

Análise lexical e análise de conteúdo: uma proposta de utilização conjugada

Lexical Analysis and Content Analysis: a Proposal for Conjugate Utilization

Adriano Roberto Afonso do Nascimento^{*, I}; Paulo Rogério Meira Menandro^{**, II}

^I Universidade Federal de Minas Gerais

^{II} Universidade Federal do Espírito Santo

[Endereço para correspondência](#)

RESUMO

O trabalho discute a utilização conjugada de um software desenvolvido para a análise textual e de um procedimento de análise de conteúdo clássica. O software ALCESTE apóia-se em cálculos efetuados sobre a co-ocorrência de palavras em segmentos de texto, buscando distinguir classes de palavras que representem formas distintas de discurso sobre o tópico de interesse da investigação. A análise de conteúdo também identifica a pluralidade temática presente num conjunto de textos, e pondera a freqüência desses temas no conjunto, permitindo, via comparação entre os elementos do corpus (palavras ou sentenças), a constituição de agrupamentos de elementos de significados mais próximos, viabilizando a formação de categorias mais gerais de conteúdo. O trabalho se vale de exemplos em que ambos os procedimentos foram efetuados a partir do um mesmo banco de dados textuais e discute possíveis limitações e vantagens em relação à utilização em separado de cada um deles.

Palavras-chave: Metodologia, análise lexical, Alceste, análise de conteúdo.

ABSTRACT

This work discusses the conjugate use of a textual analysis software and a classical content analysis procedure. The ALCESTE software is based on calculations performed on simultaneous occurrence of words in text segments, aiming to distinguish classes of words that represent

distinct forms of discourse on the topic in which the investigation is interested. The content analysis also identifies the thematic plurality present in a set of texts, and weighs the frequency of these themes in the set, allowing, by comparison between elements of the corpus (words or sentences), the constitution of groupings of elements of closest signification, thereby making possible the formation of more general content categories. This work uses examples in which both procedures were carried out upon the same textual data base and discusses possible limitations and advantages with respect to the separate use of each of them.

Keywords: Methodology, lexical analysis, Alceste, content analysis.

Análise Lexical e Análise de Conteúdo: uma proposta de utilização conjugada¹

O tratamento estatístico de dados provenientes de fontes tradicionalmente utilizadas em Ciências Humanas e Sociais, como material textual publicado ou resultante de entrevista, efetuado com o auxílio de softwares, tem se mostrado cada vez mais freqüente nos trabalhos em que o corpus a ser analisado é volumoso. Um ganho evidente de tal procedimento é a economia do tempo com tabulação e efetuação dos cálculos necessários ao tratamento do conjunto de informações coletadas. Esse "tempo poupado" pode ser utilizado na análise mais detida dos dados já organizados pelo computador ou na tarefa de confrontar o procedimento com outros no tratamento da mesma informação coletada, potencializando o entendimento do material de interesse.

O presente texto discute a possibilidade de uso conjugado de um programa computacional desenvolvido para a análise lexical (ALCESTE), que se vale de cálculos efetuados sobre co-ocorrências de palavras em segmentos de texto, com um procedimento de análise de conteúdo clássica, ambos sobre o mesmo banco de dados, avaliando possíveis avanços em relação à utilização em separado dos mesmos¹.

Antes de avaliação da possível interlocução de resultados provenientes desses dois procedimentos, é necessário caracterizá-los, ainda que de forma geral. Para exemplificações, tanto em Alceste como em Análise de Conteúdo, utilizamos resultados do trabalho de Nascimento (2004), que identificou e analisou os principais elementos discursivos em letras de canções populares disponíveis a partir de 1927 sobre a saudade da infância.

ALCESTE (*Analyse Lexicale par Context d'un Ensemble de Segments de Texte*)²

Desenvolvido por Max Reinert (1998), esse software conjuga uma série de procedimentos estatísticos aplicados a bancos de dados textuais, como entrevistas, obras literárias, artigos de jornais e revistas, entre outros. Na base do funcionamento do programa encontra-se a idéia de relação entre contexto lingüístico e representação coletiva ou entre unidade de contexto e contexto típico (REINERT, 1990). Unidade de contexto é entendida como uma espécie de representação elementar, um sentido ou um enunciado mínimo em um discurso. Tal enunciado é

postulado como uma idéia de um indivíduo psíquico, sendo que essa idéia se refere a um objeto, mas também, ao mesmo tempo, ao próprio sujeito. É essa dupla referência que permite a formação da representação de um objeto. Por extensão, a regularidade de representações entre indivíduos pode promover a existência de um determinado contexto típico de um grupo, uma representação coletiva, um "mundo". Segundo Reinert (1990, p.32):

As representações locais, múltiplas, imediatas, associadas aos enunciados de um corpus se organizam em função de leis específicas identificando tipos de 'mundos'. Seus traços na língua só podem ser revelados através de um grande número de enunciados semelhantes, de um determinado ponto de vista. A representação que podemos ter desses mundos através do corpus estudado pode ser freqüentemente arcaica [rudimentar], pois que ela não está diretamente anunciada no corpus e que nós a revelamos através de uma análise estatística bastante grosseira, por uma classificação.

A regularidade de um vocabulário específico indica a existência de certo "campo contextual", um espaço semântico específico. Reinert (2001) propõe a extensão desse espaço à idéia de "fundo associativo" ou "fundo tópico", revelado através da co-ocorrência das chamadas "palavras plenas", que devem ser entendidas como algo que excede os significados anotados nos dicionários, uma vez que se inscrevem na história dos falantes, pois são utilizadas como uma atualização do próprio sujeito e de seu campo de referência aos objetos num aqui e agora do discurso.

Observe-se, entretanto, que, em seu aspecto operacional, o objetivo do Alceste não é o cálculo do sentido, mas a organização tópica de um discurso ao colocar em evidência os 'mundos lexicais'. No Alceste, o vocabulário de um enunciado constitui um traço, uma referência, um 'funcionamento', enfim, uma intenção de sentido do sujeito-enunciador. Para esse método, que não esconde seu fundamento lexicométrico, os traços mais fortemente característicos desse 'funcionamento' do léxico são os 'mundos lexicais' (KALAMPALIKIS, 2003, p. 151).

