou alocações e escritos ou textos, ou seja, documentações em palavras.

É importante que o pesquisador em ciências sociais possa, de forma científica, analisar estes materiais ou fontes e não se contente em ter uma simples impressão. A novidade do procedimento moderno da análise de conteúdo consiste em substituir o "impressionismo" (jogo de opinião e muito pessoal) por procedimentos mais padronizados, tendendo a quantificar, em todo caso 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 ele contém, estas últimas sendo escolhidas em razão de sua relação com o objetivo de nossa pesquisa ou questão investigada (por exemplo: "quantas vezes tal partido utilizou o termo 'justiça social' durante a campanha eleitoral?").

BERELSON (in: GRAWITZ, 1976, pág.588) define assim a **análise de conteúdo**: "**é 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".

### 2.2 - Qualidades: objetiva, sistemática e quantitativa

- **Objetiva:** a análise deve proceder segundo as regras pré-estabelecidas, 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 que eles concordaram 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 frequência, etc. Esta condição não é indispensável visto que certas análises de cunho qualitativo são buscadas mais os temas do que a sua exata medida ou importância. Não são muitas as ferramentas que disponibilizam recursos de análise de conteúdo, associando o qualitativo ao quantitativo. No caso específico deste artigo, foi utilizado o sistema Sphinx® que oferece diversas maneiras de se navegar o texto tentando seja identificar temas seja definir novas variáveis oriundas de critérios subjetivos, as quais podem ser tratadas quantitativamente e cruzadas com outras mais objetivas, possibilitando maior riqueza e precisão nas inferências e conclusões. Também disponibiliza ferramenta para a codificação de textos, através da qual pode-se identificar dimensões de análise e mesmo tipologias a partir da base de texto original.

### 2.3 - 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, TV, etc), ou ainda em outras, atividades podendo ser decompostas, como a análise de uma reunião, da gestão e uso do tempo de um profissional qualquer;
- todo material especialmente gerado para a pesquisa psico-social: discussão de grupos, entrevistas de profundidade e relatos de reuniões, etc.

### 2.4 - Tipos de análise de conteúdo

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

Deve-se fazer uma clara distinção, própria às pesquisas em ciências sociais: a distinção entre a análise de documentos tendo 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 consiste em explorar, buscando a definição de hipóteses, a qual não se pode padronizar visto que ela faz apelo à intuição e à experiência.

Quando desejamos classificar certos elementos significativos em um texto ou documento, buscando uma quantificação, deve-se seja saber o que se está tentando elaborar, seja obedecer à sua 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, não sem ter ciência do risco de deixar de lado elementos essenciais (e imprevistos) de fora do escopo do estudo ou análise.

#### 2.4.2 - Análise quantitativa, análise qualitativa

Uma diferença é essencial, segundo GEORGE (in: GRAWITZ, 1976, pág. 592): a análise qualitativa se baseia na presença ou ausência de uma dada característica, enquanto a quantitativa busca identificar a frequência dos temas, palavras ou símbolos considerados. Outras oposições qualitativo-quantitativo: impressionismo versus sistematização, hipótese ao invés de verificação, e flexibilidade e liberdade contra rigidez.

A noção de importância deve ser clara em cada uma destas análises: o que aparece seguido, o "número de vezes" é o que importa para a análise quantitativa, 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 regрупar deliberadamente os dados num pequeno

número de categorias, não sem sacrificar informação talvez essencial, a qual ficará perdida no resultado final.

#### 2.4.3 - 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; já 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. Este tipo de informação (escolha das palavras, ritmo do discurso, etc) pode ser muito reveladora. 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 é algo amparado apenas no qualitativo, ela pode perfeitamente se apoiar num conteúdo quantificado.

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

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

#### 2.4.4 - Tendências...

A técnica de análise de conteúdo ganhou flexibilidade e adaptou-se a **objetivos** variados e a **etapas** diferentes da 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 persegue o jornal e da clientela à qual ele se destina.

**Etapas diferentes:** por falta de hipótese a priori e de objetivo preciso de medida de tal ou tal fator, o analista começará por se familiarizar com a sua documentação de base, relendo-a diversas vezes. Esta 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, mas sim se completam:** a análise qualitativa fornece inicialmente as categorias necessárias a uma análise de conteúdo quantitativa e rigorosa.

### 3. A aplicação e uso da análise de conteúdo

#### 3.1 - Estudo da comunicação no contexto da relação emissor-receptor

Qualquer que seja a comunicação, estaremos em 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? Este é um esquema simples proposto por LASSWELL (in: GRAWITZ, 1976, pág. 598) para os estudos da comunicação, o qual parece útil para a análise de conteúdo, as questões propostas recobrimo o conjunto dos problemas.

