

Anise A. G. Ferreira

anise.a@gmail.com

As possibilidades de uso de ferramentas da informática: software para análise de dados qualitativos e ISD¹

Applications of I.T. tools: software for qualitative data analysis and SDI¹

RESUMO – O objetivo deste artigo é identificar e discutir possibilidades de uso de software de análise de dados qualitativos no quadro dos procedimentos propostos pelo ISD (interacionismo sociodiscursivo) com ênfase nos softwares livres e/ou em versões gratuitas. Faz-se uma revisão da literatura sobre o uso de software para análise de dados qualitativos na área das ciências sociais e humanas, com foco especial na área de estudos da linguagem. Destacam-se as funções oferecidas por algumas ferramentas entre elas, WeftQDA, MLCT, Yoshikoder e Tropes. Dentre esses, detalha-se o Tropes por suas funções relativas a análise semântica e de elementos lingüísticos, tais como tipos de verbo, modalizações, etc. Embora a tentativa de automatizar uma análise inteiramente baseada nos procedimentos do ISD seja inviável em termos absolutos, os programas mostram-se poderosos auxiliares para a análise de questões específicas. Parece importante ainda conhecer e utilizar diferentes programas para se obter uma visão diversificada sobre os dados analisados. Cabe ao pesquisador ser crítico da análise produzida pelo computador.

Palavras-chave: análise de dados qualitativos, *software*, análise de texto, análise de discurso, interacionismo sóciodiscursivo, Tropes, CAQDA.

ABSTRACT – The objective of the present article is to identify and discuss the possibilities of using qualitative data analysis software in the framework of procedures proposed by SDI (socio-discursive interactionism), emphasizing free distributed software or free versions of commercial software. A literature review of software for qualitative data analysis in the area of social sciences and humanities, focusing on language studies is presented. Some tools, such as: WeftQDA, MLCT, Yoshikoder and Tropes are examined with their respective features and functions. The software called Tropes is examined in more detail because of its particular relation with language and semantic analysis, as well as its embeded classification of linguistic elements such as, types of verbs, adjectives, modalizations, etc. Although trying to completely automate an SDI based analysis is not feasible, the programs appear to be powerful helpers in analyzing specific questions. Still, it seems important to be familiar with software options and use different applications in order to obtain a more diversified vision of the data. It is up to the researcher to be critical of the analysis provided by the machine.

Key words: qualitative data analysis, software, text analysis, discourse analysis, sociodiscursive interactionism, CAQDA.

O objetivo deste artigo é identificar e discutir possibilidades de uso de software de análise de dados qualitativos² no quadro dos procedimentos propostos pelo ISD com ênfase nos softwares livres e/ou em versões gratuitas e, em especial, no chamado Tropes (Acetic, 1994-2004). Espera-se que tal automatização facilite o trabalho do pesquisador na análise de um corpus textual grande. Em relação ao

ISD, seria possível que facilitasse a verificação da infraestrutura geral do texto, por exemplo, identificando temas, tipos de discurso, ou mesmo mecanismos enunciativos?

A análise assistida por computador (*computer assisted analysis*, C.A.A.) dirigida particularmente à pesquisa qualitativa³ é uma prática que começou há mais de dezessete anos, com a revolução do computador no armaze-

¹ Trabalho adaptado da oficina “Problemas da análise dos tipos de discurso e de sua transposição didática: as possibilidades de uso de ferramentas da informática” coordenado por Dinorá Fraga e Anise Ferreira no I Encontro Internacional do Interacionismo Sociodiscursivo, na PUCSP, São Paulo, em 2006. A versão final contou com a preciosa leitura e contribuições de Anna Rachel Machado, a quem gostaria de agradecer.

² A chamada análise de dados qualitativos convencionalmente inclui análises quantitativas dos dados qualitativos que podem estar presentes na pesquisa qualitativa, porém distingue-se tanto da pesquisa qualitativa, quanto da quantitativa (Pereira, 1999).

³ A distinção entre análise de dados qualitativos e a pesquisa qualitativa está no fato de que a primeira pressupõe uma descrição tal dos dados qualitativos que seja possível quantificá-los e efetuar cálculos estatísticos e correlacionais (Pereira, 1999). A pesquisa qualitativa, por princípio, pode prescindir de quantificação, mas não prescinde da necessidade de organização e sistematização de dados.

namento, pesquisa, recuperação, exibição de dados e construção de conceitos, em lugar do trabalho manual usado na década anterior, das anotações de campo, das fotocópias, marcas de canetas coloridas, fichamentos e análise datilografada (Weitzman e Miles, 1995). Com o advento da Internet, refletindo principalmente o universo internacional dos cientistas sociais, inúmeras publicações on-line se dedicam ao tema (Fielding, 1993; Coffey *et al.*, 1996; Barry, 1998; Koenig, 2005; Lewins e Silver, 2006; Gibbs *et al.*, 2006; Drisko, 1997-2006)⁴.

Para os linguistas, o portal da conhecida lista eletrônica de correspondência acadêmica *Linguists List* (2006)⁵ divulga softwares organizando-os nas seguintes categorias: tradução auxiliada por computador, exibição de diagrama, pesquisa de campo (descrição de língua/linguagem), reconstrução histórica, léxicos, análise morfológica, processamento de linguagem natural, outras ferramentas (aplicativos diversos), análise gramatical (*parsers*), análise fonética, reconhecimento e síntese de fala, etiquetador, transcrição, concordância, diretórios de software, análise de fala (incluindo análise de fala na clínica). Nota-se uma grande difusão de softwares na área da fonética, daqueles voltados para a análise da fala⁶ (Boersma e Weenink, 2006; Huckvale, 2007; Cassidy, 2001-2008; University of Michigan Regents, 1998; Manta *et al.*, 1998-2006) e tradicionalmente dentro da Linguística de Corpus (Berber Sardinha, 2004) para análise lexical automática. Porém, quando a análise do linguista recai sobre a interpretação de unidades maiores do texto, tais como fluxos de interação, ou tipos de discurso, a aplicação do software vai se assemelhar àquela feita na área de ciências sociais, a dos CAQDAs (*Computer-Assisted Qualitative Data Analysis*), em que códigos⁷ ou anotações são atribuídas às partes dos registros que forem relevantes à pesquisa. Suas propriedades variam desde possuir mecanismos de busca e de codificar o texto até construir uma teoria baseada no código e uma criação de rede conceitual.