Kalampalikis (2003) acrescenta que tais "mundos lexicais" só podem ser entendidos num esquema maior de relações: só há a possibilidade de significação de um "mundo lexical", se há um outro "mundo lexical" que o defina por contraste ou complementaridade. De forma mais direta,

o pressuposto de Alceste é que pontos diferentes de referência produzem diferentes maneiras de falar, isto é, o uso de um vocabulário específico é visto como uma fonte para detectar maneiras de pensar sobre um objeto. O objetivo de uma análise com Alceste, portanto, é distinguir classes de palavras que representam diferentes formas de discurso a respeito do tópico de interesse (KRONBERGER; WAGNER, 2002, p. 427).

Estão descritas a seguir algumas das principais operações realizadas nas quatro etapas do procedimento Alceste. Inicialmente, a definição de alguns termos (REINERT, 1998):

UCI: Unidade de Contexto Inicial - Unidade a partir da qual o programa efetuará a fragmentação inicial. Pode ser um capítulo de livro, uma matéria jornalística, uma resposta de entrevista. Nos exemplos utilizados no presente texto trata-se de uma letra de música.

UCE: Unidade de Contexto Elementar - Definida segundo critérios de tamanho do texto (número de palavras analisadas) e pontuação. "É a partir do pertencimento das palavras de um texto a uma UCE, que o programa Alceste vai estabelecer as matrizes a partir das quais será efetuado o trabalho de classificação" (REINERT, 1998, p. 17).

UC: Unidade de Contexto - Agrupamento de UCE sucessivas dentro de uma mesma UCI, até que o número de palavras diferentes analisadas (contidas nessa unidade de contexto) seja superior ao

limiar fixado na análise. (...) Esse limiar é fixado proporcionalmente ao número de palavras analisadas. As UC devem ter uma homogeneidade semântica e devem ser de tamanho comparável. Ou seja, são unidades estatísticas de pesos idênticos. Elas estão na base de todas as estatísticas efetuadas (REINERT, 1998, p. 18).

Classe: pode ser definida como um agrupamento constituído por várias UCE de vocabulário homogêneo.

“Lematização”: operação de substituição de certas palavras por uma forma reduzida, que permite ao programa considerar equivalentes palavras com radical comum, que significam aproximadamente a mesma coisa, mas que diferem quanto ao gênero, ao número ou ao fato de serem substantivos, adjetivos ou advérbios. Por exemplo: AMOR, AMORES, AMOROSOS, são substituídas no texto pela forma AMOR+.

AFC: Análise Fatorial de Correspondência. Cruzamento entre o vocabulário (considerando a frequência de incidência de palavras) e as classes, gerando uma representação gráfica em plano cartesiano, na qual são vistas as oposições entre classes ou formas.

Quatro etapas de procedimento compõem o método Alceste. A primeira (A) refere-se a uma “leitura” do texto e aos cálculos dos dicionários. O programa gera inicialmente a listagem em ordem alfabética de todo o vocabulário do corpus. Dessa lista deriva uma segunda, composta pelas formas reduzidas e palavras representadas por essas formas reduzidas, ou seja, um dicionário de formas reduzidas. Uma outra lista é produzida com as formas reduzidas mais frequentes. Adjetivo, substantivo, pronomes, artigos são algumas das chaves categoriais utilizadas pelo programa nessa etapa para a primeira classificação do vocabulário. Deve-se observar que o programa possui dicionários acoplados (inclusive de português), o que permite identificar diferentes conjugações verbais, bem como diferenciar palavras com função puramente sintática (pronomes, artigos, advérbios) daquelas com “conteúdo” (verbos, substantivos, adjetivos). Tal procedimento é relevante, pois é com essa segunda classe de palavras, sobretudo, que o programa vai trabalhar.

Na etapa B são selecionadas as formas reduzidas com frequência maior ou igual a 04. Nessa etapa, também, é que são definidas as UCE, o que se dá, como já foi visto, segundo o critério de tamanho do texto ou de pontuação (por exemplo, ponto final = fim de uma UCE). Nessa etapa o Alceste realiza o cálculo de três cruzamentos (Classificação Hierárquica Descendente - CHD): a) Todas as UCE x todas as formas reduzidas; b) UC tamanho 1 x formas reduzidas selecionadas; c) UC tamanho 2 x formas reduzidas selecionadas. É formada uma matriz para cada um desses cruzamentos, na qual os valores 0 e 1 indicam, respectivamente, ausência ou presença de determinada palavra em uma UCE ou UC. A lógica do uso de dois tamanhos de UC é esclarecida adiante.

O conjunto total de unidades contextuais na matriz de indicadores inicial (unidades contextuais por palavras) constitui a primeira classe. O objetivo do próximo passo é conseguir uma divisão dessa classe em duas, que separem, da maneira mais nítida possível, as classes resultantes, de tal modo que as duas classes contenham diferentes vocabulários e, no caso ideal, não contenham nenhuma palavra sobreposta (KRONBERGER; WAGNER, 2002, p. 429).

O procedimento é continuamente efetuado até que não resulte em novas classes. Na análise standard, os tamanhos diferenciados de UC são definidos pelo programa (podendo também ser alterados pelo pesquisador, assim como vários outros parâmetros do Alceste) e se referem ao encadeamento de UCE sucessivas até que o número de palavras selecionadas seja maior ou igual ao valor estipulado. O objetivo de cálculos com dois valores de UC é avaliar se o tamanho do texto a ser considerado pode interferir na formação das classes.

Na etapa C os valores obtidos nas duas últimas CHD são comparados, e apenas a parte "estável" dos resultados é considerada na determinação dos perfis de classe, definidos pelo Q^2 de associação das formas reduzidas às suas respectivas classes. Relacionada a essa etapa está a Análise Fatorial de Correspondência que efetua o cruzamento entre as formas reduzidas com frequência maior do que 08 e as classes formadas.

A etapa D destina-se aos cálculos complementares e nela são formadas as listas de formas reduzidas associadas a contextos, que correspondem às classes anteriormente formadas e possibilitam a identificação das UCE características de cada uma das classes. São identificados ainda os segmentos repetidos, isto é, os trechos de frases mais recorrentes em cada classe. Por fim, realiza-se a Classificação Hierárquica Ascendente que consiste no cruzamento entre as UCE das classes e as formas reduzidas características da mesma classe.

