

EMPODERAMENTO DO PESQUISADOR NAS CIÊNCIAS DA SAÚDE



HELOISA DE CARVALHO TORRES
ILKA AFONSO REIS
ADRIANA SILVINA PAGANO
(Organizadoras)

**EMPODERAMENTO
DO PESQUISADOR NAS
CIÊNCIAS DA SAÚDE**

HELOISA DE CARVALHO TORRES
ILKA AFONSO REIS
ADRIANA SILVINA PAGANO

(Organizadoras)

**EMPODERAMENTO
DO PESQUISADOR NAS
CIÊNCIAS DA SAÚDE**

FALE/UFMG

Belo Horizonte-MG, 2015

Conselho Editorial

Célia Maria Magalhães (FALE/UFMG)

Fábio Alves da Silva Junior (FALE/UFMG)

Giacomo Patrocínio Figueredo (ICHS/UFOP)

Janice Sepúlveda Reis (IEP-Santa Casa-BH)

Maria Odete Pereira (EE /UFMG)

Sueli Aparecida Mingoti (ICEX/UFMG)

Valeria de Oliveira (UFSJ)

Projeto gráfico, diagramação e capa

Rita Motta - Editora Tribo da Ilha

1ª Edição – 2015 – Belo Horizonte-MG

Ficha catalográfica elaborada pelos Bibliotecários da Biblioteca FALE/UFMG

E55	<p>Empoderamento do pesquisador nas ciências da saúde / Heloisa de Carvalho Torres, Ilka Afonso Reis, Adriana Silvina Pagano (Organizadoras). - Belo Horizonte : FALE/UFMG, 2015. 250 p.</p> <p>Inclui referências.</p> <p>ISBN:978-85-7758-264-8 (online).</p> <p>1. Ciências médicas – Pesquisa – Metodologia. 2. Métodos e técnicas de pesquisa educacional. 3. Estatística. 4. Interdisciplinaridade. 5. Análise do discurso. I. Torres, Heloisa de Carvalho. II. Reis, Ilka Afonso. III. Pagano, Adriana Silvina. IV. Título.</p> <p style="text-align: right;">CDD : 610</p>
-----	--

FALE – Faculdade de Letras da UFMG

Av. Antônio Carlos, 6627 – Campus Pampulha

Belo Horizonte – Minas Gerais – Brasil – 31270901

www.lettras.ufmg.br

SUMÁRIO

PREFÁCIO	VII
1 O EMPODERAMENTO NAS PRÁTICAS EDUCATIVAS ORIENTADAS À AUTONOMIA NO CUIDADO EM SAÚDE.....	1
<i>Heloisa de Carvalho Torres</i>	
2 A LINGUAGEM NA CONSTRUÇÃO DAS PRÁTICAS EDUCATIVAS NAS CIÊNCIAS DA SAÚDE	19
<i>Adriana S. Pagano</i>	
3 DADOS, INFORMAÇÃO, CONHECIMENTO NAS CIÊNCIAS DA SAÚDE: DO NÃO ESTRUTURADO AO ESTRUTURADO	37
<i>Ilka Afonso Reis</i>	
4 O QUESTIONÁRIO COMO TÉCNICA DE COLETA DE DADOS.....	49
<i>Fernanda Figueredo Chaves</i> <i>Júlia Santos Nunes Rodrigues</i>	
5 TRANSCRIÇÃO DE DADOS VERBAIS	65
<i>Kícila Ferreguetti</i> <i>Júlia Santos Nunes Rodrigues</i>	
6 ANÁLISE DE DADOS NÃO ESTRUTURADOS – MINERAÇÃO DE TEXTOS.....	79
<i>Rodrigo Araújo e Castro</i> <i>Sumaya Giarola Cecílio</i>	
7 ANÁLISE DE DADOS NÃO ESTRUTURADOS COLETADOS DE GRAVAÇÕES DE ÁUDIO E VÍDEO	99
<i>Arthur de Melo Sá</i> <i>Rodrigo Araújo e Castro</i> <i>Sumaya Giarola Cecílio</i> <i>Wanderley Gonçalves dos Santos Junior</i>	

8	ANÁLISE ESTATÍSTICA DESCRITIVA DE DADOS ESTRUTURADOS	115
	<i>Edna Afonso Reis</i>	
	<i>Ilka Afonso Reis</i>	
9	BOAS PRÁTICAS PARA A PESQUISA NAS CIÊNCIAS DA SAÚDE.....	139
	<i>Adriana S. Pagano</i>	
	<i>Heloisa de Carvalho Torres</i>	
	<i>Ilka Afonso Reis</i>	
	<i>Arthur de Melo Sá</i>	
	<i>Edna Afonso Reis</i>	
	<i>Fernanda Figueredo Chaves</i>	
	<i>Júlia Santos Nunes Rodrigues</i>	
	<i>Kícila Ferreguetti</i>	
	<i>Rodrigo Araújo e Castro</i>	
	<i>Sumaya Giarola Cecílio</i>	
	<i>Wanderley Gonçalves dos Santos Junior</i>	
	SOBRE OS AUTORES	143
	ÍNDICE REMISSIVO.....	147

PREFÁCIO

No livro *101 tips for behavior change in diabetes education* [101 dicas para a mudança de comportamento na educação em diabetes], Robert Anderson e seus colegas tecem comentários sobre questionamentos frequentes de educadores em diabetes que os autores foram recolhendo ao longo de sua vivência com esses profissionais. Uma das perguntas que eles respondem, reproduzida a seguir, diz respeito a dúvidas do profissional da área da Saúde sobre seu papel no processo de empoderamento do usuário com diabetes:

Eu deixo que ele decida tudo ele próprio e faça os ajustes de sua medicação. Isso quer dizer que minha prática está baseada no empoderamento?
(ANDERSON *et al.*, 2002, p. 4)¹

No seu comentário, Anderson e colegas, como bons analistas da linguagem, percebem uma contradição na fala do profissional: “eu deixo” sugere que o profissional tem controle sobre o comportamento do usuário; todavia, empoderamento diz respeito a correponsabilidade e autonomia no desenvolvimento de

habilidades e atitudes no cuidado em saúde. Além disso, os autores ressaltam, esse suposto controle que o profissional acredita ter é mera ilusão. Quando o usuário vai embora do Centro de Saúde, é ele quem tem todo o controle para decidir, em última instância, se vai seguir ou não as orientações recebidas.

O comentário de Anderson e colegas é sempre esclarecedor sobre os aspectos da educação em diabetes, mas também se aplica à educação e ao empoderamento em geral, seja de usuários, seja de profissionais, seja de pesquisadores.

Este volume, como o título anuncia, está direcionado aos pesquisadores das Ciências da Saúde e visa explorar o impacto do empoderamento, projetado desde uma instância pontual – o pesquisador – até chegar no usuário do serviço de Saúde, passando pelo profissional da área da Saúde, peça-chave que intermedia a interação dentro e fora do âmbito acadêmico. Parte-se do pressuposto de que o empoderamento do pesquisador conduz à obtenção de resultados mais esclarecedores que poderão informar as práticas educativas do profissional da área da Saúde, corresponsável pela autonomia do usuário do serviço de Saúde.

¹ ANDERSON, Robert *et al.* (Orgs.). **101 tips for behavior change in diabetes education**. Alexandria: American Diabetes Association, 2002.

É nossa convicção que o desenvolvimento da autonomia do pesquisador, no percurso rumo à expertise acadêmica, pode ser propiciado pelo trabalho conjunto e a partilha de conhecimentos, tanto entre colegas de curso como entre pesquisadores de campos disciplinares distintos.

O volume está composto por nove capítulos. Os três primeiros capítulos abordam os aspectos teóricos do processo de empoderamento do pesquisador, do profissional da área da Saúde e do usuário do serviço de Saúde, promovendo a contribuição da Linguística Aplicada para esse processo e da Estatística para aprimoramento das pesquisas. Os capítulos de 4 a 7 abordam aspectos metodológicos relativos à construção de instrumentos para a coleta de dados, à transcrição de textos e ao processamento e análise dos dados deles coletados, e ao tratamento de dados não estruturados coletados de gravações de áudio e vídeo. O capítulo 8 apresenta os princípios básicos da análise estatística descritiva de dados estruturados. Por fim, o capítulo 9 sintetiza, por meio de uma coleção de dicas, boas práticas nas pesquisas nas Ciências da Saúde.

O público alvo deste livro é, primeiramente, pesquisadores que estão iniciando suas

carreiras nas Ciências da Saúde, que desenvolverão toda sua competência e expertise para se tornarem futuros pesquisadores e educadores ou para atuarem como profissionais em contato direto com os usuários dos serviços de Saúde. Embora sejam as Ciências da Saúde o foco das discussões dos capítulos deste volume, temos certeza de que pesquisadores de outras áreas também irão se beneficiar dos conteúdos apresentados.

Interdisciplinaridade e inovação tecnológica são os dois pilares deste livro. Nesse sentido, convidamos os leitores a se engajarem nas suas pesquisas tendo esses dois conceitos sempre em mente. Compartilhando a visão de empoderamento de Anderson e seus colegas, sabemos que, ao ler a última página deste livro, é do leitor a decisão de aprender e se capacitar para ter plena autonomia ao gerir seu percurso acadêmico. Como pesquisadores e educadores, oferecemos aqui subsídios teóricos e recursos metodológicos para empoderar nosso leitor.

Uma boa leitura a todos!

As organizadoras

O EMPODERAMENTO NAS PRÁTICAS EDUCATIVAS ORIENTADAS À AUTONOMIA NO CUIDADO EM SAÚDE

Heloisa de Carvalho Torres

Objetivos do capítulo

- Descrever a abordagem do empoderamento na formação do pesquisador nas Ciências da Saúde;
- Apresentar uma abordagem educativa pautada na busca de sinais da efetividade, aliada a uma metodologia inovadora e tendo por base a construção coletiva do conhecimento sobre os principais aspectos da condição crônica em diabetes;
- Propor ao pesquisador novato a elaboração de um plano de metas para que ele se mantenha no controle das ações e desenvolva a capacidade de inovar na busca do conhecimento ao incorporar novas tecnologias de coleta e análise de dados.

1 Introdução

Este capítulo apresenta a noção de empoderamento sob a perspectiva da formação do pesquisador e educador nas Ciências da Saúde, partindo-se do pressuposto de que a capacitação do pesquisador e educador pode alavancar as pesquisas, gerando resultados que contribuam para a formação do profissional da área da Saúde, com consequente impacto no empoderamento do usuário do Centro de Saúde,

por meio de práticas educativas orientadas para a autonomia do cuidado.

O empoderamento do pesquisador e educador está pautado pelo desenvolvimento de técnicas de pesquisa aprimoradas, baseadas na interdisciplinaridade dos saberes e na inovação tecnológica. Já o empoderamento do usuário do Centro de Saúde está pautado pela educação e o cuidado centrados nele próprio, em um processo resultante de corresponsabilização juntamente com o profissional da área

da Saúde e construído por meio de diferentes estratégias educativas. Este processo requer que o profissional e o usuário reconheçam suas tarefas específicas na gestão da condição do diabetes e se comprometam com elas. Para isso, devem-se levar em consideração os aspectos psicossociais, comportamentais e clínicos que envolvem a gestão do cuidado em saúde. Se as pessoas com diabetes gerirem eficazmente a sua condição, podem reduzir o risco de complicações no longo prazo e aumentar a qualidade e a duração de suas vidas (ANDERSON; FUNNELL, 2010).

Small *et al.* (2013, p. 2) definem empoderamento como:

Um processo de desenvolvimento de capacidades, ou o resultado dele, suscitado pela interação com o profissional da área da saúde e pelo compartilhamento de informação sobre a doença, que propicia o sentimento de autocontrole do paciente, sua habilidade para lidar com sua condição e conseguir efetuar mudanças nela.¹

No caso do diabetes Mellitus, entendemos que é preciso mudar o paradigma tradicional, a começar pelo Sistema de Saúde, que não deve tratar o usuário como doente, mas como pessoa que possui uma condição diferenciada que demanda um plano de autocuidados. O usuário do Centro de Saúde com a condição de diabetes não está doente; ele pode ser uma pessoa saudável, a quem basta seguir um plano alimentar e praticar atividade física, manter seu peso e seguir o tratamento com o profissional da área da Saúde, podendo levar uma vida ativa e satisfatória e reduzir substancialmente os riscos de complicações do diabetes.

A Figura 1 apresenta o modelo conceitual do empoderamento na formação

do pesquisador e educador, do profissional da área da Saúde e do usuário. Vemos nela que a capacitação do pesquisador para conduzir uma pesquisa demanda a mobilização de conhecimentos e habilidades que possibilitarão a obtenção de resultados passíveis de serem incorporados na formação do profissional da área da Saúde. Já o profissional desenvolverá habilidades e atitudes necessárias para a concretização das práticas educativas a serem realizadas junto ao usuário do Centro de Saúde com condição de diabetes, visando a efetividade da educação do cuidado em saúde. O profissional empoderado, por sua vez, é capaz de conduzir práticas educativas que efetivamente auxiliarão o usuário a elaborar e cumprir um plano de metas, reconhecendo suas habilidades e melhorando os aspectos comportamentais, psicossociais e clínicos da condição de diabetes. O usuário empoderado adquire autonomia, o que favorece o cuidar de si no controle sobre sua condição.

O empoderamento do pesquisador se dá no âmbito acadêmico, por meio de sua interação, não somente com outros pesquisadores, mas também com profissionais da área da Saúde e usuários, sendo sua formação acadêmica, suas relações interpessoais, experiências, e prática do dia-a-dia importantes no desenvolvimento de sua expertise como pesquisador. Outro aspecto importante é sua disposição para a interdisciplinaridade e a inovação tecnológica, o que mobilizará outros atores para trabalharem nas práticas educativas, compartilhando as responsabilidades e implicações de suas ações. Essa disposição muitas vezes está manifesta em comportamentos como a disponibilidade de escutar e aceitar opiniões diferentes e a capacidade de inovar e construir conhecimentos, bem como de modificar atitudes de maneira conjunta na resolução de problemas. O empoderamento é um processo pelo qual os indivíduos experimentam um forte sentimento de controle e autoeficácia, sendo uma capacitação psicológica definida como um processo motivacional, entrelaçada com a construção da autonomia (BABENKO-MOUL, 2015; SMALL *et al.*, 2013).

¹ Nossa tradução de: “*An enabling process or outcome arising from communication with the health care professional and a mutual sharing of resources over information relating to illness, which enhances the patient’s feeling of control, self-efficacy, coping abilities and ability to achieve change over their condition.*”

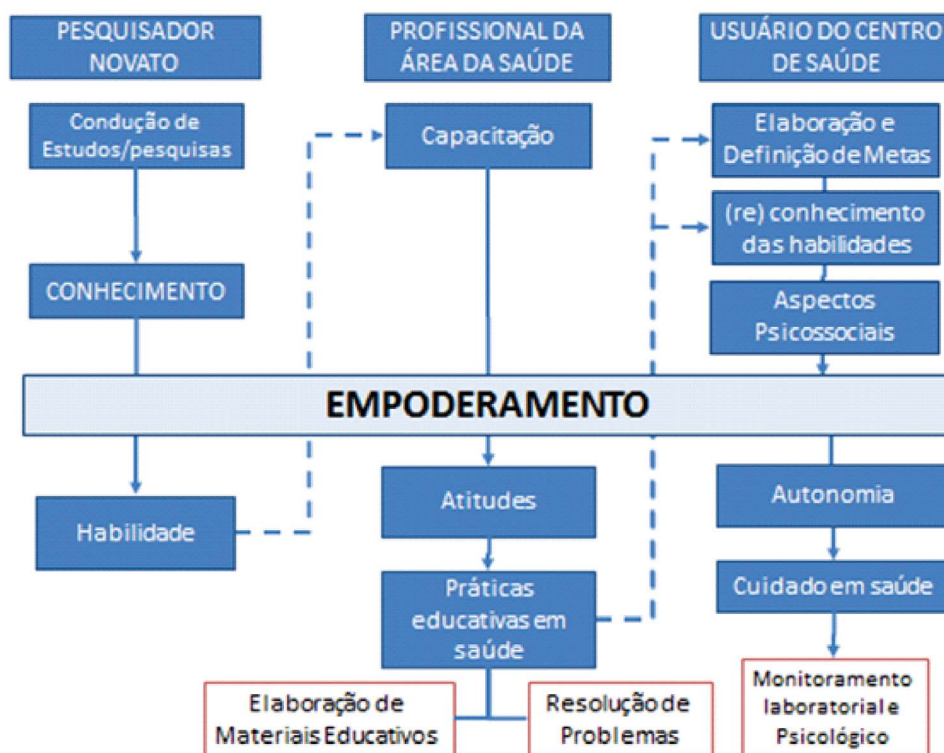


Figura 1 - Modelo conceitual do empoderamento: pesquisador – profissional da área da Saúde – usuário

Fonte: Elaborada para fins deste estudo

Nessa perspectiva, o empoderamento do pesquisador nas Ciências da Saúde está ligado a um processo que capacita o pesquisador novato no curso de graduação e pós-graduação para o desenvolvimento de expertise por meio do desenvolvimento de leitura crítica e o aprendizado de métodos e técnicas. Sob a supervisão do educador, o pesquisador novato empreende a revisão de literatura e ganha capacidade para planejar um desenho robusto de pesquisa com sua respectiva coleta e análise de dados. O pesquisador desenvolve, ainda, habilidades necessárias à escrita acadêmica e à elaboração de instrumentos para avaliar as práticas educativas (SANTOS *et al.*, 2007; MINAYO, 2008).

O desenvolvimento da escrita acadêmica está intrinsecamente relacionado, como veremos no Capítulo 2, com o desenvolvimento da percepção do pesquisador novato sobre o papel fundamental da linguagem na vida acadêmica, em geral, e nas práticas educativas,

em particular. A elaboração de instrumentos, como veremos no Capítulo 4, requer um trabalho conceitual que envolve a definição cuidadosa dos significados e o alcance dos termos a serem utilizados.

Ao planejar o procedimento de coleta de dados, o pesquisador deve pensar em métodos que garantam indicadores confiáveis. Na área da Saúde, um número crescente de questionários e escalas está atualmente disponível com este objetivo (KEBIAN *et al.*, 2012; MINAYO, 2008; MONTEIRO *et al.*, 2005; TORRES *et al.*, 2010; TOL *et al.*, 2013). A participação de especialistas na análise de instrumentos pretende apurar e aperfeiçoar o conteúdo dos itens e o formato do construto como um todo. O papel desse comitê de juízes é avaliar todas as versões do instrumento e desenvolver o que será considerado na versão pré-final para testes de campo (TIMM; RODRIGUES; MACHADO, 2013; TOL *et al.*, 2013).

Instrumentos são elaborados com o intuito de guiar as práticas educativas em saúde e torná-las exequíveis, além de conduzir intervenções, visando o autocuidado e empoderamento do usuário. A incorporação de novas tecnologias é desejável, como, por exemplo, a utilização de uma tecnologia de consulta por via eletrônica, por meio de questionário on-line, apontada como ferramenta eficaz e apropriada nas práticas de intervenção em saúde pública para a avaliação do autocuidado e empoderamento antes e depois das intervenções educativas (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2005; ANDERSON; FUNNELL, 2010; TORRES; PEREIRA; ALEXANDRE, 2011). Essas informações poderão prover subsídios para planejamento, execução e avaliação das práticas educativas na condição crônica do diabetes.

Para melhorar a qualidade do cuidado e garantir a padronização dos dados coletados, a fim de compará-los de forma segura, há a necessidade de utilização de instrumentos que permitam explorar com clareza o fenômeno a ser investigado. No entanto, no Brasil, faltam instrumentos apropriados à nossa realidade que possam ser empregados nas práticas educativas dos profissionais da área da Saúde. A elaboração de instrumento não é tarefa fácil, por requerer atenção na construção de itens, que devem ser claros e abrangentes, além de se ter a necessidade de adequá-los aos novos contextos culturais (ALEXANDRE; COLUCI, 2011).

Como veremos no Capítulo 2, a linguagem é um dos sistemas semióticos que dá expressão aos significados que configuram a cultura e, para examinar esses significados,

precisamos considerar o contexto no qual a linguagem opera. Por isso, uma abordagem interdisciplinar que incorpora subsídios da Linguística Aplicada elucidada aspectos do uso da língua importantes para se obter, não apenas a adaptação cultural de instrumentos (GUILLEMIN; BOMBARDIER; BEATON, 1993), como também a adequação dos textos aos distintos contextos de uso. Além disso, como veremos no Capítulo 4, estudos têm demonstrado que, durante o processo de elaboração de um instrumento, devem-se identificar os domínios, formar itens, construir o instrumento, bem como criar enunciados representativos desse conhecimento.

As seções seguintes trarão mais detalhes sobre o empoderamento do pesquisador nas Ciências da Saúde, profissional da área da Saúde e usuário do Centro de Saúde.

2 Empoderamento do pesquisador nas Ciências da Saúde

O empoderamento do pesquisador novato nas Ciências da Saúde, no período de sua formação em nível de graduação, pode ser propiciado por meio de sua participação em trabalhos acadêmicos, projetos de extensão e pesquisa. Para ilustrar este processo, descrevemos a experiência de um pesquisador novato na execução de seu trabalho de conclusão de curso, o qual envolveu a elaboração e adaptação de um questionário sobre conhecimento em Doença Falciforme (GARCIA, 2015). A Figura 2 sintetiza as etapas percorridas pelo pesquisador na elaboração do instrumento.

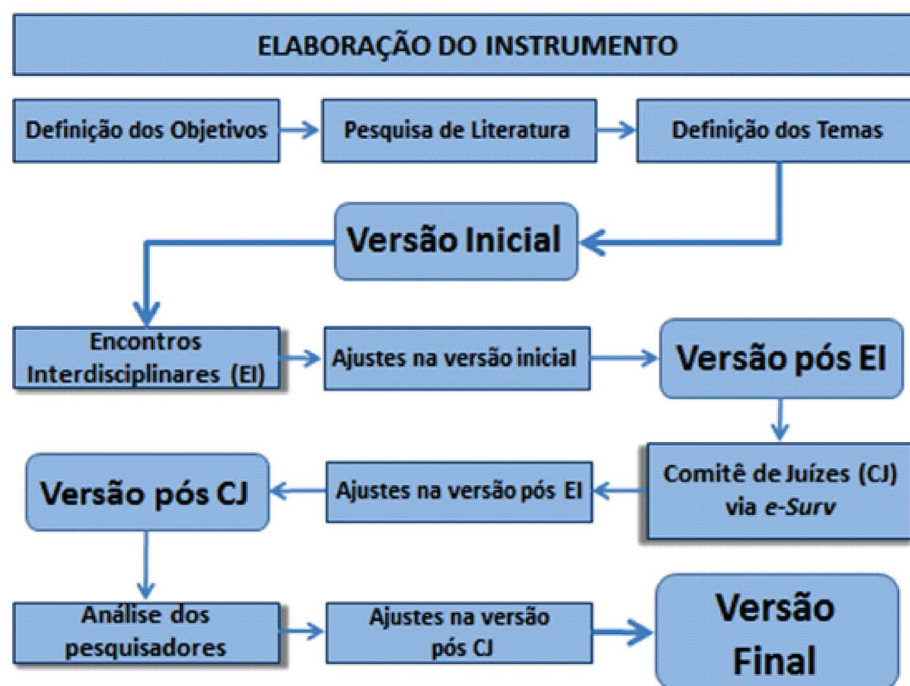


Figura 2 – Fases da elaboração de instrumento em saúde, 2015

Fonte: Elaborada para fins deste estudo

Cada uma das fases envolveu as seguintes etapas:

- 1) Elaboração de instrumento – Realização de uma revisão integrativa da literatura, buscando-se conhecer os fatores condicionantes e específicos da doença falciforme.
- 2) Encontros interdisciplinares – Realização de encontros presenciais e online com a finalidade de discutir, avaliar e ajustar o instrumento.
- 3) Comitê de Juízes- Seleção de uma amostra de conveniência de profissionais da área da Saúde com experiência no atendimento a pessoas com doença Falciforme para avaliar a versão elaborada.
- 4) Versão Síntese Final – Realização de encontros interdisciplinares para a análise das sugestões e elaboração da versão síntese final do instrumento.

A capacitação do pesquisador em conduzir a investigação no processo de elaboração e adaptação do instrumento, a mobilização

dos conhecimentos e habilidades adquiridas na leitura e escrita do trabalho, a incorporação de novas tecnologias de coleta e análise de dados, todos estes componentes favorecem o desenvolvimento de expertise do pesquisador. Os encontros interdisciplinares, em particular, proporcionam o compartilhamento do conhecimento das áreas afins, como a Enfermagem, a Linguística Aplicada e a Estatística, e as implicações de suas ações e opiniões diferentes na capacidade de inovar e construir os conhecimentos. O empoderamento do pesquisador novato estimula a aprendizagem, a compreensão da leitura, a aquisição de habilidades metodológicas da pesquisa e o exercício da escrita, além da capacidade de disseminação dos resultados por meio de apresentações em eventos científicos e publicação. Estudos comprovam que as escolas que preparam os alunos para a investigação durante a graduação apresentam resultados positivos no processo seletivo de pós-graduação (BABENKO-MOUL, 2015; EPSTEIN et al., 2013).

A Figura 3 é uma representação do papel do empoderamento na formação do pesquisador nas Ciências da Saúde. Vemos nela a integração entre pesquisador novato do curso de graduação, mestrado e doutorado que, ao longo do percurso, têm a oportunidade de desenvolver suas habilidades de leitura, escrita e apresentação de trabalhos, assim como ampliar sua capacidade de inovar na busca do conhecimento ao incorporar novas tecnologias de coleta e análise de dados. No caso dos cursos de graduação e pós-graduação em Enfermagem

da UFMG, o empoderamento do pesquisador nas Ciências da Saúde pode ser propiciado por meio de sua participação nas atividades do Núcleo de Gestão, Educação e Avaliação em Saúde (NUGEAS). Pesquisadores e docentes das áreas de Enfermagem, Linguística Aplicada e Estatística trabalham a interdisciplinaridade e inovação tecnológica, por meio da integração de graduandos, orientandos de iniciação científica e extensão, mestrandos e doutorandos. Isso favorece a aquisição de habilidades e incentiva o trabalho de pesquisa.



Figura 3 – Empoderamento do pesquisador nas Ciências da Saúde: integração e interdisciplinaridade

Fonte: Elaborada para fins deste estudo

A metodologia interdisciplinar na formação do pesquisador se pauta por critérios de reciprocidade e complementaridade entre os pesquisadores nos diferentes estágios da carreira.

A reciprocidade envolve o planejamento e a execução de sessões interativas online e atividades de trabalho, as quais propiciam o intercâmbio de ideias e referenciais teóricos, bem como de experiências com metodologias e recursos tecnológicos. A integração entre

os pesquisadores e docentes das áreas de Enfermagem, Linguística Aplicada e Estatística no projeto Empoder@ – *Protótipo conceitual e metodológico para avaliação de intervenções orientadas ao autocuidado em diabetes*, desenvolvido pelo NUGEAS, em parceria com o Laboratório Experimental de Tradução (LETRA) da Faculdade de Letras e o Departamento de Estatística da UFMG, propicia o desenvolvimento de trabalhos de monografia, iniciação científica, mestrado e doutorado; e capacita

profissionais através de workshops e cursos de iniciação e disciplinas de graduação e pós-graduação. Ao longo dos últimos anos, o NUGEDAS vem incorporando uma perspectiva interdisciplinar e desenvolvendo uma metodologia pioneira e inovadora para a investigação do empoderamento nas práticas de autocuidado e meta-reflexão da pessoa com diabetes sobre seu comportamento cotidiano, associado não somente à dieta e à atividade física, mas também à percepção de seu papel no planejamento de hábitos e condutas que propiciem seu autocuidado.

Babenko-Mould (2015) destaca que a abordagem interdisciplinar na pesquisa encoraja o pesquisador novato na prática profissional, tornando-o um participante ativo no processo de investigação. O avanço do pesquisador requer a integração de teoria, prática profissional e pesquisa. Para alcançar essa integração, faz-se necessário o trabalho interdisciplinar entre as áreas afins do objeto a ser estudado. O trabalho em conjunto desenvolve habilidades e incentivo para a pesquisa, parceria com outros pesquisadores, além de outras visões do mundo. Consequentemente, o empoderamento dos participantes da pesquisa inclui habilidades nas questões de comunicação. Os participantes devem estabelecer objetivo no curto e longo prazo, horário de encontros rotina estabelecidos na agenda. Todos os membros da equipe devem participar dos encontros e rotina as principais tecnologias de comunicação (videoconferência, ligações por plataformas web, entre outros).

A reciprocidade na formação do pesquisador envolve o planejamento e a execução da prática educativa, as quais propiciam o intercâmbio de ideias e referenciais teóricos, bem como de experiências com metodologias e recursos tecnológicos (TORRES *et al.*, 2009; SANTOS; TORRES, 2012). A utilização de tecnologias inovadoras como veremos nos

capítulos de 5 a 7, apresentando a coleta de dados via plataforma *web*, anotação eletrônica de textos, vídeo e áudio e mineração de textos, juntamente com a congregação de profissionais de diferentes áreas, potencializa a análise e interpretação dos dados.

A competência do pesquisador no processo de investigação, em especial, nas práticas em saúde, pode ser entendida como a capacidade que o pesquisador tem de realizar investigação, além de saber agir com responsabilidade, de maneira reconhecida, implicando na mobilização de conhecimentos e habilidades, agregando, portanto, valor à organização e ao educador (BABENKO-MOULD, 2015). Pode ser descrita como “tomar iniciativa” e “assumir responsabilidade”, por parte do educador, nas diversas situações de investigação, caracterizando-se como um entendimento prático de situações, apoiado nos conhecimentos adquiridos na trajetória profissional, passíveis de mudanças conforme se alteram as situações.

3 Empoderamento do profissional da área da Saúde

A formação acadêmica tem incorporado a abordagem do empoderamento e a competência do profissional da área da Saúde para atuar nas práticas educativas orientadas à autonomia no cuidado em saúde, buscando habilidades e atitudes necessárias para a concretização das ações educativas a serem realizadas junto aos usuários. No caso do curso de Enfermagem da Universidade Federal de Minas Gerais (UFMG), a disciplina do curso de graduação intitulada “Práticas educativas em doenças crônicas: interdisciplinaridade e inovação” tem a finalidade de debater o empoderamento em condições crônicas com base na educação do usuário. O profissional estimula o usuário a contar suas histórias de vida,

buscando explorar os problemas, mantém uma escuta qualificada e faz perguntas que ajudam o usuário a entender como é viver com a condição de diabetes. Somente depois que um problema é totalmente explorado e esclarecido é que se deve passar para a identificação de estratégias ou metas para sua resolução. Para ser efetiva, a prática educativa em saúde não deve somente focar um problema, mas tem que se adequar à pessoa que tem esse problema (YANG; HSUE; LOU, 2015).

No Programa de Pós-Graduação em Enfermagem da UFMG, a disciplina “Educação em Saúde: comunicação, saúde e educação para o autocuidado” explora as novas perspectivas da educação em saúde no cuidado-de-si, com ênfase na abordagem do empoderamento, o que facilita a relação entre o usuário e profissional da área da Saúde baseada na confiança e compreensão. Tal relação irá melhorar a comunicação entre usuário e profissional e facilitará o compartilhamento da tomada de decisões. Ao se envolver na sua investigação em nível de mestrado e doutorado, o pesquisador desenvolve conhecimentos que terão impacto na formação dos profissionais da área da Saúde para ajudar os usuários a identificar e gerir os desafios emocionais de sua condição, como

redução da depressão, angústia do diabetes e outras condições psicológicas.

No âmbito do projeto Empoder@, anteriormente citado, foi realizada a tradução, adaptação cultural e validação de um instrumento que examina o empoderamento do profissional da área da Saúde: o *Diabetes Attitudes Scale – third version*, (DAS-3). O instrumento DAS-3 permite conhecer as atitudes do profissional da área da Saúde na prática educativa em diabetes realizada junto ao usuário do Centro de Saúde, além de identificar as necessidades de capacitação do profissional, verificar o impacto psicossocial da condição do diabetes na vida da pessoa, avaliar as atitudes referentes à autonomia da pessoa, verificar as atitudes do profissional em relação à pessoa com diabetes como o principal tomador de decisão a respeito do cuidar de si (FUNNELL; ANDERSON, 2004).

Conhecer as atitudes do profissional proporciona o planejamento e a sistematização da prática educativa em diabetes, integrando os conceitos de autonomia da pessoa durante o tratamento. As atitudes do profissional podem ser avaliadas pelas respostas dadas pelo profissional diante dos problemas elencados simulando afirmações do usuário conforme ilustrado no Quadro 1:

Quadro 1 – Respostas do profissional da área da Saúde aos problemas apresentados pelo usuário

Como você, profissional da área da Saúde, responderia as afirmações dos usuários do Centro de Saúde abaixo.

1. **Usuário:** “Eu tive muita dificuldade em seguir o plano alimentar saudável durante os feriados.”

Profissional:

2. **Usuário:** “Eu sei que preciso tomar minha insulina e sei o que acontece quando eu não tomo, mas, às vezes, eu simplesmente decido não tomá-la.”

Profissional:

3. **Usuário:** “Às vezes, parece que minha vida inteira se resume em cuidar da minha condição do diabetes.”

Profissional:

5. **Usuário:** “Eu realmente odeio testar meu sangue 4 ou 5 vezes por dia.”

Profissional:

6. **Usuário:** “Há certos dias em que eu sinto que simplesmente não vou conseguir continuar cuidando da minha condição do diabetes.”

Profissional:

A condição do diabetes exige cuidados contínuos, o que pode representar um desafio, quer para o usuário que convive com essa condição, quer para o profissional da área da Saúde envolvido na orientação. Sabe-se que as estratégias psicossociais e comportamentais com enfoque na colaboração mútua são as chaves para colaborar com esse usuário no desafio diário com os cuidados de saúde.

O desenvolvimento de competências por parte do profissional, tais como: empatia, saber escutar, não expressar julgamentos de

valores, ser curioso, saber perguntar, conduzir o diálogo de forma dinâmica e interativa é fundamental para chegar ao ponto chave do problema e para conduzir o usuário ao processo de mudança comportamental por meio da motivação. O exemplo a seguir, extraído e adaptado do roteiro para fortalecimento das práticas de autocuidado baseadas na obra de (ANDERSON *et al.*, 2002), auxiliará o profissional da área da Saúde na competência da sua prática em saúde.

Quadro 2 – Roteiro para fortalecimento das práticas de Autocuidado

1) Como é para você viver com a condição do diabetes?

