

## PROGRAMAS DE COMPUTADORES PARA ANÁLISE DE DADOS QUALITATIVOS

JOSÉ AZEVEDO<sup>162</sup>

### *Introdução*

Na última década desenvolveram-se de forma acentuada um conjunto de programas de computadores para análise de dados qualitativos, existindo hoje mais de uma dezena<sup>163</sup>. Todavia, e como refere Tesch (1993) mesmo a maioria dos investigadores que utilizam métodos de análise textual ainda desconhece a existência de tais programas e as suas potencialidades.

Todos os programas que vamos referir foram desenvolvidos especificamente com o propósito de analisar dados qualitativos. Daí que nos permitam trabalhar quer com textos muito estruturados (por exemplo, respostas a perguntas abertas de um questionário), quer com textos pouco estruturados (por exemplo, transcrições de grupos de discussão, notas de campo, biografias ou textos literários).

### *Características dos programas*

É importante começar por referir que embora a designação mais comum para estes programas seja a de “software para análise qualitativa de dados”, os procedimentos que estes efectuem não são verdadeiramente analíticos mas somente facilitadores de rotinas necessárias à análise.

Salvo em obras de ficção, o computador não pode analisar os dados, somente ajudar. O tipo de ajuda depende, como é evidente, do tipo de software em questão. Podemos identificar dois tipos básicos de programas com aplicação nas ciências sociais<sup>164</sup>.

---

<sup>162</sup> Faculdade de letras da Universidade do Porto (Sociologia).

<sup>163</sup> Não temos a pretensão de rever aqui todos os programas existentes, tanto mais que nesta área estão constantemente a surgir novidades. Para uma actualização das versões disponíveis sugere-se a consulta na Internet das páginas de cada um dos programas ou de centros que proporcionam a possibilidade de importar versões de demonstração dos programas existentes (por ex. <http://www.soc.surrey.ac.uk/caqdas/>).

<sup>164</sup> Um outro tipo de programas mais básico tem igualmente sido identificado por diversos autores. (por ex. Weitzman & Miles, 1995). Consideramos, todavia, que esses programas não são relevantes para o tipo de investigação qualitativa geralmente produzido nas ciências sociais.

TIPO I - programas recuperadores de texto<sup>165</sup> (exemplos “Kwalitan”, “The ethnograph”)

Estes programas são particularmente dirigidos à realização de análises de conteúdo simples. Neste sentido estes programas efectuem com eficácia as tarefas de dividir o texto em segmentos, colocar codificações nos segmentos, e depois encontrar e apresentar todos os extractos dentro de uma categoria ou combinação de categorias.

Pode-se, por exemplo, pedir todos os extractos em que surge a expressão “classe social” ou todos os extractos codificados na categoria “classe social” (onde a expressão não tem obviamente de estar presente).

TIPO II programas para construção de uma teoria<sup>166</sup> (exemplos ATLAS-ti, NUD.IST)

Estes programas possuem todas as funções descritas para o Tipo I, mas têm igualmente funções específicas que auxiliam os investigadores no desenvolvimento da teoria. Por exemplo, permitem que se estabeleçam conexões entre categorias, que se desenvolvam esquemas categoriais com vários níveis, que se formulem esquemas conceptuais e se teste a sua adequação aos dados. Estes programas estão frequentemente organizados em torno de um sistema de regras ou baseados numa lógica formal, oferecendo uma análise Booleana completa.

#### *Crítérios para a selecção de um programa*

Não existe uma resposta simples à questão “qual é o melhor programa?”. As considerações que apresentamos em seguida procuram simplesmente ajudar a clarificar a relação entre as necessidades do investigador e as características dos vários tipos de software disponível.

Para se tomar uma decisão sobre se devemos ou não considerar o uso de um programa de computador para analisar os nossos dados qualitativos, e se sim que tipo de programa, vários aspectos parecem relevantes: a) estruturação e tipo de dados do nosso projecto; b) o suporte dado à codificação; c) as facilidades para elaborar notas sobre os textos ou sobre as codificações; d) as facilidades de procura e apresentação dos dados; e) as facilidades de estabelecer relações entre os dados.

#### *Estruturação e tipo de dados*

---

<sup>165</sup> Text retrievers

<sup>166</sup> Theory-building

Uma das primeiras razões para a escolha de um programa deriva do tipo de dados recolhidos no seu projecto. Na realidade, os programas variam nas suas capacidades de lidar com dados mais ou menos estruturados ou com projectos que comportam fontes únicas ou múltiplas. Se o projecto em questão resultar de uma recolha de fontes múltiplas (por exemplo, num estudo sobre delinquência juvenil poderemos ter entrevistas com os jovens, com técnicos de serviços de apoio aos jovens, com a polícia, a família, etc.) então os programas de tipo II são mais indicados, já que permitem estabelecer conexões entre as diferentes fontes. Uma outra facilidade proporcionada por estes programas é a de possibilitarem anexar aos extractos um “marcador” que nos vai permitir saber sempre a origem da informação. Esta facilidade é importante porque oferece a possibilidade de uma abordagem de tipo hipertexto.