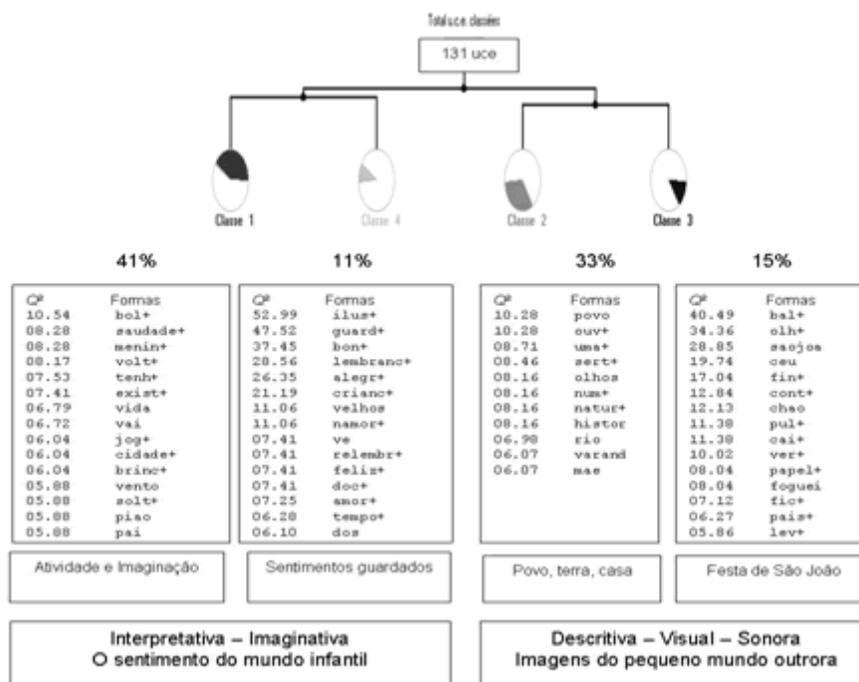
Para os resultados de algumas das operações, informações complementares: Segundo Reinert (2001, p. 34),

as classes obtidas [na análise] podem ser interpretadas de três pontos de vista: a) Como conteúdo, ou seja, observando-se a lista de palavras ou a lista de UCE que lhes são específicas; b) Como 'funcionamento', porque, por suas oposições, exprimem um certo dinamismo do percurso discursivo; c) Como representação, pois essas classes formam [um] sistema e refletem uma certa estabilização do 'funcionamento' do autor.

Esses pontos de vista não se excluem. Há em cada um deles elementos que permitem articular de forma mais geral os resultados do programa, a fim de que forneçam uma leitura mais abrangente e integrada do corpus.

A figura 01 mostra as formas reduzidas distribuídas por Classe. Nessa figura encontram-se: a) o número total de UCE classificadas (131); b) o dendrograma mostrando as relações entre as Classes. No dendrograma constante da figura 1 se pode constatar que os 02 blocos formados, respectivamente, pelas Classes 1 e 4 e pelas Classes 2 e 3 são independentes. Isso significa também dizer que a Classe 1 possui maior relação, ou proximidade, com a Classe 4 e pouca ou nenhuma com as Classes 2 e 3 ; c) a porcentagem de UCE em cada uma das Classes (observe-se que a Classe que mais agrupou UCE foi a Classe 1 – 41%); d) as formas reduzidas mais frequentes que formaram cada uma das classes, com o Q^2 de associação à sua respectiva Classe (note-se que, como exemplo, pode-se afirmar que a forma reduzida *ilus+* ($Q^2 = 52.99$) está mais fortemente ligada à Classe 4 do que a forma reduzida *bal+* ($Q^2 = 40.49$) à Classe 3); e) o nome/tema das Classes (nomes dados pelo pesquisador, analisando as listagens de formas reduzidas, as UCE características de cada Classe e a Análise Fatorial de Correspondência – serão descritas adiante tanto as UCE características quanto a AFC) e; f) os nomes/temas dos blocos formados pelas Classes 1 e 4 e pelas Classes 2 e 3, respectivamente (dados também pelo pesquisador).

Figura 01: Formas reduzidas distribuídas por Classe



O Quadro I mostra exemplos de UCE características. O programa gera uma listagem dessas UCE para cada uma das Classes, o que permite ver em que “contexto” aparece cada uma das formas reduzidas (em negrito) no interior dessas mesmas Classes. Também no Quadro I se pode ver o Q² de associação da UCE característica à sua respectiva Classe.

Quadro I: Exemplos de UCE características de cada classe.

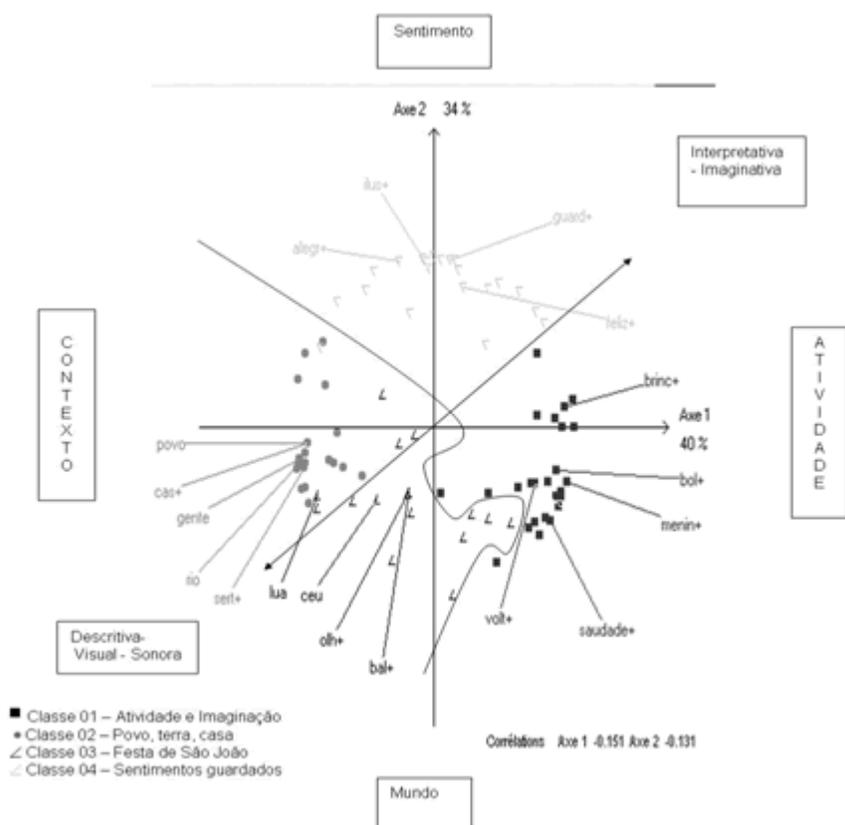
Classes	Q² de ligação à Classe	Unidade de Contexto Elementar
Classe 01	23	estou velho e a saudade da infância vai chegando e eu fico a recordar que soltava o meu barquinho na torrente e ele ia para nunca mais voltar .
Classe 02	24	quando eu sai de casa , me atropeli no espaço. Tudo era aço e todo inteiro que eu pensei era pedaço . Eu sei , eu via o mundo com os olhos presos numa segundo.
Classe 03	41	eu corria acelerado pra segurar a sua mão que, de macia, parecia papel fino de balão . E hoje, quando chega São João , eu vejo quanta coisa se acabou.
Classe 04	52	ilusão . Revendo antigos brinquedos dos velhos tempos de criança guardei você, meu boneco de palha, para sempre na minha lembrança .