Entre os programas comerciais que se mantêm no mercado e que se situam na categoria dos codificadores-recuperadores de dados, estão o HyperQual (Padilha, 2004) (que se tornou SuperHyperQual), Kwalitan (Peters, s.d.) e The Ethnograph (Qualis Research, 1985-2008). Entre os construtores de teoria baseados em códigos, estão AQUAD (Huber, s.d.), Atlas/ti (ATLAS-ti GmbH, 2002-2008), Hyper-Research (ResearchWare, 1997-2007), Nud*ist (QSR, 1981-2007) e Qualitative Comparative Analysis (QCA) que se tornou fuzzy/set QCA (Ragin *et al.*, 2005-2007). Entre os construtores de redes conceituais, estão Nvivo (QSR, 1981-

2007) (versão moderna do Nud*ist) e o Atlas/Ti.

Quanto às funções e características, os programas possuem formas diferentes de inserir e estruturar os dados; podem segmentar os dados livremente, de acordo com a vontade do usuário, ou em diferentes níveis, sejam palavras, frases, linhas, sentenças ou parágrafos. Aos segmentos são atribuídos códigos⁸ também de diferentes formas: um único ou vários para cada segmento, com um só ou múltiplos níveis de codificação, e com modos diferentes de reorganizar os códigos e recuperar os segmentos codificados. A ligação entre códigos e dados é feita, em alguns softwares, por vinculação dinâmica como a do hipertexto⁹. Com isso, eles também exibem, de formas diferentes, os resultados de pesquisa; ao processar a recuperação dos dados para o usuário, produzem relatórios diferentes, podem criar ou não gráficos e ainda alguns permitem a realização de projetos em rede e por equipes.

O pesquisador, então, depara-se com programas que oferecem diferentes formas de anotações e codificações, dando maior ou menor liberdade para se criarem códigos ou classificações. Nesse sentido, o pesquisador terá maior automação quanto mais o programa contiver categorias ou códigos prontos embutidos. Por outro lado, quanto mais houver liberdade para criar ou organizar categorias ou códigos, maior será o tempo e trabalho investido pelo pesquisador, porém, em compensação, este poderá construir sua teoria a partir dos dados, ou abordar os dados dentro de uma gama mais ampla de modelos teóricos, à medida que organiza e analisa seus dados.

A maioria dos programas que têm mantido uma distribuição estável é comercial e os mais sofisticados, obviamente, são mais caros, chegando aos milhares de dólares. Um investimento desse porte se justifica quando a necessidade analítica é compartilhada por uma grande quantidade de usuários. Minha experiência dentro da L.A. tem mostrado considerável diversidade de processamento de dados, criando necessidades diferentes de análise. Um software apenas, então, não seria suficiente. Tal fato inibe grandes investimentos em softwares comerciais, levando o pesquisador ou estudante a buscar alternativas mais simples, de acesso gratuito ou de custo consideravelmente reduzido.

Passo, então, a apresentar e discutir como alguns softwares podem ajudar o pesquisador que lida com textos, preferivelmente, de uma perspectiva do interacionismo sociodiscursivo (ISD).

Na categoria genérica CADQS, *Computer-Assisted Qualitative Data Analysis Software*, estão os softwa-

⁴ Cf. sites de recursos acadêmicos que divulgam software sob a rubrica *Computer-Assisted Qualitative Data Analysis* (CAQDAS) ou métodos informatizados de dados qualitativos.

⁵ *The Linguist List* (software).

⁶ Entre eles, Praat, WASP, EMU, *Soundsciber* e *Transcriber*.

⁷ Código aqui é um termo genérico para designar qualquer sistema de classificação de conteúdo que o pesquisador adote.

⁸ Ver nota acima.

⁹ Os códigos e dados se apresentam na forma de links.

res que organizam, exibem, processam, resumizam e “interpretam” os dados, i.e., emitem relatórios ou gráficos. Neles, o pesquisador “etiqueta/rotula/classifica” partes dos documentos segundo seus próprios parâmetros teórico-metodológicos e a categorização é parte do processo de interpretação. Para os pesquisadores do ISD, tais programas permitiriam ao pesquisador nomear, organizar e recuperar passagens relevantes dos textos analisados. Dentre esses podemos mostrar a interface de um programa muito simples e *freeware*, chamado WeftQDA (Fenton, 2006) (Figura 1). Basicamente, sua função é criar códigos de categorias, localizá-los no texto e marcá-los ou recuperá-los por meio de *queries* ou pesquisas. O pesquisador cria as categorias de valor interpretativo, marcando, no texto, as unidades empíricas representativas de cada categoria (e.g. ‘verbos_1ª_pessoa’, Figura 1).

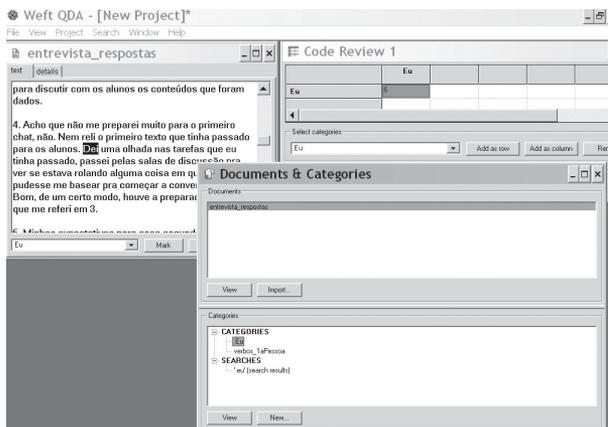


Figura 1. Janelas de gerenciamento do *freeware* WeftQDA.

Com uma versão limitada gratuita, o programa HyperResearch também proporciona a designação de partes do documento, criação de categorias e atribuição das categorias às partes. Sua vantagem é a autocodificação de trechos padronizados em todo o texto de uma vez. Também opera com fontes não-textuais: vídeo e sons, de modo que partes do material sonoro ou em vídeo podem ser também selecionadas para atribuição dos códigos. Para a classificação e categorização de vídeos e suas transcrições, podemos mencionar o Transana¹⁰ (Fassnacht e Woods, 2002-2006), que permite observar o vídeo, efetuando sua transcrição e, ao mesmo tempo, criar categorias e associá-las aos trechos classificados. Em pesquisas em que são observados, por exemplo, os gestos dos sujeitos associados às modalizações na transcrição de sua fala, ou em que o registro é exibido ao sujeito no intuito de eliciar

uma reação ou comentário, a capacidade de analisar gravações em vídeo é um recurso importante.