- Você poderia me falar um pouco mais sobre isso (explorar o que for falado; positivo ou negativo).
- Qual o seu entendimento sobre condição?
- Você pode me explicar um pouco mais isso?

OBJETIVO: Discutir as percepções e preocupações do usuário a respeito de viver e cuidar da condição em seu cotidiano.

2) Como as situações descritas acima (as dificuldades e as insatisfações com a condição) fazem você se sentir?

- Você se sente triste, com raiva, confuso, preocupado, frustrado, depressivo, curioso, esperançoso... etc.?
- Em que momentos você tem estes sentimentos? Pode me explicar melhor?

OBJETIVO: Abordar os sentimentos que podem auxiliar e/ou prejudicar o fortalecimento das práticas de autocuidado.

3) Como você acha que pode modificar esta situação para que se sinta melhor?

- Em quanto tempo você gostaria de modificar esta situação?
- O que você acha que irá acontecer com você se não fizer nada para modificar esta situação?
- Como você irá sentir se as coisas não mudarem?

OBJETIVO: Auxiliar o usuário a identificar como as coisas seriam caso ele optasse por modificar comportamentos.

4) Você está disposto(a) a realizar alguma ação para melhorar esta situação?

- Quanto esta situação que deseja modificar é importante para você?
- Você se lembra de alguma mudança que teve que fazer para mudar sua condição do diabetes?
- Você obteve resultados?

OBJETIVO: Auxiliar o usuário a estabelecer prioridade e a corresponsabilizar pela mudança que almeja.

5) Quais são os passos que você pode dar para alcançar o que deseja?

- O que você pode fazer para resolver este problema?
- Você acha que existe alguma barreira que possa atrapalhá-la(o) a alcançar seus resultados?
- Existem outras pessoas que possam ajudá-la(o) a seguir suas metas?

OBJETIVO: Ajudar o usuário a desenvolver um plano de autocuidado específico para sua realidade e objetivos.

De forma complementar, o recurso de empoderamento no contexto da condição crônica de saúde refere-se à autonomia do usuário do Centro de Saúde em ter a habilidade de tomar as decisões diárias para seu cuidado, reconhecendo-se a importância da responsabilização, na afirmação de que usuário é o responsável pelas escolhas e também pelas consequências no cuidado da saúde, acreditando-se na viabilidade de que pode ser uma pessoa saudável convivendo com essa condição crônica de saúde (ANDERSON; FUNNELL, 2010; MANTWILL *et al.*, 2015). A corresponsabilização é decorrente de um resultado da comunicação entre o profissional da área da Saúde e usuário, e uma partilha mútua de recursos sobre informações relativas à condição crônica, o que aumenta a sensação de controle, autoeficácia do usuário, habilidades de enfrentamento e capacidade para alcançar a mudança sobre sua condição de saúde (BANDURA, 2004; SMALL *et al.*, 2013).

A maneira como esse profissional da saúde recebe a pessoa, ou seja, o acolhimento desse usuário, a sua criatividade na condução desse processo, assim como a flexibilidade, são atitudes que podem definir a condução da prática educativa para torná-la efetiva (SANTOS; TORRES, 2012). O reconhecimento da corresponsabilização, assim como da necessidade de desenvolvimento de autonomia e protagonismo da pessoa com diabetes, a partir do estabelecimento de vínculos solidários entre profissional e usuário, apresenta o caminho para o desenvolvimento de atitudes profissionais para uma prática educativa prazerosa.

As práticas pautadas no empoderamento são direcionadas aos profissionais e usuários do Centro de Saúde e discutem os temas sobre as experiências e dificuldades do autocuidado, os sentimentos relacionados ao viver com a condição do diabetes, a elaboração de metas para o autocuidado selecionadas pelos

usuários, sendo um convite para um compromisso com a resolução de problemas. É importante que o profissional da área da Saúde esteja empoderado para as práticas educativas, e o usuário a gestão do cuidado em saúde.

4 Empoderamento do usuário na gestão do cuidado em saúde: práticas participativas e inovadoras

As práticas orientadas ao cuidado em saúde no diabetes Mellitus podem ser baseadas no instrumento Protocolo de Mudança de Comportamento (FUNNELL; ANDERSON, 2004; CHAVES, 2014), que contempla passos para estimular a reflexão dos usuários por meio de uma sequência lógica de questões não estruturadas: 1) definição do problema; 2) identificação e abordagem dos sentimentos; 3) definição de metas; 4) elaboração do plano de cuidados para conquista da(s) meta(s) (Meu Plano Inteligente); 5) Avaliação e experiência do usuário sobre o plano de cuidados. Nas práticas apresentadas neste capítulo, foram mencionadas algumas dessas estratégias como resolução de problemas com plano individualizado para alguns casos e definição de metas com avaliação do plano de cuidados. Todas as ações são acompanhadas presencialmente e por ligação telefônica e por telefone, sempre com a mesma ideia de que a meta a ser realizada é escolha do usuário. Ressalta-se que esse plano pode ser usado em outras condições crônicas.

O estabelecimento de metas é fundamental e não se refere apenas a metas clínicas, mas também psicológicas, pois são as metas daquilo que é o mais importante para o usuário lidar no momento. Estabelecer metas colabora para que o usuário se sinta mais independente, uma vez que essas são elaboradas por eles em parceria com os profissionais da

área da Saúde, o que ajuda a manter a motivação pela avaliação periódica dessas metas e pode aumentar a competência sobre a resolução de problemas (ANDERSON *et al.*, 2002; KEBIAN *et al.*, 2012). Esse estabelecimento de metas auxilia no processo de aumentar a autonomia por fornecer um *feedback* sobre seu comportamento de cuidados com a saúde e manutenção da motivação. É importante ressaltar que, no processo de definição de metas, o foco não é o sucesso ou a falha, mas a aprendizagem que ocorre como um resultado da experiência vivida (MANTWILL *et al.*, 2015; BODENHEIMER; HANDLEY, 2009; ANDERSON; FUNNELL, 2005).

A utilização do protocolo é viabilizada e potencializada quando se utilizam processos educacionais baseados no diálogo, na escuta qualificada e na problematização, visando-se ao auto reconhecimento de necessidade de mudanças de comportamento pelo usuário que se encontra com uma condição crônica (FREIRE, 2002; ANDERSON; FUNNELL, 2005; TORRES *et al.*, 2010; HOLMAN; LORIG, 2000). A problematização baseia-se no uso das experiências vividas ao longo da vida, nos diálogos e na construção das metas. As práticas são executadas por meio de (i) interlocução coletiva, (ii) visita domiciliar e (iii) ligação telefônica por meio de software de comunicação pela Internet, encorajando os usuários a aderirem ao autocuidado, mediante a educação recebida, que está relacionada aos conhecimentos e atitudes psicológicas e à mudança de comportamento para o cumprimento de metas.

Interlocução Coletiva: os usuários compartilham suas experiências e aprendem uns com os outros, o que proporciona a eles a capacidade de identificar e gerir os desafios emocionais de sua condição. Consiste em atividades em pequenos grupos que discutem as

necessidades dos usuários em seu cotidiano em relação aos sentimentos que podem prejudicar as práticas de autocuidado. Para discutir os temas, são realizados de um a três encontros, com duas horas de duração, em média, para cada encontro. Em todos os encontros, há a presença de dois profissionais, por exemplo, uma enfermeira e uma nutricionista, que conduzem o processo educativo de forma interdisciplinar, contribuindo para a interação e o reforço de conteúdos e levando o usuário a se beneficiar de uma mudança de comportamento e a se conscientizar de que suas ações fazem diferença no tratamento. Os temas podem ser apresentados por meio de dinâmicas lúdicas e interativas.

Os profissionais da área da Saúde (médico, enfermeiro, fisioterapeuta, educador físico, terapeuta ocupacional e nutricionista) são orientados quanto à postura, linguagem e forma de comunicação. Profissionais motivados e capacitados apresentam condições de ajudar os usuários a transpor as barreiras que impedem a melhora do autocuidado e disponibilizam condições para o aprendizado.

Os temas abordados nos encontros (Quadro 3) objetivam problematizar o conhecimento e atitudes associados à condição do diabetes. Dessa forma, os profissionais da área da Saúde se tornam facilitadores durante a realização dos encontros, estimulando os usuários a compartilharem suas experiências e a refletirem sobre as práticas de autocuidado. São desenvolvidas dinâmicas lúdicas e interativas visando à participação crítica e reflexiva dos usuários. As vertentes como alimentação saudável, prática de atividade física e planejamento de metas são inseridas nas estratégias educativas por meio de técnicas facilitadoras da abordagem com os usuários, como ilustrado no Quadro 3.

Quadro 3 – Descrição das atividades de Interlocução Coletiva, Belo Horizonte, 2015

	Objetivos	Técnicas facilitadoras
INTERLOCUÇÃO COLETIVA	<ul style="list-style-type: none"> • Discutir os sentimentos e emoções associados com a condição crônica do diabetes <i>Mellitus</i>; • Perceber as sensações e emoções na descoberta de características/cuidados em comum; • Levar o usuário a ter o conhecimento e o cuidado de si. 	<ul style="list-style-type: none"> • Dinâmicas: “Qual figura me identifica”; • “Como me sinto hoje”; “Transporte do açúcar”; “caixinha de surpresas”; “Fotos de artistas que são diagnosticados com condição crônica”; “Brincadeira da verdade”.
	<ul style="list-style-type: none"> • Orientar sobre o planejamento alimentar. 	<ul style="list-style-type: none"> • Dinâmicas: “Construindo cardápios para cada refeição”; “Vestindo a camisa”.
	<ul style="list-style-type: none"> • Incentivar o cuidado em saúde: alimentação saudável, prática da atividade física, uso correto da medicação e monitoramento da glicose. 	<ul style="list-style-type: none"> • Dinâmica: Guia de orientação sobre as práticas de autocuidado;
	<ul style="list-style-type: none"> • Definir a meta: O que; quando; onde; e como vai fazer; • Descrever o que pode atrapalhar a conquista da(s) sua(s) meta(s) e a confiança para alcançar a(s) sua(s) meta(s); • Discutir com o usuário como foi seguir o plano, experiência e verificar o ele faria diferente na próxima vez. 	

Atendimento Domiciliar: valoriza a troca de experiências e saberes, aumentando o vínculo entre usuário e profissional e possibilitando orientações a partir da compreensão do contexto de vida e a abordagem complicações emocionais experimentadas pelo usuário. A estratégia atendimento domiciliar permite levar educação para usuários com dificuldade de acesso ao Centro de Saúde. O método dialógico facilita o entendimento do usuário e seu envolvimento nas discussões (FREIRE, 2002). A sistematização do atendimento, por sua vez, visa contemplar as necessidades do usuário, promover a sua autonomia durante o processo de ensino-aprendizagem para o cuidado em

saúde e conduzir a um atendimento integral e humanizado.

As orientações para o usuário seguem um roteiro de reforço das práticas de autocuidado (Quadro 4). Os objetivos são: 1) discutir as necessidades enfrentadas pelos usuários no dia-a-dia com relação a sua condição; 2) abordar os sentimentos que podem prejudicar as práticas de autocuidado. Cada atendimento tem a duração de 45 minutos, sendo de um a dois atendimentos para cada usuário, que são previamente agendadas por meio de ligações telefônicas, conforme o interesse dos usuários e executadas por profissionais da área da Saúde (enfermeiros, fisioterapeuta e nutricionista).

Quadro 4 – Descrição das atividades de atendimento domiciliar, Belo Horizonte, 2015

	Objetivos	Técnicas facilitadoras
ATENDIMENTO DOMICILIAR	<ul style="list-style-type: none"> • Compreender as barreiras para a prática do autocuidado; • Discutir os sentimentos e emoções associados com a condição crônica; como o corpo se ajusta com e sem a condição crônica; • Orientar o planejamento alimentar; • Elaborar e cumprir as metas para o autocuidado: O que; quando; onde; e como você irá fazer? • Descrever o que pode atrapalhar a conquista da(s) sua(s) meta(s) e a confiança para alcançar a(s) sua(s) meta(s); • Discutir com o usuário como • foi seguir o plano, experiência e verificar o que ele faria diferente na próxima vez. 	<ul style="list-style-type: none"> • Roteiro de fortalecimento das práticas de autocuidado e plano de ação para cuidado; • Dinâmica “Compreendendo o cuidado em saúde na condição crônica” e “Alimentação saudável e atividade física”.

Ligação telefônica por meio de software de comunicação pela Internet: A interação mediada pelo computador é um veículo capaz de fornecer orientação ao usuário do Centro de Saúde sobre medidas de prevenção e controle da condição crônica do diabetes, como incentivar o autocuidado e a prática de hábitos saudáveis.

Observa-se que o usuário possui a necessidade de falar sobre as suas dificuldades nas práticas de autocuidado e aproveita o momento da ligação telefônica para expressar seus sentimentos de angústia e desafio para a realização das práticas de atividade física e seguir o plano alimentar saudável. A valorização dos sentimentos vivenciados e do comportamento dos usuários em relação à condição do diabetes é uma resposta importante para modificação das atitudes relacionadas ao autocuidado.

Em síntese, a interação mediada pelo computador é capaz de atingir boa performance, em termos de educação e comunicação para o autocuidado. Entretanto, é preciso definir

melhor a aplicabilidade, a sustentabilidade, a continuidade, e a aceitação do usuário e dos órgãos públicos e privados. A ampliação das práticas educacionais e comportamentais tem por meta envolver e mobilizar, desenvolver habilidades e fortalecer as atividades educativas para o autocuidado requerido pelo usuário.

A efetividade da proposta é dada em grande parte pela escolha teórico-metodológica, nesse caso, centrada em uma prática educativa dialógica, em que se incluam outros conhecimentos e saberes para além do conhecimento técnico e normativo dos profissionais. Esse componente segue um roteiro estruturado abordando questões relacionadas ao plano alimentar, às atividades físicas e aos sentimentos, barreiras e o cumprimento de metas. As ligações são feitas por um profissional da área da Saúde com duração média de 25 minutos. Os dados são registrados em uma planilha eletrônica para consulta futura. Os temas abordados e técnicas utilizadas na estratégia educativa seguem detalhadas no Quadro 5.

Perguntas facilitadoras
<ul style="list-style-type: none">• Você consegue dizer como se sente sobre a sua saúde?• Como você se sente com essa situação de ter de cuidar da sua saúde?• Definição da meta: O que; quando; onde; e como você irá fazer?• O que você acha que pode atrapalhar a conquista da(s) sua(s) meta(s)?• Como você vê sua confiança para alcançar a(s) sua(s) meta(s)? Como foi seguir o plano?• O que você faria diferente na próxima vez?• Você terminou o plano. E agora, o que vai fazer?

A ideia de se combinar diversas estratégias educativas possibilita maior comodidade e opção de envolvimento do usuário, contato alternativo fora do expediente, fora do horário de trabalho e evita a vinda do usuário ao Centro de Saúde. Assim, associar diversas ferramentas educativas pode reforçar o diálogo e trazer ganhos cognitivos diante das questões que são tratadas.

Em geral, observa-se que, com a aplicação do protocolo para mudança de comportamento, o usuário propõe metas que contribuiriam para colaborar no controle da condição do diabetes, sendo que a maioria consegue cumpri-las total ou parcialmente. A comunicação efetiva por meio da dialógica evidencia o cumprimento de metas por parte do usuário, podendo ser uma forma de atuação dos profissionais da área da Saúde nas práticas orientadas a autonomia do cuidado em saúde favorecendo a corresponsabilização.

5 Empoderamento do pesquisador novato: proposta de plano de metas

O empoderamento do pesquisador novato a partir da elaboração de um plano de metas a serem alcançadas surge da motivação e da reflexão sobre a pesquisa nas Ciências da Saúde. O plano de metas permitirá ao pesquisador a descoberta e utilização do conhecimento e habilidades em pensar criticamente,

agir autonomamente, usar o saber para solucionar problemas e tornar-se uma expertise responsável na investigação.

O plano de metas estimula a reflexão do pesquisador por meio de uma sequência lógica de questões não estruturadas que são distribuídas em cinco passos visando o empoderamento do pesquisador: 1) definição do fenômeno a ser estudado; 2) identificação e abordagem metodológica; 3) definição de metas; 4) elaboração do plano de metas; 5) avaliação e experiência do pesquisador sobre o plano de metas.

O educador deve solicitar ao pesquisador que preencha uma lista de ações que poderá executar para conseguir alcançar a meta identificada. Uma vez realizada a lista, pode pedir ao pesquisador que retire da lista as ações que não possui a intenção de fazer e, priorizar a meta que considera que irá realizar. Posteriormente, o pesquisador poderá escolher uma destas ações para especificá-la em relação ao tempo, frequência, lugar e duração. O educador possui importante papel para ajudar o pesquisador a monitorar e avaliar a efetividade das estratégias escolhidas.

É importante encorajar o pesquisador a elaborar um plano de metas que o mantenha no controle das ações como: a leitura, escrita e habilidade de apresentação de trabalhos, entre outros. Para tanto, a meta escolhida deve ser simples e clara, deve ser algo que ele realmente consiga fazer, deverá ter um tempo de realização, com começo e fim para que possa ser avaliada. A construção e a avaliação do plano de

metas são importantes para o desenvolvimento da autonomia do pesquisador em tomar decisões informadas, controlar as decisões, implementar decisões e ter experiência suficiente para avaliar a efetividade de suas ações.

6 Considerações finais

Pesquisador, profissional da área da Saúde e usuário empoderados favorecerão a melhoria dos resultados dos programas educativos, controle glicêmico e a qualidade de vida dos usuários. Os instrumentos permitirão avaliar as práticas educativas em saúde, se têm tomado a direção de empoderar o usuário, além da utilização de uma metodologia inovadora de adequação cultural de instrumentos destinados à promoção do autocuidado em diabetes.

O usuário empoderado é capaz de tomar decisões informadas, adequadas à sua saúde e contexto de vida, possui melhor autocuidado, maior habilidade de lidar com o estresse, maior satisfação com o tratamento e, assim, pode alcançar um melhor controle glicêmico, que é o preditor positivo de cuidados de saúde no caso do diabetes. Reconhecendo o mérito do empoderamento como um caminho bastante promissor para a abordagem de usuário do Centro de Saúde é um cenário favorável para

as intervenções mobilizadoras, inovadoras e participativas envolvendo pesquisador, profissionais da área da Saúde e usuário.

Os capítulos seguintes permitirão aprofundar as inovações tecnológicas para o empoderamento do pesquisador nas Ciências da Saúde, sob as premissas de que:

- 1) A inovação e a interdisciplinaridade no trabalho conjunto com pesquisadores de outras áreas, como Enfermagem, Linguística Aplicada e Estatística, fortalece a noção do empoderamento na formação do pesquisador nas Ciências da Saúde, comprovando que o trabalho interdisciplinar rende produtos aplicados nas práticas de educação em saúde e na pesquisa;
- 2) O empoderamento do pesquisador novato no trabalho interdisciplinar na elaboração, adaptação e validação de questionário como técnica de coleta de dados favorece a orientação das práticas educativas em saúde, tornando-as exequíveis;
- 3) O empoderamento do pesquisador pode ser em grande parte propiciado pelo contato com metodologias de inovadoras de coleta e análise de dados automáticas e semiautomáticas, possibilitando a exploração do potencial da complementariedade homem-máquina.

Síntese do capítulo

- As práticas orientadas ao cuidado em saúde na condição crônica do diabetes evidenciam melhoras dos resultados do controle glicêmico e apontam a educação em saúde sob uma perspectiva de promoção, prevenção e controle.
- O trabalho interdisciplinar com o pesquisador novato desenvolve competência profissional, não somente na pesquisa, mas em rede com outros parceiros de investigação, ampliando sua visão de mundo.
- A interdisciplinaridade entre as áreas da enfermagem, linguística aplicada e estatística favorece o empoderamento de pesquisadores, profissionais da área da Saúde e, futuramente, também dos usuários, constituindo-se em uma inovação no campo de práticas educativas em condições crônicas.

Referências

- ALEXANDRE, N. M. C.; COLUCI, M. Z. O. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. **Ciência & Saúde Coletiva**, v. 16, n. 7, p. 3061–3068, 2011.
- ANDERSON, B.; FUNNELL, M. M. **The art of empowerment: stories and strategies for diabetes educators**. 2nd ed ed. Alexandria, Va: American Diabetes Association, 2005.
- ANDERSON, R. et al. (Eds.). **101 tips for behavior change in diabetes education**. Alexandria: American Diabetes Association, 2002.
- ANDERSON, R. M.; FUNNELL, M. M. Patient empowerment: myths and misconceptions. **Patient Education and Counseling**, v. 79, n. 3, p. 277–282, 2010.
- BABENKO-MOULD, Y. Empowerment. In: SMITH, M. J.; CARPENTER, R.; FITZPATRICK, J. J. (Eds.). **Encyclopedia of nursing education**. New York: Springer Publishing Company, 2015. p. 123–125.
- BANDURA, A. Health promotion by social cognitive means. **Health Education & Behavior: The Official Publication of the Society for Public Health Education**, v. 31, n. 2, p. 143–164, 2004.
- BODENHEIMER, T.; HANDLEY, M. A. Goal-setting for behavior change in primary care: an exploration and status report. **Patient Education and Counseling**, v. 76, n. 2, p. 174–180, 2009.
- CHAVES, F. A. **Tradução e adaptação cultural do “Behavior Change Protocol” para a língua portuguesa-Brasil**. Dissertação (Mestrado em Enfermagem)—Belo Horizonte: Universidade Federal de Minas Gerais. Escola de Enfermagem, 2014.
- EPSTEIN, J. et al. Cross-cultural adaptation of the Health Education Impact Questionnaire: experimental study showed expert committee, not back-translation, added value. **Journal of Clinical Epidemiology**, v. 68, n. 4, p. 360–369, 2013.
- FREIRE, P. **Pedagogia da autonomia: saberes necessários à prática educativa**. 24. ed. São Paulo: Paz e Terra, 2002.
- FUNNELL, M. M.; ANDERSON, R. M. Empowerment and Self-Management of Diabetes. **Clinical Diabetes**, v. 22, n. 3, p. 123–127, 2004.
- GARCIA, T. R. **Elaboração e Adaptação de um Instrumento sobre Conhecimento em Doença Falciforme**. Monografia (Graduação em Enfermagem)—Belo Horizonte: Universidade Federal de Minas Gerais. Escola de Enfermagem, 2015.
- GUILLEMIN, F.; BOMBARDIER, C.; BEATON, D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. **Journal of Clinical Epidemiology**, v. 46, n. 12, p. 1417–1432, 1993.
- HOLMAN, H.; LORIG, K. Patients as partners in managing chronic disease. **BMJ : British Medical Journal**, v. 320, n. 7234, p. 526–527, 2000.
- KEBIAN, L. V. A. et al. As práticas de saúde de enfermeiros na visita domiciliar e a promoção da saúde. **Revista de APS**, v. 15, n. 1, 2012.
- MANTWILL, S. et al. EMPOWER-support of patient empowerment by an intelligent self-management pathway for patients: study protocol. **BMC medical informatics and decision making**, v. 15, p. 18, 2015.
- MINAYO, M. C. DE S. **O desafio do conhecimento**. 11. ed. São Paulo / Rio de Janeiro: HUCITEC / ABRASCO, 2008.

- MONTEIRO, C. A. et al. Monitoramento de fatores de risco para doenças crônicas por entrevistas telefônicas. **Revista de Saúde Pública**, v. 39, n. 1, p. 47–57, 2005.
- ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Envelhecimento ativo: uma política de saúde**. Brasília: OPAS, 2005.
- SANTOS, M. A. DOS et al. Grupo operativo como estratégia para a atenção integral ao diabético. **Rev. enferm. UERJ**, v. 15, n. 2, p. 242–247, 2007.
- SANTOS, L.; TORRES, H. DE C. Práticas Educativas em Diabetes Mellitus: compreendendo as competências dos profissionais de saúde. **Texto e Contexto - Enfermagem**, v. 21, n. 3, p. 574–580, 2012.
- SMALL, N. et al. Patient empowerment in long-term conditions: development and preliminary testing of a new measure. **BMC health services research**, v. 13, p. 263, 2013.
- TIMM, M.; RODRIGUES, M. C. S.; MACHADO, V. B. Adherence to Treatment of type 2 Diabetes Mellitus: a systematic review of randomized clinical essays. **J Nurs UFPE online**, v. 7, n. 4, p. 1204–1215, 2013.
- TOL, A. et al. Empowerment assessment and influential factors among patients with type 2 diabetes. **Journal of Diabetes and Metabolic Disorders**, v. 12, n. 1, p. 1–5, 2013.
- TORRES, H. DE C. et al. Evaluation of group and individual strategies in a diabetes education program. **Revista de Saúde Pública**, v. 43, n. 2, p. 291–298, 2009.
- TORRES, H. DE C. et al. Capacitação de profissionais da atenção primária à saúde para educação em Diabetes Mellitus. **Acta Paulista de Enfermagem**, v. 23, n. 6, p. 751–756, 2010.
- TORRES, H. DE C.; PEREIRA, F. R. L.; ALEXANDRE, L. R. Evaluation of the educational practices in promoting self-management in type 2 diabetes mellitus. **Revista da Escola de Enfermagem da USP**, v. 45, n. 5, p. 1077–1082, 2011.
- YANG, S.; HSUE, C.; LOU, Q. Does Patient Empowerment Predict Self-Care Behavior and Glycosylated Hemoglobin in Chinese Patients with Type 2 Diabetes? **Diabetes Technology & Therapeutics**, v. 17, n. 5, p. 343–348, 2015.

A LINGUAGEM NA CONSTRUÇÃO DAS PRÁTICAS EDUCATIVAS NAS CIÊNCIAS DA SAÚDE

Adriana S. Pagano

Objetivos do capítulo

- Apresentar uma teoria linguística que explique como funciona a linguagem nas diferentes práticas de intervenção orientadas para o autocuidado em saúde e leve o pesquisador a formular perguntas sobre elas;
- Introduzir os conceitos de *contexto de cultura* e *contexto de situação* para situar as diversas práticas de uso da linguagem;
- Explorar as variáveis prototípicas de tipos de texto comumente utilizados nas diferentes práticas;
- Discutir os pressupostos da tradução e adaptação cultural de instrumentos, com base na teoria apresentada;
- Aplicar os conceitos apresentados na análise de exemplos de textos produzidos nas práticas de intervenção.

1 Introdução

O papel da linguagem na práxis de médicos, enfermeiros, nutricionistas e outros profissionais afins é objeto de estudo, tanto nas Ciências da Saúde, como na área de Letras, mais especificamente na Linguística Aplicada. Em 2003, Candlin e Candlin fizeram uma revisão do estado da arte, à época, de pesquisas

desenvolvidas sobre esse tópico e apontaram alguns dos principais desafios para o diálogo entre as áreas. Nas pesquisas em Linguística Aplicada, os autores observam que o tipo de interação mais frequentemente estudado é a interação médico-paciente, havendo uma carência de estudos enfocando outros tipos de interação. Os estudos revisados pelos autores apontam, também, problemas de comunicação

como sendo grandes obstáculos na interação de especialistas e pacientes, bem como a escassa formação, por parte dos profissionais das Ciências da Saúde, sobre aspectos da linguagem em uso nas situações vivenciadas no exercício de suas profissões. Uma maior colaboração entre as áreas, os autores concluem, requer mudanças nas práticas de pesquisa. Por um lado, as Ciências da Saúde precisam adotar uma teoria linguística consistente e abandonar o pressuposto de que a linguagem é um mero código ou veículo de transmissão de significados. Por outro, a Linguística Aplicada precisa ampliar o escopo dos seus estudos e oferecer aos pesquisadores das Ciências da Saúde um arcabouço teórico que permita compreender a linguagem em seu contexto de uso, com base em dados coletados em situações reais de interação.

Nesse sentido, uma teoria abrangente da linguagem e com potencial de aplicação para a resolução de problemas é a Linguística Sistêmico-Funcional (HALLIDAY; MATTHIESSEN, 2004, 2014), a qual pode oferecer às Ciências da Saúde direcionamentos claros sobre como lidar com a linguagem nos contextos de uso da prática profissional.

Ao longo deste Capítulo, veremos alguns dos pressupostos básicos da teoria e como suas categorias possibilitam caracterizar os contextos de produção de significados. Procuraremos mostrar como a conscientização e reflexão por parte do pesquisador novato sobre questões da linguagem em uso são fundamentais, não somente para a o desenho da pesquisa, a coleta e análise de dados e a interpretação de resultados, mas também, para a própria formação do pesquisador rumo à expertise acadêmica. Exemplos extraídos de estudos desenvolvidos conjuntamente entre pesquisadores da Escola de Enfermagem, da Faculdade de Letras e do Departamento de Estatística da UFMG ilustram as categorias e aspectos abordados.

2 A linguística sistêmico-funcional

A linguística sistêmico-funcional (LSF) foi desenvolvida por M.A.K. Halliday a partir da década de 1950 e, desde então, encontra-se em contínua expansão e elaboração. Seu principal pressuposto diz respeito ao papel da linguagem na constituição da realidade humana: a linguagem não “reflete” a realidade; ela é o próprio agente de sua construção. Assim, não há nada anterior à linguagem, nada dentro ou fora dela. Todo e qualquer significado é ativamente construído *na* linguagem e *pela* linguagem, principal componente da cultura:

As categorias e conceitos da nossa existência material não constituem algo “dado”, anterior à sua expressão na linguagem. São construídos pela linguagem, na interseção do material e do simbólico [...] a gramática cria o potencial dentro do qual podemos agir e encenar nossa existência na cultura. (HALLIDAY, 2003, p. 145)¹.

Entendemos por *cultura* um conjunto orgânico de conhecimentos que nos permitem interpretar nossa experiência e estruturar nosso comportamento (Keith Basso apud HALLIDAY, 2003). Nossa *cultura* constitui o potencial de recursos simbólicos por meio dos quais podemos criar e intercambiar significados.

Nessa perspectiva, a *cultura* é um complexo de significados, um grande sistema de informação, constituído por sistemas de signos, isto é, sistemas semióticos. Dentre os diversos sistemas semióticos que nos oferecem recursos para gerar significados, temos sons, gestos, movimentos e, como dissemos, o prin-

¹ Minha tradução para: “*The categories and concepts of our material existence are not “given” to us prior to their expression in language. Rather, they are construed by language, at the intersection of the material with the symbolic... grammar creates the potential within which we act and enact our cultural being*”.

cial de todos, a linguagem. Ela nos permite criar o mundo em que vivemos.

A linguagem é o conjunto de comportamentos possíveis em nossa existência como seres sociais. Para estudar a linguagem, a LSF nos oferece uma teoria do contexto. Contexto é um termo técnico que difere da palavra “contexto”, utilizada por nós no nosso cotidiano. Para a LSF, contexto é um sistema semiótico no qual está imbricada a linguagem. O contexto é modelado em dois níveis de abstração: *contexto de cultura* e *contexto de situação*.

O *contexto de cultura* é o conjunto de todos os aspectos da cultura que são relevantes para a produção da linguagem. No escopo do *contexto de cultura*, o *contexto de situação* são os aspectos relevantes para a produção de uma

instância linguística numa dada *situação*. *Situação* é um termo técnico da teoria e se refere a uma unidade contextual. Ao estudar uma instância linguística, podemos observar as variáveis do contexto em que essa instância está imbricada, isto é, seu *contexto de situação*. A *situação* está vinculada a um *tipo de situação*, por sua vez imbricado numa *instituição*, a qual é parte da *cultura*. Vejamos um exemplo.

O Exemplo 1 a seguir é a transcrição de um excerto retirado de uma das entrevistas feitas por pesquisadores do Núcleo de Pesquisa em Gestão, Educação e Avaliação em Saúde (NUGEAS) durante uma das visitas domiciliares a usuários dos Centros de Saúde com a condição crônica de diabetes.

Exemplo 1

Entrevistado: falta disciplina para fazer atividade física, isso que pra mim é o que mais me coisa, eu faço, por exemplo, eu tô fazendo desde segunda feira pra cá, eu tô caminhando todo dia de manhã

Entrevistador: ah!! O senhor tá caminhando todo dia de manhã

Entrevistado: é só que é assim, eu caminho um mês, aí o corpo enjoa, aí arruma uma motivação qualquer pra parar, e eu sei que precisa, entendeu? Mas essa inconstância minha em manter uma disciplina, o resto eu faço numa boa sim

Entrevistador: tá tranquila a alimentação?

Entrevistado: a alimentação eu como em casa. É, dificilmente eu como alguma coisa fora. Eu não tenho o costume de comer salgadinho na rua, eu só como em casa. E em casa a gente come bastante fruta, eu compro muita fruta, nós temos muita fruta em casa. Pão, eu sou tarado em pão, mas aí o que nós estamos fazendo, eu compro pão integral e eu vou mesclando, um dia o pão integral, no outro aquele pãozinho delicioso.²

No Exemplo 1, a interação face-a-face, registrada com gravador pelos pesquisadores, pode ser reconhecida como prototípica de uma entrevista. A entrevista pressupõe uma interação dialógica na qual entrevistador e entrevistado alternadamente tomam a palavra para falar (*turno*), como pode ser visto no Exemplo 1. Essa interação é uma instância linguística na qual é gerado um texto vinculado à denominação entrevista, por sua vez, imbricado dentro de práticas vinculadas a uma *instituição*, neste caso, a instituição acadêmica e, mais especificamente, aquela relativa ao campo disciplinar da Enfermagem. É no escopo dessa *instituição*,

que por sua vez faz parte da cultura brasileira, que se desenvolve a práxis de um pesquisador das Ciências da Saúde que possui um roteiro de perguntas para conduzir a entrevista. Esse roteiro, no caso específico do Exemplo 1, é o Protocolo de Mudança de Comportamento, versão traduzida e adaptada do *Behavior Change Protocol* (FUNNELL; ANDERSON, 2004) para o português brasileiro.

Como veremos no Capítulo 5 deste volume, o Exemplo 1 é a transcrição de um texto

² Entrevista face-a-face. Banco de dados do Núcleo de Pesquisa em Gestão, Educação e Avaliação em Saúde (NUGEAS), Escola de Enfermagem, Universidade Federal de Minas Gerais.

falado, ou seja, o registro por escrito da fala, o qual obedece a parâmetros definidos pelo pesquisador, tendo em vista um objetivo específico. Neste caso, a transcrição foi reproduzida acima para fins de ilustrar um texto, e sinais de pontuação e convenções da escrita foram utilizados, bem como foram indicados os responsáveis pela troca de *turnos* com as palavras “entrevistado r” e “entrevistado”. Como será introduzido no Capítulo 3 e explorado de forma mais elaborada nos Capítulos 6 e 7, transcrições de textos falados, em gravações de áudio e/ou vídeo, podem ser utilizadas como fonte de dados que receberão a denominação de não estruturados e sua análise demandará métodos específicos.