* Em negritas palavras que contribuíram para a formação da Classe.

Como mencionado acima, a Análise Fatorial de Correspondência (AFC) permite verificar as relações entre as classes num plano gráfico, apontando, segundo a "localização" dessas classes, a sua interação. Ainda que apresente limitações, como a não contextualização (o "como" e o "porquê") da fala e a não consideração de que a ordem das sentenças altera o resultado final do sentido do discurso (LORENZI-CIOLDI, 1997), a AFC permite visualização geral articulada dos agrupamentos de vocábulos presentes no discurso.

Para a leitura da Figura 02 é preciso lembrar que a "lógica" da Análise Fatorial de Correspondência é de que, quanto mais distantes os elementos dispostos no plano, menos eles "falam" das mesmas coisas. É importante observar que a disposição de agrupamentos em pólos opostos no plano dos eixos não necessariamente indica relação de oposição semântica desses mesmos agrupamentos (como indicado acima, um mundo lexical só é possível a partir da existência de outros e essa relação entre os mundos também pode ser de complementaridade). Como resultado, o Alceste oferece os eixos, a disposição das formas reduzidas e Classes no plano e a informação sobre qual dos eixos compõe mais fortemente a disposição dos elementos (no caso, Eixo 1 - 40%). A partir desse resultado e da leitura dos resultados apresentados acima, é possível para o pesquisador nomear os eixos e seus respectivos pólos (no exemplo o Eixo 1 (x) é formado pelos pólos Contexto e Atividade e o Eixo 2 (y) pelos pólos Sentimento e Mundo). A partir da informação da existência de 02 blocos (Classes 1 e 4 e Classes 2 e 3) vistos na figura 01, é possível propor a existência na AFC de um terceiro "eixo" que apresente os pólos Interpretativa-Imaginativa e Descritiva-Visual-Sonora.

Figura 02: Análise Fatorial de Correspondência (Classes, formas reduzidas e eixos).



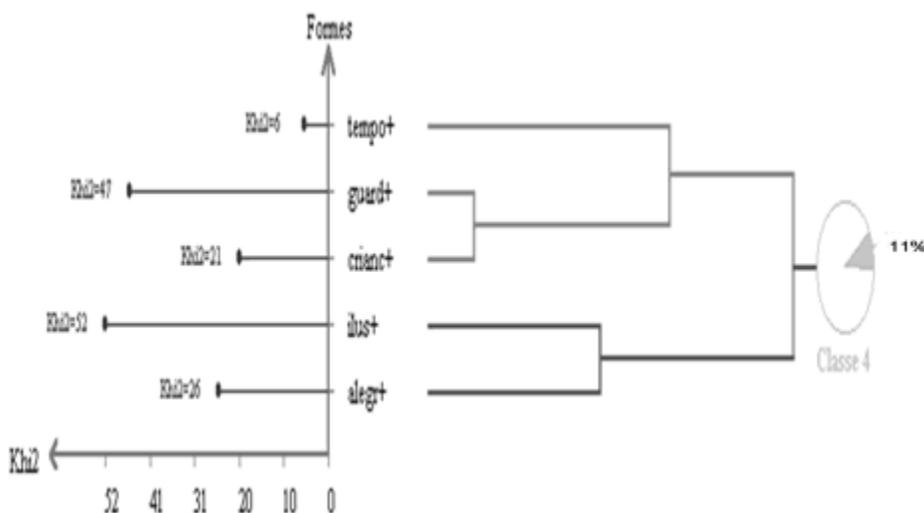
A Classificação Hierárquica Ascendente (CHA) realizada pelo Alceste cruzando as UCE da classe selecionada e as formas reduzidas específicas dessa mesma classe permite a visualização de

ESTUDOS E PESQUISAS EM PSICOLOGIA, UERJ, RJ, ANO 6, N. 2, 2º SEMESTRE DE 2006

“laços de vizinhança ou sinônimos” (REINERT, 1998, p. 33) indicativos de contextos ou “núcleos” que devem ser considerados como tal e que podem auxiliar tanto na leitura da formação das classes, quanto na avaliação da inter-relação entre elas.

Na Figura 03 pode-se ver o exemplo do Dendrograma da CHA da Classe 4. Há dois subgrupos de formas reduzidas que podem ser considerados centrais para a Classe 4: o primeiro é constituído pelas formas tempo+, guard+ e crian+; o segundo é constituído pelas formas ilus+ e alegr+. Também podem ser vistos os Q^2 de associação das formas reduzidas à Classe 4 e a porcentagem de UCE classificadas nessa mesma Classe (11%), conforme já visto na Figura 01. No trabalho utilizado para a extração dos exemplos (Nascimento, 2004), tal dendrograma sintetiza as idéias de infância como um tempo guardado ou do qual se guarda algo (primeiro subgrupo) e de alegria, ou como sinônimo, ou como indissociável de ilusão (segundo subgrupo). É ainda possível, na Figura, identificar que as formas guard+ e crian+ guardam uma maior proximidade entre si do que as formas ilus+ e alegr+, uma vez que quanto mais próximo do eixo das formas estiver o traço de ligação entre duas formas, mais recorrentemente essas mesmas formas aparecem juntas nas UCE classificadas (2).