**Quem fala?...** ou estudo do emissor: duas situações - (1) o emissor reage a um estímulo mais ou menos controlado pelo observador, como numa entrevista dirigida; ou, caso mais freqüente, o analista não tem nenhuma relação com o emissor e busca com base no conteúdo as informações emitidas, podendo mesmo por vezes buscar identificar o autor ou 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 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 os relatos de uma reunião ou discussão em grupo, possibilitando uma sistematização.

**A quem?...** ou estudo do receptor: determina-se quem é o receptor, a quem se dirige a mensagem; podendo ser interessante identificar mesmo o que é latente e não somente explícito claramente (o estudo de uma mensagem de tal dirigente se endereça à sua Nação, mas de fato ele deseja passar uma mensagem 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 esta 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 deseja-se, com os elementos certos e da melhor forma, influenciar um dado público num certo sentido; ou ainda, caso as estratégias do emissor sejam claras, prever o que ele busca e seus objetivos implícitos. Ou seja, por vezes é explicativo ou elucidativo, por vezes é mais preditivo (ou indicativo).

### 3.2 - 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 'trivial'), levaram a considerar na análise, além das questões acima, a influência do ambiente, tanto sobre o emissor quanto sobre o receptor, para medir o efeito de uma comunicação particular. A análise de conteúdo pode levantar ou mesmo resolver grandes problemas na área da sociologia e da psicologia social (GRAWITZ, 1976, pág. 603).

## 4. Etapas técnicas da análise de conteúdo

Como uma pesquisa, a primeira parte está associada à idéia do que se deseja buscar ou investigar e no exercício de precisar seu objetivo. O meio de se obter uma resposta, é colocar uma questão. Numa 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.

### 4.1 - 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 que ignoramos no início dos trabalhos. As categorias, quando não se tem idéia precisa, devem surgir a com base no conteúdo. Na enquete de exploração as categorias constituem elas próprias o quadro, objeto ou contexto de análise (permitindo selecionar os dados no texto); enquanto que na busca de verificação de uma hipótese pode-se ter categorias pré-definidas. GRAWITZ (1976, pág. 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 se originar seja do documento objeto da análise, seja de um certo conhecimento geral da área do conhecimento ou de atividade no qual ele se insere. Das respostas, caso se trate de uma entrevista; dos objetivos, intenções, crenças do emissor, no caso de um texto; não sem considerar 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 legitimidade das categorias de análise. É o objetivo perseguido que deverá pautar a escolha ou definição do que deve ser quantificado. As categorias devem ser exaustivas (percorrer todo 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 com os objetivos perseguidos e com o conteúdo tratado).

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

**Excessos a evitar:** (1) impor um esquema de análise muito rígido desde o início, o que impedirá a apreensão da complexidade do conteúdo; (2) elaborar o esquema de forma superficial, centrando-se no explícito e esquecendo o conteúdo latente; (3) escolher categorias muito detalhadas e muito numerosas, reproduzindo quase que o texto pensando em "não perder nada", ou então (4) estabelecer categorias muito 'grosseiras', as quais não distinguirão suficientemente os elementos que elas regрупam. Deve-se estabelecer um nível de medida adequado.

**Padronização:** seria interessante se dispor de um certo número de categorias aplicáveis a diversos casos (idade, sexo, profissão, religião, nacionalidade, nível de instrução, são alguns exemplos). Mas, de resto, não é fácil se adotar um padrão. Certas áreas se organizam para tal, como os pesquisadores de/em marketing (organizados na AMA - American Marketing Association ou na ESOMAR - Associação Européia com a mesma finalidade, por exemplo). Para serem satisfatórios, os esquemas de análise devem convir ou respeitar um sistema conceitual geral aceito por todos, mas também a um conteúdo específico, o qual muda com cada nova pesquisa.

#### 4.2- Segunda etapa: os problemas oriundos da quantificação do conteúdo

A dificuldade de análise é associada à ambigüidade, à variedade e à complexidade das fontes...

**O problema da amostra:** este problema não existe nas ocasiões em que estudamos um texto na sua totalidade ou uma série de discursos, mas quando se trata de algo vasto como vários anos de um jornal, ou muito variados como as propagandas, deve-se forçosamente delimitar a pesquisa, escolhendo um certo tipo, uma certa rubrica, um certo setor (somente os editoriais de um jornal, por exemplo). Ou então pode-se, ao invés de limitar o tema, limitar uma amostra que seja representativa do todo.

Regras para constituir uma amostra: (1) delimitar o universo ao qual se aplica a generalização; (2) assegurar-se que cada unidade deste universo tem uma probabilidade conhecida de estar contida na amostra; (3) proceder a uma amostragem que seja independente de toda correlação entre as unidades do universo; (4) escolher uma amostragem relativamente ampla para que comporte somente um reduzido provável erro de amostragem.