Retomando nossos conceitos de *contexto de cultura* e *contexto de situação*, podemos entender o trecho da entrevista gravada, reproduzido no Exemplo 1, a partir de escolhas feitas dentre as possibilidades existentes no ambiente da cultura e da situação em particular. Essas escolhas de significados estão materializadas na linguagem.

Assim, dentre inúmeros aspectos, observamos no excerto trocas de informações, com a demanda por parte do entrevistador, realizada pelo modo interrogativo, na forma de perguntas, e as respostas por parte do entrevistado, realizadas no modo declarativo, na forma de afirmativas. O entrevistador interpela o entrevistado como “o senhor”, termo que em português brasileiro é escolhido como forma polida (cf. FIGUEREDO, 2011). Observamos, também, que o domínio experiencial ou campo da experiência humana que está representado no texto é o da condição diabetes e, mais especificamente, cuidados relativos à alimentação e atividade física, bem como sentimentos vivenciados pelo entrevistado em relação a sua condição. Observamos, ainda, que o entrevistador utiliza duas formas de se referir à condição crônica: “a diabete” e “o diabetes”,

sugerindo formas concorrentes, cada uma delas associada a escolhas em tipos de texto diferentes: “a diabete” mais provável em conversas espontâneas, não necessariamente imbricadas na instituição acadêmica, e “o diabetes” com maior frequência de ocorrência em textos da instituição acadêmica.

O Exemplo 1 ilustra um dos textos mais frequentes nas pesquisas nas Ciências da Saúde que contemplam métodos de elicitação e coleta de dados quantitativos e qualitativos. Como veremos nos Capítulos 5 e 6, a compreensão das características deste texto é importante, não apenas para a tomada de decisão sobre como preparar os dados coletados de modo a analisá-los de forma estruturada ou não estruturada, mas também para poder interpretar os resultados obtidos. Uma observação adicional e de suma importância numa abordagem do texto sob a perspectiva da Linguística Sistêmico-Funcional pode ser feita em relação ao Exemplo 1. A visão prevalecente no nosso cotidiano, atrelada, por sua vez, à educação formal na escola, é a da linguagem como um sistema de regras que prescrevem o que pode ou não ser expresso. Diferentemente dessa visão, a LSF concebe a linguagem como um sistema de recursos que nos possibilitam gerar significados. Não há para a LSF instâncias “agramaticais” ou “exceções a uma regra gramatical”, formas “certas” ou “erradas”; toda instância de linguagem é gerada por um sistema e pode ser explicada em termos de sua maior ou menor probabilidade de ocorrência de acordo com seu contexto de uso. No Exemplo 1, temos instâncias de uso da linguagem frequentes na conversação espontânea, tais como, “isso que pra mim é o que mais me coisa” ou “a alimentação eu como em casa”. Longe de ser consideradas formas “erradas” ou “incorretas” são, sob a perspectiva da LSF, formas gramaticais e com alto grau de probabilidade

de ocorrência, que realizam funções fundamentais para os significados que os falantes estão construindo na sua interação. O fato de essas formas não serem as esperadas em um outro texto, como o artigo acadêmico, não pode gerar interpretações como a de essas formas serem menos relevantes no sistema linguístico. A atribuição de valor às formas linguísticas constitui um outro aspecto do sistema da cultura e de suas instituições e não têm lugar no âmbito de pesquisas acadêmicas. Assim, como veremos no Capítulo 5, não constitui uma prática acadêmica aceitável corrigir ou alterar instâncias linguísticas encontradas nos dados coletados, bem como rotulá-las de “incorretas” ou “agramaticais”.

Para uma melhor compreensão de muitos dos pontos abordados nesta seção, veremos na seção seguinte algumas das categorias utilizadas para se caracterizar o contexto de situação das instâncias linguísticas e como essa caracterização pode ser útil para o pesquisador.

3 A linguagem e seus distintos contextos de uso

Como vimos, o Exemplo 1 ilustra um tipo de variação linguística – entrevista –, que como a LSF postula, é uma das variações dentro de um grande sistema que contempla muitas variações possíveis. Ainda, dentro de uma mesma instância linguística, como a do Exemplo 1, verificamos uma variação que responde a usos da linguagem mais ou menos frequentes no texto sob foco, como “a diabete” e “o diabetes”.

Para explicar essa variação, podemos lançar mão do conceito de *contexto de situação* e as categorias estabelecidas para seus distintos parâmetros, conforme explicitadas a seguir.

Como Halliday (1978) explica, a linguagem varia de acordo com o que estamos falando numa dada *situação* e o que estamos

fazendo enquanto falamos. Não se trata meramente do assunto ou tópico da fala, mas do objetivo para o qual nossa fala está direcionada. Esta variável do contexto de situação recebe o nome de *campo* (*field*) e abrange tanto o *domínio experiencial* (i.e. a parte de nossa experiência humana que é objeto da fala), como a *atividade sócio-semiótica* na qual estamos engajados, seja ela uma ação ou uma reflexão.

As *atividades sócio-semióticas* prototípicas do ser humano são oito, sendo uma delas primária e sete delas secundárias. Utilizamos a linguagem primeiramente para *fazer coisas* (*doing*), interagindo com outros, seja no sentido de colaborar numa tarefa ou comandar a mesma. Esta atividade é primária e essencial à vida humana em sociedade e não necessariamente demanda a linguagem verbal para sua consecução. As atividades secundárias são intrinsecamente simbólicas e requerem, em maior ou menor grau, a linguagem verbal para sua consecução. São elas: *habilitar* outros a fazerem coisas por meio de instruções ou regulando seu comportamento (*enabling*); *recomendar* por meio do aconselhamento ou da promoção de um comportamento (*recommending*); *explorar* uma posição ou ponto de vista por meio da argumentação e da crítica (*exploring*); *explicar* algum aspecto da realidade por meio de taxonomias ou categorizações e *examinar* suas relações lógicas (*expounding*); *relatar* experiências ou mapear e documentar fatos (*reporting*); *recriar* experiências passadas, reais ou imaginárias (*recreating*); e *compartilhar* experiências e valores (*sharing*).

A linguagem também varia de acordo com os papéis assumidos nas interações com os outros. Esta variável do contexto de situação recebe o nome de *sintonia* (*tenor*) e especifica o tipo de relação social que é estabelecida pelos interlocutores. Esta relação pode ser caracterizada de acordo com: o *papel institucional* dos falantes, isto é, o papel atribuído pela instituição na qual está imbricada situação e seu contexto (presidente, diretor, professor, aluno,

etc.); o *papel social* ou relação de poder entre os falantes em termos de idade, gênero, expertise, classe social (experto, novato, profissional, leigo, etc.); o *papel discursivo* ou papel criado pela linguagem através do sistema de tomada de turnos (interpelador, interpelado); o *papel sociométrico* ou envolvimento do falante numa situação de fala em termos de comportamentos cooperativos ou dissociativos em relação aos outros interlocutores (comportamento cooperativo, não cooperativo); a *distância social* ou grau de proximidade entre os falantes (desconhecido, conhecido, familiar, íntimo); e a *valoração* ou atribuição de avaliação (neutro, positivo, negativo).

Por último, há uma terceira variável do contexto de situação, de acordo com a qual temos variação linguística: o *modo* (*mode*), que diz respeito ao *meio* (escrito, falado); o *canal* (fônico, gráfico, eletrônico); o *modo retórico* (argumentativo, descritivo, narrativo, instrucional); a *divisão de tarefas* entre a linguagem e outros sistemas semióticos (imagem, som); a *orientação* da linguagem para o campo (visando explicar um domínio experiencial) ou para a sintonia (visando regular o comportamento social); e o *turno* ou modo da interação (monológico ou dialógico).

A Figura 1 sintetiza as três variáveis do contexto de situação.

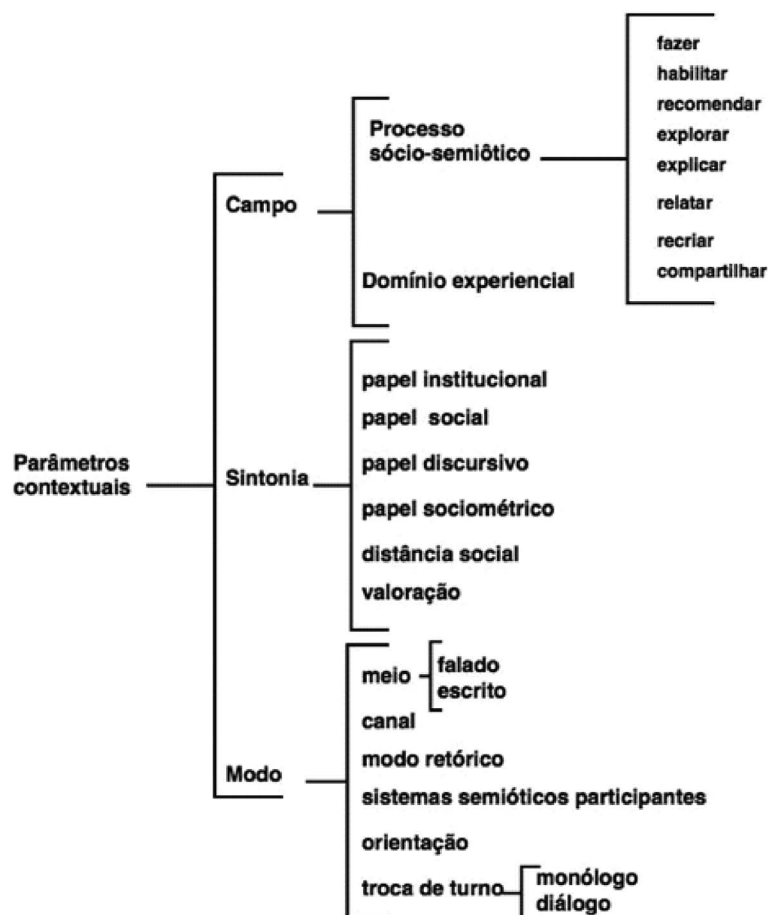


Figura 1 – Variáveis do contexto de situação

Fonte: Elaborada para fins deste estudo

Halliday e Matthiessen (2014) apresentam as atividades sócio-semióticas numa tipologia/topologia pautada pelos parâmetros

contextuais. A Figura 2 apresenta uma versão traduzida e adaptada da ilustração original.

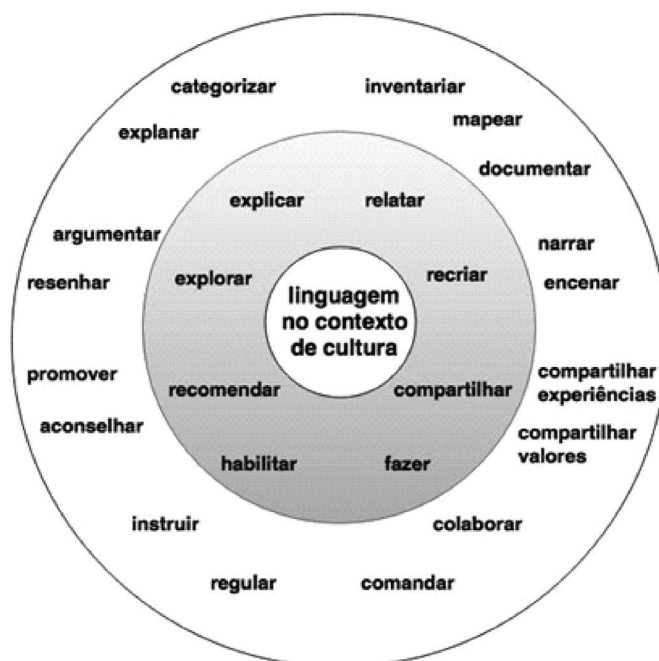


Figura 2 – Tipologia/Topologia de processos sócio-semióticos

Fonte: Adaptada de Halliday e Matthiessen (2014, p. 37)

A Figura 2 mostra a tipologia de *atividades sócio-semióticas* organizada em círculos concêntricos, numa escala crescente de especificidade. Seu ponto de partida é o núcleo central – o potencial do *contexto de cultura* –, o qual é subespecificado em oito atividades ou *processos sócio-semióticos*, como vimos acima. Estes são, por sua vez, subespecificados. As *atividades sócio-semióticas* também estão dispostas topologicamente, isto é, numa relação de contiguidade espacial, pela qual as atividades obedecem a uma sequência circular baseada na maior proximidade em termos de variáveis contextuais. Assim, *fazer* está próximo de *habilitar* e *compartilhar*, enquanto *explorar* e *explicar* são atividades próximas uma da outra, porém distantes de *fazer*, por exemplo.

O Exemplo 1 acima analisado é um excerto de uma entrevista, isto é, um texto falado, dialógico. Pode ser classificado como um texto vinculado principalmente à *atividade sócio-semiótica* relatar, na qual as interações têm por objeto recontar acontecimentos, neste caso, o entrevistado é interpelado para que conte como foi sua experiência com um plano de cuidados para o manejo do diabetes. Ainda, como já dissemos, o Exemplo 1

constitui uma transcrição, para a língua escrita, da entrevista falada e registrada por meio de gravação de áudio. Apesar de ter sido transcrita e ter sido convertida do *canal* fônico para o gráfico, pelo fato de sua origem ser uma instância linguística do *meio* falado e o *canal* ter sido originalmente fônico, a transcrição da entrevista é classificada, nas análises linguísticas, como um texto falado.

Somando-se à complexidade do Exemplo 1, como também apontamos, a entrevista foi conduzida tendo como roteiro as perguntas previstas num questionário escrito – o *Protocolo de Mudança de Comportamento* – a adaptação brasileira do *Behavior Change Protocol*. O questionário também pode ser classificado como um texto vinculado à *atividade sócio-semiótica* relatar, diálogo; contudo, diferentemente da entrevista, trata-se de um texto escrito.

O Exemplo 1 ilustra claramente que a especificação do *meio* em escrito e falado possui opções mais específicas. Gregory (1967) explorou essas subespecificações numa taxonomia, que traduzimos e adaptamos na Figura 3, adicionando exemplos, sempre que disponíveis, de textos frequentemente utilizados nas pesquisas das Ciências da Saúde.

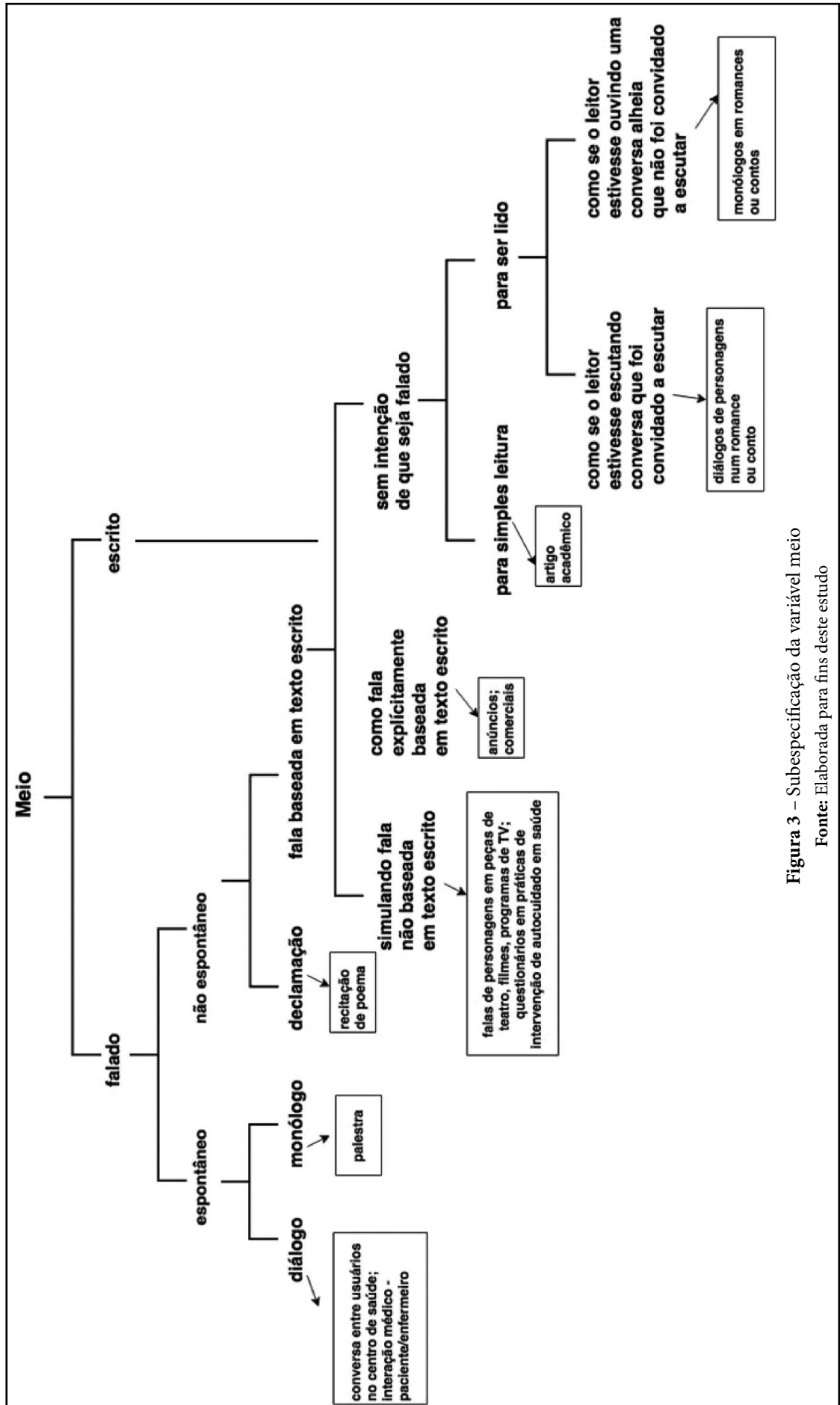


Figura 3 – Subespecificação da variável meio

Fonte: Elaborada para fins deste estudo

A Figura 3 nos permite localizar o Exemplo 1 como transcrição de uma fala não espontânea (uma entrevista agendada e pautada por um roteiro e objetivo específico), baseada num texto escrito (o Protocolo de Mudança de Comportamento), simulando ser uma fala não baseada em texto escrito (a entrevistadora

lê as perguntas dentro do fluxo da conversa, com um comentário inicial para introduzir a mesma). No Quadro 1, a seguir, vemos a pergunta tal qual consta no questionário escrito e a pergunta formulada pela entrevistadora na entrevista falada:

Quadro 1 – Diferenças entre questionário escrito e entrevista

Texto escrito	Texto falado
Item do Protocolo de Mudança de Comportamento	Pergunta correspondente ao item do Protocolo de Mudança de Comportamento formulada na entrevista
<i>1º Passo: Definição do problema</i> <i>Qual é a sua maior dificuldade para controlar o diabetes?</i>	Com a diabete assim, já que o senhor participou do grupo, qual é a maior dificuldade para controlar o diabetes?

A observação dos valores selecionados para cada variável do *contexto de situação*, como dispostas na Figura 1 e a subespecificação da subvariável *meio* na Figura 3 nos permitem desenvolver uma percepção mais acurada da variação linguística em cada texto, possibilitando extrairmos padrões ou comportamentos regulares que caracterizam as interações. Essa percepção propicia, não apenas a apreensão dos significados construídos e negociados na interação pelos interlocutores, mas também sua análise mais aprofundada para complementar a análise de dados quantitativos que possam fazer parte do mesmo banco de dados. Se, como vimos no Capítulo 1, a abordagem de condições crônicas como o diabetes demanda a corresponsabilidade de todos os atores envolvidos – pesquisadores das Ciências da Saúde,

profissionais da área da Saúde e usuários dos Centros de Saúde –, é fundamental explorarmos o dado textual extraído das interações para podermos elucidar aspectos de práticas educativas mais e menos bem-sucedidas. Como veremos no Capítulo 7, a *análise de sentimentos* é uma das técnicas que se perfilam como produtivas para tal elucidação.

A percepção mais acurada das características dos distintos *tipos de texto* também propicia, no caso do pesquisador novato, uma aprendizagem constante, à medida que lê e toma contato com diferentes *tipos de texto*.

Os Exemplos 2, 3, 4 e 5 a seguir ilustram como a análise textual baseada na detecção dos parâmetros do *contexto de situação* evidencia características prototípicas dos textos e da variação da linguagem em cada um deles.

Exemplo 2

Se não me cuidar posso ter hipoglicemia que é quando o “açúcar no sangue” está baixo. Se isso acontece eu sinto fraqueza, tremor, tontura, muita fome. Fico confuso, não consigo pensar direito e posso desmaiar e até entrar em coma.

³ TORRES, H.; PAULA, D. Conhecendo o diabetes mellitus. Cartilha elaborada pelo NUGEAS/UFMG.

Exemplo 3⁴

Os sinais e sintomas da hipoglicemia são, na maioria, neurológicos decorrentes da privação de glicose à atividade cerebral e, incluem sensações de fraqueza, fadiga, dificuldade para pensar, confusão, mudanças comportamentais e labilidade emocional. Ressalta-se que a hipoglicemia está primeiramente relacionada com a administração de insulina.⁴

Exemplo 4

Entrevistado: Aí eu tomo o remédio e como alguma coisa, porque não pode ficar sem alimentar. Não pode tomar o remédio sem alimentar.

Entrevistador: Por que senão o que acontece?

Entrevistado: Eu é que sei o que acontece. Dá uma suadeira na gente, mas eu não sei se isso é a causa do remédio ou da glicemia.

Entrevistador: Aí é a glicemia que abaixa.

Entrevistado: Então você acha que isso aí então é por causa do problema que abaixa a glicemia?

Entrevistador: É a glicemia abaixa porque tomou o remédio. Aí tem que alimentar para a glicemia aumentar.⁵

Exemplo 5⁶

Sintomas comuns que o aluno apresenta quando tem hipoglicemia: () Sudorese () Tremores () Taquicardia () Confusão mental () Agressividade () Sonolência () Fome () Formigamento () Dor de cabeça () Visão embaçada (..) ⁶

Vemos que nos textos dos quatro exemplos há um *domínio experiencial* comum: a hipoglicemia. *Domínio experiencial*, contudo, é apenas uma das seleções dentro da variável *campo*, como ilustrado na Figura 1. Há ainda uma seleção disponível que diz respeito à *atividade sócio-semiótica*, distinta em cada um dos textos.

O texto do Exemplo 2 está vinculado principalmente à *atividade sócio-semiótica habilitar*. Trata-se de uma cartilha com orientações para capacitar o usuário a se responsa-

bilizar pelo cuidado da sua condição de diabético. Adicionalmente, a cartilha está vinculada às *atividades sócio-semióticas* de *explicar*, ao apresentar uma taxonomia de sintomas e de relações entre si, e *relatar*, pelo fato de a personagem da cartilha narrar sua vivência pessoal. Já o texto do Exemplo 3 – um excerto de artigo acadêmico – está vinculado primeiramente às *atividades sócio-semióticas* *explicar* e *explorar*, uma vez que, como toda a linguagem da ciência, tem por objeto apresentar uma taxonomia de fatos e suas relações causais e também persuadir o leitor a aderir à argumentação e ao ponto de vista apresentado. Ademais, vinculamos o artigo acadêmico às *atividades sócio-semióticas* *de habilitar* e *relatar*, sobretudo nas seções de metodologia e revisão da literatura.

O texto do Exemplo 4, por sua vez, semelhantemente ao do Exemplo 1, é um excerto de entrevista falada, principalmente vinculada à *atividade relatar*. Por fim, o texto do Exemplo

⁴ MACHADO SR, RODRIGUES ALCC, SILVA SC, ALVIM RO, SANTOS PCJL. Intervenção nutricional padronizada em pacientes hipoglicêmicos hospitalizados. *Rev Gaúcha Enferm.* Porto Alegre, v. 33, n.1, p. 64-8, 2012.

⁵ Entrevista face-a-face. Banco de dados do Núcleo de Pesquisa em Gestão, Educação e Avaliação em Saúde (NUGEAS), Escola de Enfermagem, Universidade Federal de Minas Gerais

⁶ Plano Médico para Manejo do Diabetes na Escola (PMMDE). Versão traduzida e adaptada do *Diabetes Medical Management Plan*, da *American Diabetes Association*.

5 é um excerto de questionário escrito, também vinculado à *atividade relatar*.

Ao passarmos para as variáveis de *sintonia* e *modo*, observamos características distintas de cada um deles e que respondem por escolhas

linguísticas diferenciadas. O Quadro 2, a seguir, ilustra os principais valores de cada seleção em cada uma das três variáveis do *contexto de situação*.

Quadro 2 – Valores selecionados para as subvariáveis de *campo*, *sintonia* e *modo* nos textos dos Exemplos 2, 3, 4 e 5

Variável	Subvariável	Exemplo 2 cartilha	Exemplo 3 artigo acadêmico	Exemplo 4 entrevista	Exemplo 5 questionário
Campo	atividade sócio-semiótica	habilitar	explicar - explorar	relatar	relatar
	domínio experiencial	hipoglicemia	hipoglicemia	hipoglicemia	hipoglicemia
Sintonia	papel institucional	instrutor - aluno	pesquisador - pares	entrevistador - entrevistado	demandante - respondente
	papel social	experto - leigo	experto - experto	experto - leigo	experto - profissional treinado
	papel discursivo	fornecedor de informações	fornecedor de informações	demandante de informações (entrevistador) - fornecedor de informações (entrevistado) (interpelado)	demandante de informações (autor) - fornecedor de informações (respondente)
	papel sócio-métrico	cooperativo	cooperativo	cooperativo	cooperativo
	distância social	próximo	distante	próximo	distante
	avaliação	avaliatividade	neutralidade	avaliatividade	neutralidade
Modo	meio	escrito	escrito	falado	escrito
	canal	gráfico	gráfico	fônico - contato visual	gráfico
	modo retórico	narrativo - argumentativo	argumentativo	narrativo	expositivo
	sistemas semióticos participantes	verbal e visual	verbal	verbal e gestual	verbal
	orientação	sintonia	campo	campo	campo
	turno	monológico	monológico	dialógico	dialógico

O pesquisador novato tende a adotar uma posição ingênua em relação à linguagem e a caracterizar textos como os dos exemplos acima em termos de serem mais ou menos “formais” ou “informais”, rótulos não técnicos que pouco revelam sobre a linguagem em uso. Em contrapartida, uma teoria linguística

abrangente como a LSF nos permite perceber distinções mais fundamentadas, como as evidenciadas pelo Quadro 2. Assim, dependendo da subvariável enfocada, os textos podem ser agrupados de forma distinta. A subvariável *turno* agrupa a cartilha e o artigo acadêmico, por um lado, e a entrevista e o questionário,

por outro. Já a subvariável *canal* separa a entrevista dos outros três tipos de texto. Um agrupamento que parece compartilhar o maior número de valores selecionados para as subvariáveis é o artigo acadêmico e o questionário, com impacto, sobretudo, do *papel social* ou relação de poder entre os interlocutores (experto para experto e experto para profissional treinado), *distância social* entre eles e *avaliação* neutra. Entretanto, o *papel discursivo* e o *turno* são diferenciados, bem como a *atividade sócio semiótica*, estreitamente relacionada com o *modo retórico: explicar/explorar* e modo argumentativo no artigo acadêmico, e *relatar* e modo narrativo no questionário.

Além disso, o impacto das variáveis contextuais pode ser claramente observado ao compararmos distintas formulações para significados análogos. “Suadeira” na entrevista é “sudorese” no questionário”. “Açúcar no sangue” na cartilha; “glicose” no artigo acadêmico; e “glicemia” na entrevista são formas de construir significados semelhantes, porém pautados por variáveis do contexto diferentes. No caso de “glicemia”, termo técnico incorporado pelo leigo na sua fala, temos uma clara evidência de que termos técnicos não são o que basicamente caracteriza um texto numa sintonia entre expertos. O leigo também utiliza determinados termos técnicos da sua condição. A língua opera de forma integrada o léxico e a gramática. Os itens lexicais são mais propensos a serem incorporados pelo leigo; o mesmo não acontece com determinadas características gramaticais da fala do experto. Por exemplo, a hipoglicemia é explicada por meio de orações simples tanto na cartilha como na entrevista, ambas pressupondo um interlocutor leigo: “o ‘açúcar no sangue’ está baixo” e “a glicemia abaixa”. Já o artigo acadêmico utiliza uma *nominalização* típica da linguagem acadêmica: “privação de glicose à atividade cerebral”, isto é, uma forma

compacta de expressar significados por meio de substantivos abstratos que condensam um processo. Essas *nominalizações* demandam, para sua interpretação, a compreensão desses processos subjacentes a essa formulação: quem priva de glicose a quem?

A utilização de *orações* pode ser observada na cartilha; já as *nominalizações* são mais frequentes no artigo acadêmico e no questionário: “eu sinto fraqueza”, “sensações de fraqueza”; “fico confuso”, “confusão”, “confusão mental”; “não consigo pensar direito”, “dificuldade para pensar”.

Além das diferenças observadas na forma em que cada texto representa aspectos da condição do diabético, observamos também características prototípicas do texto dialógico falado, como é o caso da entrevista, e de falas inseridas num texto monológico escrito, como é a cartilha, que simula um diálogo da personagem da cartilha com o leitor. São estas a utilização da primeira pessoa (“eu” e “a gente”) e da segunda pessoa do discurso (“você”).

A LSF, como expomos logo no início deste Capítulo, é uma teoria da linguagem que, além de nos possibilitar explicar como a linguagem funciona no seu contexto de uso, nos permite fazer observações sobre regularidades no uso de determinados itens e formularmos perguntas que nos levem a indagar quais significados estão sendo construídos e como podem ser interpretados no âmbito das instituições às quais uma instância linguística se vincula.

Retomando o Exemplo 1 deste Capítulo, podemos observar na fala do entrevistado a ocorrência de diversas escolhas linguísticas pelas quais o entrevistado representa sua vivência do diabetes e encena a troca de informações com o entrevistador.

Se analisarmos cada uma das orações da fala do entrevistado, sobretudo no que diz respeito a como o entrevistado constrói sua representação da vivência do cotidiano com

o diabetes, com base em Figueredo (2011), observamos nas escolhas gramaticais que há instâncias nas quais o entrevistado representa sua vivência utilizando o pronome de primeira pessoa singular do discurso “eu”; em outras, utiliza formas impessoais, tais como “a gente” e

“ nós”; e há ainda na sua fala formas nas quais o entrevistado atribui ao corpo algumas ações e utiliza expressões impessoais sem atribuição específica de a quem se referem. O Quadro 3 classifica as diferentes instâncias encontradas na fala do entrevistado.

Quadro 3 – Classificação das orações do Exemplo 1 de acordo com sua realização gramatical

“Eu”	“A gente / nós”	Outras formas
<i>eu caminho um mês</i>	<i>em casa a gente come bastante fruta</i>	<i>é só que é assim</i>
<i>eu sei que precisa</i>	<i>nós temos muita fruta em casa</i>	<i>aí o corpo enjoa,</i>
<i>o resto eu faço numa boa sim</i>	<i>aí o que nós estamos fazendo</i>	<i>aí arruma uma motivação qualquer pra parar</i>
<i>eu como em casa</i>		
<i>difícilmente eu como alguma coisa fora</i>		
<i>eu não tenho o costume de comer salgadinho na rua</i>		
<i>eu só como em casa</i>		
<i>eu compro muita fruta</i>		
<i>eu sou tarado em pão</i>		
<i>eu compro pão integral</i>		
<i>e eu vou mesclando</i>		

A escolha do usuário para representar a experiência parece indicar que no caso de comportamentos condizentes com seu plano de metas para o controle do diabetes, ele representa sua vivência, majoritariamente, por meio de “eu” e utiliza “a gente” e “nós”, quando parece se referir a algum suporte que ele tem na sua casa. Diferentemente, quando os comportamentos não são aqueles condizentes com o seu plano, a representação da experiência se dá atribuindo ações ao seu corpo e sem atribuição clara como em “arruma uma motivação qualquer para não andar”. Cumpre lembrar que o entrevistado se encontra numa interação que, apesar de parecer espontânea, é planejada, uma vez que há um roteiro de perguntas e há um domínio experiencial pelo qual se pautam as perguntas: o plano de metas para o

autocuidado da condição diabetes, no escopo da instituição acadêmica (o pesquisador atua como entrevistador) e da práxis do serviço de Saúde (o entrevistador atua como profissional da área da Saúde). Há também uma expressão – “Mas essa inconstância minha em manter uma disciplina” – na qual o usuário utiliza uma *nominalização* que compacta o significado “eu não tenho constância”. A *nominalização*, como dizemos anteriormente, condensa significados e faz com que, na fala do caso do entrevistado no Exemplo 1, a responsável pela não manutenção da disciplina seja a “inconstância” e não o “eu”.

Aspectos como os acima apontados podem ser explorados no escopo de uma análise textual baseada na LSF. Esse tipo de análise pode ser prospectivo, no sentido de se analisar

primeiro pequenos excertos para detectar tendências nas falas dos entrevistados e posteriormente verificar se as mesmas são observadas num conjunto maior de falas. Ou, como veremos no Capítulo 6, podemos computar o número de ocorrências de determinada palavra de um texto e com base nos resultados observar quais escolhas são mais ou menos frequentes, uma informação que pode ser importante numa análise de entrevistas pelo que elas podem revelar quanto ao posicionamento do falante no que diz respeito à sua condição de diabético e sua corresponsabilidade no cuidado de si.

As observações sobre variação linguística depreendidas dos exemplos analisados confirmam o impacto das variáveis contextuais na produção de significados em diferentes tipos de texto. Este é um fator a ser levado em consideração na hora da análise de dados bem como da produção textual. No caso do pesquisador novato, é importante que ele desenvolva sua percepção sobre variação linguística e incorpore o apreendido na redação seus próprios textos ao longo do seu percurso acadêmico, como foi destacado no Capítulo 1. Também, é relevante considerar a variação linguística no processo de tradução e adaptação cultural de instrumentos, muito frequente nas pesquisas das Ciências da Saúde. A seguir veremos alguns exemplos.