Figura 03: Dendrograma CAH Classe 04



Análise de Conteúdo

A Análise de Conteúdo (BARDIN, 1977; BAUER, 2002; VALA, 2003) também pode ser caracterizada como procedimento de fragmentação de textos com o objetivo de identificar regularidades. Possibilita tratar todo o material textual, diferentemente do Alceste, que, no corte inicial, só analisa formas reduzidas com frequência maior ou igual a 04, o que pode representar um problema por descartar sinônimos menos utilizados, identificando-os como palavras pouco frequentes.

Buscando identificar a pluralidade temática presente num conjunto de textos, ao mesmo tempo em que pondera a frequência desses temas dentro do mesmo conjunto, a Análise de Conteúdo pode proporcionar, numa comparação entre os elementos do corpus (palavras ou sentenças), a constituição de agrupamentos de elementos de significados mais próximos, possibilitando a formação de categorias mais gerais de conteúdo.

Esse processo pode ser efetuado da seguinte forma:

1- Primeira leitura do corpus. Anotação de impressões gerais sobre possíveis elementos recorrentes nas informações coletadas; 2- Segunda leitura. Pré-categorização. Contagem parcial dos elementos identificados na primeira leitura, visando confirmação ou não de categorias gerais e identificação de possíveis novas categorias; 3- Terceira leitura. Construção de grades de categorias contendo tema geral, palavras e frases relacionadas a esse tema; 4- Agrupamento de elementos (palavras ou frases) da categoria geral em subcategorias; 5- Contagem da frequência final dos elementos nas subcategorias e nas categorias gerais.

A Análise de Conteúdo pode ser descritiva, mostrando, por exemplo, a variedade de elementos presente no conjunto de dados, ou analítica, procurando regularidades de avaliações nesses mesmos dados. Assim, uma categorização menos "rígida" dos elementos nas subcategorias pode estar diretamente relacionada às características específicas do corpus analisado, o que não impede ver, no interior das categorias gerais, outros possíveis agrupamentos de dados. Em situações nas quais algo está dito de duas ou mais maneiras em diferentes textos, deve-se manter essas diferentes versões. Na fala, em geral, também importa, em algo que é dito, o como foi dito. Nessa linha tênue entre categorias mais abertas e rigor de classificação, deve-se privilegiar, pelo menos num primeiro momento, o critério de se considerar somente o que explicitamente estiver representado nas respostas. Assim, nesse procedimento não há subentendidos, não há a ausência de determinada informação sobre um elemento considerada como dado em si.

As categorias gerais resultantes do procedimento apontam linhas gerais de temas e avaliações recorrentes nos dados analisados, incluindo identificação de elementos e sua avaliação. Por fim, chega-se a uma listagem de elementos significativos (Quadro II) que, entretanto, indica ainda de forma pouco clara as ligações entre as categorias formadas por esses mesmos elementos.

Quadro II: Categorias iniciais da Análise de Conteúdo

Categorias	Total de itens
Atividades	43
Brincadeiras, Jogos, Brinquedos	26
Elementos Naturais	23
Concepções de Infância	20
Elementos Musicais	16
Ontem	15
Hoje	12
Sociabilidade	11
Lugares	09
Lugares específicos	08
Vínculo com o lugar	08
Comidas/ Bebidas	07
Interlocutor	07
Conseqüências do passar do tempo	06
Elementos Religiosos	06
Cidades, Estados	05
Características do lugar	03
Festas	03
Atividade/ Profissão	03
Saudade & Coração	03
Elementos Amorosos	02
Ilusões / Sonhos	01
Ilusões Perdidas/ Sonhos Mortos	01
Ontem Feliz – Hoje Sofrer	01
Quem dera voltar	01
Não volta mais	01
Lembrar-Sofrer	01

Alceste & Análise De Conteúdo

Segundo Bauer (2002, p.194),

a análise de conteúdo pode reconstruir 'mapas de conhecimento' à medida que eles estão corporificados em textos. As pessoas usam a linguagem para representar o mundo como conhecimento e auto-conhecimento. Para reconstruir esse conhecimento, a AC pode necessitar ir além da classificação das unidades do texto, e orientar-se na construção de redes de unidades de análise para representar o conhecimento não apenas por elementos, mas também em suas relações.

Nessa direção, pode-se efetuar, na construção de uma estrutura inicial de rede, a junção de algumas categorias gerais, levando-se em conta aspectos confluentes na categorização dos dados apontados tanto pela análise do Alceste como pela Análise de Conteúdo, ambas realizadas antes dessa fase (Quadro III).

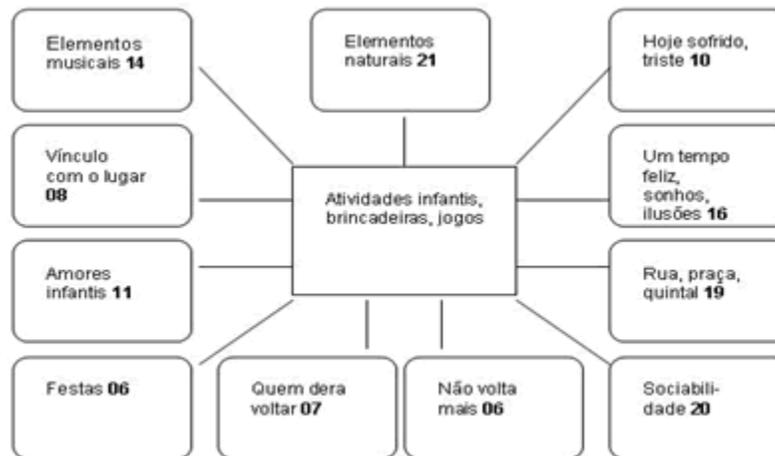
Quadro III: Agrupamento de categorias da Análise de Conteúdo para a formação dos nós das Redes de Conteúdos.