Na categoria genérica de análise textual automática¹¹, podemos mostrar a funcionalidade e limitações dos gratuitos Multi-Lingual Concordance Toolkit (MLCT) (Piao, 2007) e Yoshikoder (Love, 2006). O MLCT (Figura 2) permite operar com duas áreas de texto, exibindo o texto na janela e o resultado da contagem e ordenação de linhas e contagem de *tokens* na janela da direita. Executa operações de busca de expressões regulares. As suas ferramentas lexicais fornecem listas de uma a seis palavras e de colocados pela frequência de co-ocorrências. Extrai listas de concordância do texto exibido na janela da esquerda, mostrando diferentes formas de ordenação. Na Figura 2, o MLCT exibe as frequências de grupos de três palavras (à direita), a concordância do pronome ‘me’ com destaque da palavra a sua direita; e o próprio texto analisado (à esquerda). É especialmente interessante para o levantamento automático de padrões lexicais e sintagmáticos e para busca de concordâncias, o que pode auxiliar, por exemplo, na verificação de sintagmas nominais, das posições das vozes expressas no texto, auxiliares de modalização (poder), frases impessoais (é lógico que), etc.

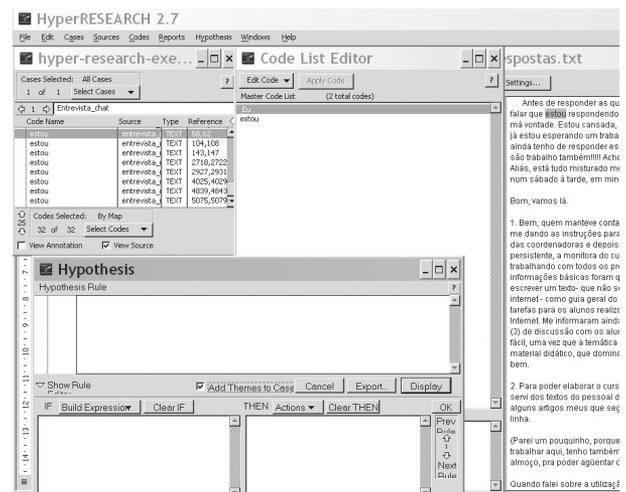


Figura 2. Interface do MLCT exibindo lista de frequência de padrões de 3 palavras, concordância do pronome ‘me’ com destaque da palavra a sua direita; e o texto analisado.

Já o Yoshikoder (Figura 3) combina operações simples ao mesmo tempo de um concordanciador, como o MLCT, e de um codificador, como o WeftQDA. O pesquisador pode criar um dicionário de categorias e, a cada categoria, podem ser adicionadas várias entradas chamadas ‘pa-

¹⁰ Originalmente criado por Chris Fassnacht e atualmente desenvolvido e mantido por David K. Woods do *Center for Education Research*, da *University of Wisconsin-Madison*. A partir de 2007, as novas versões não serão mais gratuitas. Uma alternativa ao Transana é o Anvil, de Michael Kipp (2000-2007), do *German Research Center for Artificial Intelligence* (DFKI).

¹¹ Em março de 2007, foi lançada a última versão do gratuito AntConc de Lawrence Antony (2007), para as funções de criação de listas de palavras, concordâncias, colocados, lista de *key words*, etc.

drões' que são itens lexicais selecionados do texto. Se o pesquisador localiza itens relevantes para a identificação do mundo discursivo em seu documento, pode agrupá-los sob uma categoria na construção do que é denominado, no programa, 'dicionário'. O programa tem a limitação de não efetuar concordâncias ou de criar 'padrões' compostos de mais de uma palavra, como o MLCT, mas também gera um relatório de frequência das categorias e padrões levantados, os pronomes de primeira pessoa "eu" e "me" da Figura 3. Este programa ajudaria o pesquisador do ISD no levantamento e estimativa de seus dados.

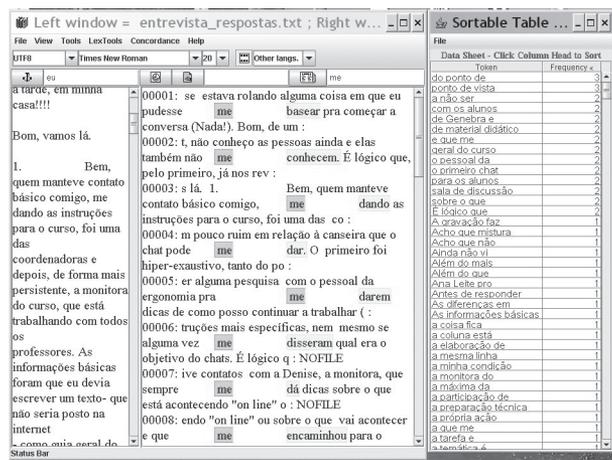


Figura 3. Interface do freeware Yoshikoder, com a janela de códigos (esq), do documento (centro), de contagem (superposta à dir.) e de concordância (quadro inferior).

Entre todos os programas disponíveis na rede, o que parece avançar nas categorias semânticas de análise para auxiliar o pesquisador na perspectiva do ISD é o Tropes-Zoom (Acetic, 1994-2004), que possui uma versão limitada gratuita. Ele se propõe a realizar uma análise textual e semântica de corpora de textos, seguindo, segundo seus autores, uma orientação teórica cognitivo-discursiva (Ghiglione *et al.*, 1998; Ghiglione *et al.*, 1995), com base na abordagem gramatical descrita por Charaudeau (1992).

O programa faz diferentes cálculos estatísticos dividindo o texto em itens lexicais, orações e frases, e contabilizando as co-ocorrências¹²; possui classificações para categorizar os itens significativos, conforme uma nomenclatura de descrição estabelecida *a priori*, bem como uma biblioteca de universos de referência semântica pré-construída ou alimentada pelo pesquisador. Do ponto de vista

da interface de usuário, esse programa é um dos que apresentam os resultados por meio de *links* através dos quais o usuário vê diferentes aspectos dos dados. Tais propriedades ficarão mais claras adiante com a descrição de algumas funções e respectivas figuras.