4 Tradução e adaptação cultural de instrumentos

Nas Ciências da Saúde, a tradução e adaptação cultural de instrumentos está pautada por procedimentos metodológicos que buscam garantir equivalência de conteúdo e validade (GUILLEMIN; BOMBARDIER; BEATON, 1993; WILD et al., 2005) (2. Os

procedimentos são amplamente descritos na literatura desde a década de 1990 e, mais recentemente, vêm sendo revisados em função de problemas detectados e necessidade de aprimoramento (EPSTEIN et al., 2013; EPSTEIN; SANTO; GUILLEMIN, 2015). Dentre os principais questionamentos levantados, encontra-se o do papel da retrotradução (*back translation*), considerada uma das principais etapas no processo. Tida como teste inequívoco da equivalência entre um instrumento original e sua versão traduzida, a retrotradução nem sempre se mostra produtiva, principalmente diante de outros procedimentos com maior impacto, dentre eles a participação do comitê de especialistas ao qual se submete a versão traduzida e o pré-teste dos instrumentos traduzidos em amostras da população alvo.

Como apresentado no Capítulo 1, o projeto *Empoder@ - Protótipo conceitual e metodológico para avaliação de intervenções orientadas ao autocuidado em diabetes* - vem realizando um trabalho conjunto com pesquisadores das Ciências da Saúde, da Estatística e da Linguística Aplicada. Um de seus objetivos é desenvolver uma metodologia de tradução e adaptação cultural, considerando-se os problemas e recomendações relatados na literatura e ensejando-se uma adequação às instituições sociais da cultura-alvo. O projeto vem desenvolvendo estudos visando minimizar problemas de compreensão do instrumento por parte do público-alvo e garantir o que denominamos adequação cultural. Entendemos como adequação cultural o processo de garantir a plena funcionalidade de um instrumento para sua utilização num público-alvo, adotando-se uma perspectiva interdisciplinar para a resolução de problemas decorrentes do uso da linguagem registrados quando da aplicação teste do instrumento junto à população alvo. A necessidade de se garantir a adequação

cultural emerge diante das dificuldades observadas na tradução e adaptação cultural, não somente aquelas apontadas nas etapas de retrotradução e avaliação pelo comitê de especialistas, mas também e sobretudo aquelas relativas à compreensão do instrumento pela população-alvo: um fator que intervém na funcionalidade do instrumento e compromete sua utilização bem-sucedida para os fins para os quais foi desenhado.

A metodologia desenvolvida foi aplicada à tradução e adaptação cultural de diversos instrumentos, dentre eles, o *Protocolo de Mudança de Comportamento*, tradução do *Behavior Change Protocol* (FUNNELL; ANDERSON, 2004); a *Versão resumida da Escala de Empoderamento em Saúde*, tradução do *Diabetes Empowerment Scale-Short Form - DES-SF* (ANDERSON et al., 2003); a *Escala de atitudes dos profissionais em relação ao Diabetes Mellitus* (EAP-DM), tradução do *Diabetes Attitude Survey – third version* (DAS-3); e o *Plano médico para manejo do diabetes na escola* (PMMDE), tradução do *Diabetes Medical Management Plan (DMMP)* (AMERICAN DIABETES ASSOCIATION, 2015). Todos eles foram objeto de estudo visando sua *adequação cultural*.

A *adequação cultural* de um instrumento visa produzir um texto que possa ser usado na cultura alvo para elicitare uma interação entre pesquisador e usuário, de forma tão espontânea e natural, como o faz o instrumento original na sua respectiva cultura. Sob a perspectiva da Linguística Sistêmico-Funcional, a tradução é uma operação que visa a se atingir equivalência de funções gramaticais em contextos análogos (HALLIDAY, 1992), isto é, produzir um texto-alvo no qual “as formas linguísticas operem na língua alvo com efeito análogo, num *contexto de situação* análogo, àquele das formas linguísticas na língua

fonte” (HALLIDAY, 2005, p. 9)⁷. Essa visão de tradução demanda que o tradutor possua expertise como analista de textos e desenvolva sua percepção do contexto de situação na língua-fonte para emular um *contexto de situação* na língua-alvo. Como vimos na Figura 3, há configurações do texto que demandam maior atenção por parte do analista de textos e do tradutor, como é o caso dos questionários aplicados por um pesquisador, pois são textos escritos que serão lidos como se fossem textos falados, numa interação simulando uma conversa espontânea. O questionário é, assim, um roteiro para a entrevista, que é um texto falado não espontâneo.

O grande desafio para o pesquisador é perceber que o texto escrito, traduzido, é um texto diferente daquele que será encenado na entrevista e ainda levar em consideração a subvariável *canal* do *contexto de situação*, uma vez que um mesmo *canal* fônico possui especificidades, dependendo de se a interação é face-a-face ou mediada pelo computador, no caso de ligações telefônicas via internet.

O Quadro 4 a seguir apresenta um item do *Diabetes Empowerment Scale-Short Form (DES-SF)* (ANDERSON et al., 2003), traduzido e adaptado no âmbito do Projeto Empoder@, sua versão para o português brasileiro aprovada pelo comitê de especialistas, uma versão adequada para sua aplicação presencial, face-a-face e por ligação telefônica e a fala gravada durante a entrevista presencial na qual foi usado o instrumento.

⁷ Minha tradução de: “*the form in Language 2 operates with identical effect in the identical context of situation as the form in Language 1*”

Quadro 4 – Variação linguística em distintas versões de um mesmo item do DES-SF

Texto original em inglês	Versão traduzida aprovada pelo comitê de especialistas	Versão adequada para aplicação presencial, face-a-face e por ligação telefônica	Fala gravada durante a aplicação do instrumento de forma presencial, face-a-face
<i>In general, I believe that I know enough about myself as a person to make diabetes care choices that are right for me.</i>	<i>Em geral eu acredito que me conheço bem para fazer as escolhas certas para mim</i>	<i>Em geral, o senhor/ a senhora acha que se conhece bem para fazer as escolhas certas para o senhor/ a senhora.</i>	O (a) senhor (a) sabe bem como é que o (a) senhor (a) é, não sabe? Então, dá para o (a) senhor(a) escolher direitinho o que vai dar certo para o(a) senhor(a) cuidar da sua saúde.”

No Quadro 4, vemos que o original em inglês e a versão do instrumento em português, aprovada pelo comitê de especialistas, são textos escritos que podem ser utilizados como questionários escritos autoaplicáveis: o respondente lê as afirmativas e marca “sim” ou “não” em resposta para cada um dos itens do questionário. Assim, trata-se de orações declarativas e não interrogativas.

Na versão a ser aplicada pelo entrevistador, em vez de uma oração declarativa, temos uma interrogativa, ou seja, uma pergunta a ser feita pelo entrevistador. A versão adequada para aplicação presencial, face-a-face, e por ligação telefônica gerarão um texto falado, no qual haverá troca de *turnos* em tempo real. Logo, em vez da primeira pessoa do discurso (“eu”) que é utilizada no questionário escrito, temos a interpelação do entrevistado por meio da forma polida “o senhor / a senhora”.

Já na gravação que registrou a aplicação do questionário observamos que a interação se inicia com um outro *turno*, uma pergunta anterior àquela do item do questionário, que a entrevistadora insere para manter o fluxo da conversa e introduzir a pergunta em foco. Também observamos que a gramática do texto falado evidencia uma tendência à utilização de expressões de modalidade objetiva, como “dá para o senhor/a senhora escolher”, em vez de modalidade subjetiva: “o senhor / a senhora consegue escolher”. Vemos, ainda, o uso de expressões avaliativas (“escolher direitinho”).

Como veremos nos Capítulos 6 e 7, transcrições de entrevistas em práticas educativas orientadas para o autocuidado, sejam estas gravadas em áudio ou vídeo, podem constituir um dado a ser analisado pela Mineração de Textos ou Mineração de Dados e podem revelar frequências e regularidades na fala dos entrevistados, que podemos interpretar para obtermos subsídios sobre o percurso bem-sucedido ou não do usuário dos Centros de Saúde. Como os Capítulos 6 e 7 mostrarão, é neste tipo de análise que os avanços tecnológicos nos oferecem recursos para fazermos transcrições, armazená-las, analisá-las de forma automática ou semiautomática e fazermos o tratamento estatístico dos dados. E é nesse sentido que a complementaridade homem-máquina se apresenta como uma configuração essencial nas pesquisas do século XXI.

A seguir teceremos algumas considerações finais deste Capítulo.

5 Considerações finais

A análise textual das diferentes versões de um instrumento, conforme visto acima, mostra a relevância de uma abordagem sensível à variação linguística, para se garantir a *adequação cultural* de um instrumento e a elicitación de dados produtivos para a pesquisa. No Capítulo 3, veremos que o texto, seja ele escrito ou transcrição de uma fala, gera um

tipo de dado particular, diferente do dado demográfico ou do dado quantitativo produto de uma mensuração em pesquisa. Veremos, também, que ele pode ser analisado de diferentes maneiras com metodologias distintas e para fins distintos.

O sucesso de uma pesquisa demanda a atenção a uma série de fatores que, muitas vezes, são desconhecidos para o pesquisador novato. Ademais, se a pesquisa envolve a tradução e adaptação cultural de um instrumento, a *adequação cultural* é requisito fundamental. Ainda como parte do processo de tradução e adaptação cultural, a participação do comitê de especialistas é muito importante e, como

veremos no Capítulo 4, a elaboração de questionários de consulta ao comitê precisa estar pautada por princípios que garantam a eficácia da mesma. Uma vez aplicado um questionário e conduzida uma entrevista, a transcrição da fala também precisa se pautar por princípios consistentes, como será elucidado no Capítulo 5. Além disso, nos Capítulos 6 e 7, veremos detalhes de métodos e técnicas para examinar o dado obtido do texto falado, seja em gravação de áudio ou vídeo. Por último, a transformação do dado não estruturado (texto) em estruturado nos permite, como veremos no Capítulo 8, dar a ele tratamento estatístico, juntamente com dados demográficos e medições.

Síntese do capítulo

- A visão ingênua da linguagem como veículo de transmissão ou conjunto de regras precisa ser substituída por uma visão informada por uma teoria linguística abrangente, que possibilite examinar o funcionamento da linguagem nas diferentes práticas de intervenção orientadas para o autocuidado;
- Conceitos como *contexto de cultura* e *contexto de situação* no âmbito da Linguística Sistêmico-Funcional permitem uma análise textual aprofundada, capaz de situar as diversas práticas de uso da linguagem;
- O pesquisador novato precisa conhecer e desenvolver sua percepção sobre as variáveis prototípicas dos tipos de texto comumente utilizados nas diferentes práticas;
- A tradução e adaptação cultural de instrumentos demanda uma avaliação mais acurada dos parâmetros de cada tipo de texto, assim como reflexões e reformulações no sentido de se garantir a *adequação cultural*.

Referências

AMERICAN DIABETES ASSOCIATION. **Diabetes Medical Management Plan**. Disponível em: <<http://www.diabetes.org/living-with-diabetes/parents-and-kids/diabetes-care-at-school/written-care-plans/diabetes-medical-management.html>>. Acesso em: 1 out. 2015.

ANDERSON, R. M. et al. The Diabetes Empowerment Scale-Short Form (DES-SF). **Diabetes Care**, v. 26, n. 5, p. 1641–1642, 2003.

CANDLIN, C. N.; CANDLIN, S. Health Care Communication: a problematic site for

applied linguistics research. **Annual Review of Applied Linguistics**, v. 23, 2003.

EPSTEIN, J. et al. Cross-cultural adaptation of the Health Education Impact Questionnaire: experimental study showed expert committee, not back-translation, added value. **Journal of Clinical Epidemiology**, v. 68, n. 4, p. 360–369, 2013.

EPSTEIN, J.; SANTO, R. M.; GUILLEMIN, F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. **Journal of Clinical Epidemiology**, v. 68, n. 4, p. 435–441, 2015.

FIGUEREDO, G. P. **Introdução ao perfil metafuncional do português brasileiro:**

- contribuições para os estudos multilíngues.** Tese de Doutorado—Belo Horizonte: Universidade Federal de Minas Gerais. Faculdade de Letras, 2011.
- FUNNELL, M. M.; ANDERSON, R. M. Empowerment and Self-Management of Diabetes. **Clinical Diabetes**, v. 22, n. 3, p. 123–127, 2004.
- GREGORY, M. Aspects of varieties differentiation. **Journal of Linguistics**, v. 3, n. 02, p. 177–198, 1967.
- GUILLEMIN, F.; BOMBARDIER, C.; BEATON, D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. **Journal of Clinical Epidemiology**, v. 46, n. 12, p. 1417-1432, 1993.
- HALLIDAY, M. A. K. **Language as social semiotic: the social interpretation of language and meaning.** Baltimore: University Park Press, 1978.
- HALLIDAY, M. A. K. Language Theory and Translation Practice. **Rivista internazionale di tecnica della traduzione**, p. 15-25, 1992.
- HALLIDAY, M. A. K. **On language and linguistics.** London: Continuum, 2003.
- HALLIDAY, M. A. K. Linguistics and Machine Translation: STUF - Language Typology and Universals. In: HALLIDAY, M. A. K. (Ed.). **Computational and quantitative studies.** London: Continuum, 2005.
- HALLIDAY, M. A. K.; MATTHIESSEN, C. **An Introduction to Functional Grammar.** 3. ed. London: Hodder Education, 2004.
- HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. **Halliday's introduction to functional grammar.** 4. ed. Milton Park, Abingdon, Oxon: Routledge, 2014.
- MACHADO, S. R. et al. Standardized nutritional intervention in hypoglycemic inpatients. **Revista Gaúcha de Enfermagem**, v. 33, n. 1, p. 64–68, 2012.
- TORRES, H. DE C.; PAULA, D. **Conhecendo o diabetes mellitus.** Belo Horizonte: NUGEAS/UFMG, 2013.
- WILD, D. et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. **Value in Health: The Journal of the International Society for Pharmacoeconomics and**

DADOS, INFORMAÇÃO, CONHECIMENTO NAS CIÊNCIAS DA SAÚDE: DO NÃO ESTRUTURADO AO ESTRUTURADO

Ilka Afonso Reis

Objetivos do capítulo

- Definir e diferenciar os conceitos de dado, informação e conhecimento no âmbito das Ciências da Saúde;
- Apresentar os dados dos tipos estruturado e não-estruturado;
- Discutir as vantagens e desvantagens de cada tipo de dado.

1 Introdução

Nos dias de hoje, é frequente nos depararmos com a afirmação de que estamos expostos a uma imensa quantidade de informação ou ainda que produzimos uma grande quantidade de dados. As palavras “dado” e “informação” são, muitas vezes, usadas indistintamente como se fossem sinônimas. O mesmo ocorre com “informação” e “conhecimento”.

Se no cotidiano as palavras “dado”, “informação” e “conhecimento” são usadas de forma imprecisa, a definição e o estudo dos

conceitos de dado, informação e conhecimento, como termos técnicos, são bastante importantes para áreas como Tecnologia da Informação e Ciência da Computação, tamanha é a quantidade de trabalhos e encontros científicos que discutem esses termos e os aspectos relacionados a eles. Para Schumaker (2011), por exemplo, dados são adquiridos a partir do estímulo e da inspeção cuidadosa do mundo ao nosso redor. Podemos medir (inspecionar) a altura de um indivíduo ou observar seu grau de escolaridade a partir de um estímulo, ou seja, uma questão sobre seu grau

de instrução formal. Por si só, um *dado* seria apenas um símbolo usado para registrar nossa observação; todavia, sua transformação nos daria *informação* (SCHUMAKER, 2011). A agregação de informação relacionada nos daria *conhecimento* (BARLAS; GINART; DORRITY, 2005).

Segundo Davenport e Prusak (1997), os conceitos de dado, informação e conhecimento não são fáceis de serem separados na prática. Ainda assim, os autores tentam distinguí-los, afirmando que informação é “a conexão entre dado bruto e o conhecimento que se pode descobrir nele”. A medida de altura de um indivíduo, assim como o registro do seu grau de escolaridade, são exemplos de dados brutos, ou seja, dados sem tratamento.

Os conceitos de dado e informação, e também o seu grau de abstração, variam conforme a área de estudo. Neste capítulo, vamos trabalhar sob a ótica de estudos desenvolvidos nas Ciências da Saúde. Na seção seguinte, vamos definir e diferenciar esses conceitos, além de explicitarmos seu papel na construção do conhecimento. Como veremos, os dados são a base para a construção do conhecimento. Sendo assim, as seções 3 e 4 tratarão dos tipos básicos de dados: estruturados e não estruturados.

2 Dado, informação e conhecimento

Dado pode ser definido como um símbolo ou conjunto de símbolos que são usados para registrar características de um indivíduo. O dado pode ser um número (ex., a idade de uma pessoa), uma categoria (ex., o sexo de um voluntário ou sua profissão), um texto (ex., a fala de um paciente sobre suas dificuldades para seguir o tratamento), uma

imagem (ex., uma fotografia mostrando a expressão de um indivíduo ao receber uma notícia ou assistir a uma determinada cena). O dado sozinho caracteriza apenas o indivíduo ao qual ele foi atribuído. Se o nosso objetivo for estudar apenas um indivíduo, necessitamos apenas de seus dados: saberemos sua idade, seu sexo, sua profissão e sua condição de trabalho, por exemplo. No entanto, a imensa maioria dos estudos na área das Ciências da Saúde tem o objetivo de estudar um grupo ou população de indivíduos. Nesse caso, uma coleção de dados sobre vários indivíduos de um grupo não é capaz, por si só, de caracterizar esse grupo. Essa coleção de dados precisa ser tratada, isto é, organizada, resumida, analisada e, desse tratamento, surge a informação.

Sendo assim, podemos definir a informação como o produto gerado pelo tratamento de uma coleção de dados. Neste Capítulo, consideramos que o tratamento dos dados será feito por meio do emprego de técnicas estatísticas.

Uma vez obtida a informação, sua interpretação no contexto do objeto de estudo leva à produção de conhecimento. Segundo Whitney (2007), conhecimento pode ser definido como “uma combinação de informação, experiência e intuição que pode beneficiar o indivíduo ou a organização (empresa)”.

A Figura 1 ilustra, de maneira geral, o processo para geração de conhecimento a partir de dados com o auxílio da Estatística. Inicialmente, são postuladas perguntas ou hipóteses, e a partir de delas, são planejados estudos, que, por sua vez, geram dados. Em sua forma bruta, os dados não são muito úteis para responder às perguntas. No entanto, o processamento desses dados por meio de técnicas estatísticas gera informação, que pode levar a novos conhecimentos quando interpretada no contexto do objeto de estudo.

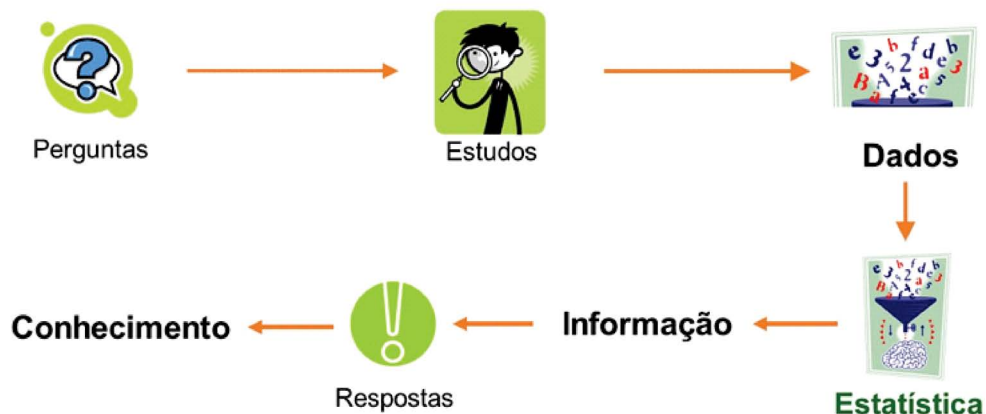


Figura 1 – Processo para geração de conhecimento a partir de dados com o auxílio da Estatística

Fonte: Elaborada para fins deste estudo

Exemplo 1

Considere um grupo de profissionais da área da Saúde que esteja planejando um programa de educação para indivíduos com determinada condição crônica. Uma parte do programa envolve o estabelecimento de metas para a mudança de estilo de vida, que são definidas em um primeiro contato com o indivíduo. Essas metas podem ser, por exemplo, “beber dois litros de água por dia”, “ingerir menos sódio”, “praticar atividade física diariamente”, entre outras. Uma das perguntas que o

grupo de profissionais gostaria de responder é: “O sexo do indivíduo interfere no sucesso do cumprimento das metas estabelecidas?” Sendo assim, desenham um estudo, que acompanha 150 mulheres e 90 homens. Num primeiro encontro, os participantes estabelecem as metas que se propõem a alcançar. Num segundo encontro, os profissionais registram se os participantes cumpriram ou não as metas. As respostas dos participantes geram *dados* como os apresentados na Figura 2.

ID	01	02	03	04	05	06	07	08	...	235	236	237	238	239	240
Sexo	M	M	F	M	F	F	F	M		M	F	F	M	F	F
Cumpriu meta?	Não	Não	Não	Sim	Sim	Sim	Sim	Não		Sim	Não	Sim	Não	Não	Sim

Figura 2 – Dados gerados pelo estudo do Exemplo 1. As letras F e M representam, respectivamente, o sexo feminino e o sexo masculino (dados fictícios).

Fonte: Elaborada para fins deste estudo

O simples exame dos dados brutos, como os da Figura 2, não permite ao grupo de profissionais responder à questão inicial (“O sexo do indivíduo interfere no sucesso do cumprimento das metas estabelecidas?”). É preciso *processar e organizar* esses dados, o que pode ser feito por meio da técnica *estatística* de construção de tabelas de contingência, como a que é apresentada na Tabela 1. Uma tabela

de contingência é uma técnica estatística utilizada para verificar a existência de associação entre duas características como sexo e o cumprimento de metas, no caso desse exemplo. Ela organiza a distribuição do total de indivíduos segundo seus atributos registrados simultaneamente. A construção e o uso das tabelas de contingência serão detalhados no Capítulo 8.

Tabela 1 – Distribuição de frequências dos indivíduos do estudo do Exemplo 1 segundo o sexo e o cumprimento de metas

Sexo	Cumpriu a meta?		Total
	Sim	Não	
Feminino	105	45	150
Masculino	41	49	90
Total	146	94	240

Com os dados organizados na Tabela 1, é possível extrair as seguintes *informações*: a) 70% das 150 mulheres (105/150) cumpriu a meta estabelecida no primeiro encontro; b) entre os homens, o percentual de indivíduos que cumpriram a meta foi de 45,6% (41/90).

As *informações* extraídas da Tabela 1 podem ser interpretadas em relação com estudos anteriores (para efeitos deste Exemplo, fictícios), que mostram que, em geral, mulheres são mais cuidadosas com sua saúde do que os homens. Assim, as informações extraídas da Tabela 1 confirmam os estudos prévios. Essa confirmação possibilita estabelecer o *conhecimento* de que “mulheres com condições crônicas tendem a cumprir suas metas de mudança de estilo de vida mais frequentemente do que os homens na mesma condição”. Sendo assim, a resposta à pergunta inicial (“O sexo do indivíduo interfere no sucesso do cumprimento das metas estabelecidas?”) é “Sim”. Este conhecimento pode ser útil ao grupo de profissionais, uma vez que ele poderá estabelecer estratégias de educação diferenciadas para homens e mulheres.

Sendo os dados a base para a construção do conhecimento, eles devem ser muito bem coletados e processados. Na seção seguinte, apresentaremos e descreveremos os dois tipos básicos de dados existentes.

3 Dados estruturados e dados não estruturados

Os dados coletados em um estudo podem ser do tipo estruturado (categóricos/qualitativos ou numéricos/quantitativos) ou não estruturado, como ilustra a Figura 3. Dados estruturados são aqueles que podem ser organizados dentro de um formato rígido, com regras bem definidas (ALMEIDA, 2012), por exemplo com o formato de uma tabela (linha/coluna), também chamado matriz: cada linha representa um sujeito do estudo e os dados das características dos sujeitos são dispostos em colunas (uma para cada característica). Dados não estruturados são aqueles que não possuem um formato pré-definido (por exemplo, linha/coluna), como textos e imagens.

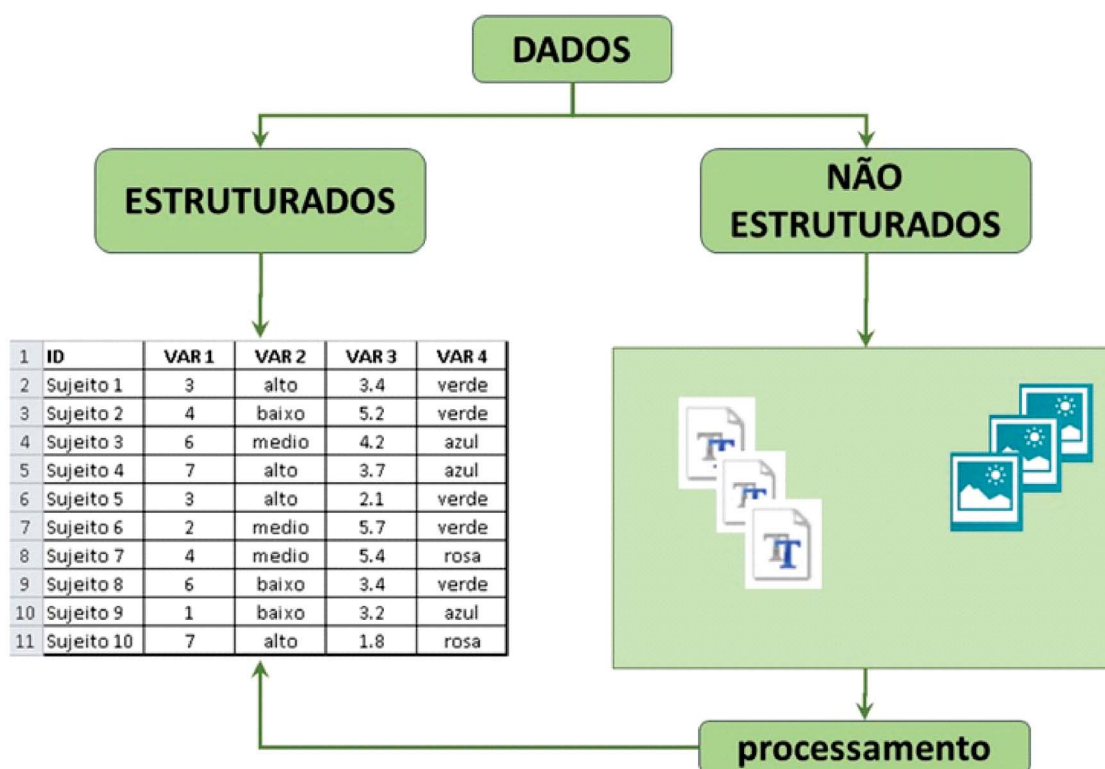


Figura 3 – Tipos de Dados

Fonte: Elaborada para fins deste estudo

É importante ressaltar que essas definições para dados estruturados e não estruturados são do domínio da Ciência da Computação, para a qual é importante saber como armazenar os dados (em bancos de dados pré-definidos ou não) e que tipo de consulta pode ser feita a esses dados armazenados. Por exemplo, com os dados estruturados apresentados na Figura 3, podemos fazer as seguintes consultas: “qual é o número de sujeitos cuja coluna 2 (VAR2) vale ‘alto’?” e “qual é o número de sujeitos com ‘baixo’ na coluna 2 e ‘verde’ na coluna 4?”. Quando os dados são do tipo não estruturado, como textos e imagens, esse tipo de consulta não é possível justamente pela falta da estrutura linha/coluna no modo como os dados são armazenados.

O dado não estruturado é frequentemente associado ao que se chama texto. Como vimos no Capítulo 2, a palavra “texto” no nosso cotidiano pode ser usada para se fazer referência ao que o linguista denomina instância linguística, a qual está vinculada a uma atividade sócio-semiótica, em particular, a um tipo de texto com variáveis contextuais específicas.

Embora “texto” seja comumente associado com uma instância linguística escrita, há também textos falados. De fato, a variável *modo* do *contexto de situação* diz respeito ao *meio* escrito ou falado e ao *canal* gráfico ou fônico. Como também vimos no Capítulo 2, *meio* escrito e falado são dois valores dessa subvariável, os quais podem ser subespecificados em distintas combinações. Há ainda que se considerar que a linguagem é apenas um dos sistemas semióticos e que uma instância linguística pode estar integrada a outros sistemas semióticos, como imagem e som. Assim, o que se entende por dado não estruturado ou do tipo textual nas Ciências da Computação abrange o dado extraído tanto de textos escritos como falados, textos registrados por meio do canal gráfico e do canal fônico e podem estar acompanhados de dados coletados de outros registros não verbais, como imagem e som. Como será discutido no Capítulo 4, questionários são instrumentos bastante utilizados para a coleta de dados estruturados e não estruturados na área das Ciências da Saúde. Para a coleta de dados

não estruturados, as gravações de áudio e vídeo são ferramentas muito úteis. Como será visto nos capítulos 5 e 7, o processamento dessas gravações gera tanto dados não estruturados (transcrições de textos/falas) quanto dados estruturados (anotações de categorias atribuídas à fala e/ou ao som, gestos e outros).

Uma vez coletados, os dados precisam ser analisados para gerarem informação. Atualmente, a análise estatística de dados estruturados conta com um conjunto de técnicas bastante desenvolvido, tanto para a parte descritiva e exploratória quanto para a parte inferencial, que é a extensão dos resultados obtidos em uma amostra para a população da qual ela foi retirada.

A análise de dados não estruturados, por sua vez, exige que eles passem por algum tipo de processamento, no qual podem ser extraídos outros dados para serem dispostos no formato estruturado (Figura 3). É o caso de textos escritos, por exemplo, romances traduzidos, dos quais podemos extrair o número de palavras, de frases, se houve ou não uso de expressões negativas, entre outros atributos, para que sejam comparados com os dados dos mesmos

atributos extraídos dos romances em sua versão original. Em outros casos, no entanto, os dados não estruturados podem ser analisados sem a necessidade de pré-estruturação, como é o caso das linhas de concordância, que serão estudadas no Capítulo 6.

3.1 Organização de dados estruturados

A forma básica de organizar dados estruturados é utilizar o que denominamos *planilha de dados*. A Figura 4 apresenta parte de uma planilha, que foi construída para organizar os dados de sujeitos participantes de um estudo com pessoas diagnosticadas com Diabetes Mellitus, no qual foram registrados a identificação do sujeito (ID), o grupo ao qual ele pertence, seu sexo, sua idade (em anos completos), sua escolaridade, seu estado civil, se trabalha fora de casa ou não, número de filhos, seu nível de glicohemoglobina no início do estudo (ghb1, em %), sua altura (em cm), seu peso (em Kg), sua circunferência abdominal (em cm) e o seu o nível de glicohemoglobina no fim do estudo (ghb2, em %).

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	grupo	sexo	idade	escolaridade	estadocivil	trabalho	filhos	ghb1	altura	peso	circ.abdom	ghb2
2	ID78	C	F	55	Fundamental	Casado	SIM	3	3.77	161.4	72.7	79	1.62
3	ID5	A	F	54	Superior	Viuvo	NAO	2	4.98	155.8	72.3	86.6	2.4
4	ID19	A	M	73	Superior	Separado	NAO	4	5.65	167.3	95.4	105.4	2.7
5	ID28	A	F	60	Fundamental	Separado	NAO	4	6.58	158.8	71.3	93.2	2.77
6	ID2	A	M	68	Analfabeto	Solteiro	SIM	1	2.84	170.6	80.3	106.9	3.16
7	ID14	A	F	54	Fundamental	Separado	NAO	4	6.86	158.4	65.5	83	3.73
8	ID32	B	M	76	Fundamental	Solteiro	NAO	2	7.87	169.6	79.8	118.3	3.91
9	ID3	A	F	57	Medio	Casado	NAO	1	5.86	168.1	74.3	94.5	4.00
10	ID85	C	M	60	Medio	Separado	SIM	3	8.93	178.2	74.1	104.2	4.31
11	ID59	B	F	59	Fundamental	Viuvo	SIM	3	4.17	163	75.1	89.2	4.64
12	ID37	B	F	53	Superior	Casado	SIM	5	6.51	163.4	65.9	91.3	4.96
13	ID6	A	M	80	Medio	Casado	NAO	3	10.24	179.3	69.5	101.7	5.04
14	ID25	A	M	68	Fundamental	Solteiro	SIM	0	8.05	175.4	73.9	103	5.11
15	ID51	B	F	57	Fundamental	Solteiro	NAO	2	7.56	157.3	71.3	94.3	5.13
16	ID67	C	F	54	Fundamental	Casado	NAO	0	6.8	162.7	72.00	92.6	5.79
17	ID18	A	M	69	Fundamental	Casado	SIM	4	9.51	172.9	83.4	110.1	6.00
18	ID17	A	M	61	Fundamental	Separado	SIM	1	6.88	164.8	84.6	106.9	6.13
19	ID33	B	F	49	Medio	Viuvo	SIM	2	4.62	159.3	69.9	98.8	6.17
20	ID39	B	F	64	Medio	Solteiro	NAO	1	7.32	155.6	66.4	84.5	6.17
21	ID44	B	F	66	Analfabeto	Separado	SIM	3	9.98	154.1	71	87	6.2
22	ID64	C	F	50	Medio	Separado	SIM	2	6.49	167.9	69.3	76.2	6.21
23	ID1	A	F	64	Medio	Solteiro	NAO	3	7.18	158.9	69.3	77.9	6.24
24	ID15	A	F	48	Fundamental	Separado	NAO	2	5.82	160.9	71	92.2	6.25
25	ID73	C	F	61	Medio	Casado	NAO	3	7.4	157.6	66.3	90.4	6.28
26	ID36	B	F	66	Superior	Viuvo	SIM	0	7.6	164	64.1	89.7	6.38
27	ID4	A	M	69	Medio	Casado	SIM	0	7.45	176.2	82.5	103.1	6.62

Figura 4 – Formato básico de uma Planilha de Dados

Fonte: Elaborada para fins deste estudo

Uma planilha de dados é uma estrutura no formato de matriz, constituída por linhas e colunas, como destacado na Figura 5. O cruzamento de uma linha e de uma coluna é denominado de célula. A primeira linha da planilha é reservada para o nome das variáveis

e primeira coluna geralmente recebe os nomes dos sujeitos ou, por questões de sigilo, um identificador numérico ou alfanumérico (letras e números), como é o caso do identificador utilizado na planilha da Figura 4.

O diagrama mostra uma planilha de dados com 10 colunas (A-I) e 7 linhas (1-7). A primeira linha (linha 1) contém os cabeçalhos das variáveis: ID, grupo, sexo, idade, escolaridade, estadocivil, trabalho, filhos, ghb1. A quarta linha (linha 4) contém os dados para o indivíduo ID19. A quarta coluna (coluna D) contém os dados para a variável 'idade'. A célula na interseção da linha 4 e coluna D (contendo o valor 73) está destacada com um retângulo verde. Uma seta verde aponta para esta célula com o rótulo 'Célula'. O rótulo 'Linha' está à direita da linha 4, e o rótulo 'Coluna' está abaixo da coluna D.