<ul style="list-style-type: none">* Atividades infantis (atin_) = atividades + brincadeiras, jogos, brinquedos (maior quantidade de elementos reunidos no Alceste na Classe 01 – Atividades);* Um tempo feliz, sonhos, ilusões (feli_) = concepções (positivas) de infância + sonhos + ilusões + ontem (maior quantidade de elementos reunidos no Alceste na Classe 04 – Sentimento);* Vínculo com o lugar (vinc_) = saudade também do lugar (ainda que indefinido) + vínculo com o lugar + "bairro onde nasci/ terra onde nasci" da categoria Lugares;* Hoje sofrido, triste (trist_) = hoje + ilusões perdidas/ sonhos mortos;* Rua, praça, quintal (rua_) = "Rua" + "Praça/ Largo" + "Quintal/ terreiro" da categoria Lugares;* Sociabilidade (soci_) = Sociabilidade – "primeiro amor/ meninas/ mocinhas";* Amores (amor_) = elementos amorosos + "primeiro amor/ meninas/ mocinhas" da categoria Sociabilidade;* Elementos musicais (elmu_) = elementos musicais;* Festas (fest_) = festas (maior quantidade de elementos reunidos no Alceste na Classe 03 – São João);* Quem dera voltar (volt_) = quem dera voltar;* Não volta mais (nvol_) = não volta mais;* Elementos naturais (elna_) = elementos naturais.

Utilizando as categorias mais freqüentes, provenientes da Análise de Conteúdo, é possível realizar nova leitura dos dados, desta vez identificando em quais categorias um mesmo item, no caso letra de música, teve elementos (palavras e sentenças) classificados. O resultado desse processo é a representação gráfica da rede de relações entre os elementos (por exemplo, Atividades infantis, brincadeiras, jogos) e a avaliação tanto do contexto relacionado à interação desses elementos (Um tempo feliz, sonhos, ilusões), quanto da situação na qual se dá a avaliação (Hoje sofrido, triste). Tal procedimento aponta, no mínimo, para uma caracterização geral do corpus como formado por textos eminentemente multi-referenciais. A figura 04 exemplifica um dos nós que compõem o que está sendo chamado de "Rede de Conteúdo".

Podem ser vistos na figura os 12 nós resultantes do Quadro III, bem como a freqüência de co-ocorrência (número de letras de músicas para um n = 70) de cada um dos 11 nós relacionados ao nó que é central - nessa figura o nó Atividades infantis, brincadeiras, jogos. Observe-se que, por se tratar de uma rede, o resultado gráfico final deveria comportar, no caso, os 12 nós ao mesmo tempo, assim como todos os seus valores de co-ocorrência. Por motivos práticos selecionou-se um nó como central para a apresentação do exemplo. Assim, no caso de opção pela apresentação de todos os nós, cada um deles faria a vez de nó central em outras figuras.

Figura 04: Nó "Atividades infantis, brincadeiras, jogos".



Outra possibilidade é utilizar os nós das Redes de Conteúdos como variáveis na caracterização das UCI a serem consideradas pelo Alceste (Quadro IV). Deve-se, entretanto, observar que a consideração de categorias da Análise de Conteúdo como variáveis para o programa não substitui as Redes de Conteúdos na busca de co-ocorrências: as redes tratam de significados, o Alceste de vocabulário.

Quadro IV: Letra com os nós da Rede de Conteúdo como variáveis para o procedimento Alceste.

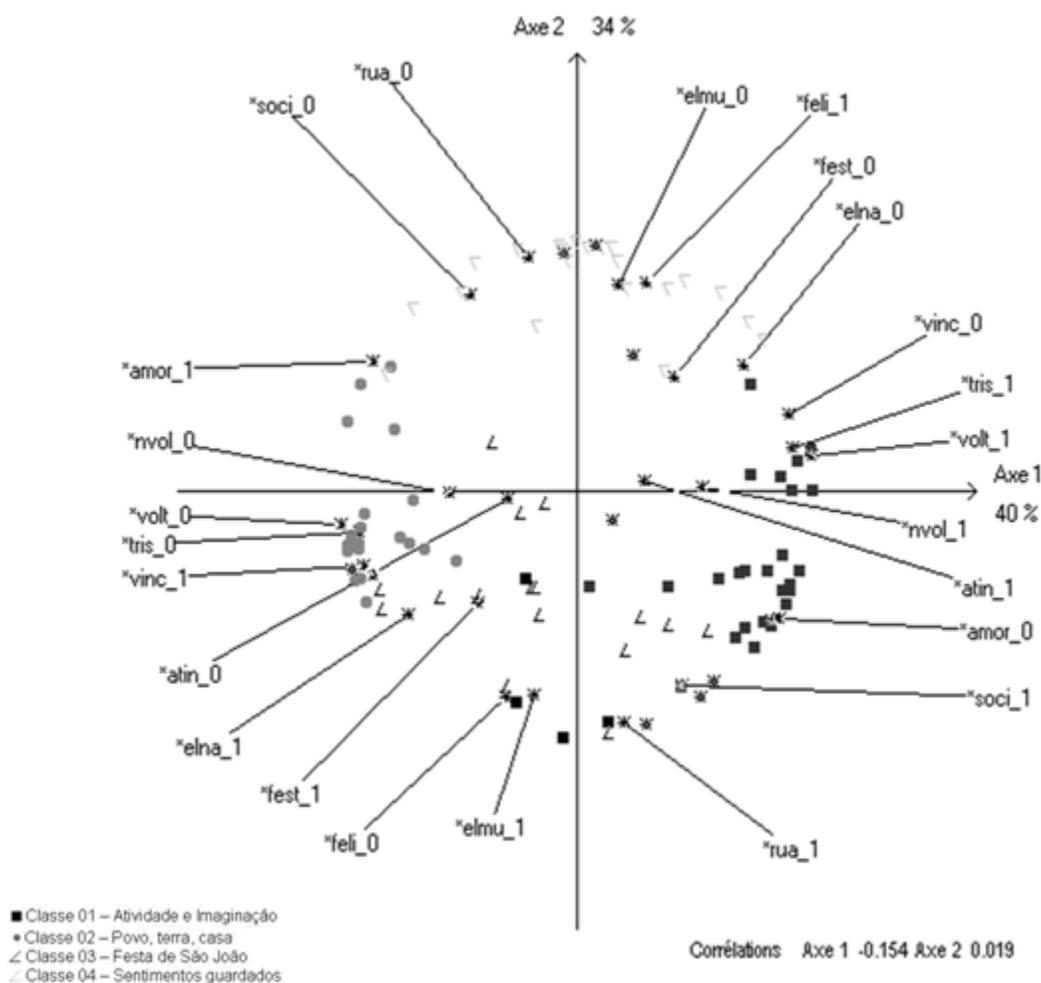
```

    ***MUS01 *atin_1 *feli_1 *vinc_0 *trist_1 *rua_0 *soci_0 *amor_0 elmu_0 *fest_0 volt_1 nvolt_1
    *elna_0
    A CANÇÃO DA SAUDADE (OSVALDINHO); 1966.
    Nas horas tristes, meus pensamentos, tanta lembrança do meu tempo de criança. Brincar de
    roda, amarelinha. Inda não tinha as ilusões do amor. Minha boneca, onde andará você? Queria
    vê-la e abraçá-la como outrora fiz. Doce lembrança da minha infância, querido tempo que nunca
    mais terei. Vou caminhando e procurando o que fazer para esquecer meu sofrimento, meus
    desejos, tudo, tudo e voltar àquele tempo.
  