O que parece interessante para nossas análises são as classificações lexicais do Tropes, que são, seguindo a nomenclatura do programa: a dos **verbos** (factivos, estativos, declarativos e performativos); a dos **conectores** (conjunções de coordenação e subordinação, locuções conjuncionais: de condição, causa, finalidade, adição, disjunção, oposição, tempo e lugar); a dos **pronomes pessoais**, separados por número e pessoa; a das **modalizações** (advérbios ou locuções adverbiais: tempo, lugar, modo, afirmação, dúvida, negação e intensidade); a dos **adjetivos qualificativos**: objetivos, subjetivos e numéricos; a dos **pronomes pessoais**, e as **classes de equivalentes**¹³. A partir da quantidade e o peso de cada elemento, o próprio programa classifica o **estilo geral do texto**, que pode ser: argumentativo, narrativo, enunciativo ou descritivo; e relativo a quatro **tipos de enunciação**: dinâmica e ativa, com predominância de verbos de ação; ancorada no real, com verbos de existência e de posse; de narrador subjetivo, com predomínio de pronomes de primeira pessoa do singular; e de narrador objetivo, com verbos que permitem fazer declaração sobre um estado, uma ação. Isso não quer dizer que o pesquisador que trabalhe no quadro do ISD vá aceitar essa classificação, mas ela pode lhe ser uma primeira pista para a sua própria classificação, de modo que as marcas dos estilos propostos pelos autores do programa, por exemplo, possam ser comparados aos indicadores de classificações do quadro do ISD que compõem os tipos de discurso.

Se os resultados do Tropes precisam ser reinterpretados à luz das categorias de análise empregadas na perspectiva do ISD, quais os elementos no Tropes poderiam auxiliar a distinguir os tipos de discurso: interativo, teórico, relato-interativo e narração? Ou as modalizações: epistêmico, deontico, pragmático e apreciativo? Ou os tipos de agir: situado, canônico, evento passado, experiência, prescritivo, fonte ou decorrente? Ou a distinguir os papéis sintático-semânticos: instrumental, objetivo, beneficiário e factivo?

Vejamos a seguir que unidades podem ser buscadas pelo Tropes para nos auxiliar em relação à identificação do tipo de discurso interativo: por exemplo, o software localiza pronomes de 1ª e 2ª pessoas, de verbos modais,

¹² Segundo os autores, o programa efetua os seguintes cálculos estatísticos: o total de frequência de ocorrências das categorias e subcategorias de palavras, a co-ocorrência e a taxa de conexão de classes de equivalentes e de categorias de palavras; a probabilidade das palavras ocorrerem em feixes ou rajadas (*bundles*); geometricamente analisa os feixes de episódios; e realiza uma análise cognitivo-discursiva para caracterizar o texto.

¹³ As classes de equivalentes são classes semânticas atribuídas a agrupamentos de referências, sejam estas designadas por nomes próprios ou comuns, que ocorrem com frequência no texto e possuem atributos semânticos comuns. O exemplo fornecido pelos autores é a da classe "família" que agrupa os termos "pai" e "mãe" na análise de equivalência.

conectores e modalizações (na terminologia adotada no programa) de tempo e lugar. Entretanto, o Tropes não fornece a relação entre o tamanho da sentença e o total de sentenças; para essa finalidade, seria preciso adotar outro software¹⁴. Em relação às unidades típicas do discurso teórico, o software pode indicar a baixa ocorrência dos pronomes pessoais de 1ª e 2ª pessoas, bem como dos conectores de tempo e lugar, e a maior presença de pronomes de 3ª pessoa, incluindo ‘-se’.

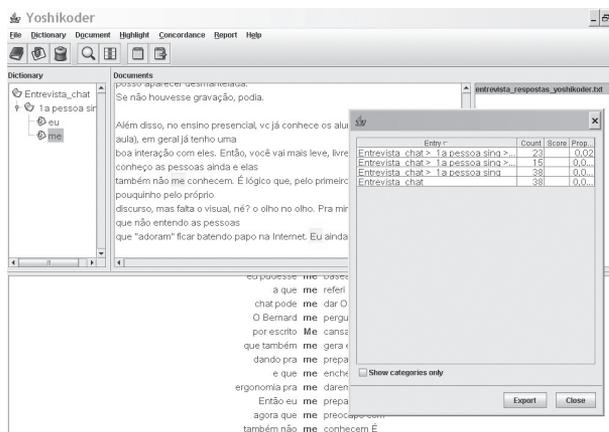


Figura 4. Tropes destacando o primeiro resultado da análise, referente ao estilo dominante: índice das análises (janela sup. esq), itens da análise (jan. centr. esq.) e texto com itens destacados (jan.dir.).

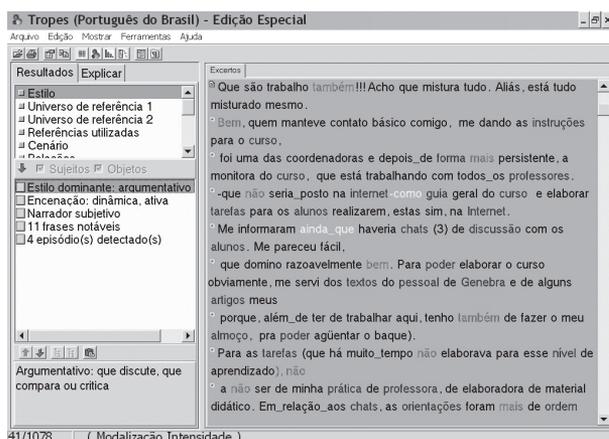


Figura 5. Tropes exibindo o Episódio 1. Lista (alto, à esq.), rajadas (centr.esq.); gráfico (dir. inf.); parte do texto (alto, à dir.).

Exemplifiquemos uma aplicação do programa a parte do corpus de Abreu-Tardelli (2006)¹⁵ composto de res-

postas a perguntas sobre o trabalho do professor atuando em *chat*, com a identificação, pelo programa, de seu estilo, de acordo com a classificação do próprio programa. Para lermos adequadamente a Figura 4, devem ser notadas as janelas: os tópicos de análise são mostrados na janela superior esquerda; os itens analisados referentes ao estilo são mostrados na janela central esquerda; e o texto com suas palavras classificadas são exibidos na janela principal da direita. O programa interpreta o estilo dominante como argumentativo, porque “compara, discute, ou critica”. As cores das palavras do texto foram configuradas por nós e indicam as diferentes análises do programa: azul, as classes de equivalentes dentro do universo de referência; marrom - os verbos; amarelo, os conectores; rosa, as modalizações; verde claro, os adjetivos; e azul turquesa, os pronomes.