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

Talvez pudéssemos defini-la em função das hipóteses formuladas, mas seria pouco científico e não compreenderia todos os fatores em jogo. BERELSON (in: GRAWITZ, 1976, pág. 613) nota que o problema de amostragem implica **3 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, ele deverá ser capaz de formular as razões que o levaram a definir um dado universo de conteúdo); número e data das mensagens (qual a quantidade de mensagens a reter para que seja representativa a amostra? qual o período a considerar? o tema pode determinar a época, mas não o período, atenção... a unidade escolhida do universo a estudar deve ter probabilidade conhecida de estar contida na amostra e de ser independente de toda correlação entre as unidades deste universo); e conteúdo observado (a amostra da amostra, difícil de definir, dada a variedade de objetivos; o que nos conduz seguido a usar métodos bem 'flexibilizados', os quais aceitam num mesmo estudo tipos de amostra levantados a partir de fontes, períodos e conteúdos diferentes).

**Diversas unidades de quantificação:** desejando quantificar, deve-se escolher índices a reter para categorizar e decidir o tamanho (ou comprimento) dos elementos segundo os quais vai-se decompor o conteúdo; o que envolve alguns conceitos:

- **a unidade de registro:** é a primeira unidade de análise (e a mais estreita), de tamanho variável (**palavra**, menor unidade, pode-se contar quantas vezes tal palavra foi utilizada e em que contexto ela foi usada; **tema**, fragmento significativo correspondendo à idéia que recobre uma das categorias, é um elemento menos preciso que a palavra, pode ter 3 linhas ou 4 páginas, sendo difícil obter resultados de grande fidelidade devido à grande diferença de

percepção de um analista a outro para um mesmo conteúdo; **item**, é um conteúdo total, um livro, um filme, um discurso);

- **a unidade de contexto:** ela é mais sutil e flexível e não se ampara numa quantificação rigorosa; é o mais amplo segmento de conteúdo ao qual nos referimos quando desejamos compreender a unidade de registro; por exemplo, caso a unidade de registro seja a palavra, a unidade de contexto pode ser a frase;

- **a unidade de numeração:** trata-se da forma pela qual se vai contar, a forma de se medir 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 estes 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 3 linhas e outro tratado 2 vezes em 100 linhas, é necessário ter uma unidade de numeração (o número de linhas) distinta da unidade de registro (número de vezes onde aparece o tema). As unidades de numeração se referem ao espaço ou ao tempo: parágrafos, linhas, centímetros, minutos de registro. Elas devem ser passíveis de medida objetiva, fiel e verificável, prestar-se a manipulações matemáticas (desde a adição).

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

## 5. 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 criatividade para escolher as categorias. Ao mesmo tempo deve ter disciplina e perseverança, rigor ao decompor um conteúdo ou contabilizar resultados ou análises.

**A 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 de ponto de vista entre os analistas. Mas este é um problema antigo nas ciências sociais: *vale mais à pena uma medida exata ou dados ricos?* A fidelidade não pode então ser encarada da mesma forma segundo se trate de uma análise quantitativa de um conteúdo claro e manifesto ou segundo se trate de uma análise de tipo mais qualitativo onde se busca identificar intenções latentes, onde vale mais a presença ou ausência de um elemento e não a sua frequência. Neste caso, vai-se ser conduzido a minimizar a questão da fidelidade. Já na análise quantitativa, o analista vai tratar a princípio da frequência dos elementos, não se importando tanto com a validade de tal ou tal dado isolado.

**A validade lógica:** será que o instrumento mede o que ele se propõe a medir? Uma análise é válida quando a descrição quantificada que ela oferece do conteúdo (objeto do estudo) é significativa para o problema originalmente colocado e reproduz fielmente a realidade dos fatos que ela traduz. Naturalmente, é condição essencial de partida a representatividade da amostra e que as condições técnicas inerentes a cada etapa sejam satisfatoriamente observadas. 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.

**A inferência:** atenção especial, pois por vezes uma mesma expressão tem dupla interpretação, ou então interpretação negativa ou positiva, de acordo com seu contexto no conteúdo analisado.

**A validade empírica e a medida da validade:** a predição inerente é justa ou precisa? Questão de difícil resposta... Ao invés de convicções, indicações... recomenda-se prudência e humildade nas conclusões. Entretanto, a experiência e capacitação do analista referendarão o valor de suas análises, acreditamos.

## 6. Colocando em evidência certas potencialidades, aplicações e limites das ferramentas de análise de dados textuais

Nosso 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)

envolvido num debate. O trabalho sobre um texto organizado ("lematizado") permite utilizar categorias gramaticais: substantivo, verbo, adjetivo, de forma a organizar a primeira apreensão ou impressão do conteúdo do corpo do texto.