	A	B	C	D	E	F	G	H	I
1	ID	grupo	sexo	idade	escolaridade	estadocivil	trabalho	filhos	ghb1
2	ID78	C	F	55	Fundamental	Casado	SIM	3	3.77
3	ID5	A	F	54	Superior	Viuvo	NAO	2	4.98
4	ID19	A	M	73	Superior	Separado	NAO	4	5.65
5	ID28	A	F	60	Fundamental	Separado	NAO	4	6.58
6	ID2	A	M	68	Analfabeto	Solteiro	SIM	1	2.84
7	ID14	A	F	54	Fundamental	Separado	NAO	4	6.86

Figura 5 – Linhas, colunas e células em uma planilha de dados

Fonte: Elaborada para fins deste estudo

É importante destacar que, em uma planilha de dados, cada linha deve conter os dados de apenas um sujeito e cada coluna deve conter os dados de apenas uma variável. Assim, a coluna *estadocivil*, por exemplo, registra apenas as categorias solteiro, casado, separado e viúvo, independentemente de ser o sujeito um homem ou uma mulher. O dado sobre o sexo do indivíduo é registrado na coluna *sexo*. Consultando as colunas *sexo* e *estadocivil* na linha 3, por exemplo, saberemos que o sujeito ID5 é uma mulher viúva. Se quisermos saber quantos homens solteiros participam do estudo, devemos cruzar os dados das colunas *sexo* e *estadocivil* e gerar uma tabela de contingência, o que será abordado no Capítulo 8. Outro registro que gera bastante confusão é o de respostas múltiplas a uma mesma questão. Nesse caso, cada opção de resposta deve ser registrada em uma coluna separada e o dado do sujeito será “sim” somente nas colunas que se referem às opções de resposta que ele escolheu e será “não” em todas as outras colunas.

Há ainda vários outros equívocos que são cometidos na organização de dados

estruturados e que podem dificultar ou invalidar o tratamento desses dados. O capítulo final deste livro (Boas Práticas) traz uma lista de pontos a serem observados na construção de uma planilha de dados.

3.2 Organização de dados não estruturados

Organizar dados que não têm uma estrutura significa processá-los para deles extrair outros dados, estes, sim, estruturados, como ilustra a Figura 3. Na transcrição da fala de uma pessoa diagnosticada com Diabetes Mellitus sobre sua convivência com sua condição, por exemplo, podemos preparar o texto por meio da separação dele em palavras e contar o número de vezes em que ela emprega os pronomes da primeira pessoa do singular (eu, me/mim, meu/minha, comigo). O resultado dessa contagem é um dado estruturado, que pode ser disposto em uma planilha (linha/coluna) e cujo tratamento gerará informação que pode nos ajudar a entender a atitude da pessoa frente à sua condição, por exemplo.

O processamento de dados não estruturados deve ser direcionado pela pergunta/hipótese do estudo que gerou esses dados ou vai utilizá-los. Um mesmo conjunto de dados não estruturado pode ser processado de maneiras diferentes, gerando estruturas diferentes.

Exemplo 2

O conjunto de cenas de um filme pode ser visto com um conjunto de dados não estruturados. Um pesquisador A gostaria de verificar se as cenas que contêm violência são maiores do que as cenas sem violência. Já o pesquisador B gostaria de verificar se as cenas em ambiente externo são mais frequentes do que as cenas em ambiente interno. O pesquisador A vai processar os dados (cenas) classificando-as em “com violência” ou “sem violência”, além de anotar a duração das cenas (em segundos), gerando dois dados estruturados (um categórico e outro numérico) para cada cena. Já o pesquisador B vai classificar as cenas em “externas” ou “internas”, gerando um dado estruturado para cada cena. Além disso, a definição do que é uma cena dependerá do objetivo do estudo, pois, para o pesquisador A, uma cena que comece sendo filmada em ambiente interno, mas termine na parte externa pode ser considerada uma só cena. No entanto, o pesquisador B pode considerar que há duas cenas, uma externa e outra interna. Assim, o número de cenas consideradas também vai depender do objetivo de cada estudo.

A tarefa de processar dados não estruturados é geralmente exaustiva, tanto pelo fato de que dados sem estrutura são unitariamente volumosos (textos e imagens) quanto pela grande quantidade que pode ser produzida desse tipo de dados. Considerando ainda a sua natureza não estruturada, o processamento desse tipo de dados pode se beneficiar das técnicas de mineração. Segundo Han, Kamber e Pei (2012), “mineração de dados é o processo

de descobrir padrões interessantes e conhecimento a partir de grandes quantidades de dados”. No caso de dados não estruturados, os padrões a que se referem os autores podem significar a estrutura que gostaríamos de dar a esses dados.

Com o uso do computador e de softwares especializados, o processamento de dados não estruturados torna-se bem mais rápido e, de certa forma, mais eficaz, uma vez que a máquina não considerará certos padrões pré-definidos que a mente humana rejeitaria e, desse modo, pode encontrar estruturas mais interessantes, mas que dificilmente despertariam a atenção de um ser humano. No entanto, como o processamento dos dados deve ser direcionado pela pergunta/hipótese do estudo, a orientação humana nessa tarefa é muito importante e a complementaridade homem-máquina é fundamental para o sucesso da estruturação de dados não estruturados.

Os capítulos 6 e 7 tratam com mais detalhes dos processos computacionais para a análise de dados não estruturados quando estes são extraídos de textos e falas gravadas em áudio/vídeo, respectivamente. No Capítulo 6, é exemplificado o procedimento de gerar listas com as palavras mais frequentes em um conjunto de textos escritos e linhas de concordância para a obtenção de padrões ou regularidades no uso da linguagem; o Capítulo 7 mostra como anotar falas classificando-as em categorias pré-definidas usando o software ELAN. A partir de dados não estruturados (textos escritos no Capítulo 6 e textos falados no Capítulo 7), são gerados conjuntos de dados estruturados (conjunto de palavras e conjunto de categorias, respectivamente), que podem ser dispostos no formato linha/coluna (planilha de dados): uma coluna com os identificadores dos textos de onde vêm as palavras ou categorias, outra coluna para dispor as palavras ou categorias, sendo que cada palavra ou categoria

ocupa uma linha. Os dados assim dispostos em uma planilha podem ser resumidos e organizados para gerarem informação, o que é feito por meio das listas de frequência e linhas de concordância de palavras (Capítulo 6) e das tabelas de frequência de categorias exportadas pelo ELAN (Capítulo 7).

4 Como coletar os dados: estruturados ou não estruturados?

Uma dúvida comum no planejamento de um estudo é sobre o tipo de instrumento de coleta de dados a ser adotado: estruturado (questões fechadas) ou não estruturado (questões abertas). Há ainda a opção da versão semiestruturada, que mescla questões de ambos os tipos.

Questões fechadas geram naturalmente dados estruturados como, por exemplo, sexo, grau de escolaridade, estado civil, uso de medicamentos (sim ou não) e o grau de concordância com alguma afirmativa. Questões abertas, mas com somente uma opção de resposta, geram dados estruturados (ex. “quantos anos você tem?”, “quantos filhos você teve?”, “há quanto tempo você tem esta doença?”, “qual é o seu peso?” e “qual é sua altura?”). Questões abertas e sem opções de respostas pré-definidas geram dados não estruturados (ex. “quais as dificuldades o(a) sr.(a) enfrenta para seguir sua dieta?”, “como o(a) sr.(a) se sente em relação à sua doença?”, “como é a sua rotina no trabalho?”).

Como já vimos, dados estruturados serão organizados em uma planilha de dados. Os dados não estruturados no formato de texto podem ser armazenados em uma planilha, mas nem por isto estarão prontos para serem analisados da mesma forma como os dados estruturados serão analisados. Antes disso, é preciso processá-los de acordo com o objetivo

da análise. Então, surge a dúvida: é melhor formular a questão para coletar dados já no formato estruturado ou formulá-la para coletar dados não estruturados?

Limitar as opções de respostas a uma questão de modo a coletar dados estruturados facilita o processo de organização e análise desses dados. No entanto, exige do pesquisador um conhecimento sólido sobre o universo de respostas possíveis assim como uma certeza sobre o papel desses dados na construção da resposta à sua pergunta/hipótese de pesquisa. Além disso, categorias pré-definidas podem induzir a resposta do sujeito, como é o caso de questionários auto aplicados.

Por outro lado, ao optarmos pela coleta de dados não estruturados, deixamos o sujeito livre para expressar sua opinião ou dar o seu relato. Isto é importante quando não se conhece muito sobre o objeto de estudo e há a intenção de explorar melhor suas características. Contudo, corremos o risco de coletarmos dados nulos no sentido de que não gerarão dados válidos quando preparados para análise, principalmente quando a questão não for bem compreendida pelo sujeito. Como exemplo, podemos citar a questão “como você se sente em relação ao diabetes?”, para a qual esperaríamos respostas que englobassem expressões do tipo “revoltado”, “com raiva”, “conformado”, “bem”, mas que pode ter respostas do tipo “sinto um formigamento no pé”, “não estou me sentindo bem hoje” ou “sinto tonturas às vezes”. Além disso, já sabemos que os dados não estruturados necessitam passar por uma etapa adicional de processamento antes de serem analisados.

Sendo assim, a decisão sobre que tipo de dados coletar deve levar em conta os objetivos do estudo, o grau de conhecimento sobre as características sendo investigadas, a forma como serão coletados os dados (questionários auto aplicados ou entrevistas) e a capacidade de processamento dos dados no caso de serem coletados no formato não estruturado.

Ainda sobre dados não estruturados, é importante ressaltar que procurar padrões onde eles aparentemente não existem é uma ação intuitiva quando queremos analisar dados. Por exemplo, quando um profissional da área da Saúde entrevista uma pessoa à procura de indícios de depressão, por exemplo, o que ele está fazendo? A partir de um dado não estruturado (fala), ele tenta identificar categorias de sentimentos (desânimo, tristeza sem motivo aparente, pessimismo) e queixas (insônia, cansaço exagerado, falta de apetite e de concentração), geralmente associados a um quadro depressivo. O que esse profissional faz é procurar por padrões (categorias) na fala da pessoa entrevistada, sendo orientado pelo objetivo da entrevista (identificar sintomas de depressão). Se há um conjunto de falas da mesma pessoa em tempos diferentes ou de várias pessoas de um grupo, o profissional pode anotar a recorrência dos sentimentos ou queixas em cada fala, organizando esses dados, agora estruturados, em uma planilha, e, posteriormente, utilizar de técnicas estatísticas para extrair informações desses dados.

Dados não estruturados são o principal tipo de dado coletado nas pesquisas qualitativas, bastante utilizadas nas Ciências da Saúde (TURATO, 2005; KNAUTH; LEAL, 2014). A discussão das vantagens e desvantagens desse tipo de pesquisa foge ao escopo deste livro. No entanto, é certo que os dados não estruturados coletados em uma pesquisa qualitativa passarão

por algum tipo de processamento antes de serem interpretados. Turato (2005, p. 511) deixa claro que o “tratamento/análise dos dados” se dá pela “categorização por relevância teórica (...) realizada pelo pesquisador” e que a “discussão dos resultados” é feita pela “interpretação de dados categorizados, simultaneamente à apresentação destes”. Como vimos anteriormente, a “categorização por relevância teórica” é uma das formas de dar estrutura aos dados não estruturados e a “interpretação de dados categorizados” pode ser feita pelo uso de técnicas de estatística descritiva¹, como ilustra o exemplo no início dessa seção. Quanto a ser “realizada pelo pesquisador”, nada deveria impedi-lo de ter a ajuda de métodos computacionais, transferindo para a máquina o esforço de procurar categorias e deixando para o ser humano as tarefas mais nobres de definir quais são as categorias de maior “relevância teórica” e dar aos achados da máquina “uma interpretação com base em um referencial teórico” (KNAUTH; LEAL, 2014).

É preciso compreender que o pré-processamento de dados não implica em “perda de informação”. Ao contrário, e no contexto da nossa discussão sobre dado/informação/conhecimento, o pré-processamento de dados não estruturados pode facilitar a extração de informação relevante por meio de técnicas estatísticas, permitindo que essa informação seja interpretada para gerar conhecimento.

¹ Como a amostra de uma pesquisa qualitativa é selecionada “intencionalmente” e o tipo de generalização pretendida não é a estatística (inferencial), mas a “conceitual” (TURATO, 2005), utilizar técnicas de estatística inferencial não seria adequado.

Síntese do capítulo

- Dado, informação e conhecimento possuem conceitos distintos, porém interligados.
- Dados estruturados são aqueles que podem ser organizados no formato linha/coluna, enquanto dados não estruturados não se encaixam nesse formato.
- Dados não estruturados podem ser processados para gerar dados estruturados.
- O processamento de dados não estruturados deve ser direcionado pela pergunta/hipótese de pesquisa.

Referências

ALMEIDA, M. B. Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares. **Ciência da Informação**, v. 31, n. 2, p. 5–13, 2002.

BARLAS, I.; GINART, A.; DORRITY, J. L. **Self-Evolution in Knowledgebases**. Proceedings of IEEE Autotestcon 2005. Orlando, FL: IEEE, 2005. p. 325-331.

DAVENPORT, T. H.; PRUSAK, L. **Information Ecology: Mastering the Information and Knowledge Environment**. New York: Mastering the Information and Knowledge Environment, 1997.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. 3. ed. Waltham, MA, USA: Elsevier Inc., 2012.

KNAUTH, D. R.; LEAL, A. F. A expansão das Ciências Sociais na Saúde Coletiva: usos e abusos da pesquisa qualitativa. **Interface**, Botucatu, 18, 2014. 457-67.

SCHUMAKER, R. P. From Data to Wisdom: The Progression of Computational Learning in Text Mining. **Communications of the International Information Management Association**. 2011. 39-48.

TURATO, E. R. Métodos qualitativos e quantitativos na área: definições, diferenças e seus objetos de pesquisa. **Revista de Saúde Pública**, 39, 2005. 507-14.

WHITNEY, H. How to define data, information and knowledge. **TechTarget**, November 2007. Disponível em: <<http://searchdatamanagement.techtarget.com/feature/Defining-data-information-and-knowledge>>. Acesso em: 04 de Junho de 2015.

O QUESTIONÁRIO COMO TÉCNICA DE COLETA DE DADOS

Fernanda Figueredo Chaves
Júlia Santos Nunes Rodrigues

Objetivos do capítulo

- Apresentar as etapas para a construção de um questionário;
- Descrever as vantagens e desvantagens dos questionários impressos e eletrônicos;
- Ilustrar o uso de uma plataforma da web para construção de questionários eletrônicos.

1 Introdução

O questionário é um instrumento de estudo aplicado a uma população-alvo para colher dados sobre essa população. Pode ser aplicado em estudos que utilizam metodologia qualitativa ou quantitativa, sendo uma técnica de levantamento de dados capaz de gerar resultados válidos em todas as áreas do conhecimento (GIL, 2010). O questionário pode também ser um instrumento para avaliar um outro questionário, como é o caso de pesquisas das Ciências da Saúde, em que questionários

são traduzidos, adaptados e, posteriormente, avaliados por meio de um questionário que os avaliadores preenchem.

Como apontado no Capítulo 2, a aplicação do questionário pode ser feita por um aplicativo, que lê em voz alta o texto e registra as respostas do entrevistado ou pode ser lido em silêncio pelo próprio respondente. Além disso, os questionários podem ser apresentados de maneira impressa ou eletrônica.

Com o avanço da tecnologia, tem sido cada vez mais frequente a utilização de questionários construídos e aplicados pela *internet*,

por meio de plataformas e/ou *sites* desenvolvidos para esse fim, auxiliando na condução de estudos e possibilitando ampliar a representatividade da amostra utilizada (RUBENFELD, 2004; FORZA, 2002)

Os pesquisadores precisam ficar atentos às adequações que devem ser feitas durante a elaboração de questionários, pois esse processo implica o detalhamento de técnicas e informações que podem prejudicar a compreensão por parte dos participantes e, conseqüentemente, os resultados do estudo (EYSENBACH, 2004).

Este capítulo abordará a concepção e aplicação de questionários como metodologia para a coleta de dados. Apontaremos também as vantagens e desvantagens dos questionários impressos e eletrônicos, bem como as técnicas que podem ser usadas para a criação de questionários *online*. Para finalizar, apresentaremos

um estudo de caso com o objetivo de ilustrar os tópicos descritos no decorrer do capítulo.

2 O questionário como técnica de coleta de dados

O questionário é composto por um conjunto de questões sobre determinado tópico para coletar dados sobre as opiniões, os interesses ou as características dos entrevistados. Deve ser construído a partir das seguintes etapas: desenvolvimento de um modelo teórico, caracterização da população-alvo e elaboração das questões (MARCONI; LAKATOS, 1999; GIL, 2010).

Os questionários podem ser classificados segundo o tipo de questões, seu preenchimento, o meio de divulgação e o tipo de aplicação, como ilustra a Figura 1:

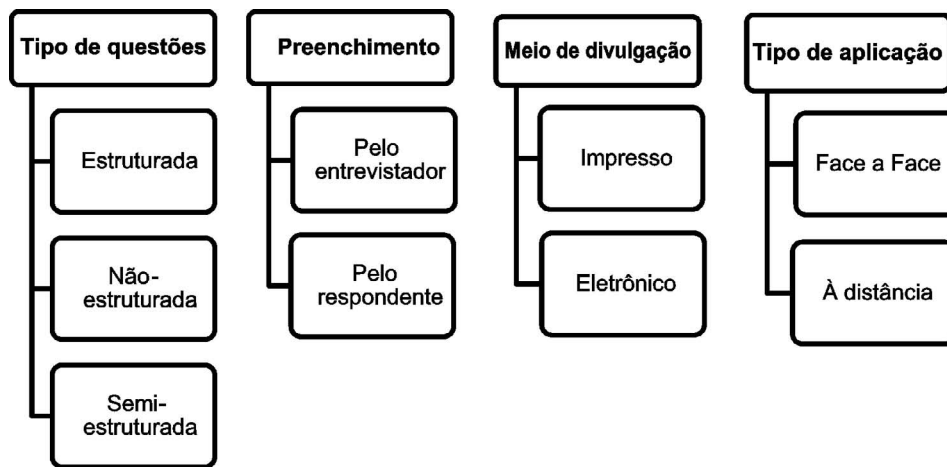


Figura 1 – Classificação dos questionários

Fonte: Elaborado para fins deste estudo, baseado nos conceitos apresentados em Gil (2010).

As combinações das categorias de classificação da Figura 1 configuram diversos tipos de questionários; por exemplo, um questionário impresso pode ser do tipo estruturado, preenchido pelo entrevistador ou pelo respondente e pode ser aplicado por entrevista face a face ou à distância (via telefone, por exemplo). O mesmo pode acontecer com um questionário eletrônico, excluídas, algumas combinações de categorias, como é o caso de “eletrônico,

preenchido pelo respondente e aplicado face a face”, pois, se é o próprio respondente quem preenche o questionário, não é necessária a presença de um entrevistador, que está implícita na aplicação face a face.

A seguir, veremos alguns aspectos dos questionários como instrumento de coleta de dados, desde a etapa de elaboração até a etapa de aplicação.

Etapa 1: *Elaboração do questionário*

Questionários podem ser aplicados de forma presencial ou à distância; ou podem ser autoaplicáveis (EVANS; MATHUR, 2005; HOSS; CATEN, 2010). Antes da sua aplicação, algumas orientações devem ser consideradas, tendo em vista sua qualidade (ROSAL; CARBONE; GOINS, 2003; EYSENBACH, 2004), tais como:

- Adequar os termos e as expressões de acordo com a população-alvo, evitando o uso de linguagem técnica;
- Redigir as questões com uma sequência lógica e objetiva, a fim de não prejudicar

a compreensão dos participantes e facilitar o preenchimento;

- Certificar-se de que o público-alvo escolhido para participar de um estudo *online* tenha familiaridade com o ambiente virtual, caso o questionário seja eletrônico.

No caso de questões com respostas estruturadas, recomenda-se o uso de uma escala de medição que permita uma avaliação rápida e precisa, visando tornar a análise dos dados mais fácil. Os níveis de mensuração que podem ser utilizados em uma questão estruturada são discutidos em (TRIOLA, 2013) e são apresentados no Quadro 1, a seguir:

Quadro 1 – Níveis de Mensuração das Respostas às Questões Estruturadas

Nível de Mensuração	Definição	Exemplos
Nominal	As respostas são classificadas por nomes ou rótulos, sem o uso de ordenação.	<ul style="list-style-type: none">sexo (masculino, feminino);estado civil (solteiro, casado, viúvo, separado);região da cidade (norte, sul)uso de medicamento (sim, não)
Ordinal	As categorias de respostas ou valores numéricos podem ser ordenados, mas, no caso de valores numéricos, a operação de diferença não faz sentido. Pode-se usar símbolos que caracterizam as categorias de resposta, como: <i>emoticons</i> , estrelas e setas.	<ul style="list-style-type: none">escolaridade (nenhuma, primário, ensino médio, universitário);estágio da doença (0-ausente, 1-inicial, 2-intermediário, 3-terminal).<i>escala Likert</i> de sete, cinco ou três pontos, que apresenta possibilidades de resposta que vão da “completa concordância” até a “completa discordância” sobre determinada questão.
Intervalar	As respostas são valores numéricos que podem ser ordenados e a operação de diferença faz sentido. No entanto, o valor zero é apenas uma posição arbitrária na escala, de modo que não faz sentido escrever um valor como múltiplo de outro.	<ul style="list-style-type: none">temperatura (em °C), como medida do calor de um corpo: o valor 0°C é apenas um ponto arbitrário na escala e não significa “nenhum calor”; a diferença entre dois valores de temperatura faz sentido (a temperatura de hoje está 2° C mais baixa do que a de ontem), mas um corpo com temperatura de 20° C não está duas vezes mais quente do que um corpo a 10° C.
Razão	As respostas são valores numéricos que podem ser ordenados e o valor zero (absoluto) é uma quantidade (nenhuma). A operação de diferença faz sentido e, além disso, pode-se calcular a razão entre dois valores, tal que um valor pode ser escrito como múltiplo de outro.	<ul style="list-style-type: none">idade (em dias, meses ou anos completos);número de filhos (0,1,2, ...);altura (em centímetros);peso (em quilogramas);valor da glicohemoglobina, em %;distância a percorrer diariamente por um maratonista durante um mês, em quilômetros.

Fonte: Elaborado para fins deste estudo, baseado nos conceitos apresentados em Triola (2013).

Como exposto no Capítulo 3, as respostas às questões estruturadas também podem ser classificadas em categóricas ou qualitativas (nominais ou ordinais) e quantitativas (discretas, para respostas que admitem apenas valores inteiros; e contínuas, para respostas que admitem valores não-inteiros). Respostas do tipo qualitativa nominal e ordinal utilizam os níveis de mensuração nominal e ordinal, respectivamente. No entanto, as respostas do tipo quantitativa discreta e contínua podem utilizar tanto o nível de mensuração intervalar quanto o nível razão. Por exemplo, a medida de temperatura corporal é uma resposta do tipo contínua e com nível de mensuração intervalar. Já o peso é uma resposta também do tipo contínua, mas com nível de mensuração razão. Como veremos no Capítulo 8, a classificação das questões de um instrumento quanto ao tipo de respostas que elas geram é importante para determinar que ferramentas devem ser utilizadas para a análise estatística

dessas respostas, ou seja, dos dados coletados pelo instrumento.

Ainda no Capítulo 3, vimos que questionários com respostas não estruturadas viabilizam que o participante utilize suas próprias palavras; porém dificultam a tabulação e análise estatística dos dados, sendo necessária uma técnica de pré-processamento, que pode ser a mineração de textos, tema que será apresentado no Capítulo 6.

Etapa 2: *Treinamento dos Aplicadores*

Para assegurar a eficiência do estudo, recomenda-se a realização de treinamento para todos aplicadores, a fim de que a aplicação do questionário saia como o previsto e os resultados fiquem coerentes com o estudo, evitando, portanto, conclusões errôneas do questionário (MOYSÉS; MOORI, 2007). A Figura 2 delinea algumas orientações para o preenchimento do questionário, quando este preenchimento for feito por um aplicador.

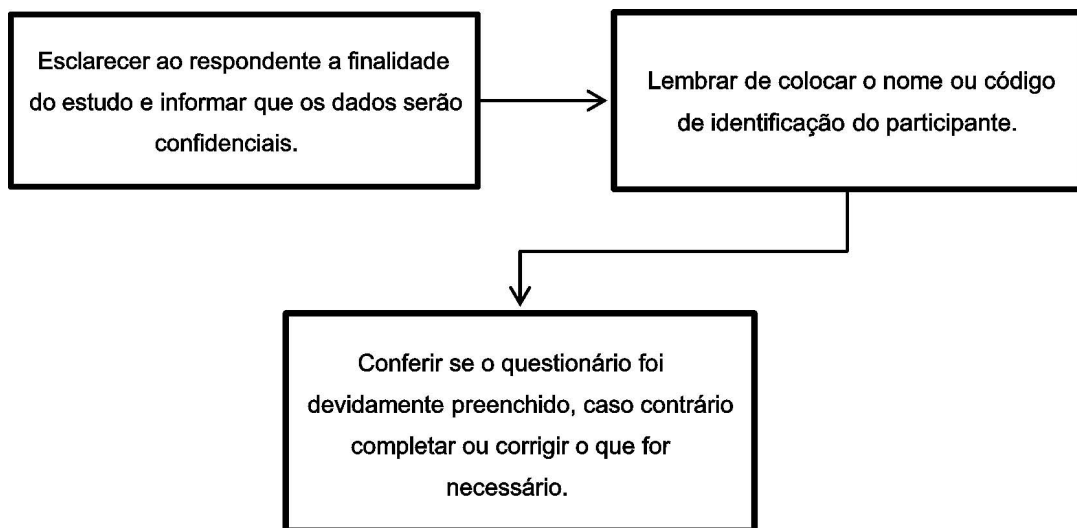


Figura 2 – Orientações para o preenchimento de questionários

Fonte: Elaborado para fins deste estudo, baseado em Marconi e Lakatos (1999).

Etapa 3: *Pré-teste do questionário*

Após o pré-teste, recomenda-se que encontros interdisciplinares com especialistas

das áreas da linguística, estatística e de outras áreas importantes para o estudo sejam realizados para discutir e efetuar os ajustes de

acordo com os comentários dos participantes do pré-teste. Esta etapa já é realizada dentro do projeto Empoder@ e pode ser repetida até que se chegue à versão final do instrumento (EPSTEIN et al., 2013; CHAVES, 2014). É importante lembrar que esses encontros interdisciplinares também devem ser feitos em questionários desenvolvidos pelos pesquisadores, não somente em questionários traduzidos.

Etapa 4: Aplicação do questionário

Os questionários impressos podem ser autoaplicáveis ou aplicados face a face pelo entrevistador. Já os questionários eletrônicos podem ser aplicados via ligação telefônica, na qual o entrevistador preenche o questionário a partir da conversa com o participante, ou pode ser *online*, quando o *link* do questionário é enviado pelo correio eletrônico e o respondente preenche o questionário. Neste caso, para evitar que a mensagem com o convite não seja considerada um spam, sugere-se que o tema do *e-mail* a ser enviado seja revelado na própria caixa de assunto. Além disso, recomenda-se que no corpo do *e-mail* apareça o logotipo da instituição e telefone para confirmação da pesquisa. E, por último, aconselha-se fazer contato prévio via mensagem de celular, dessa forma, após todas essas precauções, espera-se que o respondente tenha condições de atestar que não se trata de uma pesquisa falsa ou de um *spam* (FREITAS et al., 2000).

3 Questionários impressos versus eletrônicos – vantagens e desvantagens

A partir do avanço tecnológico na área da informática, geralmente os questionários impressos são elaborados no computador, com o auxílio de um editor de texto, depois impressos e distribuídos para os indivíduos responsáveis pela avaliação do instrumento e/ou para as pessoas que responderão o questionário. Os questionários eletrônicos são, em geral, elaborados com auxílio de plataformas/sites disponíveis na *internet* previamente criados para esse fim. Em outras palavras, todo o processo de avaliação e aplicação do questionário pode ser feito totalmente *online* (STARR, 2012; EPSTEIN et al., 2013). No entanto, existem plataformas para a elaboração de questionários eletrônicos que não necessitam de conexão com a *internet* para serem aplicados, como é o caso do *EpiData Entry* (LAURITSEN, 2000), ou seja, os questionários podem ser preenchidos mesmo se o aplicador e/ou respondente estiver *offline*.

No que diz respeito às vantagens e as desvantagens desses dois tipos de questionários, o Quadro 2 apresenta a comparação entre os questionários impressos e eletrônicos:

Quadro 2 – Comparação entre os questionários impressos e eletrônicos

Aspecto	Questionário Impresso	Questionário Eletrônico
Despesas com Material e Serviços	É preciso ter recursos financeiros para a impressão de cópias dos questionários e para o envio dos questionários via correio convencional, se for o caso.	Evitam-se gastos com papel, tinta e envio pelo correio convencional, pois não há necessidade de imprimir várias cópias do questionário e nem de enviá-los para a residência do participante.
Digitação dos dados	Os dados devem ser digitados e podem ocorrer erros de digitação, perda de dados, entre outros.	Esta etapa é desnecessária, pois os dados já são armazenados em uma planilha eletrônica, desde que tenham sido implementados numa plataforma para questionários.
Análise Quantitativa	As respostas dos respondentes ou avaliadores não estão no formato adequado para serem inseridas em <i>softwares</i> estatísticos, a fim de que a análise quantitativa possa ser feita.	As respostas dos respondentes ou avaliadores podem ser direcionadas para um software estatístico sem necessidade de modificações, pois a grande maioria dos arquivos são compatíveis com os softwares, já que estão em formato .txt, .csv ou xls (cf. capítulo 3).
Alcance da população-alvo	Caso os questionários impressos forem enviados via correio convencional, é necessário que os pesquisadores tenham um banco de dados com as informações das residências dos respondentes, o que pode ser mais difícil de se conseguir do que os endereços eletrônicos.	Como são eletrônicos, podem ser enviados via correio eletrônico a um custo menor que se o correio convencional fosse utilizado e também podem ser disponibilizados em <i>sites</i> , blogs e redes sociais.
Tempo de resposta	No caso de levantamento de dados à distância, o tempo total de resposta inclui, além do tempo despendido com o preenchimento do questionário, o tempo de entrega e de retorno do questionário via correio convencional ou outro serviço de entrega.	No caso de levantamento de dados à distância, o tempo total de resposta inclui somente o tempo gasto com o preenchimento do questionário, visto que os tempos de entrega e de retorno são praticamente nulos quando se emprega o correio eletrônico.
Organização das respostas e/ou avaliações	Os pesquisadores responsáveis pelos questionários precisam ser organizados e estabelecerem uma metodologia para o armazenamento das respostas e/ou avaliações que recebem, pois como se tem uma grande quantidade de papel com o questionário respondido e/ou avaliado por diferentes pessoas é possível misturar ou perder dados importantes.	As respostas e/ou avaliações feitas pelo comitê de juízes ou respondidas pelo público-alvo são armazenadas na plataforma onde o questionário foi construído. Assim, os pesquisadores não correm o risco de se perderem em tantos papéis e respostas/avaliações diferentes.
Peculiaridades para construção dos questionários	O pesquisador tem maior autonomia para escolher como o questionário será desenvolvido, seja por meio de editor de texto, seja manualmente, usando papel e caneta.	Muitas das plataformas e/ou <i>sites</i> disponíveis na <i>internet</i> estão em inglês, portanto os pesquisadores responsáveis por criarem os questionários precisam ter conhecimento básico nessa língua para desenvolver o questionário de maneira correta e conseguir atingir os propósitos da pesquisa.
Dependência da máquina	Não há dependência da máquina e, consequentemente, nem de energia elétrica para responder e/ou avaliar o questionário, pois esta etapa é feita a mão.	Caso o questionário eletrônico for aplicado <i>online</i> , os pesquisadores precisam verificar se todos os possíveis respondentes e/ou avaliadores têm acesso a um dispositivo com <i>internet</i> , caso contrário deverão providenciar um lugar adequado para que eles possam preencher o questionário.

Fonte: Elaborado para fins deste estudo, baseado nos resultados de Toledo et al. (2015).

Assim, com base nas vantagens e desvantagens descritas no Quadro 2, podemos afirmar que cada tipo de pesquisa exige um tipo de questionário diferente, eletrônico ou impresso. No entanto, vale ressaltar ainda que essa decisão é muito importante e não deve ser tomada somente com base nos aspectos relacionados à construção do questionário. Ela também deve ser pautada nos objetivos da pesquisa e na forma de avaliação, presencial ou à distância. E, por fim é necessário ter em mente que a relação-homem-máquina para construção de questionários deve sempre atentar-se para o empoderamento do pesquisador, do profissional da área da Saúde e do usuário do serviço de Saúde como proposto no Capítulo 1.

4 Construção do questionário em uma ferramenta online

A máquina, isto é, as plataformas e/ou os sites desenvolvidos para a criação de questionários *online*, servem para que o processo de montagem do questionário seja otimizado, pois os pesquisadores despendem menos tempo para criar o questionário. A probabilidade de erros de digitação de respostas e contabilização de dados é praticamente inexistentes. Indivíduos de diferentes lugares do mundo podem ser convidados a responder o questionário, pois todas as etapas do estudo são realizadas de forma *online*.

Entretanto, se por trás das plataformas e/ou sites de criação de questionários eletrônicos não houver um pesquisador humano, os questionários *online* não seriam tão úteis,

porque a máquina por si só não é capaz de verificar se a linguagem do questionário está adequada, ou seja, se o respondente será capaz de compreender corretamente o que os pesquisadores querem avaliar. Além disso, a máquina também não é capaz de interpretar os resultados que os questionários preenchidos geram. Portanto, cabe aos pesquisadores escolher pela plataforma e/ou site de desenvolvimento de questionários que mais trará benefícios de acordo com os objetivos da pesquisa. Somente o ser humano é considerado capaz de montar questionários eletrônicos que extraíam tudo que o respondente tem a acrescentar ao aprimoramento do instrumento que está sendo avaliado. E apesar de algumas plataformas e alguns sites contarem com a ferramenta de ajuda para auxiliar na criação de questionários *online*, é necessário que os pesquisadores saibam distinguir o que favorece e o que desfavorece o questionário que está sendo montado.