O que é denominado de “encenação” pelo programa se refere aos tipos predominantes de verbos, de acordo com a classificação semântica do programa. No caso, a predominância é de verbos de ação. Note-se que, mesmo que não haja concordância quanto à classificação do estilo geral do texto, se colocarmos este texto em comparação a outro, poderemos obter resultados interessantes, pois poderemos identificar diferentes tipos de textos (ou gêneros) que se distinguem, por exemplo, pela maior ou menor ocorrência do que Bronckart chama de “reconfiguração do agir humano” nos e pelos textos (Bronckart, 2004; Bronckart *et al.*, 2004), abrindo-se possibilidades múltiplas de pesquisas, até mesmo com objetivos diferenciados.

A análise por meio desse programa indica que o narrador é subjetivo, devido à forte presença dos pronomes pessoais. Ainda de acordo com os autores do programa, as “frases notáveis” são aquelas que caracterizam mais o texto e elas são definidas pela sua importância segundo uma pontuação¹⁶ dada pelo programa às funções de análise.

O programa detecta também os episódios, que são partes estatisticamente significativas no texto de acordo com a cronologia, o que poderia corresponder à identificação do plano geral do texto, de acordo com o ISD, o que deveria ser testado. Na classificação dos autores do programa, os episódios são compostos por rajadas (*bundles*, em inglês). Ao clicar em “4 episódios” da Figura 5, abrem-se as janelas da Figura 6, especificando os episódios. Na Figura 5, o programa exhibe na janela inferior direita as rajadas que compõem cada episódio. No caso, o episódio 1 é composto por rajadas que contêm “modalização tempo”, “vida escolar”, “atividade escolar”, “documento” e “modalização negação”. Na janela inferior direita, o programa exhibe a divisão dos quatro episódios em retângulos. O

¹⁴ O MLCT fornece a frequência de cada comprimento de sentença e o total de sentenças.

¹⁵ Texto extraído de corpus usado na tese de doutorado de Abreu-Tardelli (2006), fornecido pela autora que efetuou a análise manualmente, sem usar recursos automáticos.

¹⁶ Tal pontuação não é indicada no manual do programa.

que aparece em destaque na janela da esquerda é mostrado na cor verde na janela da direita. As rajadas são mostradas por linhas tracejadas horizontais. Na margem inferior direita, há uma barra de *status* que indica que esse episódio representa 22,3% de todo o texto. Os termos que descrevem as rajadas se constituem tanto de categorias de palavras analisadas pelo programa (modalização tempo, negação), quanto de classes pertencentes a um universo de referência semântico embutido no programa (ex.: ‘vida escolar’, ‘documento’, ‘atividade escolar’, etc.). O episódio em destaque, portanto, apontando para as classes de palavras de maior relevo no primeiro quarto do texto, parece exibir as modalizações apreciativas do autor em relação às referências semânticas do “chat” e da “atividade escolar”.

Na Figura 6, podemos ver um gráfico relativo às referências semânticas da ‘modalização negação’. Os termos que estão à esquerda da modalização negativa aparecem antes no texto e, neste exemplo, são “chat” e “tarefa”, que se relacionam negativamente às classes de equivalência semântica da direita. Como as classes semânticas dos universos de referência ou cenário do programa são pré-estabelecidas, todas as classes precisam ser verificadas no texto, clicando-se sobre elas. Em princípio, o gráfico reflete o universo semântico associado às modalizações negativas.

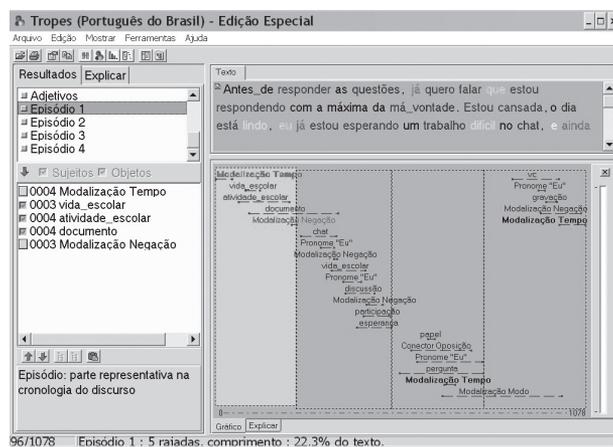


Figura 6. Exibição do Episódio 2, seus componentes, gráfico estrela e texto correspondentes.

Entretanto, a nosso ver, talvez o gráfico visualmente mais impressionante do Tropes seja o gráfico em “esfera” (Figura 7), cujo tamanho é proporcional à frequência de cada classe de referência semântica e cuja distância é proporcional à do texto. No gráfico da Figura 7, há quatro esferas: a que representa a referência semântica central “tarefa” e as que representam as referências relacionadas a ela “vista”, “papel” e “vida escolar”. Pelo tamanho e pela proximidade, é possível comparar sua dimensão e distância no texto. Os gráficos em esfera tratam das referências semânticas do texto em três níveis: (a) o uni-

verso de referência 1, em nível hierárquico mais alto, por exemplo, pode indicar se o texto refere-se mais amplamente à vida humana, ou à educação e ensino, ou emprego e trabalho, etc.; (b) o universo de referência 2, mais específico indicaria, por exemplo, se se trata de organização do ensino, de condições e organização do trabalho, de discurso, hardware, informação, etc e (c) o universo de referência utilizado, que reúne todas as classes selecionadas pelo programa para classificar semanticamente o texto. Se o programa não conseguir atribuir ao texto uma classe de referência semântica pré-existente em sua biblioteca, a palavra não agrupada em referências é colocada como uma nova referência. No texto que analisamos, as palavras “chat”, “tarefa”, “papel”, entre outras, por exemplo, foram acrescentadas automaticamente como referências semânticas, por não terem sido associadas a categorias pré-estabelecidas na biblioteca existente no programa. Esse tipo de classificação semântica global dos textos auxiliaria o pesquisador a, por exemplo, delimitar seu corpus de análise, a distinguir classes e subclasses de textos no interior de um corpus previamente selecionado.

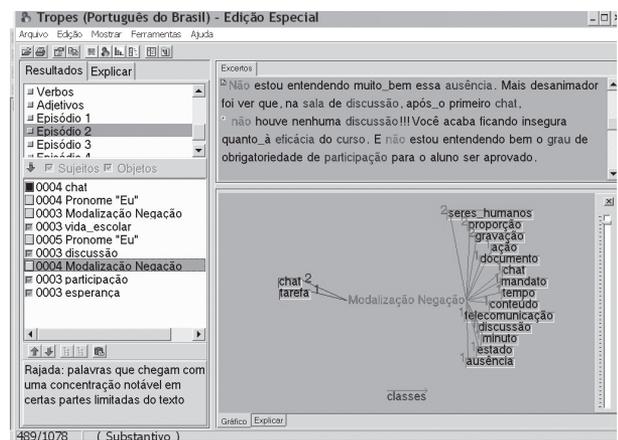


Figura 7. Exibição do gráfico e texto com a referência semântica à “tarefa” no programa Tropes.