Os programas diferem ainda na sua capacidade de lidar com diferentes níveis de estruturação dos dados. Dados muito organizados (por ex. questões abertas de um inquérito) são geralmente bem tratados por programas de tipo I, enquanto dados menos estruturados requerem programas mais flexíveis (tipo II).

### *Codificação*

A codificação é um aspecto central na análise qualitativa. Com frequência necessitamos de distribuir os dados por categorias através de indexação de códigos a segmentos de texto, de forma a verificarmos que segmentos representam cada categoria. Todos os programas permitem pesquisar o texto com base em palavras-chave que o investigador considere relevantes. É possível assim localizar palavras, expressões ou frases e combinações destas, bem como fazer uma contagem e apresentação das palavras chave em contexto.

O resultado desta pesquisa é geralmente o início da codificação. Os investigadores que necessitem de uma codificação com um elevado grau de complexidade (muitas categorias com vários níveis) são aconselhados a escolher programas de tipo II. Nestes, as possibilidades de codificação e a complexidade destas são consideravelmente mais elaboradas. Os programas de tipo I devem ser escolhidos se o objectivo é mais quantificador e se o investigador está mais centrado sobre a mera listagem de palavras-chave nos seus contextos e para determinado grupo de sujeitos.

Em resumo, estes programas são extremamente eficazes nos procedimentos de segmentação dos textos (por exemplo, em palavras, linhas, parágrafos ou qualquer outro tipo de segmento) bem como na atribuição de codificações a cada um dos segmentos.

A categorização pode ser efectuada de diversas formas, atribuindo uma ou mais categorias a cada segmento. Muitos programas permitem uma categorização hierárquica (por exemplo, “fumar” pode ser categorizado sucessivamente nos níveis mais altos de “comportamento aditivo” e “comportamento”), e alguns têm particularidades que permitem a modificação da categorização ou do nível de categorização de uma forma extremamente simples.

Nos programas de tipo II é geralmente também possível que a apresentação dos segmentos seja acompanhada de uma indicação sobre a proveniência desse mesmo segmento (de que entrevista, quem a efectuou, etc.). Vários programas oferecem ainda a possibilidade de editar a lista completa das categorias utilizadas.

### *Notas*

Alguns programas (Tipo II) permitem que se escrevam notas sobre o significado de cada categoria, ou sobre qualquer outro aspecto relacionado com os dados ou com a codificação. A vantagem desta facilidade é que nos permite editar um manual de codificação em que as categorias são apresentadas conjuntamente com os critérios para a sua classificação, bem como com exemplos de segmentos classificados sob esta categoria. Esta facilidade é particularmente importante quando se necessita de treinar pessoas para codificarem um determinado conjunto de dados qualitativos.

Programas como o NUD\*IST e o ATLAS/ti fazem muito mais do que simplesmente codificações, eles permitem, por exemplo, ligar notas e outras considerações analíticas a segmentos específicos do texto. O objectivo é incorporar muitas das tarefas centrais da “teoria enraizada”<sup>167</sup> em programas de software.

### *Procura e apresentação dos dados*

Os programas variam consideravelmente nas suas capacidades de procura e recuperação de texto.

Todos os programas permitem processos de procura segundo variáveis determinadas, tais como idade, género ou classe social. Com esta procura dirigida, os investigadores podem por exemplo comparar as atitudes face a um dado problema dos homens e das mulheres de diferentes grupos etários e de diferentes classes sociais. As procuras por palavra-chave e por palavra-chave em contexto são igualmente generalizadas.

---

<sup>167</sup> A designação “Grounded theory” foi introduzida por Glaser & Strauss (1967) e significa “a descoberta da teoria a partir dos dados”, neste sentido consideramos que a expressão “teoria enraizada” era a que melhor traduzia o sentido a que os autores aludiam. Desenvolvimentos recentes desta abordagem podem ser encontrados em Glaser (1992) e Strauss & Corbin (1990).

As diferenças encontram-se em buscas mais complexas como sejam: encontrar regularidades na sequência com que surgem determinados extractos, ou informação sobre a co-ocorrência de determinadas categorias num dado documento. As co-ocorrências podem ser definidas de várias formas: indicadas pela sobreposição de determinadas categorias ou pela inserção de segmentos de texto numa dada categoria.

#### *Estabelecimento de ligações entre os dados*

As ligações que se podem estabelecer entre os dados processam-se a vários níveis.

Num primeiro nível, a ligação consiste na inserção de uma dada categoria numa categorização mais geral, ou a subdivisão de uma categoria em subcategorias mais refinadas.

Num segundo nível, já só acessível a alguns programas (Tipo II), as ligações consistem em estabelecer-se conexões entre vários tipos de dados, tais como notas de campo, categorias, entrevistas e mesmo outro tipo de dados como vídeos que, não estando na nossa base de dados podem conter dados que queremos relacionar. Os programas mais poderosos neste aspecto permitem ao investigador “navegar” facilmente entre estes vários tipos de dados e ainda visualizar um esquema gráfico que mostra a relação entre os vários componentes.