Como foi reiterado, de acordo com Freitas *et al.* (2004), a *internet* vem agregando novas tecnologias, com novos recursos, cada vez mais interativos. Os pesquisadores têm a possibilidade de utilizar recursos que não seriam possíveis em um processo tradicional de coleta de dados em uma pesquisa que usa questionários impressos. O respondente, por sua vez, recebe estímulos (visuais, sonoros, etc.) que o incentivam a participar. Pesa a favor do pesquisador também a facilidade com que tudo isto é feito, e a favor do respondente a liberdade de participar quando lhe for mais conveniente. A Figura 3 ilustra algumas dicas para a criação de questionários *online*.

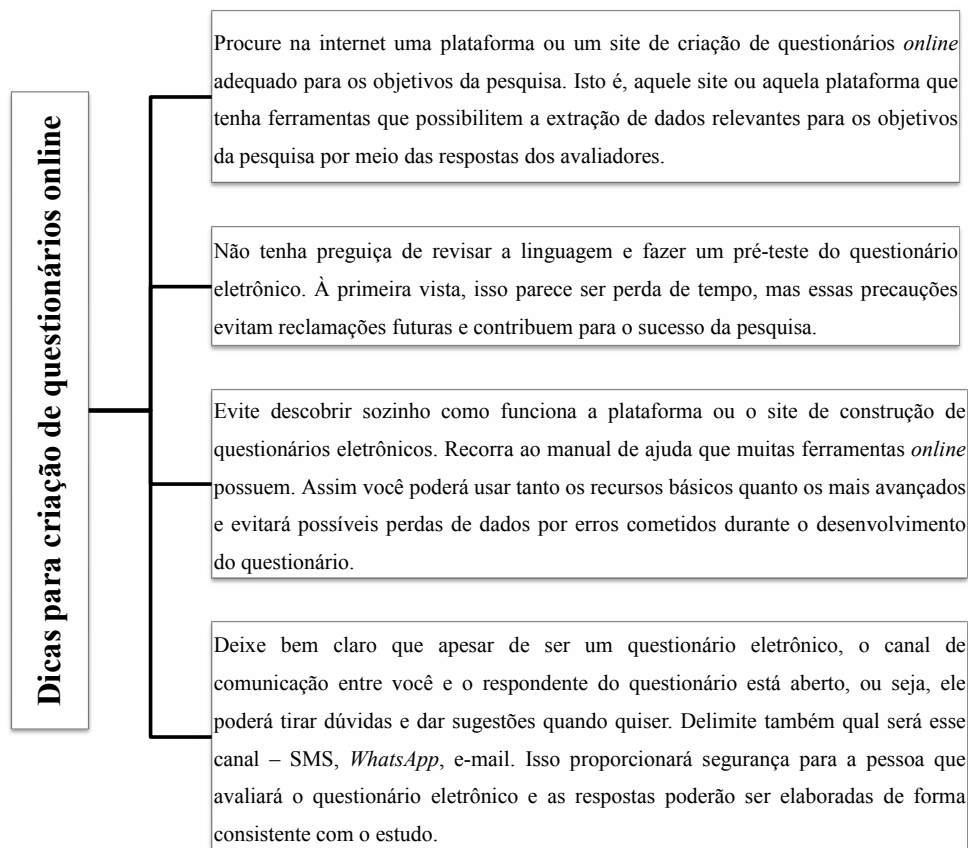


Figura 3 – Dicas para a criação de questionários eletrônicos

Fonte: Elaborado para fins deste estudo, baseado em Moysés e Moori (2007).

5 Estudo de caso / exemplo – construção passo a passo

A seguir, serão apresentadas as diferentes etapas para a construção de um questionário *online* no contexto de tradução e adaptação cultural de instrumentos e sua posterior validação. O estudo foi desenvolvido pelo projeto Empoder@ - Protótipo conceitual e metodológico para avaliação de intervenções orientadas ao autocuidado em diabetes (Escola de Enfermagem/UFMG, Faculdade de Letras/UFMG e Departamento de Estatística/ICEX/UFMG) e aplicado junto a profissionais da área da Saúde e indivíduos com diagnóstico de Diabetes Mellitus tipo 2, partindo de um estudo que teve por objetivo traduzir e adaptar culturalmente o “*Diabetes Empowerment Scale-Short Form*”, visando a sua aplicação no contexto

cultural brasileiro, bem como validar a sua versão para a língua portuguesa brasileira. (ANDERSON et al., 2000).

Nesse estudo, a plataforma *web* eSurv (ESURV.ORG, 2015) foi utilizada em três etapas: na avaliação da versão traduzida por um comitê de juízes (adaptação cultural) (GUILLEMIN; BOMBARDIER; BEATON, 1993; BARROSO, 2015); na aplicação da versão adaptada a uma amostra da população-alvo a fim de identificar erros e problemas de compreensão (adequação cultural); e na validação da versão adequada culturalmente (teste e reteste com a população-alvo). A plataforma eSurv foi escolhida por se tratar de um *site* gratuito para criação de questionários *online*, por apresentar várias ferramentas que dinamizam o questionário, por disponibilizar uma função que permite que

o criador envie o *link* da pesquisa por *e-mail* para os respondentes, por ter uma função que envia um *e-mail* ao pesquisador que montou o questionário, quando a pessoa já terminou de responder, por permitir que o respondente tenha flexibilidade, ou seja, ele pode começar a responder e terminar quando for conveniente, por fazer análises estatísticas básicas de forma automática e, por fim, por ser uma plataforma desenvolvida por universidades renomadas¹.

Na etapa de adaptação cultural, uma carta convite foi enviada aos profissionais da área da Saúde via correio eletrônico contendo os objetivos e a metodologia do estudo, a justificativa do processo de tradução e adaptação do instrumento, e a solicitação da participação na pesquisa como juiz avaliador, por meio do acesso a uma plataforma da *web* - eSurv (Figura 4).

Atenção: Esta parte do DES-SF é uma instrução apenas para o aplicador do questionário!

Diabetes Empowerment Scale-Short Form (DES-SF)
The 8 items below constitute the DES-SF. The scale is scored by averaging the scores of all completed items (Strongly Disagree = 1, Strongly Agree = 5).

Versão curta da Escala de Empoderamento em Diabetes (DES-SF).
Os 8 itens abaixo fazem parte da versão curta da Escala de Empoderamento em Diabetes. Para cada item há cinco opções de resposta, que serão pontuadas de 1 (Não estou de acordo de jeito nenhum) a 5 (Estou muito de acordo). A avaliação final é feita pela média da pontuação das respostas.

1. Você acha que o texto em português está de acordo com o do inglês?

Sim
 Não
 Em parte

2. Se você respondeu NÃO ou EM PARTE, como você reformularia o texto em português?

Anterior Próximo

Create your own survey for free with esurv.org

Figura 4 – Tela capturada ilustrando o questionário de avaliação da versão traduzida construído na plataforma da *web* eSurv

O comando no eSurv de receber por *e-mail* as respostas finalizadas de cada juiz foi ativado, para controlar quantos profissionais da área da Saúde já haviam avaliado o instrumento. E, caso o convidado não

respondesse em sete dias, um novo correio eletrônico e uma mensagem de texto pelo celular (se disponível) eram enviados, pelo pesquisador responsável, para reforçar o convite, já que a plataforma não é capaz de enviar automaticamente um novo *e-mail* para aquele que não respondeu.

Na fase de adequação cultural, a versão adaptada foi implementada no eSurv e utilizada em um pré-teste com uma amostra da

¹ The Ohio State University; The University of Edinburgh, UK; Indiana University; University of Tennessee, Knoxville; University of Provence lettres et sciences humaines, France e University of Toronto, Canadá.

população alvo, a fim de detectar eventuais deficiências no instrumento, de modo que a versão final fosse de fácil entendimento para os respondentes de acordo com o português brasileiro coloquial. O pré-teste foi feito via ligação telefônica com um grupo de onze indivíduos diagnosticados com Diabetes Mellitus tipo 2.

Os quadros de respostas para as questões abertas foram padronizados, através da inserção de espaços delimitados para a escrita da resposta. Foram ainda inseridas fontes em negrito ou em azul e linhas separando o texto original e o texto traduzido para não prejudicar a leitura ao longo do questionário. As questões foram numeradas e apresentadas em uma rolagem de página única para facilitar a visualização do questionário e facilitar o seu preenchimento.

No canto direito superior do eSurv, há a logomarca das instituições envolvidas e na última página o *e-mail* de contato dos pesquisadores com objetivo de reforçar a credibilidade do estudo, caso alguém desejasse confirmar a veracidade da mesma.

O questionário implementado com o uso do eSurv era composto pelas seguintes seções:

Seção 1: *Instruções para o entrevistador*

O entrevistador foi orientado a acolher o usuário, explicar o estudo, em especial os objetivos, obter sua anuência para participar e fornecer as respostas do instrumento, e ainda solicitar para que a ligação telefônica pudesse ser gravada (Figura 5).

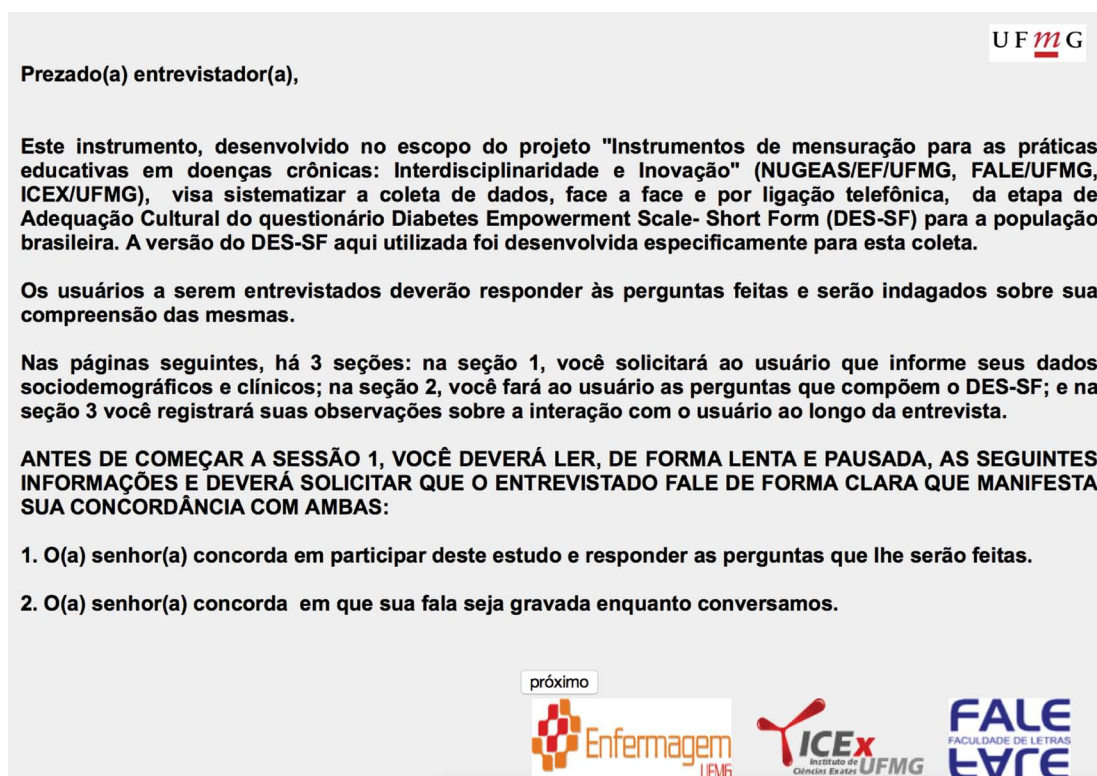


Figura 5 – Tela capturada ilustrando a página no eSurv - Instruções para o entrevistador – Questionário utilizado na adequação cultural

Seção 2: *Dados sociodemográficos e clínicos do usuário*

Preenchimento de nome, sexo, idade, estado civil, ocupação, duração da doença, tipo de tratamento e se apresentava algum problema

de saúde. Na parte superior do questionário, há uma indicação do tema que as questões abordam, bem como os botões de navegação, que permitem avançar ou retroceder (Figura 6).

Adequação Cultural do Diabetes Empowerment Scale- Short Form (DES- SF) para a população brasileira

UFMG

SEÇÃO 1 :

IDENTIFICAÇÃO DO USUÁRIO

*
Nome:

*
Idade:

Telefone:

Endereço:

Centro de saúde:
 Horto
 Paraíso
 Pompéia
 São José Operário
 Boa Vista

* Duração da doença (anos):

Problema de saúde:
 Sim
 Não
Se sim, qual?

* Escolaridade:
 Analfabeto
 Fundamental Completo
 Fundamental Incompleto
 Médio Completo
 Médio Incompleto
 Superior Completo
 Superior Incompleto




Figura 6 – Tela capturada ilustrando a página no eSurv - Dados sociodemográficos e clínicos do usuário – Questionário utilizado na adequação cultural

Seção 3: *Afirmativas da versão brasileira do instrumento*

Realização da leitura de cada afirmativa do instrumento pelo entrevistador, seguida da

apresentação das opções da escala do tipo *Likert* e a análise da compreensão dos usuários do serviço de Saúde (Figura 7).

Adequação Cultural do Diabetes Empowerment Scale- Short Form (DES- SF) para a população brasileira

UFMG

SEÇÃO 2:

QUESTIONÁRIO

Eu (profissional de saúde) vou falar algumas frases sobre como o(a) senhor (a) está cuidando do diabetes e ai o(a) senhor(a) me fala se está de acordo, não está de acordo ou não tem opinião. Combinado?

*

1) O(a) senhor (a) sabe que coisas tem que fazer para cuidar da sua saúde, mas não gosta de fazer.

O(a) senhor (a):

- Não está de acordo de jeito nenhum COM ISSO QUE FALEI
- Não está de acordo COM ISSO QUE FALEI
- Não tem opinião SOBRE ISSO QUE FALEI
- Esta de acordo COM ISSO QUE FALEI
- Esta muito de acordo COM ISSO QUE FALEI

*

Compreensão do usuário:

- Usuário não teve nenhuma dificuldade de entender
- Usuário teve dificuldade de entender
- Usuário solicitou a repetição da frase mais de uma vez
- Usuário não respondeu

Sugestão/Comentário:

Figura 7 – Tela capturada ilustrando a página no eSurv - Afirmativas do instrumento – Questionário utilizado na adequação cultural

Seção 4: Observações do entrevistador

Descrição do tempo de duração da aplicação, um espaço para o entrevistador pontuar se o participante teve alguma dificuldade em

responder o questionário, e agradecimento pela participação e colaboração no o estudo (Figura 8).

Adequação Cultural do Diabetes Empowerment Scale- Short Form (DES- SF) para a população brasileira

UFMG

SEÇÃO 3:

OBSERVAÇÕES DO ENTREVISTADOR

* Tempo de duração da aplicação:

- 5 - 10 min
- 10 - 15 min
- 15 - 20 min
- 20 - 25 min
- 25 - 30 min

Comentário/ Sugestão:

Agradecer o usuário pela participação e colaboração com o estudo.

Figura 8 – Tela capturada ilustrando a página no eSurv - Observações do entrevistador – Questionário utilizado na adequação cultural

Essa versão foi aplicada via ligação telefônica em 100 usuários do serviço de Saúde com Diabetes Mellitus tipo 2 em dois momentos (teste e reteste), com um intervalo de 15 dias. Tanto a primeira quanto a segunda entrevistas foram realizadas pelo mesmo entrevistador, nos mesmos horários, a fim de priorizar a padronização das respostas e não haver influência de outro entrevistador.

Todos os dados obtidos foram, então, codificados e armazenados anonimamente em uma planilha eletrônica de dados (Figura 9), que foi importada no ambiente de programação estatística R (R CORE TEAM, 2015) para a análise descritiva dos dados, caracterizando a amostra segundo as variáveis sociodemográficas e clínicas.

Código do usuário	Idade	Telefone	Educationalidade	Ocupação	Renda mensal	Estado civil	Duração da doença	Em anos	Se < que 1 ano, favor
390-ILS	65	34031605	Fundamental Incompleto	Inativo	2 salários	com companheiro	1		
893-ISMV	67	34665043	Fundamental Completo	Inativo	1 salário	sem companheiro	6		
901-MAFC	70	34664450	Fundamental Incompleto	Inativo	1 salário	com companheiro	15		
388-MFS	70	34056499	Fundamental Completo	Inativo	1 salário	sem companheiro	10		
903-MAC	56	34037175	Fundamental Completo	Inativo	1 salário	sem companheiro	3		
906-IavC	56	34037477	Fundamental Incompleto	Inativo	2 salários	com companheiro	10		
217-VERH	65	3403-5551	Analfabeto	Ativo	1 salário	com companheiro	4		
226-JAG	72	34034523	Analfabeto	Inativo	1 salário	sem companheiro	6		
367-LCA	56	34031304	Fundamental Incompleto	Inativo	1 salário	com companheiro	2		
373-CHA	67	34033266	Fundamental Completo	Inativo	1 salário	sem companheiro	10		
315-CFC	63	3405-6735	Fundamental Completo	Ativo	2 salários	sem companheiro	6		
276-EGF	73	34059675	Fundamental Completo	Inativo	2 salários	sem companheiro	7		
283-ORA	63	93573486/34858990	Fundamental Completo	Ativo	< 1 salário	com companheiro	24		
300-MCS	75	34057060	Fundamental Completo	Inativo	2 salários	sem companheiro	5		
299-MFS	69	34057793	Fundamental Completo	Inativo	2 salários	com companheiro	5		
813-ARSM	94	9337-6499/ 3463-1098	Fundamental Incompleto	Ativo	2 salários	com companheiro	20		
760-APP	73	3405-9395	Analfabeto	Inativo	2 salários	com companheiro	10		
799-JA	61	3468-2432	Fundamental Completo	Ativo	1 salário	com companheiro	3		
236-ACS	71	3468-2605	Fundamental Incompleto	Inativo	1 salário	sem companheiro	30		
322-DGM	62	34056091	Superior Completo	Ativo	2 salários	com companheiro	5		
799-SGS	73	34070794	Fundamental Completo	Inativo	2 salários	com companheiro	15		
379-MCB	68	34039392	Analfabeto	Inativo	2 salários	sem companheiro	7		
	60		Médio Completo	Inativo	2 salários	com companheiro	8		
892-ON	65	34032096	Fundamental Incompleto	Inativo	< 1 salário	com companheiro	20		
137-CFD	67	3466-3459	Médio Completo	Inativo	2 salários	com companheiro	15		
118-JAC	69	3408-1039	Médio Completo	Inativo	1 salário	com companheiro	4		

Figura 9 – Tela capturada ilustrando a planilha eSurv de armazenamento de dados

Podemos concluir que o estudo de caso acima descrito, trata-se de um questionário eletrônico em que o respondente teve total autonomia, ou seja, não houve necessidade de um aplicador, responsável para fazer as perguntas. Além disso, o desenvolvimento do questionário foi realizado de maneira eficaz, pois a plataforma disponibiliza a ferramenta de ajuda (*help*), para que o pesquisador não tenha

dúvidas na hora da criação. Vários respondentes de diferentes estados brasileiros puderam avaliar o instrumento que estava em pauta, pois não houve necessidade de encontros presenciais para tal avaliação. Por fim, uma nova versão do instrumento foi desenvolvida levando em conta as contribuições dadas pelos avaliadores que participaram do questionário *online* acima descrito.

Síntese do capítulo

- Um questionário pode servir como instrumento de coleta de dados e como instrumento de avaliação de outro questionário;
- Um questionário pode ser impresso ou eletrônico, aplicado face a face ou à distância, preenchido pelo entrevistador ou pelo próprio respondente e ter questões estruturadas, não-estruturadas ou uma mistura de ambas;
- A escolha da escala a ser utilizada em cada questão é muito importante e deve ser feita com cuidado;
- As vantagens e desvantagens dos questionários eletrônicos e impressos devem ser consideradas de acordo com o estudo a ser realizado;
- Questionários eletrônicos são mais econômicos, pois evitam gastos com impressão, digitalização das respostas e envio pelo correio convencional;
- As plataformas e/ou os sites disponíveis *online* para criação de questionários eletrônicos permitem que os pesquisadores usem recursos como vídeos, *emoticons* e barras de progresso, o que chama a atenção do respondente;
- Os dados gerados pelas respostas dos questionários eletrônicos são mais fáceis de armazenar e analisar, pois os sites ou plataformas que constroem questionários *online* apresentam ferramentas que facilitam a análise quantitativa das respostas.

Referências

- ANDERSON, R. M. et al. The Diabetes Empowerment Scale: a measure of psychosocial self-efficacy. **Diabetes Care**, v. 23, n. 6, p. 739-743, 2000.
- BARROSO, R. A. A. **Multidisciplinaridade e inovação metodológica na dinâmica de atuação do comitê de juízes na adaptação de instrumentos: Diabetes Empowerment Scale-Short Form (DES-SF)**. Dissertação (Mestrado em Enfermagem)—Belo Horizonte: Universidade Federal de Minas Gerais. Escola de Enfermagem, 2015.
- CHAVES, F. A. **Tradução e adaptação cultural do “Behavior Change Protocol” para a língua portuguesa-Brasil**. Dissertação (Mestrado em Enfermagem)—Belo Horizonte: Universidade Federal de Minas Gerais. Escola de Enfermagem, 2014.
- EPSTEIN, J. et al. Cross-cultural adaptation of the Health Education Impact Questionnaire: experimental study showed expert committee, not back-translation, added value. **Journal of Clinical Epidemiology**, v. 68, n. 4, p. 360-369, 2013.
- ESURV.ORG. **Survey Maker, Free Online Questionnaire Creator**. Disponível em: <<http://esurv.org/>>.
- EVANS, J. R.; MATHUR, A. The value of online surveys. **Internet Research**, v. 15, n. 2, p. 195-219, 2005.
- EYSENBACH, G. Improving the Quality of Web Surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). **Journal of Medical Internet Research**, v. 6, n. 3, 2004.
- FORZA, C. Survey research in operations management: a process-based perspective. **International Journal of Operations & Production Management**, v. 22, n. 2, p. 152-194, 2002.
- FREITAS, H. et al. O método de pesquisa survey. **Revista de Administração da Universidade de São Paulo**, v. 35, n. 3, 2000.
- FREITAS, H. et al. Pesquisa via internet: características, processo e interface. **Revista Eletrônica GIANTI**, 2004.
- GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2010.

- GUILLEMIN, F.; BOMBARDIER, C.; BEATON, D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. **Journal of Clinical Epidemiology**, v. 46, n. 12, p. 1417-1432, 1993.
- HOSS, M.; CATEN, C. S. TEN. Processo de Validação Interna de um Questionário em uma Survey Research Sobre ISO 9001:2000. **Produto & Produção**, v. 11, n. 2, p. 104-119, 2010.
- LAURITSEN, J. M. **EpiData Data Entry, Data Management and basic Statistical Analysis System**. Odense Denmark: EpiData Association, 2000.
- MARCONI, M. DE A.; LAKATOS, E. M. **Técnicas de pesquisa**. São Paulo: Atlas, 1999.
- MOYSÉS, G. L. R.; MOORI, R. G. **Coleta de dados para a pesquisa acadêmica: um estudo sobre a elaboração, a validação e a aplicação eletrônica de questionário**. In: XXVII ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO. Foz do Iguaçu, 2007.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation for Statistical Computing, 2015.
- ROSAL, M. C.; CARBONE, E. T.; GOINS, K. V. Use of cognitive interviewing to adapt measurement instruments for low-literate Hispanics. **The Diabetes Educator**, v. 29, n. 6, p. 1006-1017, 2003.
- RUBENFELD, G. D. Surveys: an introduction. **Respiratory Care Journal**, v. 49, n. 10, p. 1181-1185, 2004.
- STARR, S. Survey research: we can do better. **Journal of the Medical Library Association: JMLA**, v. 100, n. 1, p. 1-2, 2012.
- TOLEDO, D. et al. Managing an online survey about influenza vaccination in primary healthcare workers. **International Journal of Environmental Research and Public Health**, v. 12, n. 1, p. 541-553, 2015.
- TRIOLA, M. F. **Introdução A Estatística - Atualização da Tecnologia**. 11. ed. São Paulo: LTC, 2013.

TRANSCRIÇÃO DE DADOS VERBAIS

Kícila Ferreguetti
Júlia Santos Nunes Rodrigues

Objetivos do capítulo

- Diferenciar linguagem falada e escrita;
- Apresentar o conceito de transcrição e os principais critérios que devem ser levados em consideração durante o processo de transcrição;
- Enfatizar a importância da complementaridade na relação entre homem e máquina para a automatização das análises em pesquisas;
- Apontar a relevância das transcrições para pesquisas que usam questionários para coleta de dados, ilustrando a metodologia usada em um estudo de caso da área das Ciências da Saúde.

1 Introdução

A transcrição surgiu da necessidade de se registrar a linguagem falada para que ela e suas diversas particularidades pudessem ser estudadas. No entanto, o fato de ser uma representação escrita de um texto falado faz com que a transcrição seja considerada um tipo de texto híbrido, o que, conseqüentemente, pode gerar diversas dúvidas quando resolvemos transcrever uma fala.

O presente capítulo aborda a transcrição, tomando como ponto de partida a distinção entre linguagem falada e escrita, como ela pode influenciar o que entendemos por transcrição e como ela deve ser feita. Em seguida, mostramos que, para se fazer transcrições, não basta somente tentar reproduzir uma fala por escrito; é necessário primeiramente definir os objetivos da pesquisa e estabelecer uma metodologia que desenhará todo o processo de transcrição, a fim de termos consistência em

todas as falas transcritas, com um padrão comum a todas elas.

Finalmente, para ilustrar como as transcrições podem ser feitas, bem como seu potencial para as pesquisas que fazem uso delas, apresentamos um estudo de caso com o *Protocolo Mudança de Comportamento em Diabetes Mellitus tipo 2*. Esse instrumento foi traduzido e adaptado culturalmente para o português brasileiro dentro de uma pesquisa realizada pelo projeto Empoder@, apresentado na Introdução deste volume.

2 Linguagem falada e linguagem escrita – características

Segundo (HALLIDAY; MATTHIENSEN, 2014), utilizamos a linguagem como um recurso para gerar significados na forma de textos falados ou escritos, que nos possibilitam tanto representar a nossa experiência do mundo quanto interagir com as pessoas com as quais compartilhamos nossa experiência.

Como explica Halliday (1989) embora muitos considerem que um mesmo texto pode ser construído de forma falada e/ou escrita, e que a principal diferença entre eles está no grau de formalidade, fala e escrita são, na verdade, meios utilizados para construir diferentes tipos de texto, cujas diferenças não são medidas em termos de formalidade ou prestígio.

Para Halliday (1989), a primeira diferença entre a fala e a escrita está na sua origem: enquanto a linguagem falada foi a primeira a se desenvolver, a escrita só foi desenvolvida muito tempo depois, quando o ser humano precisou utilizar a linguagem para gerar novos significados ou tipos de texto. Nesse sentido, a linguagem escrita não surgiu para possibilitar que aquilo que era falado pudesse ser escrito, mas sim para realizar funções que a fala não era capaz de realizar.

Com o passar do tempo, Halliday aponta (1989), cada vez mais funções consideradas importantes para a nossa vida em sociedade, como as nossas leis, o registro da nossa história, ou os livros nos quais estudamos, passaram a ser realizadas pela linguagem escrita, o que fez com que ela adquirisse um status de prestígio e fosse tida como mais importante do que a linguagem falada. No entanto, essa superioridade não existe. O que distingue a linguagem escrita da falada são as funções desempenhadas por elas: a linguagem falada constrói significado como processo (acontecimento), enquanto a linguagem escrita o faz como produto (existência).

Imagine que você está dando instruções para um colega de como chegar no centro da cidade. O diálogo entre vocês representa um processo que está acontecendo em um determinado tempo e espaço. Processo esse que é dinâmico, ou seja, você pode começar a dar as instruções, lembrar que existe um caminho melhor e reformular o que estava dizendo. Ao mesmo tempo, esse colega pode interromper a sua explicação, fazendo uma pergunta que, conseqüentemente, faz com que você reformule novamente as suas instruções. Esse processo pode continuar indefinidamente ou até que o colega manifeste ter entendido as suas instruções, por exemplo.

Por outro lado, se esse colega tivesse pesquisado na *internet* como chegar ao centro, ele verificaria que existem diversos textos prontos contendo uma ou várias instruções de caminhos que ele poderia seguir. A partir do momento em que foram escritos, esses textos se tornaram produtos ou objetos estáticos que podem ser lidos e reproduzidos a qualquer momento.

Sendo assim, enquanto a linguagem falada é considerada dinâmica e imprevisível, uma vez que não temos como prever como será um texto falado, quantos turnos de fala terá ou o

rumo que a conversa tomará, a linguagem escrita, por sua vez, é vista como estática, estruturada e, em diversos aspectos, previsível. Um exemplo disso é que ao imaginarmos como seria um texto escrito contendo instruções de como chegar no centro da cidade, não temos dúvidas sobre o vocabulário que seria utilizado e nem sobre como ele estaria estruturado.

Entretanto, como Halliday argumenta (1989), é importante ter em mente que o motivo pelo qual o texto escrito é descrito como estruturado e, até mesmo, previsível é porque só temos acesso à sua versão final, ou seja, ao produto. Em outras palavras, nós não temos acesso ao seu processo de produção e nem às várias revisões pelas quais ele passa, nas quais partes são reescritas, mudadas de lugar e/ou cortadas, até que se chegue ao produto final. Em contrapartida, o texto falado é construído à medida que uma ou mais pessoas falam e, por isso, apresenta hesitações, reformulações e, inclusive, inconsistências ou “erros”, que não podem ser revisados e editados no final.

Nesse sentido, entender a diferença entre textos falados e escritos é importante, principalmente quando nos deparamos com textos nos quais essa distinção não é tão clara, como é o caso do texto produzido a partir de uma transcrição, assunto que abordaremos na próxima seção.

3 Transcrição da linguagem falada

Como já apontamos, a transcrição surgiu para possibilitar que a linguagem falada pudesse ser registrada e, então, estudada. Por ser uma representação escrita de um texto falado, a transcrição pode ser considerada um texto híbrido, o que, à primeira vista, pode gerar confusão e dúvidas sobre o que seria uma boa transcrição e como ela deve ser feita.

Uma das principais fontes de confusão e dúvida advém do fato de a transcrição ser escrita. Isso faz com que tenhamos a impressão de que um texto transcrito deve se assemelhar a um texto escrito, ou seja, ficamos tentados a editar as falas, omitindo as hesitações e repetições dos falantes, bem como a corrigir inconsistências ou “erros” de linguagem. Todavia, como mencionado, esse não é o objetivo da transcrição: transcrevemos uma fala de forma escrita para possibilitar que ela seja analisada e estudada com todas as suas particularidades.

A seguir, listamos algumas perguntas e respostas que nos ajudam a esclarecer algumas características da transcrição.

3.1 Em que consiste a transcrição da linguagem falada?

A transcrição da linguagem falada consiste em ouvir um texto falado gravado e digitar palavra por palavra do que ouvimos, seguindo os critérios e objetivos estabelecidos previamente pelos pesquisadores envolvidos, como abordado no Capítulo 4.

3.2 Para que serve?

As transcrições servem para que os pesquisadores possam estudar algum aspecto da fala. No caso das transcrições realizadas nas pesquisas da área das Ciências da Saúde, um dos objetivos é observar, nos depoimentos coletados em entrevistas, como os entrevistados se sentem ou avaliam sua condição no escopo das práticas educativas das quais participam. Essas práticas, como vimos, geralmente envolvem instrumentos (questionários, roteiros, etc.) (cf. Capítulo 4) desenhados para coletar informações específicas de interesse do pesquisador.

Através das transcrições, podemos também verificar o que precisa ser aprimorado no instrumento para que o respondente consiga

entender o que está sendo perguntado. Os elaboradores de questionários geralmente não estão presentes durante o processo de aplicação do instrumento; o registro gravado ou filmado possibilita analisar o que está dificultando a compreensão tanto do respondente quanto do aplicador. Os problemas de compreensão podem ser detectados e, depois, solucionados; por isso a importância das transcrições para o empoderamento dos pesquisadores, profissionais da área da Saúde e usuários do serviço de Saúde, como mencionado no Capítulo 1.

3.3 Quando fazer?

O processo de transcrição de linguagem falada pode ser demorado e minucioso, dependendo do tamanho do instrumento do estudo, de quantas aplicações foram feitas e da quantidade de tempo despendida durante a conversa. Contudo, não é em toda pesquisa que envolve a aplicação de instrumentos que transcrições são necessárias. Dessa forma, é importante que os pesquisadores tenham em mente, de maneira clara, quais são os objetivos da pesquisa que estão fazendo. Assim, depois da análise dos objetivos, eles podem decidir se a transcrição é fundamental como dado a ser analisado.

3.4 Como fazer?

Uma vez definido que a transcrição será um processo necessário na pesquisa, o próximo passo é partir para a execução da transcrição. O objetivo do trabalho é um dos principais fatores que determinam como as transcrições serão feitas, porque dependendo de quais são os objetivos, algumas partes das conversas não precisam ser transcritas por não serem relevantes para a pesquisa. Além disso, como exposto no Capítulo 2, o pesquisador também deve observar qual o tipo de texto que está sendo transcrito (entrevista, depoimento),

e se é um monólogo ou um diálogo, pois cada tipo de texto requer a tomada de decisão sobre características específicas da transcrição.

Outro aspecto importante que deve ser considerado é o planejamento dessas transcrições, principalmente se forem feitas por mais de uma pessoa. Assim, é necessário que uma metodologia de transcrição seja estabelecida pelos pesquisadores para que as pessoas responsáveis adotem critérios comuns e as transcrições tenham um padrão comum. Isso favorecerá a consistência dos dados coletados, tanto estruturados como não estruturados, abordada no Capítulo 3.

Além disso, um fator relevante e que deve ser levado em conta é a relação tempo versus produtividade, pois, após horas e horas ouvindo e digitando falas, a atenção tende a diminuir e estamos mais propensos a cometer erros de digitação. Então, é muito importante que os responsáveis pelas transcrições parem por alguns minutos de tempos e tempos, a fim de evitar equívocos, que podem prejudicar a efetividade da pesquisa e do processo de transcrição.

Finalmente, é relevante que os pesquisadores tenham em mente que possíveis problemas podem acontecer, já que estão lidando com a linguagem falada, a qual é mais imprevisível que a escrita. Por isso, é necessário saber como solucionar problemas que não estavam presentes no planejamento inicial, como a grafia de estrangeirismos, por exemplo.