```

A canção da Saudade (Oswaldinho). Intérprete: Rosely. LP: *Um cantor por dez milhões, dez milhões por uma canção*. CBS 37450. 1966.

A AFC mostrada na Figura 05 possui as mesmas formas reduzidas e Classes já vistas na Figura 02 e apresenta as variáveis (asteriscos) resultantes do agrupamento de categorias da Análise de Conteúdo para a formação dos nós das Redes de Conteúdos (como vimos no Quadro III). Vê-se na figura a presença de Nós da Rede de Conteúdo relacionada às formas reduzidas e Classes (presença indicada, por exemplo, por feli_1, e ausência, nesse caso, indicada por feli_0). Uma sobreposição das Figuras 02 e 05 mostraria que, por exemplo, no pólo que denominamos Descritiva-Visual-Sonora é característica a presença das variáveis elna_1 (elementos naturais), elmu_1 (elementos musicais) e fest_1 (festas), e no pólo denominado Sentimento é característica a ausência de referências a lugares como a rua, a praça ou o quintal (rua_0) ou a elementos de sociabilidade (soci_0).

Figura 5: Agrupamento de categorias da Análise de Conteúdo para a formação dos nós das Redes de Conteúdos como variáveis para o ALCESTE.



Questões Metodológicas

No presente caso, que envolveu letras de músicas como fonte de informações, avaliou-se como muito interessante a possibilidade de conjugar uma análise estatística de vocabulário, um método mais tradicional de análise qualitativa, e um possível desdobramento dessas nas Redes de Conteúdos (NASCIMENTO, 2004). Objetivamente, a realização de uma leitura inter-relacionada dos resultados desses procedimentos possibilitou, na avaliação das informações coerentes e discrepantes entre eles, a visão mais abrangente dos elementos caracterizadores de uma memória saudosa da infância, bem como a própria interação entre esses elementos, nas letras analisadas. De outra forma, seria difícil apontar, por exemplo, que o vocábulo saudade encontra-se recorrentemente relacionado a palavras que caracterizam atividades e não, como se podia esperar, àquelas que tratam do sentimento (o que foi indicado pelo Alceste – Figuras 01 e 02). Também não seria fácil apontar que a representação da infância saudosa nas canções brasileiras está fortemente calcada na lembrança de compositores migrantes (o que foi viabilizado pela alta ocorrência de referências a elementos naturais, lugares e atividades características de contextos não urbanos constatadas na Análise de Conteúdo). Em complemento, a possibilidade de apontar co-ocorrências de conteúdo nas letras analisadas permitiu discutir de forma mais coerente a probabilidade de serem as memórias de atividades infantis, em boa parte, articuladoras dos

demais elementos recordados, fato que podia ser apenas vislumbrado considerando-se tanto os resultados do Alceste como os da Análise de Conteúdo.

Para quem se interessar por empreitada similar, algumas considerações mais gerais sobre a utilização conjunta do Alceste, da Análise de Conteúdo e das Redes de Conteúdo:

O Alceste

Pode servir para um mapeamento introdutório mais geral do corpus (REINERT, 1990; KRONBERGER; WAGNER, 2002).

Confirma ou infirma tendências sugeridas pelos agrupamentos lexicais na Análise de Conteúdo.

Como identifica relações de elementos segundo um critério estatístico, resulta de sua aplicação uma classificação objetiva baseada nas co-ocorrências, com o que diminui a probabilidade do pesquisador encontrar "coincidentemente" as relações que já havia suposto existirem no material analisado.

É um procedimento rápido: com o corpus devidamente preparado, o programa efetua as 04 etapas de procedimento em alguns minutos.

Na apresentação dos resultados, incorpora opções de gráficos e figuras que possibilitam selecionar os mais adequados às características do trabalho no qual é utilizado.

Não obstante o Alceste poder servir para um primeiro mapeamento do corpus a ser analisado, recomenda-se que ele seja o último procedimento realizado num estudo que queira conjugá-lo à Análise de Conteúdo. Algumas vantagens de assim proceder:

- a) Quanto melhor preparado o banco de dados para o Alceste, mais confiáveis são os resultados do programa. A Análise de Conteúdo pode permitir a identificação e o registro no banco de dados de sinônimos e a correção de eventuais erros de digitação no texto.
- b) Não conhecer o relatório do Alceste permite maior independência do pesquisador na Análise de Conteúdo, evitando que os critérios de identificação, diferenciação e agrupamento dos elementos nas categorias e subcategorias sejam influenciados pelos resultados do programa.
- c) Maior familiaridade com o corpus é importante para a análise do relatório do Alceste, e ela pode ser alcançada com o desenvolvimento da Análise de Conteúdo.

Para Análise de Conteúdo

O Alceste não possui dicionário de sinônimos, o que implica o descarte de palavras pouco frequentes, ainda que sinônimas de outras com freqüência maior. Entende-se que tal característica não é uma limitação do programa porque, como se viu, a idéia na base de sua concepção é a de que grupos diferentes se expressam, inclusive em termos de vocabulário, de formas distintas. Nesse sentido, o descarte de tais palavras indica que as mesmas não pertencem aos mundos lexicais do(s) grupo(s) em questão. Por sua vez, a Análise de Conteúdo pode, e deve,

recuperar tais sinônimos, que podem informar que há mais de um discurso sobre o tema, ainda que este discurso esteja menos presente no banco de dados.