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

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

Estes métodos são de fácil emprego graças ao progresso dos sistemas ("software") e das técnicas de análise, de navegação e estatísticas lexicais. A grande variedade de vias de investigação e de novas leituras que estes sistemas aplicativos colocam à disposição dos pesquisadores não deve entretanto mascarar ou esconder a dificuldade da interpretação final inerente a esta atividade. A dedicação e a paciência, aliados à curiosidade, devem ser características marcantes do pesquisador neste 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 estas abordagens num mesmo ambiente (ver funções principais no *Anexo*) e autoriza a qualquer instante o retorno ao texto e sua leitura, seu uso permitindo distinguir 7 (sete) níveis de análise, a seguir descritos e ilustrados sumariamente com base em MOSCAROLA (1995, pág. 7-8). De toda forma, independente do sistema a ser utilizado, o princípio e a técnica de exploração dos dados qualitativos disponíveis, os esquemas a seguir apresentados poderão contribuir para uma melhor estratégia de análise.

#### **A aproximação lexical sumária**

**Nível 1:** reduzir o texto ao alto do léxico, o exame das palavras mais frequentes dá uma idéia do seu conteúdo.

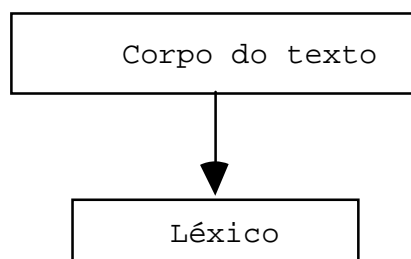


FIGURA 1 - Representação da aproximação lexical sumária.

#### **A 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.



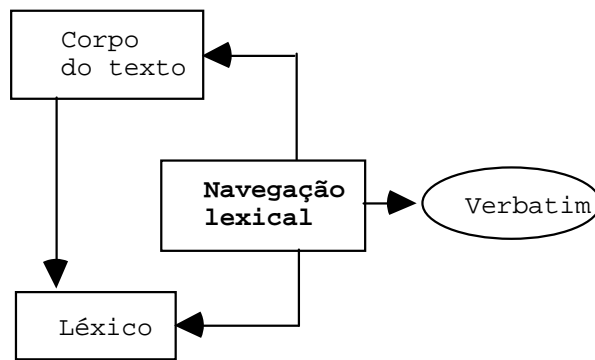


FIGURA 2 - Representação da A aproximação lexical controlada

### A aproximação lexical seletiva

**Nível 3:** trabalhar sobre um léxico reduzido após ter eliminado as palavras-ferramentas (definição daquelas palavras que pedimos "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.

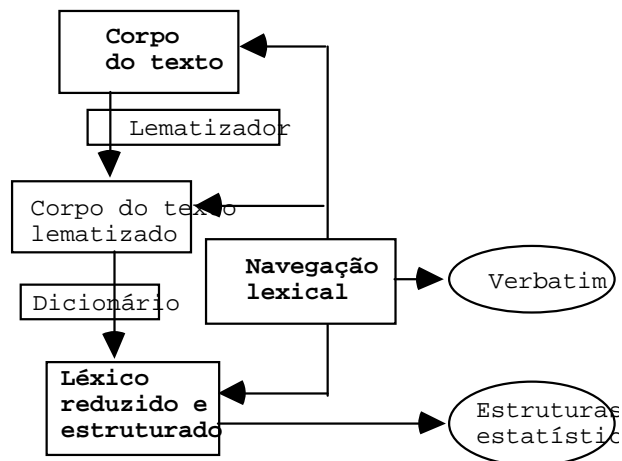


FIGURA 3 - Representação A aproximação lexical seletiva

### Estatísticas lexicais e quantificação do texto

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

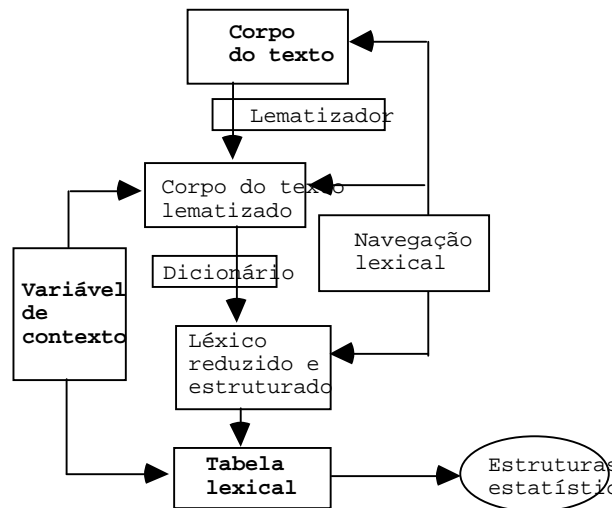


FIGURA 4 - Representação pesquisa das especificidades lexicais

**Nível 5: codificação lexical e medidas** - descrever o texto por uma variável nominal codificando a presença nas respostas de um conjunto de palavras selecionadas no léxico.

**Nível 6: a 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 nos interessam, por procedimento ou "cliques"). Cada resposta é descrita pela intensidade segundo a qual se manifestam os diferentes campos estudados.