3.5 Resumindo...

Depois que as principais perguntas foram abordadas, podemos concluir que antes de se começar o processo de transcrição, é necessário avaliar, cautelosamente, os objetivos da pesquisa e a real necessidade das transcrições. Além disso, é preciso estabelecer critérios para as transcrições. No caso das transcrições dos questionários da área das Ciências da

Saúde, por exemplo, todas devem apresentar o mesmo conteúdo, ou seja, a mesma seleção de perguntas e respostas, bem como a mesma padronização de tamanho de fonte, tipo de letra e abreviações. Por fim, tenha em mente que o processo de transcrições pode trazer grandes descobertas para a pesquisa, as quais somente serão possíveis de serem apontadas através da transcrição das falas. Para dar certo, basta somente que os pesquisadores e transcritores sejam precavidos e que estejam sempre atentos aos pequenos detalhes.

4 Preparação de textos transcritos para mineração de textos – uso de *softwares* – Qual é a participação do homem e da máquina? Como se complementam?

Os inúmeros avanços tecnológicos registrados nos últimos anos vêm contribuindo cada vez mais para uma maior automatização de diversas etapas de uma pesquisa, desde o processo de coleta e armazenamento de dados até o processamento e análise desses dados. No que tange à transcrição, inicialmente, ela só se tornou possível com o advento do gravador de áudio, que possibilitou que qualquer tipo de texto falado pudesse ser registrado e, posteriormente, transcrito e analisado.

O gravador de áudio ainda continua sendo utilizado para registrar falas, embora a utilização de câmeras para esse tipo de registro tenha crescido nos últimos anos, em função da invenção de modelos portáteis e pelo fato de a maioria dos computadores hoje em dia possuir uma câmera embutida. Como veremos no Capítulo 7, o registro de falas via vídeo permite que o pesquisador tenha acesso a uma gama ainda maior de informações, como os gestos e expressões faciais que geralmente acompanham a fala de uma pessoa.

Se bem existem *softwares* de reconhecimento de fala que registram por escrito o que nós falamos ou ditamos, a tarefa de transcrição de uma fala previamente armazenada num arquivo de áudio para um arquivo texto ainda é manual, ou seja, precisa ser feita por pessoas. Isso porque os *softwares* disponíveis atualmente que realizam essa tarefa ainda estão desenvolvendo sistemas de reconhecimento de voz capazes de garantir um percentual de acerto elevado durante a transcrição, ou de produzir um texto transcrito que não demande muita edição ou revisão por parte do pesquisador.

Apesar disso, o surgimento do computador contribuiu para facilitar esse processo, uma vez que a transcrição antes feita à mão ou datilografada numa folha de papel, hoje em dia, pode ser realizada nos diversos editores de texto, presentes em qualquer computador, o que permite não só o armazenamento seguro do conteúdo, mas também que esse mesmo conteúdo seja processado pelos diversos *softwares* utilizados em pesquisas linguísticas e/ou estatísticas.

Ao longo deste *e-book* apresentaremos e demonstraremos como utilizar diversos *softwares* gratuitos que podem auxiliá-lo durante a sua pesquisa. Neste capítulo introduziremos o AntConc (ANTHONY, 2014), um *software* concordanciador, que possui diversas ferramentas para análise textual. Nos próximos capítulos, 6 e 7, você aprenderá mais sobre o R (R CORE TEAM, 2015), um *software* e ambiente estatístico usado para mineração de textos e sobre o ELAN (MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS, 2015), uma plataforma que permite efetuar a transcrição da fala e armazená-la junto com o arquivo de áudio.

Entretanto, é importante esclarecer que, antes de serem processados e analisados com o auxílio de qualquer tipo de *software*, os textos transcritos precisam ser preparados. O

primeiro passo para a preparação dos textos transcritos começa com a escolha do editor de texto e do formato em que o texto será armazenado. Sugerimos que essa escolha, bem como a do editor de texto que será usado durante a transcrição, sejam feitas antes do início do processo de transcrição. Recomendamos a utilização de editores de texto como o Bloco de Notas, nos quais os textos já são armazenados diretamente no formato *.txt*, que é o formato exigido pela maioria dos *softwares*, sobretudo os linguísticos. Dessa forma, o pesquisador 1) evita o retrabalho de ter que salvar todos os arquivos no formato aceito pelos *softwares* que utilizará e 2) não corre o risco de ter que lidar com problemas de codificação resultantes do processo de conversão de arquivos, que não só podem demorar para serem resolvidos como podem prejudicar o andamento da análise.

O segundo passo diz respeito a quais excertos do texto transcrito são dispensáveis para facilitar tanto o trabalho dos *softwares* durante o processamento do texto, quanto a análise do pesquisador. Esse procedimento é necessário porque, ainda que as falas, na maioria das vezes, sejam transcritas em sua totalidade, nem sempre a pesquisa fará uso de todo o material transcrito.

No caso das pesquisas realizadas na área das Ciências da Saúde, como mencionado, são feitas transcrições das entrevistas realizadas com os usuários do serviço de Saúde que participam da pesquisa. Por ser uma entrevista, a transcrição contém as perguntas feitas pelos entrevistadores e as respostas dos entrevistados. Além disso, algumas delas apresentam, ainda, descrições de gestos realizados pela pessoa entrevistada.

No entanto, o pesquisador pode estar interessado em analisar apenas as respostas dos entrevistados. Nesse caso, as perguntas podem ser eliminadas do arquivo *.txt* que

será processado pelo *software*, assim como as descrições dos gestos. Essa tarefa poderia ser feita manualmente, com o pesquisador deletando pergunta por pergunta de seus arquivos; porém, ela se torna muito mais rápida com o auxílio do *software* e, na maioria das vezes, não envolve uma mudança nos arquivos mestres das transcrições, principalmente se algumas das sugestões que faremos forem seguidas durante o processo de transcrição.

Existem várias formas de se remover partes de um texto que não são consideradas relevantes para a análise com o auxílio de um *software*, mas o modo mais simples é colocar essas partes entre <parênteses angulares>. Isso porque os *softwares* utilizados para a análise de textos possuem recursos que permitem não só localizar esses parênteses como também desconsiderar tudo o que estiver entre eles. Em outras palavras, esse procedimento é considerado mais simples porque o *software* não faz nenhuma alteração no arquivo que está sendo processado, ele apenas não considera os itens que estiverem entre parênteses angulares. A Figura 1, a seguir ilustra como isso funciona no AntConc.

A Figura 1 apresenta duas telas extraídas do AntConc. A ferramenta em destaque em ambas as telas é denominada *Concordance* e possibilita que o pesquisador extraia linhas de concordância contendo as ocorrências de uma palavra buscada por ele.

Na Figura 1, a palavra que buscamos foi *diabetes* (que aparece no centro das linhas de concordância) e se você observar a tela da direita, identificará diversas ocorrências de <U:> próximas a ela. Já se observar a tela da esquerda, verificará que se trata da mesma palavra de busca e das mesmas linhas de concordância, mas as ocorrências de <U:> não estão mais visíveis.

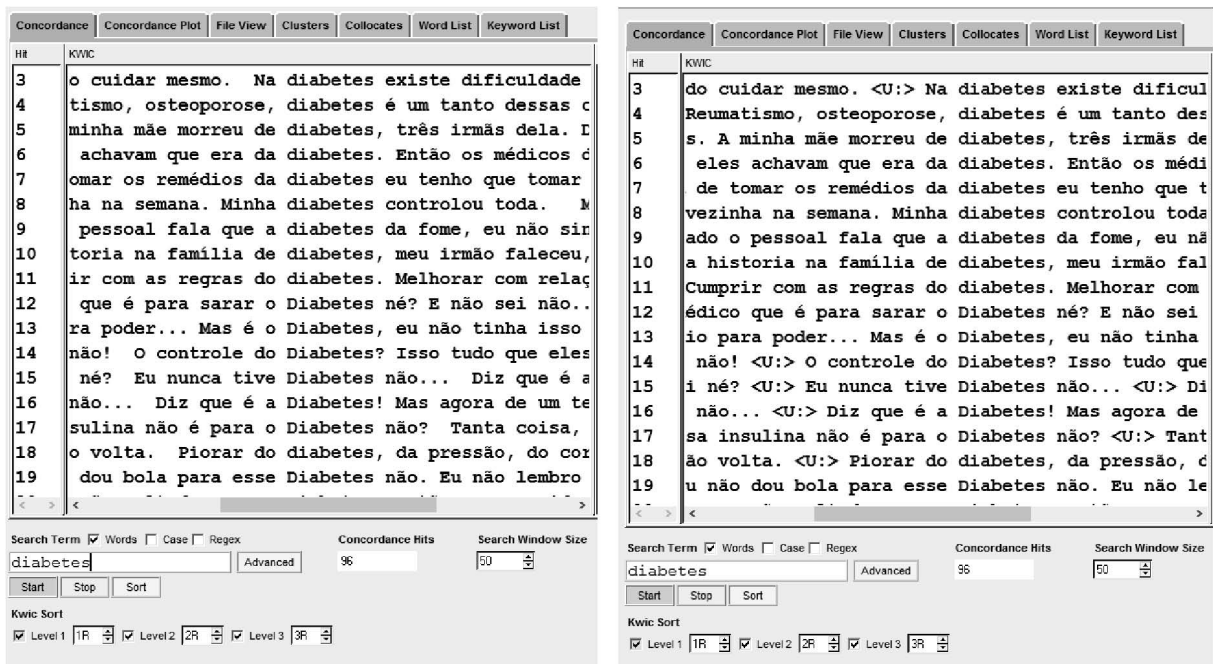


Figura 1 – O uso de parênteses angulares no AntConc

No caso das transcrições utilizadas nesse exemplo, as ocorrências de <U:> correspondem à palavra *Usuário* e foi empregada nas transcrições para indicar as falas dos entrevistados, porém como elas estão entre parênteses

angulares foi possível pedir ao programa que não as mostrasse por meio de uma configuração de preferências prévia que o pesquisador pode fazer antes de iniciar a análise. A Figura 2 demonstra como isso pode ser feito no AntConc.

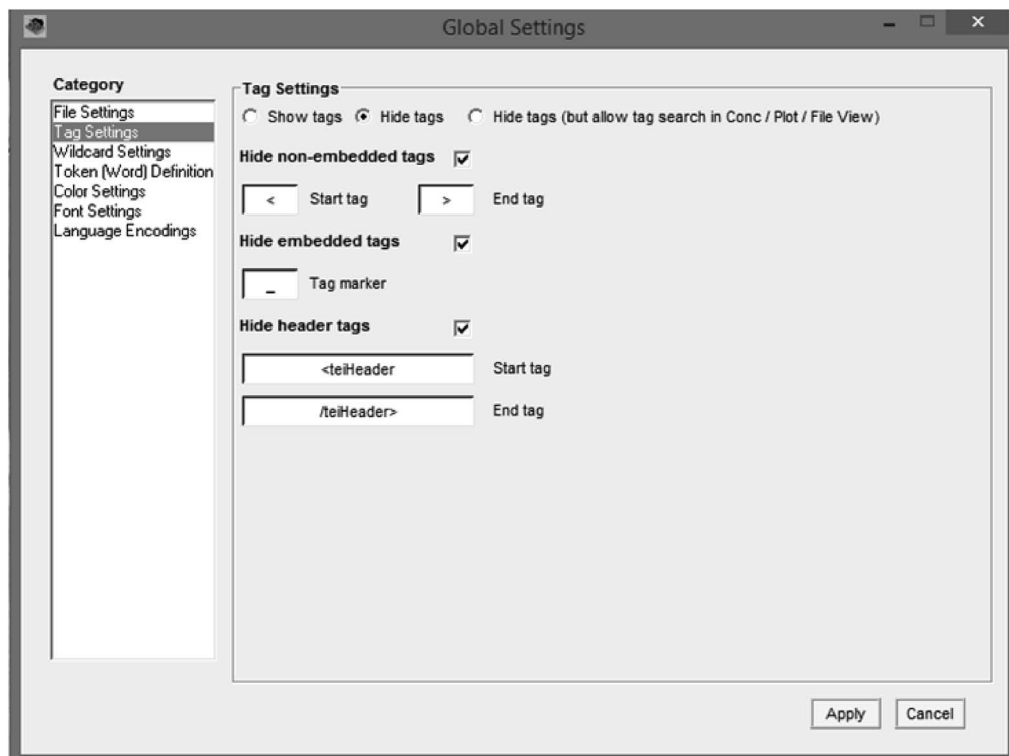


Figura 2 – Como configurar o AntConc para mostrar ou ocultar o que está dentro de parênteses angulares

Além disso, é importante esclarecer que esse recurso pode ser utilizado visando outros objetivos que não apenas eliminar partes consideradas irrelevantes para a análise. Como é possível escolher entre exibir ou ocultar algo entre parênteses angulares, o pesquisador pode utilizar esse recurso para anotar os textos que está analisando.

Vamos supor que um pesquisador resolva analisar as transcrições das entrevistas feitas com usuários do serviço de Saúde para verificar como eles se referem a sua condição, ou seja, se eles se referem a ela de forma positiva ou negativa, por exemplo. Uma forma de constatar isso, seria o pesquisador criar o que chamamos de etiquetas ou *tags* com as palavras <positiva> e <negativa> e, em seguida, analisar as transcrições, colocando essas etiquetas na frente de todas as falas dos entrevistados consideradas positivas ou negativas. Ao final dessa análise, é possível realizar uma busca no *software* pelas ocorrências das etiquetas <positiva> e <negativa>, o que possibilita estabelecer, entre outras coisas, se um usuário do serviço de Saúde se refere a sua condição de forma mais positiva ou mais negativa.

As etiquetas podem ser utilizadas, ainda, para indicar as repetições na fala de uma pessoa. A repetição é uma das características do texto falado que geralmente causa diversas dúvidas durante o processo de transcrição. Dúvidas que vão desde se ela deve ser transcrita ou não, até sobre a melhor forma de transcrevê-la. Novamente, a decisão, nesse caso, deve ser tomada com base no objetivo da pesquisa. Se a repetição não for considerada um dado relevante para a pesquisa, ela pode ser omitida durante a fase de transcrição.

Por outro lado, se o pesquisador considerar que o número de repetições, sobretudo de uma determinada palavra, pode oferecer uma informação relevante sobre o seu objeto de estudo, a transcrição desse tipo de ocorrência pode ser feita de duas formas. A primeira

delas é simplesmente transcrever a palavra todas as vezes em que ela for dita e repetida. Dessa forma, a repetição poderá ser verificada no número de vezes que essa determinada palavra foi registrada no texto. Já a segunda forma de lidar com a repetição é criar uma etiqueta denominada <repetição> e colocá-la na frente da palavra que está sendo repetida. A vantagem dessa segunda opção é que ela não só permite contar o número de repetições que ocorreram ao longo do texto falado, através da busca no *software* pelas ocorrências da etiqueta, como também possibilita identificar quais palavras foram mais repetidas em uma determinada fala. O Exemplo 1 ilustra essas duas formas de transcrever a repetição.

Exemplo 1

Transcrição de todas as palavras repetidas:
<U:> Será que vou cumprir meu Deus? Vamos! Vamos!

Uso da etiqueta <repetição>: <U:> Será que vou cumprir meu Deus? Vamos! <repetição>

O terceiro passo para a preparação dos textos transcritos está relacionado aos sinais de pontuação. Apesar de a pontuação desempenhar um papel importante na produção e leitura de um texto, muitas vezes, ela pode gerar problemas quando esse mesmo texto é lido por qualquer tipo de *software*, seja um utilizado para análise textual, como o Antconc, seja um usado para a mineração de textos, como o R. Problemas esses, em sua maioria ligados a questões de codificação e gerados durante o processo de conversão de um formato de texto para outro. Sendo assim, caso os sinais de pontuação não sejam essenciais para a análise, é sempre recomendável removê-los do texto antes que ele seja lido pelo *software*.

Utilizando novamente como exemplo as transcrições feitas nas pesquisas da área da Saúde, uma situação em que os sinais de

pontuação poderiam ser removidos do texto antes do seu processamento pelo *software*, seria se o pesquisador tivesse interesse em saber quais são os sentimentos dos usuários do serviço de Saúde. Para isso, ele precisaria apenas das falas dos entrevistados referentes aos seus sentimentos, assim a pontuação em si não desempenharia uma função relevante, podendo ser eliminada. O Exemplo 2 apresenta um mesmo trecho transcrito com e sem pontuação.

Exemplo 2

Com pontuação:

<U:> Não. Só me deixa incomodado. Não gosto de ser rotulado.

<U:> Só mesmo ter uma atividade física porque sei que isso me ajudaria, e é o que estou tentando fazer. Porque os remédios e alimentação está direito.

<U:> Mas eu já me sinto bem. A diabetes não me dá nenhum sinal, se eu não soubesse eu viveria bem, pois não sinto nada, ela não me causa nada.

Sem pontuação:

<U:> não só me deixa incomodado não gosto de ser rotulado

<U:> só mesmo ter uma atividade física porque sei que isso me ajudaria e é o que estou tentando fazer porque os remédios e alimentação está direito

<U:> mas eu já me sinto bem a diabetes não me dá nenhum sinal se eu não soubesse eu viveria bem pois não sinto nada ela não me causa nada

Com relação ao Exemplo 2, é interessante observar que, ao eliminarmos a pontuação, eliminamos também a necessidade de utilizarmos letras maiúsculas no início de uma oração ou período, o que constitui um outro aspecto a ser considerado à luz do objetivo da pesquisa e decidido antes do início do processo de transcrição.

Um outro problema que pode resultar do uso da pontuação em uma transcrição ocorre quando ela é utilizada para marcar características próprias da fala, como o uso de três pontos (...) para indicar uma hesitação, de pontos de exclamação para marcar as interjeições (ah!, ih!), ou pontos de interrogação para apontar o uso de determinadas partículas modais (né?). O uso da pontuação nesses casos pode ser problemático tanto por causa dos já mencionados problemas de codificação que dificultam o processamento do texto transcrito pelos *softwares*, quanto por exigir um maior planejamento prévio e um alto nível de atenção por parte do transcritor. Isso porque qualquer descuido que ocorra durante a transcrição pode prejudicar a análise posteriormente.

Imagine, por exemplo, que uma pesquisa resolva adotar o uso dos três pontos para marcar a hesitação do falante durante a transcrição. O transcritor, nesse caso, teria que redobrar a sua atenção ao transcrever para se assegurar de sempre digitar os três pontos cada vez que uma pessoa hesitar em sua fala, pois caso ele digite um ponto a mais (...) ou a menos (..), isso poderia dificultar a busca pelas ocorrências de hesitação nos textos transcritos, comprometendo, assim, os resultados que poderiam ser extraídos dos *softwares*.

Tendo em vista esses possíveis problemas, sugerimos que o uso da pontuação para marcar características da fala seja evitado. Em seu lugar, recomendamos, novamente, o uso das etiquetas. O Exemplo 3, demonstra como elas poderiam ser utilizadas nesse caso.

Exemplo 3:

<U:> É. É só isso aí. Doce não porque aqui em casa é a casa do doce e do salgado.

<U:> É. Eu não como. A filha faz torta doce <hesitação> Sabe como é que é <hesitação> E eu não como.

<U:> É eu não tenho vontade. Mesmo se tem vontade também <hesitação> Tem dia que dá uma vontade de comer doce, mas o que faz aqui eu não como não.

Caso o pesquisador opte por eliminar os sinais de pontuação de um texto já transcrito, isso poderia ser feito de três formas: 1) manualmente, eliminando os sinais de pontuação um por um; 2) semiautomaticamente, fazendo o uso das ferramentas de localizar e substituir presentes nos editores de texto e 3) automaticamente, utilizando, como veremos no Capítulo 6, *softwares* como o R, por exemplo, nos quais os sinais de pontuação podem ser eliminados por completo através de uma linha de comando que o *software* executa. Dentre as três opções, a terceira é a mais recomendável por ser tanto a forma mais rápida para realizar essa tarefa como por ser aquela que oferece mais garantia de que os sinais de pontuação serão eliminados por completo.

No entanto, caso o pesquisador decida que a pontuação é um elemento importante e deve estar presente no texto transcrito, ele precisa ter em mente os problemas que apon-tamos e desenvolver uma metodologia clara juntamente com um protocolo de transcrição que especifique não só quais os sinais de pontuação serão utilizados, mas também quando e como eles deverão ser utilizados. Além disso, ele precisaria alertar os transcritores sobre a necessidade de maior atenção durante a transcrição dos sinais de pontuação.

Por fim, o quarto passo na preparação diz respeito à padronização. A padronização é uma palavra chave quando falamos de transcrição, pois, como já mencionado, os textos transcritos precisam ter uniformidade, visando garantir que os dados extraídos durante a pesquisa sejam consistentes. Para além da padronização no uso de etiquetas ou das decisões tomadas acerca dos sinais de pontuação, daremos, ainda, exemplos de mais três casos em

que ela é necessária, sobretudo em pesquisas da área da Saúde.

O primeiro caso refere-se à padronização dos nomes dos usuários do serviço de Saúde que responderam às perguntas durante as entrevistas. Como nem os nomes dos usuários e nem os dos profissionais da área da Saúde que aplicam os instrumentos podem ser divulgados, por uma questão ética que pauta as pesquisas que têm a participação de seres humanos como sujeitos, normalmente, os pesquisadores e/ou transcritores optam por identificá-los por meio de siglas ou códigos que garantam o seu anonimato. Essas siglas ou códigos, por sua vez, são sempre colocadas entre parênteses angulares, o que permite, como mencionado anteriormente, a omissão deles quando as transcrições forem processadas pelos *softwares* de análise textual. Um exemplo de como isso pode ser feito já foi apresentado acima e corresponde ao uso da sigla <U:> para identificar os usuários do serviço de Saúde. Esse exemplo será retomado novamente na próxima seção.

Já o segundo caso diz respeito à padronização de como os números serão transcritos. Nesse caso existem duas possibilidades: 1) transcrever por extenso ou 2) transcrição numérica. A decisão fica a cargo do pesquisador e deve ser tomada com base nas necessidades da pesquisa. No Exemplo 4, a seguir, apresentamos como um mesmo trecho pode ser transcrito das duas formas sugeridas.

Exemplo 4

Por extenso:

<U:> Como eu tenho setenta e um completos no dia vinte e um de março.

<U:> setenta e cinco da última vez.

<U:> Eu faço quarenta minutos por dia.

Numérica:

<U:> Como eu tenho 71 completos no dia 21 de março.

<U:> 75 da última vez.

<U:> Eu faço 40 minutos por dia.

Por fim, o terceiro caso está relacionado aos gestos que os falantes fazem durante a sua fala. Esses gestos também podem ser transcritos dependendo do objetivo que o pesquisador tem na pesquisa, uma vez que podem, por exemplo, revelar alguma sensação que o respondente não consegue expressar verbalmente ou mesmo contribuir para reforçar um sentimento que o usuário do serviço de Saúde já conseguiu expor. Caso o pesquisador opte por transcrever o gesto, é importante não só indicar quando os gestos foram transcritos, o que pode ser feito através do uso de uma etiqueta como <gesto>, para facilitar a busca dessas ocorrências pelo *software*, como também realizar a descrição desses gestos entre parênteses angulares, visando evitar que essa descrição seja confundida como parte da transcrição de uma fala. Um exemplo de como isso pode ser feito será dado na próxima seção.

Como foi possível observar ao longo desta seção, muito do sucesso de uma pesquisa depende do seu planejamento, e o melhor jeito de extrair o máximo possível de um *software* em termos de resultados para serem analisados é preparando bem o texto antes que ele seja processado pelo *software*. Muitos desses passos que mencionamos aqui podem e devem ser seguidos durante o processo de compilação e transcrição dos textos, pois não só poupa tempo em etapas futuras da pesquisa como contribui para que ela possa ser conduzida sem problemas.

5 Estudo de caso / exemplo – transcrições das entrevistas

Para ilustrar os aspectos apresentados sobre a transcrição, vamos analisar as transcrições realizadas a partir das aplicações feitas por enfermeiras e nutricionistas, mestrandas ou doutorandas, pelo programa de pós-graduação da Escola de Enfermagem

da Universidade Federal de Minas Gerais em conjunto com alunos de iniciação científica e assistência à pesquisa da mesma instituição. O instrumento usado nas entrevistas foi o *Behavior Change Protocol* (FUNNELL; ANDERSON, 2004), o qual passou pelos processos de tradução para a língua portuguesa brasileira e pela fase de adaptação cultural para o contexto da população alvo, sendo, portanto, intitulado *Protocolo Mudança de Comportamento em Diabetes Mellitus tipo 2*. Esse instrumento é composto de 5 passos – definição do problema, identificação e abordagem dos sentimentos, definição de metas, elaboração do plano de cuidados para conquista de metas e avaliação e experiência do usuário do serviço de Saúde sobre o plano de cuidados. As questões que o compõem são abertas e servem para guiar o profissional da área da Saúde no manejo de práticas educativas em Diabetes Mellitus tipo 2.

Como o principal objetivo da pesquisa era fazer a adaptação do instrumento, para que ele fosse passível de ser aplicado aos usuários do Sistema Único de Saúde (SUS) da cidade de Belo Horizonte – Minas Gerais, o processo de transcrição das entrevistas foi considerado relevante pelos pesquisadores e, consequentemente, realizado. Além disso, o instrumento era composto por questões abertas, ou seja, cada respondente poderia ter uma fala e/ou entendimento diferente conforme as perguntas. E, por fim, o foco do protocolo era as atitudes e sentimentos do respondente, portanto, tratava-se de algo imprevisível e subjetivo, por isso, somente através da digitação das entrevistas era possível observar determinadas repetições de sentimentos e atitudes, contribuindo para a avaliação da situação psicológica e social do entrevistado em relação à sua condição. Dessa forma, tendo em mente, de maneira bem clara, os objetivos e as necessidades da pesquisa constatou-se que os benefícios da transcrição se sobrepunham às dificuldades

do processo, optando-se, assim, por realizar as transcrições.

A partir desse momento, um grupo de alunos de iniciação científica e apoio externo à pesquisa da Escola de Enfermagem da Universidade Federal de Minas Gerais que estavam envolvidos com a pesquisa referente ao *Protocolo Mudança de Comportamento em Diabetes Mellitus tipo 2* dentro do projeto Empoder@ foram selecionados e orientados para fazer as transcrições, a fim de que tudo saísse padronizado e uniformizado. Logo, a metodologia das transcrições foi estabelecida, a saber:

1. Escolha do *software* para fazer as transcrições e do tipo de arquivo para seu armazenamento

O Bloco de Notas foi o editor de textos escolhido, pois os arquivos eram gerados na

extensão *.txt*, a qual é aceita em grande parte dos *softwares* concordanciadores e ambientes estatísticos como o R. Além disso, trata-se de um tipo de extensão em que as alterações de formatação não são feitas. Isso permitiu que as transcrições tivessem um padrão estrutural. No entanto, foi necessário atentar-se para o tipo de arquivamento prototípico dessa extensão; assim foi estipulado que todos os arquivos deveriam ser salvos em ANSI ou UTF-8.

2. Padronização de nomes

Os nomes dos respondentes não foram revelados, por questões éticas, ficando, portanto, somente as iniciais dos mesmos, entre parênteses angulares, como ilustrado pela Figura 3:

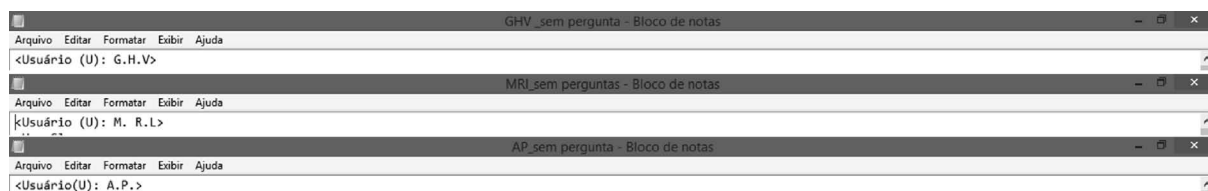


Figura 3 – Padronização de nomes por meio das iniciais dos usuários

3. Remoção das perguntas e uniformidade das respostas

As perguntas do protocolo foram omitidas para evitar que as palavras e/ou expressões presentes nas perguntas ficassem repetidas e evidenciassem uma falsa-repetição e/ou padronização. Portanto, o destaque foi dado ao que realmente era relevante para o objetivo da pesquisa, ou seja, as respostas, por exemplo:

Com pergunta:

<F:> <Qual é a sua maior dificuldade para controlar a diabetes?>

<U:> Eu acho que é comida.

<F:> < Comida?>

<U:> É. Porque eu quero misturar a comida. Porque às vezes eu cismo que quero comer batata com arroz... um pouquinho de cada uma.

Sem pergunta:

<U:> Eu acho que é comida.

<U:> É. Porque eu quero misturar a comida. Porque às vezes eu cismo que quero comer batata com arroz... um pouquinho de cada uma.

Além disso, cada resposta foi digitada em uma única linha, para que não houvesse conflito entre os excertos, sendo separadas pela inscrição <U:> que corresponde a *usuário*, como:

<U:> Como assim? Ah, ter assim um plano de cuidados?
<U:> Ter uma alimentação nas horas certas.
<U:> Tenho. Eu tenho aquela meta direitinho, eu quase não fujo.

4. Preservação de interjeições típicas da linguagem falada e de ações dos respondentes durante a aplicação do instrumento

As interjeições e algumas atitudes dos respondentes durante as entrevistas foram preservadas, já que elas poderiam revelar descobertas importantes referentes aos sentimentos e/ou atitudes, a saber:

<U:> Ih! Eu quero tudo, mas não dou conta de fazer.
<U:> Ah, não sei. Isso eu não tenho maior ideia.
<U:> Uai, está sempre lembrando da responsabilidade que tenho.
<U:> Tem. Ela. <gesto> <Ele aponta para sua filha>

5. Números escritos em forma numérica

Os números não foram escritos por extenso, mas na própria forma numérica, como:

<U:> A minha menina olhou a minha glicose há poucos dias, porque ela é enfermeira, e deu 98, né?
<U:> A caminhada, ginástica, alimentação, dormir bem. Eu durmo é 9:00 horas, no máximo 10:00 horas.
<U:> Como eu tenho 71 completos no dia 21 de março.

6. Pontuação

As falas, durante o processo de transcrição, foram pontuadas com sinais de exclamação, interrogação, reticências, ponto final e vírgula, seguindo os critérios metodológicos pré-estabelecidos antes do processo de transcrição, por exemplo:

<U:> Como é que eu faço?

<U:> Dá aquela...

<U:> Já tem três anos.

<U:> Ah, não! Foi antes da trombose.

<U:> Igual continuar com a caminhada, dosar na alimentação...

Além disso, é importante ter em mente que problemas inesperados podem ocorrer, por exemplo, o transcritor pode se deparar com palavras que foram ditas no diminutivo ou expressões próprias da fala, como evidenciado abaixo:

<U:> Tenho. Eu tenho aquela meta direitinho, eu quase não fujo.

<U:> Olha pra dizer a verdade...

E querer retirar essas características que marcam o texto falado. Contudo, como um dos principais objetivos da pesquisa era verificar a compreensão do respondente com relação a pergunta, essas evidências eram imprescindíveis para a análise e jamais poderiam ser excluídas.

Portanto, problemas inesperados podem ocorrer tanto no que diz respeito à linguagem falada, quanto com relação ao tempo despendido para as transcrições e a produtividade que também pode ser prejudicada devido à fadiga e ao cansaço dos transcritores. Por isso, para o processo de transcrição das respostas do *Protocolo Mudança de Comportamento em Diabetes Mellitus tipo 2*, os pesquisadores contaram com um grupo de responsáveis que trabalharam somente com a transcrição dos textos e que faziam pausas de tempos em tempos para evitar possíveis erros por falta de atenção e/ou cansaço.

Por fim, para ilustrar todos os passos mencionados acima, segue abaixo um trecho das transcrições em que todos os passos foram seguidos:

<U:> Ah, assim não! A gente sempre imagina estar bem... Sem não ter complicação.

<U:> Fica. Igual a minha meta é manter a minha glicose no padrão.
<U:> Igual continuar com a caminhada, dosar na alimentação...

<U:> Tem a minha esposa e a família toda.
<U:> Vantagem é que a caminhada melhora a circulação e eu achamos que a gente se sente bem melhor.

Síntese do capítulo

- Abordamos sobre o tema transcrições;
- Ressaltamos que as transcrições são fundamentais quando os pesquisadores pretendem analisar dados que podem ser revelados por meio da fala;
- Argumentamos que, para fazer as transcrições, os pesquisadores precisam ter em mente, de maneira bem clara, os objetivos da pesquisa;
- É imprescindível desenhar uma metodologia para que as transcrições fiquem padronizadas e uniformes;
- A relação máquina versus homem é amigável, já que um complementa o trabalho do outro;
- Indicamos o AntConc, um *software* livre, responsável pela automatização da análise;
- Aprendemos algumas dicas para usar os *softwares* concordanciadores;
- Verificamos um estudo de caso real em que as transcrições foram relevantes;
- Aprendemos que problemas relacionados às transcrições são inevitáveis e que temos que saber como solucioná-los sem atrapalhar o desenvolvimento da pesquisa.

Referências

ANTHONY, L. **AntConc**. Tokyo: Waseda University, 2014.

FUNNELL, M. M.; ANDERSON, R. M. Empowerment and Self-Management of Diabetes. **Clinical Diabetes**, v. 22, n. 3, p. 123–127, 2004.

HALLIDAY, M. A. K. **Spoken and written language**. Oxford: Oxford University Press, 1989.

HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. **Halliday's introduction to functional grammar**. 4. ed. Milton Park, Abingdon, Oxon: Routledge, 2014.

MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS. **ELAN Linguistic Annotator**. Nijmegen: The Language Archive, 2015.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation for Statistical Computing, 2015.

Leituras suplementares

SARDINHA, T. B. Lingüística de Corpus: histórico e problemática. **DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada**, v. 16, n. 2, p. 323–367, 2000.

VIANA, V. Linguística de Corpus: conceitos, técnicas e análises. In: VIANA, V.; TAGNIN, S. E. (Eds.). **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2011. p. 25–95.

ANÁLISE DE DADOS NÃO ESTRUTURADOS – MINERAÇÃO DE TEXTOS

Rodrigo Araújo e Castro
Sumaya Giarola Cecílio

Objetivos do capítulo

- Apresentar a Mineração de Textos como metodologia semiautomática complementar ou em substituição à abordagem tradicional da Análise de Conteúdo;
- Discutir como a Mineração de Textos pode ser utilizada em análises de dados provenientes de pesquisas na área das Ciências da Saúde.