Cabe acrescentar que o programa pode mostrar uma análise conjunta ou separada, na janela de resultados das ‘referências utilizadas’ (à esquerda), das referências que ocupam, no texto, as posições de ‘sujeito’ (antes do verbo) e de ‘objeto’ (depois do verbo). Embora a Figura 7 mostre as duas opções selecionadas (sujeito e objeto), o programa analisou como objeto apenas certas referências, como por exemplo, “chat” e “tarefa”, ao passo a referência “papel” foi analisada pelo programa apenas como ‘sujeito’. Submetidas à verificação do pesquisador, tal análise seria útil na identificação de papéis sintático-semânticos.

Podemos verificar, entretanto, que o programa se mostra mais útil na distinção de tipos de discurso, dos tipos de agir e dos mecanismos enunciativos. Por exemplo, a Figura 8 mostra, na janela inferior esquerda, o resul-

tado da análise de categorias de palavras mais frequentes executada pelo programa com destaque aos pronomes. Aqui, o programa indica a frequência desses pronomes (eu, me, mim, comigo), que são predominantes (66%) e como eles se associam às classes de equivalentes. Como sabemos, a ocorrência de pronomes de primeira pessoa é marca de ocorrência de discurso interativo, o que permitiria o tratamento de um vasto conjunto de textos para identificação de tipos de discurso, embora as ocorrências materiais ainda devessem ser verificadas no texto para se validar a análise do programa.

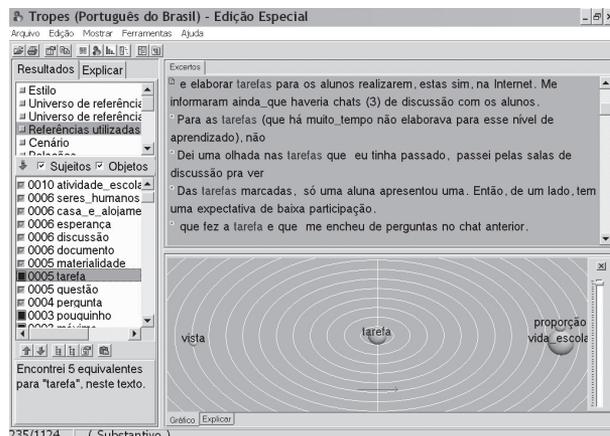
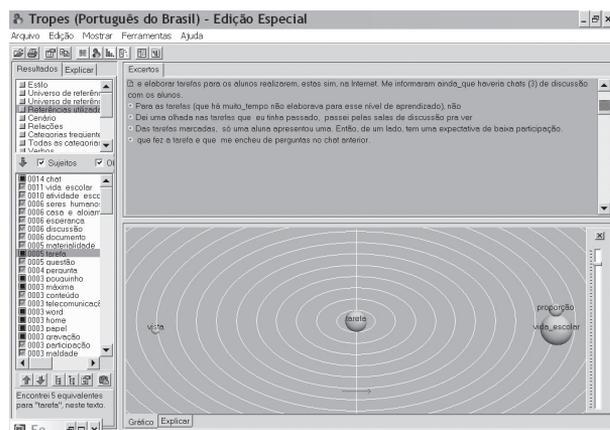


Figura 8. Categorias frequentes no Tropes (esq), destacando a categoria dos pronomes (esq.), o gráfico das classes de equivalentes correspondentes (dir.) e o texto (dir.).

A análise feita pelo programa exibe uma grande ocorrência de modalizações temporais e de negação. A Figura 9 mostra a modalização de negação, que chama a atenção por ser bastante frequente (28 casos), superior à de tempo. A avaliação negativa presente no texto de entrevista em relação ao trabalho do professor no *chat* é notada nas conclu-

sões da autora da tese (Abreu-Tardelli, 2006, p. 119). Evidentemente, o levantamento dessas ocorrências já guiaria o pesquisador na interpretação do caráter argumentativo do trecho e da existência de vozes pressupostas, de acordo com a análise da semântica argumentativa.

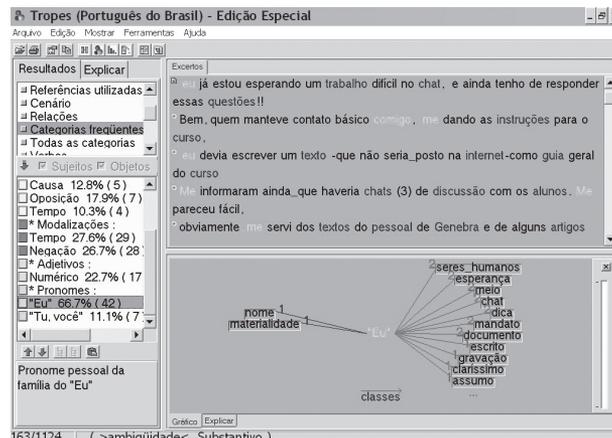


Figura 9. Modalização de negação destacada na opção ‘Todas as categorias’ no programa Tropes.

É interessante ressaltar aqui, também, de que modo o programa exibe as classes de equivalentes associadas à modalização de negação (Figura 9). Para observar as passagens no texto em que ocorre cada classe, o programa apresenta uma interface hipertextual, através da qual o pesquisador efetua a verificação no texto. Para sabermos a quantidade das categorias de palavras relevantes no presente caso, a opção ‘Todas as categorias’ está selecionada (janela superior esquerda) bem como o tipo de modalização listado (no caso, a negação). As categorias de palavras, sejam pronomes, modalizações ou conectores, são apresentadas à esquerda sempre associadas a uma estatística de frequência e porcentagem, ao mesmo tempo em que são destacadas, no texto (janela superior direita), as palavras que pertencem a cada categoria, para que o pesquisador possa verificá-las.

A Figura 10 mostra a classificação dos verbos, efetivada pelo programa quando selecionada a opção ‘Todas as categorias’. Como vemos, predominam os verbos classificados como ‘factivos’, definidos como aqueles que expressam ações. O exemplo mostra que é preciso verificar as palavras classificadas, uma vez que o programa não foi capaz de efetuar a desambiguação¹⁷ ocasionada pelo termo ‘trabalho’ (substantivo) quando este aparece sem artigo. Evidentemente, essa classificação prévia dos verbos seria de muito interesse para os pesquisadores do ISD que focam suas pesquisas na análise do agir representado nos textos.