#### *Críticas à utilização de computadores para a análise qualitativa de dados*

Coffey, Holbrook e Atkinson (1996) expressaram os seus receios de que um crescente uso de software específico poderia levar os investigadores a adoptar uma nova ortodoxia de análise qualitativa. Os autores argumentavam que isto iria contra as tendências pós-estruturalistas e pós-modernistas actuais que enfatizam a aceitação e celebração da diversidade.

Na realidade, é muito fácil desenvolver-se uma análise de dados qualitativos mecanizada e simplista. Evidentemente que esta questão não é necessariamente inerente ao software em si mesmo, encontrando-se antes na utilização que lhe é dada. Contudo a associação da análise de dados computadorizada com uma versão simplificada da “teoria enraizada” comporta certos riscos. Por exemplo Lonkila (1995) refere que a “teoria enraizada” tem sido demasiado enfatizada como substrato teórico à utilização de software para análise de dados qualitativos. Existiria assim o perigo de os investigadores em consequência da sua utilização de análises auxiliadas por computador serem

induzidos a adoptar de forma acrítica um conjunto de estratégias associados com a “teoria enraizada”. O aspecto central desta associação parece residir na importância atribuída à codificação. Na realidade a maioria dos textos sobre computadores e análise de dados qualitativos (por exemplo Weitzman & Miles (1994) dão a entender que o central da teoria enraizada é a codificação e que o principal objectivo destes programas é facilitarem os processos de codificação. Contudo, como referem Coffey, Holbrook e Atkinson (1996) não só a teoria enraizada é mais do que codificação mas também a maioria dos programas disponíveis permitem uma utilização para além da simples codificação e procura de categorias. Desta forma conclui-se que nem a identificação entre teoria enraizada e análise de dados computadorizada necessita de ser automática nem temos de restringir a análise qualitativa a procedimentos de codificação.

Um segundo grupo de críticas resulta da perspectiva que abordagens tais como as resultantes da fenomenologia ou da filosofia hermenêutica têm sobre a investigação interpretativa. Na verdade, estas perspectivas sempre enfatizaram (cf. Giddens, 1976; Denzin & Lincoln, 1994) que a ambiguidade e as questões de contexto têm de ser vistas como aspectos centrais à utilização da linguagem comum. Segundo esta posição, torna-se impossível extrair o sentido de mensagens escritas ou orais em contextos quotidianos sem um “conhecimento tácito” que não pode ser formalizado facilmente (se é que pode). Neste sentido justifica-se uma certa precaução contra a utilização de métodos que tratem a informação através de regras exactas e precisas, completamente independentes do contexto e não contendo ambiguidades. Os computadores correspondem muito facilmente a essa imagem, contudo parece-nos que este argumento só se relaciona com a possibilidade de analisar dados textuais com a ajuda de procedimentos algorítmicos (tal como os dados quantitativos são analisados com a ajuda de algoritmos estatísticos), mas não para as oportunidades de ordenar e estruturar material textual com ajuda de uma tecnologia de base de dados (que é uma das formas de se ver os programas Tipo II).

## BIBLIOGRAFIA

- COFFEY, A, HOLBROOK, B. & ATKINSON, P. (1996) 'Qualitative Data Analysis: Technologies and Representations', *Sociological Research Online*, vol. 1, no.1,  
DENZIN, N. and LINCOLN, Y. (editors) (1994) *Handbook of Qualitative Research*. Thousand Oaks: Sage.  
GIDDENS, A. (1976) *New Rules of Sociological Methods: A Positive Critique of Interpretive Sociologies*. London: Hutchinson.

- GLASER, B. and STRAUSS, A. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine de Gruyter.
- GLASER, B. (1992) *Emergence vs. Forcing: Basics of Grounded Theory Analysis*. Mill Valley: The Sociology Press.
- LEE, R. and FIELDING, N. (editors) (1991) *Using Computers in Qualitative Research*. London: Sage.
- LEE, R. M. and FIELDING, N. G. (1995) 'Users' Experiences of Qualitative Data Analysis Software' in U. Kelle (editor) *Computer-Aided Qualitative Data Analysis: Theory, Methods and Practice*. London: Sage.
- LONKILA, M. (1995) 'Grounded Theory as an Emerging Paradigm for Computer-Assisted Qualitative Data Analysis' in U. Kelle (editor) *Computer-Aided Qualitative Data Analysis: Theory, Methods and Practice*. London: Sage.
- RICHARDS, T. and RICHARDS, L. (1995) 'Using Computers in Qualitative Research' in N. Denzin and Y. Lincoln (editors) *Handbook of Qualitative Research*. Thousand Oaks: Sage.
- STRAUSS, A. and CORBIN, J. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Thousand Oaks: Sage.
- TESCH, R. (1990) *Qualitative Research: Analysis Types and Software Tools*. New York: Falmer Press.
- WEITZMAN, E. e MILES, M. (1995) *Computer Programs for Qualitative Data Analysis*. Thousand Oaks: Sage.