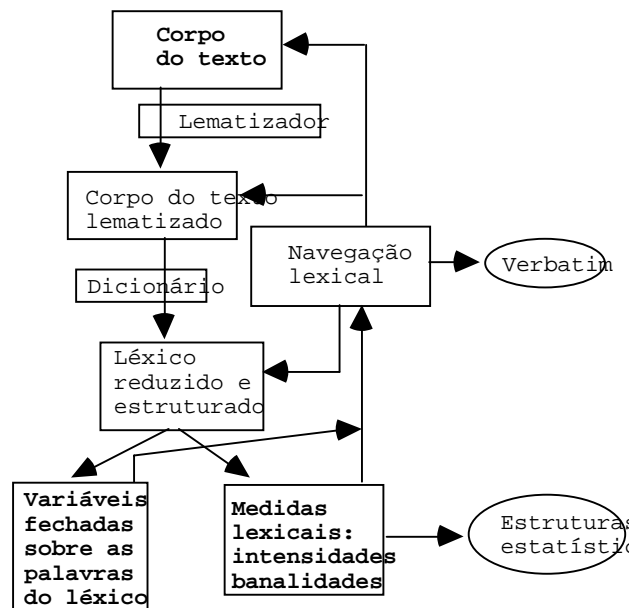


FIGURA 5 - Representação da codificação lexical e medidas e da quantificação do texto

#### A análise multivariada dos dados textuais

**Nível 7: a 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 com outras variáveis de contexto. Integra-se assim a análise de dados clássica com a análise de dados textual.

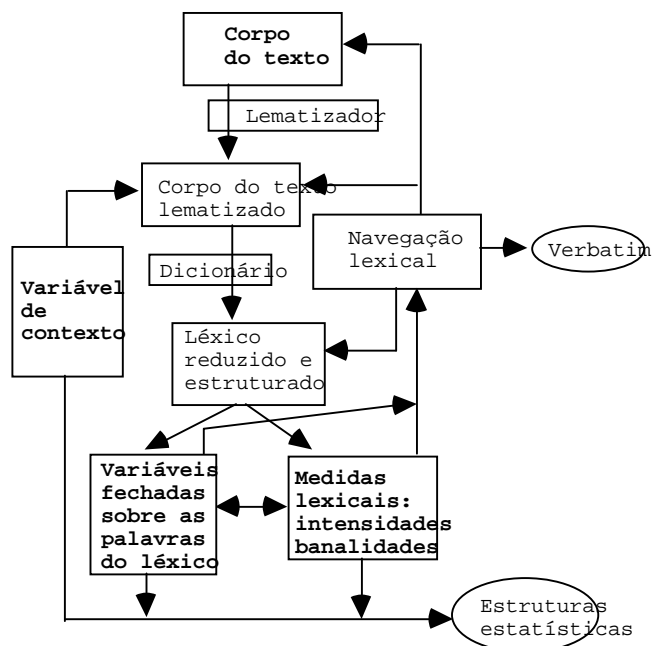


FIGURA 6 - Representação da análise de dados efetuada sobre as variáveis de origem lexical

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 não faz nada além de "aflorar" a questão do senso. Mesmo se ela pode permitir a redução do "ruído" de linguagem, a economia das longas e às vezes fastidiosas leituras, a produção bem mais rápida de conclusões e relatórios, ela conduz todos à uma leitura interpretativa um tanto quanto perigosa visto que ela é rápida e risca de ser falsamente objetiva.

## 7. Aplicação prática da análise qualitativa na área de marketing

Para ilustrar alguns dos níveis de análise de conteúdo descritos na seção 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, PIRES, FREITAS, CUNHA & ANTUNES, 1994). A pesquisa, que tinha como objetivo a avaliação dos Programas de Qualidade na indústria gaúcha, durou 4 meses compreendendo 120 empresas (30 pequenas, 30 médias e 60 grandes), com entrevistas com o diretor industrial e com o diretor da qualidade ou de recursos humanos.

A questão submetida à análise de conteúdo é relativa ao monitoramento dos ambiente em que a organização se insere. O executivo industrial era questionado sobre "quais 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**. Neste nível, o corpo do texto (formado por todas as respostas emitidas pelos 120 entrevistados), constituído de 374 palavras sendo 184 diferentes (constituindo o 'léxico'), é analisado via 'navegação lexical'. O objetivo é ter convicção do contexto no qual se inserem cada uma das palavras utilizadas em cada frase emitida. Por exemplo, a palavra 'preço' e a palavra 'preços' apareceram 19 vezes. Para compreender em que sentido esta palavra foi tão usada, 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 estas palavras em uma única que expressasse a sua idéia ('preços', no caso).