¹⁷ A correção na desambiguação automática do programa, segundo os autores, depende do tamanho do corpus: um corpus maior apresenta, teoricamente, um resultado mais preciso.

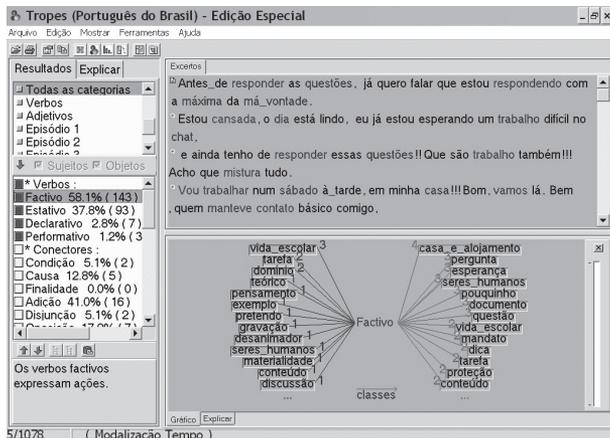


Figura 10. Verbos factivos na análise dos verbos do resultado de ‘Todas as categorias’ no programa Tropes.

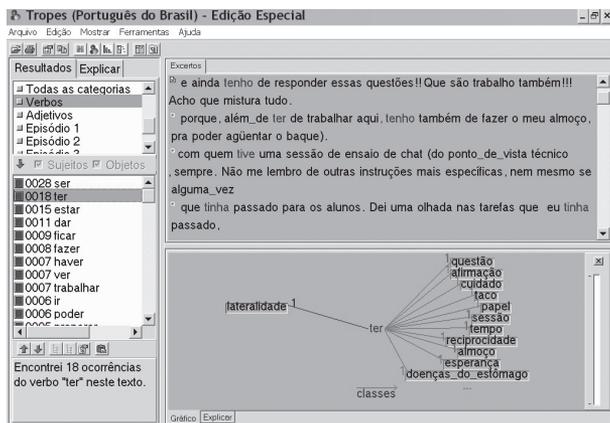


Figura 11. Lista de verbos e sua freqüência fornecida pelo programa Tropes.

A Figura 11 mostra a análise desenvolvida pelo programa sobre os verbos, com a sua listagem completa e a indicação de sua freqüência, pela qual podemos detectar as ocorrências das modalizações, tais como propostas nos procedimentos de análise do ISD, a saber, deôntica, epistêmica, pragmática, lógica e apreciativa (Bronckart, 2003).

A finalidade dos exemplos até aqui expostos foi indicar o potencial dos programas em relação ao ISD, sem a pretensão de efetuar uma análise em profundidade. Uma análise bem desenvolvida de acordo com os procedimentos propostos no quadro do ISD deveria envolver um grande conjunto de análises automáticas realizadas pelo Tropes ou por outros programas, embora idealmente, desejássemos que um único programa contemplasse todas as análises possíveis.

A tentativa de automatizar uma análise inteiramente baseada nos procedimentos do ISD parece ser um empreendimento complexo e talvez inviável em termos absolutos, já que se trata de um quadro teórico em constante construção e evolução, não possuindo categorias fixas e válidas para sempre. Apesar disso, elas têm uma relativa

estabilidade e os programas podem ser poderosos auxiliares para a análise de questões específicas, sobretudo quando se trata de um conjunto grande de dados. Nesse sentido, parece ser importante conhecer e utilizar diferentes programas para se obter uma visão diversificada e ampliada sobre os dados analisados.

De qualquer forma, acreditamos que os programas para análise de dados qualitativos (e quantitativos) podem auxiliar o olhar do pesquisador, que deverá sempre ser crítico em relação a eles e saber interpretá-los de acordo com o seu quadro de pressupostos teóricos mais amplos, mas com economia de sua capacidade perceptivo-cognitiva.

Referências

- ABREU-TARDELLI, L. 2006. *trabalhodoprofessor@chateducacional.com.br. Aportes para compreender o trabalho do professor iniciante em EAD*. São Paulo, SP. Tese de Doutorado. Pontifícia Universidade Católica de São Paulo – PUC-SP, 144 p.
- ACETIC. 1994-2004. Tropes-Zoom. Software. Distribuído por Semantic-Knowledge (em inglês, português e espanhol) e por Acetic (em francês). Disponível em: <http://www.semantic-knowledge.com/download2.htm>. Acesso em 27/11/2007.
- ANTONY, L. 2007. AntConc Homepage. Center for English Language Education in Science and Engineering. School of Science and Engineering Waseda University. Disponível em: http://www.antlab.sci.waseda.ac.jp/antconc_index.html. Acesso em 16/01/2008
- ATLAS-TI GmbH. 2002-2008. ATLAS-ti. Scientific Software Development. Disponível em: <http://www.atlasti.com/>. Acesso em 16/01/2008.
- BARRY, C.A. 1998. Choosing Qualitative Data Analysis Software: Atlas/ti and Nudist Compared, *Sociological Research Online* 3(3). Disponível em: <http://www.socresonline.org.uk/3/3/4.html>. Acesso em 27/11/2007.
- BERBER SARDINHA, A.P. 2004. *Linguística de Corpus*. São Paulo, Manole, 410 p.
- BOERSMA, P. e WEENINK, D. 2006. Praat. Doing phonetics by computer. Software. Institute of Phonetic Sciences, University of Amsterdam. Disponível em: <http://www.fon.hum.uva.nl/praat/>. Acesso em 16/01/2008.
- BRONCKART, J-P. 2003. *Atividade de linguagem, textos e discursos*. 1ª ed., São Paulo, EDUC, 353 p.
- BRONCKART, J-P. 2004. Pourquoi et comment analyser l'agir verbal et non verbal in situation de travail. In: J-P. BRONCKART e GROUPE LAF (eds.), *Agir et discours in situation de travail. Cahiers de la Section des Sciences de l'Éducation*, 103:11-144.
- BRONCKART, J-P.; BULEA, E. e FRISTALON, I. 2004. Les conditions d'émergence de l'action dans le langage. *Cahiers de linguistique française*, 26:245-369. Disponível em: <http://clf.unige.ch/num.php?numero=26>. Acesso em 27/11/2007.
- CASSIDY, S. 2001-2008. The Emu Speech Database System. Centre for Language Technology, Macquarie University. Disponível em: <http://emu.sourceforge.net/index.shtml>. Acesso em 16/01/2008.
- CHARAUDEAU, P. 1992. *Grammaire du sens et de l'expression*. 1ª ed., Paris, Hachette, 927 p.
- COFFEY, A.; HOLBROOK, B. e ATKINSON, P. 1996. Qualitative Data Analysis: Technologies and Representations. *Sociological Research Online*, 1(1). Disponível em: <http://www.socresonline.org.uk/1/1/4.html>. Acesso em 27/11/2007.
- DRISKO, J.W. 1997-2006. QDA Software, Qualitative Pages, Smith College School for Social Work. Disponível em: <http://www.smith.edu/~socialwork/qda/>