Para as Redes de Conteúdo

a) Altas frequências de menções em categorias diferentes não indicam necessariamente que tais categorias estejam relacionadas. Em informações provenientes de grupos diversos sobre o mesmo tema é possível haver interligações entre categorias que possibilitem a formação de mais de um agrupamento dessas mesmas categorias. No presente estudo, em que interessava particularmente o discurso coletivo sobre a lembrança saudosa da infância, a partir também da variabilidade das falas sobre essa mesma lembrança, só a constituição de um procedimento que possibilitasse a visualização de inter-relações entre os principais elementos presentes nesse mesmo discurso permitiria tecer considerações sobre o caráter eminentemente relacional da memória saudosa dos “tempos de menino”.

b) A formação dos nós das Redes de Conteúdos deve ser sustentada por lógica que seja condizente também com o referencial teórico do pesquisador. No caso presente, optou-se, eventualmente, pelo agrupamento de itens que se encontravam separados nas Classes do Alceste, como no caso do nó que agrega os elementos de sociabilidade (notar que mãe e pai encontram-se em duas Classes distintas). Similarmente, alguns elementos separados na Análise de Conteúdo foram unidos. Por exemplo: foi considerado como um grupo lúdico formado por atividades, jogos e brincadeira, um conjunto de itens que resultaram separados na análise de conteúdo, mas que foram apontados pelo Alceste como pertencendo a uma mesma classe (Figura 01 e Quadro III).

c) A construção das Redes de Conteúdos pode proporcionar ajustes na Análise de Conteúdo, pois não se trata apenas de contagem, mas sim de um processo de classificação. É óbvio que os resultados da Análise de Conteúdo e das Redes de Conteúdos devem ser coerentes, pois que estas se originaram da primeira. Trabalhar com categorias pré-estabelecidas é processo menos sinuoso, pois se trata de decidir se determinada menção se refere a uma categoria ou a outra, e não de agrupar elementos dispersos em vários textos partindo da idéia deles estarem ou não explicitamente relacionados ao mesmo tema. Assim, discrepâncias eventuais entre os dois procedimentos, provavelmente pelo fato das anotações terem sido feitas em momentos diferentes, podem ser corrigidas por meio de um “ajuste” na Análise de Conteúdo ou nas Redes de Conteúdo, conforme pareça mais apropriado ao pesquisador.

Referências Bibliográficas

Alba, M. (2004). El método ALCESTE y su aplicación al estudio de las representaciones sociales del espacio urbano: el caso de la Ciudad de México. **Papers on Social Representations**, cidade em que foi publicado o periódico?, v.13, 1.1-1.20?. Disponível em <<http://www.psr.jku.at>>. Acesso em: 15 jul. 2005.

Bardin, L. **Análise de Conteúdo**. Lisboa: Edições 70, 1977.

Bauer, M.W. Análise de conteúdo clássica: uma revisão. In: M. W. Bauer & G. Gaskell (Orgs.). **Pesquisa qualitativa com texto, imagem e som: um manual prático** (Tradução de Pedrinho A. Guareschi). Petrópolis: Vozes, 2002. p. 189-217.

- Camargo, B.V. ALCESTE: Um programa informático de análise quantitativa de dados textuais. In: A. S. P. Moreira (Org.). **Perspectivas teórico-metodológicas em Representações Sociais**. João Pessoa: UFPB/Editora Universitária, 2005. p. 511-539.
- Kalampalikis, N. L'apport de la méthode Alceste dans l'analyse des représentations sociales. In : J. C. Abric (Org.). **Méthodes d'étude des représentations sociales**. Paris: Érès, 2003. p. 147-163.
- Kronberger, N. E Wagner, W. (2002). Palavras-chave em contexto: análise estatística de textos. In: M.W. Bauer & G. Gaskell (Orgs.). **Pesquisa qualitativa com texto, imagem e som: um manual prático** (Tradução de Pedrinho A. Guareschi). Petrópolis: Vozes, 2002.
- Lorenzi-Cioldi, F. **Questions de méthodologie en sciences sociales**. Paris: Delachaux et Niestlé, 1997.
- Nascimento, A.R.A. **Memória dos Verdes Anos: saudade da infância na música popular brasileira – uma investigação e uma proposta de análise de dados**. 2004. 179 f. Tese (Doutorado em Psicologia) - Programa de Pós-Graduação em Psicologia, Universidade Federal do Espírito Santo, Vitória.
- Oliveira, D.C.; Gomes, A.M.T.; Marques, S.C. Análise estatística de dados textuais na pesquisa das Representações Sociais: alguns princípios e uma aplicação ao campo da saúde. In: M. S. S. Menin; A. M. Shimizu (Orgs.). **Experiência e Representação Social: questões teóricas e metodológicas**. São Paulo: Casa do Psicólogo, 2005. p. 157-200.
- Reinert, M. Alceste, une méthodologie d'analyse des données textuelles et une application: Aurelia de Gerard de Nerval. **Bulletin de Methodologie Sociologique**, v.26, p.24-54, 1990.
- Reinert, M. **Alceste**. Version 4.0 – Windows (Manual). Toulouse: Societé IMAGE, 1998.
- Reinert, M. Alceste, une méthode statistique et sémiotique d'analyse de discours; Application aux « Rêveries du promeneur solitaire ». **La Revue Française de Psychiatrie et de Psychologie Médicale**, v. 05(39), p. 32-36, 2001.
- Vala, J. A Análise de Conteúdo. In: A.S. Silva; J. M. Pinto (Orgs.). **Metodologia das Ciências Sociais**. 12 ed. Porto: Edições Afrontamento, 2003. p. 101-128.

[Endereço para correspondência](#)

Endereço: fgjan@uol.com.br; paulomenandro@uol.com.br.

Recebido em: 11/04/2006

Aceito para publicação em: 18/07/2006

Notas

* Doutor em Psicologia (UFES).

** Doutor em Psicologia (USP).

¹ A possibilidade dessa conjugação, ainda que não explorada, também foi apontada por Alba (2004).

² Em português, mais informações sobre o ALCESTE podem ser encontradas em Camargo (2005) e Oliveira, Gomes; Marques (2005).