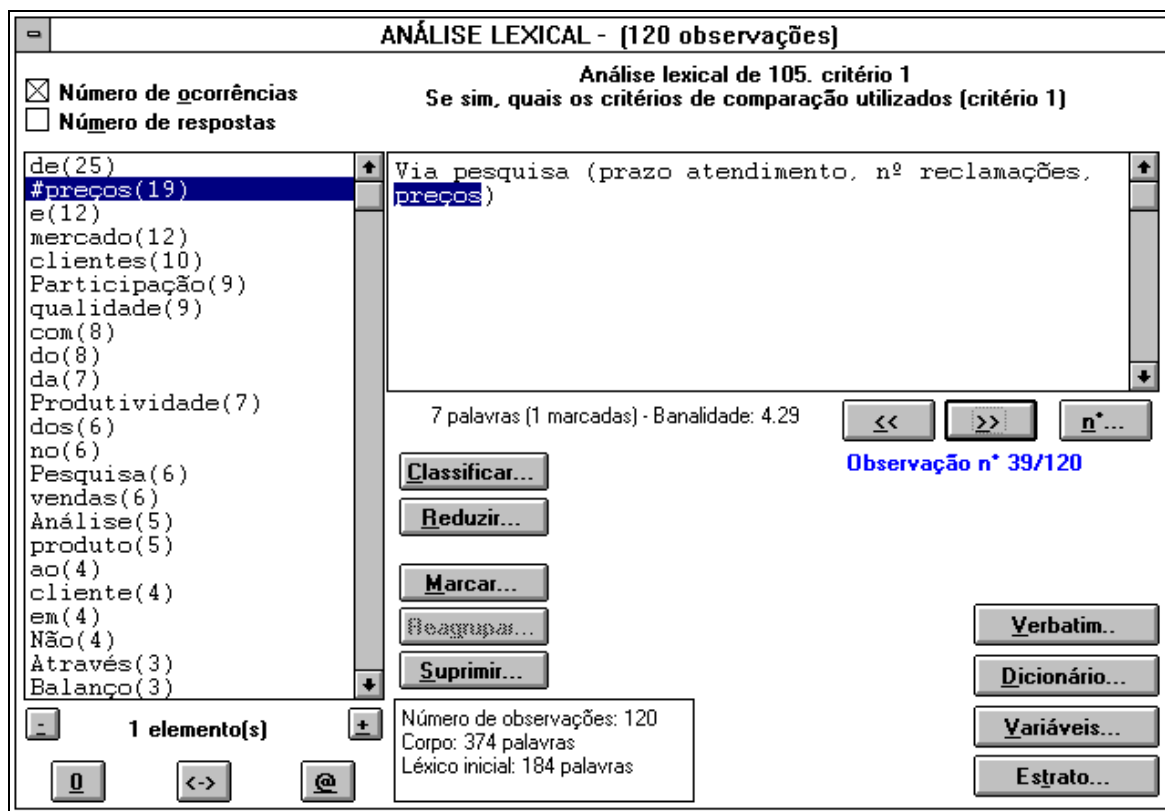


FIGURA 7 - Tela de navegação lexical. (Sphinx Plus®)

Ao realizar este 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**. Neste caso, busca-se enfocar a investigação em um número limitado de argumentos, excluindo-se, via ferramental adequado, as palavras que não contribuem para a análise em curso. Pode-se, neste Nível, marcar uma determinada palavra (ou conjunto) para se obter um 'verbatim' (qual seja o conjunto de frases dos respondentes que contenham as palavras selecionadas para análise). Por exemplo, caso marcássemos no 'léxico' a palavra 'vendedores' (reagrupamento de 'vendedores' e de 'representantes', no caso de nossa análise) e solicitássemos o 'verbatim', o resultado seria:

**# 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: -Sit. mercado e relacionamento c/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

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

#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)	

Diversos tratamentos podem se derivar desta lista de palavras-chave, importantes para a análise. Ilustraremos a seguir, o uso destas 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 de quantificação do texto, permitem a busca da compreensão das opiniões (variável que se deseja analisar) pelo simples cruzamento com uma variável externa (consideremos por exemplo o porte das organizações envolvidas: pequeno, médio e grande).

Antes disto, foi necessário criar uma variável de múltipla resposta, denominada, neste caso, de ‘critério de avaliação da concorrência’, e cujas modalidades de resposta eram as palavras da lista acima, cada qual com a identificação se SIM ou NÃO estavam presentes na resposta de cada um dos entrevistados. Feito isto, realizou-se o cruzamento bivariado entre porte e critério de avaliação da concorrência, optando-se de imediato pela representação em uma carta fatorial (AFC - Análise Fatorial das Correspondências; CUNHA Jr., FREITAS e SLONGO, 1995), visto que nosso intuito é demonstrar o interesse que se pode ter em associar técnicas qualitativas com quantitativas.