- sophia.smith.edu/~jdrisko/qdasoftw.htm. Acesso em: 16/01/2008.
- FASSNACHT, C. e WOODS, D.K. 2002-2006. Transana. Software. University of Wisconsin-Madison Center for Education Research, University of Wisconsin-Madison. Disponível em: <http://www.transana.org/support/documentation.htm>. Acesso em 27/11/2007.
- FENTON, A. 2006. WeftQDA. Software. Disponível em: <http://www.pressure.to/qda/>. Acesso em 27/11/2007.
- FIELDING, N. 1993. Analysing qualitative data by computer. *Social Research Update*, 1. Disponível em: <http://www.soc.surrey.ac.uk/sru/SRU1.html>. Acesso em 27/11/2007.
- GHIGLIONE, R.; KEKENBOSCH, C. e LANDRÉ, A. 1995. *L'analyse cognitivo-discursive*. 1ª ed., Grenoble, Presses Universitaires de Grenoble, 142 p.
- GHIGLIONE R.; LANDRÉ, A.; BROMBERG, M. e MOLETTE, P. 1998. *L'analyse automatique des contenus*. 1ª ed., Paris, Dunod, 168 p.
- GIBBS, G. R; FIELDING, N.; LEWINS, A; TAYLOR, C. 2006. Online QDA. School of Human & Health Sciences, University of Huddersfield. Disponível em: <http://onlineqda.hud.ac.uk/index.php>. Acesso em 16/01/2008.
- HUBER, G.L. [s.d.] Aquad six. Program for the analyses of qualitative data. Disponível em: <http://www.aquad.de/eng/>. Acesso em 16/01/2008.
- HUCKVALE, M. 2007. WASP, Windows Tool for Speech Analysis. Dept. of Phonetics and Linguistics. University College London. Disponível em <http://www.phon.ucl.ac.uk/resource/sfs/wasp.htm>. Acesso em 16/01/2008.
- KIPP, M. 2000-2007. ANVIL: The Video Annotation Research Tool. Disponível em: <http://www.anvil-software.de/>. Acesso em: 16/01/2008.
- KOENIG, T. 2005. Comparison of Software Packages for Qualitative Data Analysis, Assessment and Development of New Methods for the Analysis of Media Content, Department of Social Sciences, Loughborough University. Disponível em: http://www.lboro.ac.uk/research/mmethods/research/software/caqdas_comparison.html#koenig_2004. Acesso em: 16/01/2008.
- LEWINS, A. e SILVER, C. 2006. Choosing a CAQDAS Package. A working paper. *CAQDAS Networking Project*, Researcher Development Initiative. Disponível em: <http://caqdas.soc.surrey.ac.uk/ChoosingLewins&SilverV4%20june06.pdf>. Acesso em 27/11/2007.
- LINGUISTS LIST. 2006. Web resource listing. Software. Institute for Language Information and Technology (ILIT). Eastern Michigan University. Disponível em: <https://linguistlist.org/sp/Software.html>. Acesso em 27/11/2007.
- LOVE, W. 2006. Yoshikoder. Software. Identity Project. Weatherhead Center for International Affairs, Harvard U. Disponível em: <http://www.yoshikoder.org>. Acesso em 27/11/2007.
- MANTA, M.; ANTOINE, F.; GALLIANO, S. E BARRAS, C. 1998-2006. Transcriber a tool for segmenting, labeling and transcribing speech. Disponível on-line: <http://trans.sourceforge.net/>. Acesso em 16/01/2008.
- PADILHA, R. 2004. SuperHyperQual. Software For Qualitative Analysis and Concept Modeling. Disponível em: <http://home.satx.rr.com/hyperqual/>. Acesso em 16/01/2008.
- PEREIRA, J.C.R. 1999. *Análise de dados qualitativos: Estratégias Metodológicas para as Ciências da Saúde, Humanas e Sociais*. 3ª ed., São Paulo, Edusp, 160 p.
- PETERS, V. [s.d.] Kwalitan.net. Department of Research Methodology, University of Nijmegen, The Netherlands. Disponível em: <http://www.kwalitan.net/engels/>. Acesso em 16/01/2008.
- PIAO, S.S. 2007. Multilingual Corpus Toolkit. Software. Dr. Scott Songlin Piao, The University of Manchester. Disponível em: <http://personalpages.manchester.ac.uk/staff/Scott.Piao/research/Download/download.htm>. Acesso em 27/11/2007.
- QSR. 1981-2007. NUD*IST. Qualitative research software. Disponível em: http://www.qsrinternational.com/products_previous-products.aspx. Acesso em: 16/01/2008.
- QUALIS RESEARCH 1985-2008. Ethnograph. Disponível em: <http://www.qualisresearch.com/>. Acesso em 16/01/2008.
- RAGIN, C. C.; GIESEL, H.; RIHOUX, B. e YAMASAKI, S. 2005-2007. Qualitative Comparative Analysis (QCA). Disponível em: <http://www.u.arizona.edu/~cragin/fsQCA/index.shtml>. Acesso em: 16/01/2008.
- RESEARCHWARE. 1997-2007. HyperResearch. Disponível em: <http://www.researchware.com/hr/>. Acesso em 16/01/2008.
- UNIVERSITY OF MICHIGAN REGENTS. 1998. Soudscriber, MICASE project. Disponível em: <http://lw.lsa.umich.edu/eli/micase/index.htm>. Acesso em 16/01/2008.
- WEITZMAN, E.A. e MILES, M.B. 1995. *Computer Programs for Qualitative Data Analysis: An Expanded Sourcebook*. 2ª ed., Thousand Oaks, Sage, 352 p.

Submetido em: 25/10/2007
Aceito em: 10/12/2007

Anise A. G. Ferreira
LAEL/PUCSP e FCLAr/UNESP
São Paulo e Araraquara, SP, Brasil