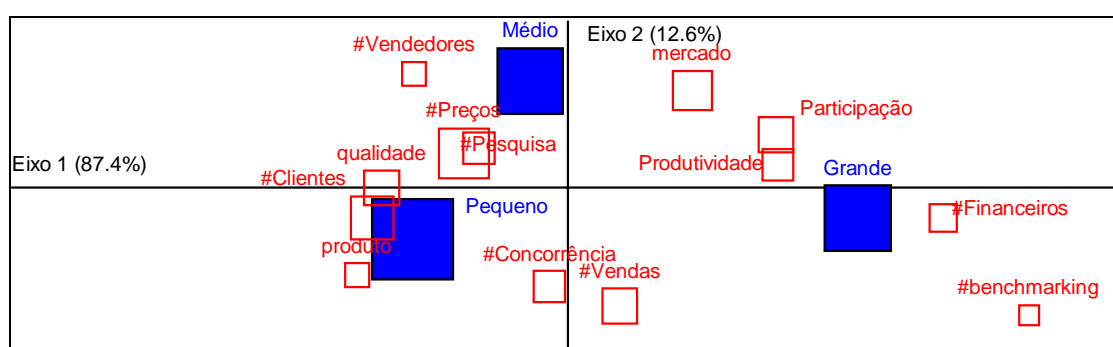


FIGURA 8 - Análise de Correspondência do cruzamento envolvendo o porte e os critérios de monitoramento do mercado.

Da análise da figura 8 depreende-se que: as empresas de grande porte utilizam-se de indicadores mais facilmente quantificáveis (produtividade, participação, financeiros,...), enquanto que as empresas de pequeno e as de médio porte utilizam-se de fontes mais relacionadas a fatores qualitativos (clientes, vendedores, concorrentes,...). Todas as análises baseadas nos textos foram realizadas sem uma prévia leitura do mesmo, demonstrando o ganho de produtividade proporcionado pela ferramenta adequada para a análise.

Os níveis 6 e 7, de estatísticas lexicais, de quantificação do texto e principalmente de análise multivariada, permitem com maior profundidade a busca da 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 se retirou 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) verificam-se 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 o número de repetições da palavra selecionada nesta mesma resposta, sendo a intensidade lexical a relação *repetição da palavra/comprimento da resposta*.

Consideremos por exemplo o ramo de atividade das organizações envolvidas: metalurgia, química, bebidas, etc. A análise da figura 9 permite-nos verificar quais os principais indicadores de avaliação do mercado utilizado nos ramos envolvidos. Os círculos permitem visualizar os agrupamentos existentes.

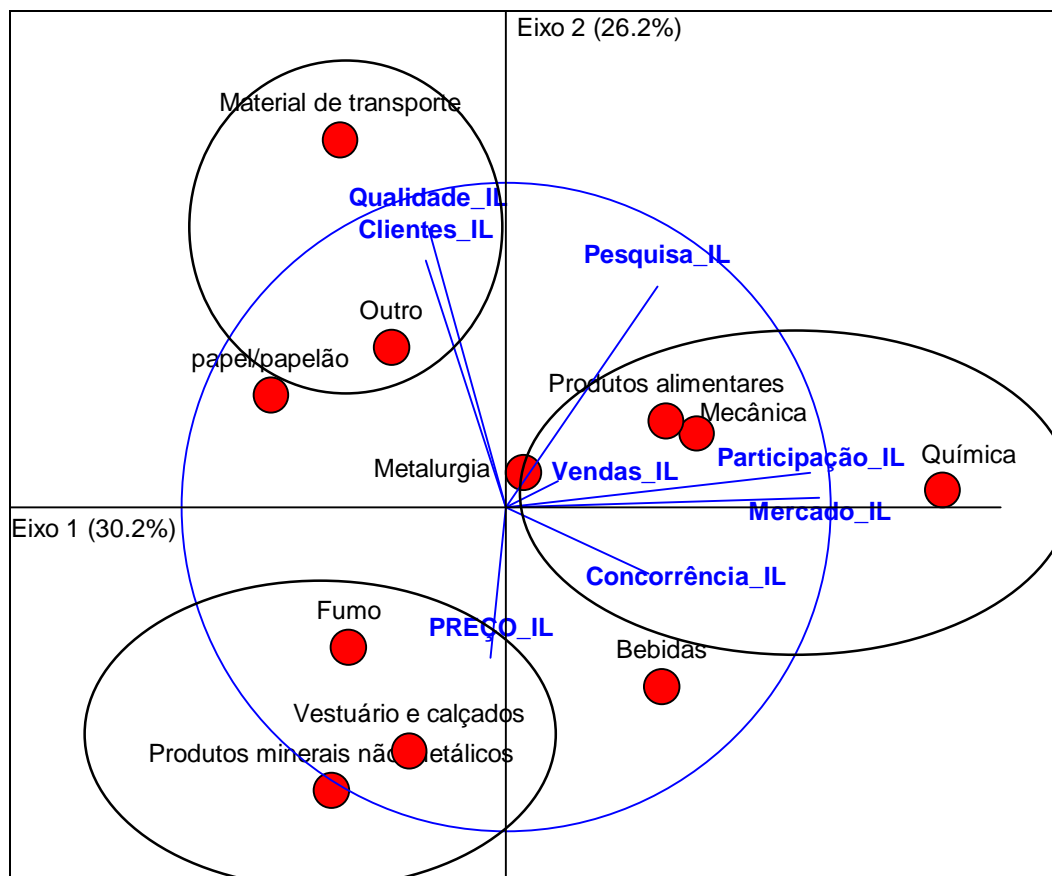


FIGURA 9 - Análise em Componentes Principais envolvendo Ramo de atividade da empresa e a intensidade lexical das palavras selecionadas.

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

## 8. 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 (in: GRAWITZ, 1976, pág. 623), ela deve se aplicar a todos os casos em que se busca atingir grande precisão e objetividade. As comparações e as evoluções formam a área de predileção da análise de conteúdo. Quando suficientemente definidas e detalhadas, elas 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 a qual ela se traduzirá 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 e elas devem corresponder ao material selecionado.

O uso de técnicas de análise QUALITATIVA sendo tema latente na comunidade científica brasileira, espera-se com este documento ter contribuído para resgatar seus conceitos e ilustrar algumas de suas potenciais aplicações.

## Referências bibliográficas:

- BACHELET (C.) e MOSCAROLA (J.).- La communication interne: diagnostic par l'analyse du discours des acteurs.- Annecy/França: IUP/GEREG, Université de Savoie, **Cahier du GERE**G, 1995, 10 pág.
- CUNHA Jr. (M.V.M.), FREITAS (H.) e 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.- Porto Alegre/RS: **caderno de pesquisa Nro. 06/95 do PPGA/UFRGS** e João Pessoa/PB: **Anais do 19o. ENANPAD**, vol. 1, Nro. 5, Marketing, 25-27 Setembro de 1995, p.109-129
- 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. **Trabalho de conclusão de curso de graduação**. Porto Alegre: DCA, Curso de administração, FCE/UFRGS, Dez. 1995.- 98 pág.
- GAVARD-PERRET (M.L.), MOSCAROLA (J.).- De l'énoncé à l'énonciation: pour une relecture de l'analyse lexicale en marketing.- Annecy/França: IUP/GEREG e CERIAM, Université de Savoie, **Cahier du GERE**G, 1995, 34 pág.
- GRAWITZ (Madeleine).- **Méthodologie des Sciences Sociales**.- Paris: Dalloz, 1976 (cap. II, pág. 586-627)
- Le SPHINX LEXICA (v. 2.0) - versão francesa em BETA teste, **manual do usuário**.- Annecy/França: Le Sphinx Développement, 1995
- MOSCAROLA (J.).- Ballardur, Chirac, Jospin, les mots d'une campagne : quelques exemples d'analyse lexicale avec le Sphinx.- Annecy/França: IUP/GEREG, Université de Savoie, **workpaper**, 1995, 8 pág.
- MOSCAROLA (J.).- Analyse de contenu et analyse de données: solutions logiciels pour une intégration quali/quantitative.- França: **Actes JADT**, 1993
- MOSCAROLA (J.).- Les actes de langage: protocoles d'enquêtes et analyse des données textuelles.- Paris: **Colloque Consensus Ex-Machina**, La Sorbonne, Avril 1994
- MOSCAROLA (J.).- La communication politique vue par l'analyse lexicale.- Annecy/França: IUP/GEREG, Université de Savoie, **Cahier du GERE**G , n° 32, 1994
- 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**, Relatório executivo, Dez. 1994q.- 68 pág.

## **Anexo - ANÁLISE DE DADOS TEXTUAIS**

### **Ferramentas de uso geral para múltiplas aplicações**

#### **1. Devem permitir acesso, leitura e análise de textos de toda natureza!**

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

#### **2. Atender a objetivos variados**

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

**3. Contemplar diversas áreas de aplicação:** a análise de dados de enquetes e de todas as formas de entrevistas; a pesquisa lingüística, a lexicologia e a terminologia; os estudos e a criação literária, o ensino literário e a lingüística; a pesquisa documentária; a vigília tecnológica e concorrencial; a criação publicitária; o jornalismo; a análise de interações de um grupo de estudos ou de executivos; a aprendizagem a partir dos discursos dos dirigentes ou políticos; a compreensão de um processo de decisão coletivo.

#### **4. 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 sendo analisado (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 compondo o índice ou 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 a 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 toda outra unidade de senso) e restrição das variáveis de contexto em função da seleção (ou recorte).

Todas estas funções devem estar disponíveis num sistema em microcomputador, *for windows*®<sup>5</sup>, de forma amigável, flexível, para que o usuário ou analista tenha a autonomia e a tranqüilidade para exercer sua curiosidade e elaborar análises e inferências em profundidade.

---

<sup>5</sup> O sistema SPHINX LEXICA® *for windows*® é um produto que atende a todas as exigências e características acima abordadas e já disponível em versão francesa. Será lançado em versão inglesa e depois portuguesa e espanhola em 1997. O sistema Sphinx Plus ® de concepção de pesquisa e de análise estatística quantitativa e qualitativa. Para mais informações, favor contactar os autores (tel/fax 051-4773610).