1 Introdução

A Inteligência Artificial é um campo da Ciência da Computação que tem como objetivo fazer com que computadores (máquinas) se comportem de maneira inteligente, ou seja, simulando o raciocínio humano. A Inteligência Artificial se relaciona com diversas áreas, como Psicologia, Biologia, Lógica, Matemática, Linguística, Filosofia, entre outras, o que a torna interessante para qualquer esfera da atividade intelectual humana e científica (GOMES,

2010). A Mineração de Dados, que incorpora técnicas da Inteligência Artificial e da Estatística (REZENDE, 2003, p. 309), se apresenta como um caminho promissor para a extração de conhecimentos independente da origem e quantidade de dados. Como exposto no Capítulo 3, o conhecimento é o produto final do tratamento e interpretação de dados obtidos por meio de pesquisas científicas. O caminho para o conhecimento contempla a coleta dos dados e seu tratamento (quer seja por meio de testes estatísticos ou análises qualitativas),

originando informação. A interpretação e discussão da informação dentro do contexto que se estuda gera o que consideramos conhecimento (GALVÃO; MARIN, 2009).

A fase de tratamento de dados, constituinte de uma das etapas para a obtenção de conhecimento, pode ser realizada por meio da Mineração de Dados e da Mineração de Textos, que possibilitam que pesquisadores de diferentes áreas, como Linguística, Estatística, Ciência Cognitiva e Ciência da Computação, atuem de maneira interdisciplinar, relacionando habilidades e construindo conhecimento de maneira conjunta (ARANHA; PASSOS, 2006). Ainda, as Ciências da Saúde podem complementar este processo de busca pelo saber na medida em que disponibilizam para análise volumes de dados com os quais lidam diariamente, quer sejam estes dados estruturados (mensuração de indicadores em saúde, tais como número de internações) ou não estruturados (dados provenientes de entrevistas em práticas educativas com usuários do serviço de Saúde, as quais geram gravações de áudio ou vídeo e suas transcrições, conforme mencionado no Capítulo 5).

Neste Capítulo, buscamos integrar teorias e métodos da Linguística Aplicada, da Estatística e da Enfermagem para apresentar a Mineração de Textos, utilizando dados não estruturados obtidos em pesquisas que tinham como objetivo explorar o processo de responsabilização entre o profissional da área da Saúde e o usuário do serviço de Saúde com a condição crônica do diabetes por meio de diferentes práticas educativas, dentre elas a interlocução coletiva e a visita domiciliar, nas quais se busca estimular que os usuários adiram ao autocuidado, mediante a educação recebida, que está relacionada aos conhecimentos e atitudes psicológicas e à mudança de comportamento para o cumprimento de metas, conforme mencionado no Capítulo 1.

Considerando que as Ciências da Saúde vêm estudando ao longo dos últimos anos

metodologias participativas e inovadoras para subsidiar o arcabouço teórico do empoderamento do usuário do serviço de Saúde, pautado pela educação e o cuidado centrados nele próprio, em um processo resultante de responsabilização e construído por meio de diferentes práticas educativas, este Capítulo visa apresentar as vantagens de se trabalhar com a Mineração de Textos com subsídios de diferentes campos disciplinares e a incorporação de novas metodologias para as pesquisas em saúde. Para facilitar a leitura, um breve glossário de alguns conceitos-chave ao entendimento deste Capítulo é apresentado ao final do mesmo.

2 O que é Mineração de Textos ou de dados não estruturados do tipo texto?

Como já mencionado anteriormente, no Capítulo 3, o campo da Inteligência Artificial, na Ciência da Computação, que lida com a exploração de dados é a Mineração de Dados. De modo geral, esta consiste no uso de técnicas computacionais, simples ou complexas, para extrair e/ou gerenciar pequenos ou grandes conjuntos de dados. A Mineração de Dados é definida como:

o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados. (REZENDE, 2003, p. 310)

Além da Mineração de Dados, que enfoca dados estruturados (em tabelas numéricas organizadas em linhas e colunas), outra técnica utilizada pelos Sistemas Inteligentes¹

¹ Sistemas Inteligentes são sistemas que possuem a “habilidade de usar conhecimento para desempenhar tarefas ou resolver problemas” e a “capacidade para aproveitar situações e inferência para trabalhar com problemas complexos que assemelham-se [sic] a problemas reais” (REZENDE, 2003, p. 7).

é a Mineração de Textos, que se diferencia da primeira por envolver o processamento de dados não estruturados do tipo texto, em outras palavras, do que, neste caso, denominam-se *strings*, ou sequência de caracteres de um texto. A Mineração de Textos consiste em “um conjunto de técnicas e processos que descubrem conhecimento inovador nos textos” (REZENDE, 2003, p. 338).

No Capítulo 2, ao explorarmos o contexto de situação como configuração que nos permite estudar a variação linguística, vimos que a variável modo diz respeito tanto ao meio escrito como o falado e ao canal gráfico e fônico. Ainda para a subvariável meio há subespecificações. Assim, o que se entende por dado não estruturado ou textual nas Ciências da Computação contempla o dado extraído tanto de textos escritos como falados, estes últimos caracterizados pela subvariável canal fônico.

Como exposto no Capítulo 3, dados estruturados podem ser definidos como dados que estão organizados de uma forma específica (geralmente em um banco de dados com formato em linhas e colunas) para possibilitar o uso de determinado banco de dados por meio de filtros, consultas, agrupamentos e extração de informações relevantes para o interesse de determinado usuário (ALMEIDA, 2002). Por exemplo, no banco de dados do Núcleo de Pesquisa em Gestão, Educação e Avaliação em Saúde (NUGEAS), encontram-se disponíveis dados relativos aos Centros de Saúde colaboradores do projeto e de seus usuários com a condição crônica do Diabetes Mellitus tipo 2, incluindo número do prontuário disponível no Centro de Saúde, nome, idade, sexo, endereço, agente de Saúde referência, resultados dos exames clínicos mais recentes, entre outros. Em contraste, dados não estruturados podem ser definidos como dados sem uma estrutura como a de um banco de dados com formato em linhas e colunas. Por exemplo, prontuários e transcrições de entrevistas resultado de intervenções educativas realizadas com usuários

do sistema de Saúde em pesquisas das Ciências da Saúde.

Para que sejam objeto de análise computacional com tratamento estatístico, é necessário que haja a preparação dos dados não estruturados, mais especificamente aqueles textuais, para análise, processo que envolve a preparação prévia dos dados. Em outras palavras, é necessário transformar o texto (a sequência de caracteres) em objetos relacionados entre si (normalmente por meio da atribuição de categorias), de acordo com uma lógica derivada da forma como o próprio texto se organiza (ARANHA; PASSOS, 2006, p. 2), por exemplo, por meio de categorias linguísticas como classe de palavra (substantivos, adjetivos, verbos) ou outras categorias.

Dessa forma, para permitir a realização de uma análise dos dados textuais, é necessário, primeiramente, fazer a preparação dos dados. A seguir, pode-se i) realizar procedimentos da Mineração de Textos que organizam a informação textual de forma que se torne passível de leitura pela máquina e geram outra representação dos dados, como a geração de linhas de concordância e listas de colocados, que serão vistas com mais detalhes mais adiante; ii) gerar um novo conjunto de dados estruturados que serão analisados por meio de técnicas da Mineração de Dados – análise que provavelmente envolverá métodos relacionados a cálculos numéricos e estatísticos. E para a representação dos resultados gerados, independentemente do método escolhido, pode-se utilizar, por exemplo, tabelas e/ou gráficos.

No caso de uma análise envolvendo entrevistas, a partir do dado não estruturado gerado por meio de coleta de uma fala ou resposta verbal e sua transcrição, devidamente planejada, como apresentado no Capítulo 5, pode ser gerado um novo conjunto de dados estruturado para a obtenção de resultados quantitativos. Essa preparação prévia diz respeito à transformação de texto em categorias que possam ser contadas e sua frequência interpretada em relação ao objeto de interesse da

análise. No escopo de uma teoria abrangente da linguagem e do contexto, categorias passíveis de serem utilizadas para estruturar o texto podem apresentar os resultados dos depoimentos do usuário do sistema de Saúde quanto às suas histórias de vida, problemas, como é viver com a condição do diabetes, observando quais itens lexicais e gramaticais são mais recorrentes na sua fala. Também podemos utilizar categorias relativas a sentimentos que podem ser atribuídas às falas e observar quais sentimentos predominam. As categorias constituem um dado estruturado que pode se mostrar esclarecedor sobre o que leva ao sucesso ou insucesso da prática educativa em saúde.

3 O que pode revelar a Mineração de Textos? Qual é a participação do homem e da máquina nesse processo? Como se complementam?

O principal benefício da Mineração de Textos está na descoberta de regularidades e

tendências presentes nos dados, os quais provavelmente não poderiam ser descobertas por meio de abordagens tradicionais executadas manualmente, como a Análise de Conteúdo (BARDIN, 2011). Isso ocorre devido à complexidade das relações entre os itens envolvidos na análise a ser realizada (neste caso as palavras dos textos) ou devido a um excesso de informação que não seria facilmente processada pelo pesquisador sem o uso de ferramentas computacionais e estatísticas.

Segundo Han e Kamber (2006) e Tan, Steinbach e Kumar (2005), a Mineração de Dados é o passo intermediário para o que se denomina Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases*), processo de análise da Inteligência Artificial que se divide em três principais etapas: pré-processamento, Mineração de Dados e pós-processamento, como pode ser visto na Figura 1:

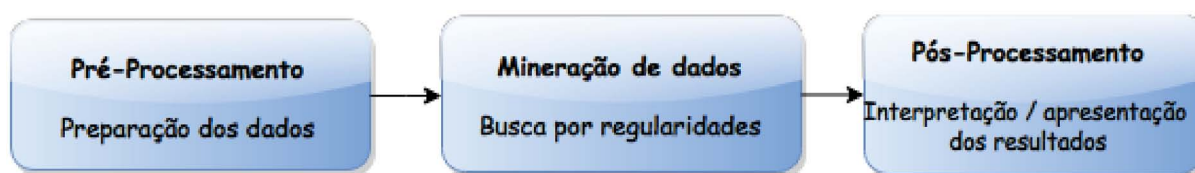


Figura 1 – Mineração de Dados no contexto do processo de descoberta de conhecimento

Fonte: Elaborado para fins deste estudo, baseado em Maciel et al. (2015).

No pré-processamento, os dados são preparados para as etapas seguintes, que consiste na conversão dos dados de uma forma bruta para um formato adequado para análise. Isso pode ser feito, por exemplo, por meio da limpeza (correção, adequação e/ou seleção de uma parte dos dados) ou da remoção de dados considerados ruidosos (os quais possuem valores que não condizem com o que está sendo

estudado). Subsequentemente, são aplicadas técnicas de Mineração de Dados para modelar os dados, visando-se uma descrição destes ou uma predição de tendências que futuros dados assumiriam. No pós-processamento, uma vez que já foi elaborado um modelo com base nos dados iniciais, este modelo é analisado, avaliado e apresentado de maneira compreensível aos usuários/pesquisadores (MACIEL et al., 2015).

Em relação à etapa de pré-processamento, embora a máquina seja mais rápida e eficiente que o humano, a máquina precisa ser preparada por um humano para tal tarefa. Acerca da análise de dados, esta pode ser realizada tanto por humanos – via métodos tradicionais (manuais) – ou pela máquina. No entanto, diversos estudos² já provaram que a máquina (por meio da Mineração de Dados/Textos) é mais eficiente que os humanos em tarefas repetitivas.

Quanto ao pós-processamento, não pode ser realizado exclusivamente pela máquina, pois ainda não está preparada para tal tipo de tarefa, visto que esta etapa envolve interpretação dos dados gerados, processo que não é facilmente programável. No entanto, a máquina pode auxiliar na elaboração de gráficos ou na filtragem dos dados.

No que diz respeito à Mineração de Textos, uma vertente da Mineração de Dados em que o objeto de estudo é o texto, uma das etapas necessárias é a tokenização do texto, isto é, a separação de um texto em *tokens*, para depois se fazer a análise da classe de palavra e discriminar dois tipos de palavras, de acordo com critérios utilizados em análises computacionais: palavras gramaticais (relativas a artigos, preposições, pronomes, conjunções) e palavras lexicais (relativas a verbos, adjetivos, substantivos e advérbios). Isso permitirá obter, por exemplo, uma lista de frequência ou uma nuvem de palavras, conforme será ilustrado mais adiante no estudo de caso apresentado neste Capítulo.

A análise automatizada se beneficia tanto da eficiência de algoritmos computacionais quanto da tomada de decisões do usuário de uma ferramenta ou de um pesquisador interessado na análise. Este tipo de análise – exe-

cutada com o auxílio do computador e não exclusivamente por ele – é feita por *softwares* denominados analisadores sintáticos (*parsers*), etiquetadores morfossintáticos (*POS taggers*) e concordanciadores (*concordancers*). As Figuras 2 e 3 a seguir ilustram analisadores sintáticos e etiquetadores morfossintáticos, com exemplos de uma das entrevistas pertencentes ao banco de dados do NUGEAS/EF/UFGM.

O analisador sintático (*parser*) é um *software* que divide uma *string* (um texto) em partes de acordo com categorias da gramática formal (sintagmas nominais, verbais, etc.), como pode ser visto na Figura 2.

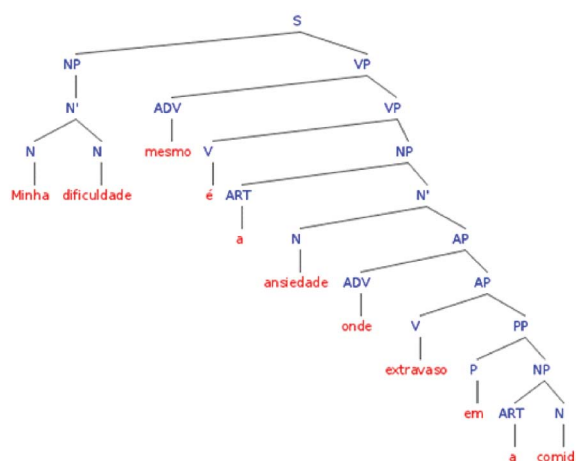


Figura 2 – Resultado gerado por um analisador sintático online de acesso livre
Legenda: Sentença (S), Sintagma nominal (NP), Sintagma Verbal (VP), Sintagma Adverbial (AP), Sintagma Preposicional (PP), Artigo (ART), Substantivo (N), Verbo (V), Advérbio (ADV), Preposição (P). N' apenas é utilizado quando o sintagma nominal possui mais de um elemento.

Fonte: Silva *et al.* (2010).

Na Figura 2, pode-se observar que as palavras da oração “Minha dificuldade mesmo é a ansiedade, onde extravaso na comida” foram separadas em classe de palavra por um critério morfossintático, utilizando categorias como artigo (ART), advérbio (ADV), substantivo (*noun* – N) e verbo (V). Elas foram agrupadas em sintagmas ou grupos por um critério sintático.

² Alguns estudos que podem ser citados são: Maciel *et al.* (2015), Han e Kamber (2006) e Tan, Steinbach e Kumar (2005).

O etiquetador morfossintático (*POS tagger*) é um *software* que classifica e/ou separa as palavras de um texto com base na classe de

palavra (substantivo, adjetivo, etc) de cada palavra, como no exemplo da Figura 3.

Enter text in Portuguese, separating paragraphs with an empty line:

Minha dificuldade mesmo é a ansiedade onde estravaso na comida

```
<p><s> Minha/POSS#fs dificuldade/DIFICULDADE/CN#fs mesmo/ADV
é/SER/V#pi-3s a/DA#fs ansiedade/ANSIEDADE/CN#fs onde/REL
estravaso/ESTRAVASAR/V#pi-1s em_/PREP a/DA#fs comida/COMIDA
/CN#fs </s></p>
```

Figura 3 – Resultado gerado por um POS tagger online de acesso livre
Fonte: Branco e Silva (2004)

Na Figura 3, pode-se observar que as palavras da mesma oração classificadas na Figura 2 foram separadas por um critério sintático, utilizando categorias como artigo (*ART*), advérbio (*ADV*), substantivo (*noun – N*) e verbo (*V*). Vale ressaltar que a forma de anotação das categorias por parte do analisador sintático segue o seguinte formato: categoria morfossintática seguida do símbolo “#” e outras categorias. Por exemplo, a classificação de “minha” é: *POSS#fs* – se refere a pronomes possessivos (*possessives*) e *F* se refere a feminino (*feminine*) e *S* se refere a singular (*singular*).

Por fim, o concordanciador (*concordancer*) é um *software* que busca uma palavra em um texto ou em um conjunto de textos e gera uma tabela com um número determinado de palavras antes e depois, determinando seu contexto, conforme será mais visto mais adiante. Para as palavras lexicais, ainda é necessária a intervenção do pesquisador na análise manual dos dados.

4 Comparação entre a Mineração de Textos e métodos tradicionais (Análise de Conteúdo)

A mineração de diferentes tipos de dados segue algumas tendências, tais como: caracterização, discriminação, associação, classificação, agrupamento e análise de similaridade dos dados (DOMINGUES, 2003). Vale a pena lembrar neste momento os tipos de dados com os quais podemos trabalhar: dados estruturados (numéricos ou categóricos – organizados em linhas e colunas) e não estruturados (obtidos por meio do processamento de áudio, vídeo, imagens e texto).

Nas Ciências da Saúde, quando somos apresentados aos conceitos da mineração de textos já citados (classificação, agrupamento e análise de similaridade) pensamos logo em

uma técnica tradicional e muito conhecida entre os pesquisadores, a Análise de Conteúdo. A Análise de Conteúdo consiste em uma técnica de exploração de dados que possui uma metodologia sistematizada com a qual é possível realizar inferências e fazer a interpretação de documentos para obter informações relevantes (BARDIN, 2011).

A Análise de Conteúdo tem suas raízes no início do século XX, nos Estados Unidos, tendo sido consagrada na literatura por Laurence Bardin. O objetivo primordial à época era tratar dados provenientes de jornais,

reportagens, cartas e discursos políticos. Hoje em dia, é um recurso metodológico utilizado em grande escala pelos pesquisadores das Ciências da Saúde que têm como objeto de estudo as comunicações em massa. De tal forma, a Análise de Conteúdo trata as mensagens dos textos para identificar significados (BARDIN, 2011).

Na Análise de Conteúdo, a análise de dados pode ser dividida em três fases: i) Pré-análise; ii) Exploração do Material e iii) Tratamento dos resultados: inferência e interpretação. As etapas estão detalhadas e esquematizadas na Figura 4:

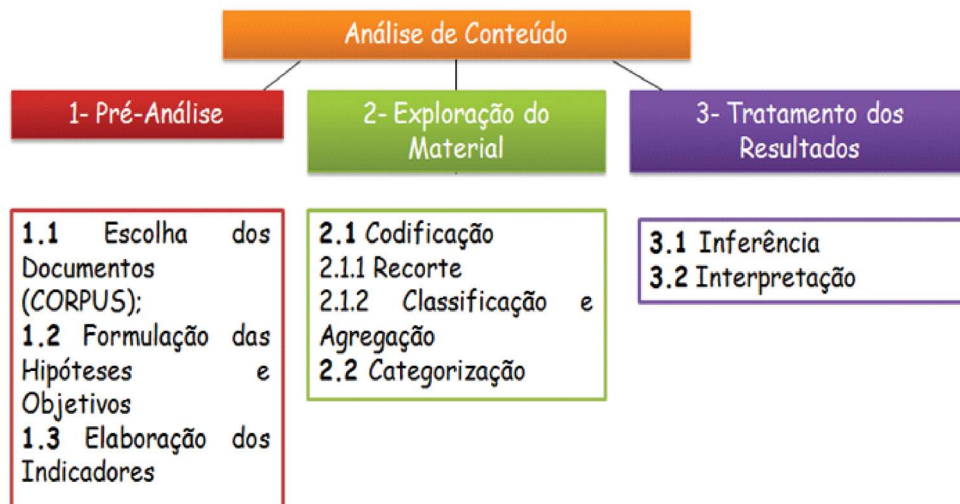


Figura 4 – Passos metodológicos da Análise de Conteúdo fundamentada por Bardin

Fonte: Elaborado para fins deste estudo, baseado em Bardin (2011).

Depois de realizada a escolha dos documentos a serem analisados, que, no caso de entrevistas, consistem em transcrições da fala dos entrevistados, conforme explicado nos Capítulos 2 e 5, parte-se para a formulação de uma hipótese, que nada mais é do que uma afirmação provisória que nos propomos verificar. Por exemplo: quando o pesquisador realiza uma entrevista com usuário do sistema de Saúde, com o objetivo de explorar barreiras que dificultem a prática de autocuidado em diabetes, espera-se encontrar dificuldades relacionadas aos aspectos comportamentais e emocionais que repercutirá

no controle glicêmico. Posteriormente com à Análise de Conteúdo, essa hipótese poderá ser confirmada ou não.

No que tange à segunda etapa da Análise de Conteúdo, a Exploração do Material, esta é a fase mais longa do processo de análise. Esta etapa, constituída pela codificação e categorização do material, consiste no processo de transformar dados brutos em dados categorizados, agregando-os em unidades de acordo com as características semelhantes dos conteúdos (BARDIN, 2011).

A última etapa da Análise de Conteúdo, denominada de Tratamento dos Resultados, é

o momento em que o pesquisador deve procurar tornar os resultados mais relevantes, por meio da inferência e interpretação (BARDIN, 2011).

De tal forma, utilizando a metodologia descrita, que não requer o uso de *softwares*/ambientes computacionais no processo, é possível adotar-se em conjunto a técnica de Mineração de Textos para subsidiar a Análise de Conteúdo. Nesse sentido, a Mineração de Textos vem ganhando espaço e, para seu crescimento como campo de estudo, foram desenvolvidos vários *softwares* e ambientes, os quais podem ser livres

ou comerciais e frequentemente incluem cálculos estatísticos. Como exemplos, temos o ambiente estatístico R, os *softwares* livres Poronto (ZAHRA; CARVALHO; MALUCELLI, 2013) e Sobek (GTECH.EDU RESEARCH GROUP, 2015) e o *software* comercial Text Mining Suite (INTEXT, 2005), entre outros.

Para efeitos de síntese, detalhamos no Quadro 1, as características da Mineração de Textos realizada por meio da máquina, comparada com a Análise de Conteúdo não automatizada.

Quadro 1 – Comparação das análises realizadas pela Mineração de Textos e pela Análise de conteúdo

Mineração de Textos	Análise de Conteúdo
Utiliza o método dedutivo	Utiliza o método indutivo
Parte-se da definição de um problema	Parte-se da definição de um problema
Deve-se avaliar o contexto para preparar a máquina com a finalidade de auxiliar em todas as fases do processo	Deve-se avaliar o contexto durante todas as fases do processo com a finalidade de fazer inferência
Analisa e localiza palavras no texto, independentemente de a frequência (se utilizada essa medida) ser baixa ou elevada	Analisa e busca palavras (índices ¹), independentemente de a frequência de ocorrência ser baixa ou elevada
Realizada por meio de programas que utilizam cálculos matemáticos e estatísticos	Realizada de maneira manual, porém com possibilidade de realização de cálculos estatísticos
Trabalha-se com a objetividade da máquina na análise e com a subjetividade do pesquisador na etapa de pós processamento	Trabalha-se com a subjetividade do pesquisador em todo o processo
Permite obter resultados de acordo com a frequência dos itens ou outras medidas, como distâncias calculadas matematicamente	Permite obter resultados de acordo com o índice de pertinência

Fonte: Elaborado para fins deste estudo e baseado em Bardin (2011) e Carvalho Filho (2014).

É interessante notar que, apesar de distintas, a Mineração de Textos, que se apoia na complementaridade homem-máquina, e a Análise de Conteúdo compartilham alguns aspectos semelhantes, como a avaliação do contexto diante dos resultados encontrados, partindo-se sempre do problema de pesquisa, além do fato de se considerarem relevantes palavras e suas frequências independentemente do número de ocorrências no corpus. Afinal, os itens menos frequentes podem ser relevantes em determinados estudos dependendo do

objetivo da pesquisa. Entretanto, na Análise de Conteúdo, sujeita a subjetividade do pesquisador, corre-se o risco de enaltecer pontos não tão relevantes ou, ainda, desconsiderar passagens importantes (BARDIN, 2011).

³ Os índices são “temas” identificados nos textos, os quais originarão os indicadores. Estes são evidências ao longo da mensagem que permitem inferir sobre a sua realidade, esteja essa implícita ou explícita. Por exemplo, ao analisar uma mensagem com verbalizações ou balbucios, como repetição de “hã” e “é”, é possível considerá-los como indicadores de dúvida (BARDIN, 2011)

Utilizando-se programas estatísticos como o ambiente R, o processo de busca das palavras mais frequentes nos textos se torna rápido e eficiente, o que pode nos indicar qual tópico deverá ser mais explorado dentro da pesquisa, respondendo rapidamente as hipóteses desenvolvidas no início da análise. Ainda, com a adoção de uma teoria abrangente da linguagem, como é a linguística sistêmico-funcional, apresentada no Capítulo 2, podemos interpretar as frequências extraídas de forma a obtermos características do contexto da interação (por exemplo, entre entrevistador e entrevistado) e examinarmos escolhas linguísticas que, seja pela sua reiteração (reforço ou repetição de determinado tópico), seja pela sua singularidade (ocorrência única de um determinado item), apontam para significados importantes construídos na interação.

Conforme será exemplificado por meio do estudo de caso deste Capítulo, a construção e a análise das listas de colocados e das linhas de concordância, juntamente com a frequência de cada item, facilitam a identificação dos resultados a serem discutidos. A nuvem de palavras também auxilia a visualização das palavras mais frequentes e permite uma interpretação por parte do leitor.

Cumprindo observar que na terceira parte do livro “Análise de Conteúdo” de Bardin (2011), publicado primeiramente em 1977, há um capítulo específico que aborda a informatização da análise. Nele, é sugerida a possibilidade de automatização da análise, por meio da contagem da frequência de palavras, unidades de registro e índice; da seleção de trechos de frases e sua localização no texto; da organização das classificações de temas com o auxílio de um processador de ideias e sua ordenação em esquema de árvore. Além disso, em Bardin (2011) são citados alguns processos informatizados específicos para a Análise de Conteúdo. Entretanto, devido à data de tais estudos, referentes à década de 80, esses são descritos de maneira pouco detalhada.

Esta seção trabalhou com a ideia de realçar as vantagens de uma Análise de Conteúdo complementada por meio da Mineração de Textos e apresentar resultados obtidos unicamente pela automatização da análise, tais como lista de colocados, linhas de concordância, nuvem de palavras e lista de frequências, potencializando a discussão dos resultados e a geração de conhecimento. Ainda, por meio da complementaridade homem-máquina, é possível promover técnicas de elicitação (geração do dado a ser coletado), análise e armazenamento de dados e resultados obtidos e explorar novas percepções que a visão humana ou da máquina de forma isolada não seriam capazes de suscitar. A próxima seção descreverá de maneira didática as etapas realizadas em um estudo de caso, para ilustrar a automatização do processo e apresentação dos resultados.

5 Estudo de caso

O estudo de caso presente neste Capítulo foi feito com base em um conjunto de dados não estruturados composto por 28 textos, transcrições provenientes de gravações realizadas no projeto Empoder@ – Protótipo conceitual e metodológico para avaliação de intervenções orientadas ao autocuidado em diabetes. As práticas orientadas ao cuidado em saúde no diabetes Mellitus foram baseadas no instrumento Protocolo de Mudança de Comportamento (FUNNELL; ANDERSON, 2004; CHAVES, 2014), como resolução de problemas, plano individualizado, definição de metas com avaliação do plano de cuidados, apresentado no Capítulo 1.

6 Metodologia

A preparação prévia dos textos foi realizada, de acordo com os passos sugeridos e

discutidos no Capítulo 5 sobre como realizar transcrições provenientes de gravações, porém com uma etapa adicional, que foi a conversão dos textos em formato *.doc* para *.txt*. A razão dessa etapa foi o fato de os arquivos de texto que serviram de base para este estudo estarem em formato *.doc*, e não em formato *.txt*. Logo, foi necessária a conversão do texto, que não precisaria ter sido feita, caso tivesse se optado pelo formato *txt*. A transcrição adequada contribui com a coleta e análise de dados estruturados e não estruturados, visto que, caso haja erros (por exemplo, de ortografia) nos dados gerados, isso pode enviesar os resultados da análise. Após a conversão de formato, os textos foram importados para o ambiente R com a finalidade de dar início à análise automatizada com base em procedimentos da Linguística de Corpus.

6.1 Razões da escolha do R como ambiente de análise

O R é ao mesmo tempo um ambiente e uma linguagem de programação livre, sendo amplamente utilizado em análises estatísticas e em representações gráficas (R CORE TEAM, 2015). Diferentemente de outros *softwares* (livres ou comerciais), como AntConc ou ParaConc, que também podem ser utilizados para realizar procedimentos da Linguística de Corpus (como listas de frequência, listas de colocados e linhas de concordância), o R é capaz de auxiliar não apenas na análise dos dados, mas em todas as etapas da pesquisa (FEINERER; HORNIK; MEYER, 2008, p. 3) a qual, neste caso, utilizou dados textuais. Em outras palavras, o R auxilia no Pré-processamento (preparação, importação, limpeza e pré-processamento geral dos dados), Associação (análise de associação, ou seja, a busca por associações para determinada palavra baseado nas contagens de frequência de co-ocorrência), Aglomeração

(agrupamento de documentos similares nos mesmos grupos) e Sumarização dos dados (sumarização de conceitos importantes de um texto, comumente as palavras de mais alta frequência).

Por fim, uma vez que o R é um *software* livre e utiliza linhas de comandos criadas ou obtidas pelo usuário de outra fonte, este permite que o pesquisador utilize praticamente quaisquer ferramentas computacionais à disposição para realizar suas análises, desde que estas estejam dentro das limitações da linguagem e do *hardware* (a máquina, fisicamente falando).

6.2 Preparação prévia dos textos

A preparação prévia do corpus consistiu, primeiramente, na transcrição das entrevistas e intervenções educativas para 56 arquivos no formato *.doc*: 28 arquivos contendo tanto as perguntas quanto as respostas e outros 28 arquivos, derivados do anterior, contendo apenas as respostas.

Em seguida, os textos foram preparados para análise no formato *.txt* de acordo com os padrões do ambiente R e dos scripts gerados para tal ambiente. Nesta adequação, os arquivos em formato *.txt* e codificação⁴ ANSI foram corrigidos quanto a erros ortográficos que pudessem ser críticos para a análise (como o item “a gente”, que muitas vezes havia sido transcrito como “agente”). Nesta correção, utilizou-se o editor de texto Notepad++⁵ versão 6.6.9, diferentemente do Bloco de Notas, sugerido no Capítulo 5, visto que o Notepad++ possui recursos adicionais para lidar com o texto, como a substituição textual em diversos arquivos em *.txt* (dentro de uma pasta) simultaneamente.

⁴ Uma codificação é uma forma utilizada pela máquina para entender os caracteres do teclado. Logo, um arquivo de texto estará em alguma codificação, como em UTF-8 ou ANSI (formato utilizado por padrão no Windows).

⁵ O Notepad++ é um editor de texto de código livre que possui suporte a diversas linguagens de programação e que funciona em ambiente MS Windows.

Além disso, houve a identificação entre parênteses angulares (“<” e “>”) de quem seria o turno do falante durante a entrevista, como sugerido no Capítulo 5, a fim de que estes pudessem ser ignorados pelo R durante o processamento dos textos utilizando parte dos scripts, que foram posteriormente elaborados. É importante mencionar que houve a conferência se todos os parênteses angulares estavam abertos e fechados, caso contrário o *software* R não saberá lidar com isso. Nos casos em que os parênteses angulares não estavam adequadamente fechados, fez-se uma substituição textual para fechá-los. Foram realizados, ainda, outros procedimentos de forma mais específica para adequar os textos à maneira como a máquina os processará, os quais tiveram o objetivo de evitar possíveis problemas durante a leitura pelo R. Esses procedimentos consistiram em: i) correção dos marcadores de fala (entre parênteses angulares); ii) substituição das aspas com a codificação do editor de texto por aquelas com a codificação do Bloco de Notas, ambos feitos utilizando o editor de texto Notepad++. Outros procedimentos do mesmo tipo, como a eliminação de acentos e de espaços extras (incluindo aqueles em início e final de linha) e a conversão das letras maiúsculas em minúsculas, foram realizados automaticamente utilizando os scripts elaborados.

7 Textos para análise

Na análise dos textos foram considerados apenas aqueles textos que continham os depoimentos dos usuários do serviço de Saúde, sem as indagações dos profissionais da área da Saúde, para que as frequências de palavras analisadas pudessem refletir apenas a percepção da autonomia no cuidar de si. Logo, os textos analisados contêm no total 38.300 itens (*tokens*), distribuídos em 28 arquivos de texto

em formato *.txt*, um para cada usuário entrevistado.

8 Procedimentos utilizando o ambiente R

A fim de que os textos pudessem ser processados e analisados no ambiente R, foi necessário levantar, em diversos sites e fóruns, pacotes do R que possuíam funções⁶ úteis para as pesquisas linguísticas. Com isso, foi possível a elaboração de scripts que permitiram não apenas a preparação dos dados para análise, mas também a realização dos seguintes procedimentos computacionais com aplicação na Linguística de Corpus e em outros campos de estudo:

- Listas de frequência
- Nuvens de palavras
- Linhas de concordância
- Listas de colocados

Com estes procedimentos, foi possível analisar os textos tanto globalmente (utilizando a lista de frequência e as nuvens de palavras) quanto especificamente (utilizando a lista de concordância e a lista de colocados) para estudar as temáticas mais frequentes citadas pelos usuários do serviço de Saúde com a condição crônica do diabetes.

Para interpretar os resultados obtidos, a linguística sistêmica funcional, apresentada no Capítulo 2, foi adotada. Esta teoria pode subsidiar a interpretação das frequências extraídas de forma automática (pela Mineração de Textos) ou manual (pela Análise de Conteúdo). A teoria possibilita explorar essas ocorrências textuais no sentido de indagar se elas poderiam

⁶ Funções, em computação, são conjuntos de instruções (dentro de algum *script*) organizadas em blocos de linhas de comando de determinada linguagem de programação visando atingir determinado objetivo.

apontar para características da fala do entrevistado (no caso do uso de entrevistas) acerca dos temas que são o objeto da pesquisa (sejam aspectos do autocuidado, como a alimentação, ou sentimentos explícitos no discurso).

A partir das 28 entrevistas pré-processadas e utilizando os scripts do R, foi possível gerar uma lista de frequência, linhas de concordância, listas de colocados (à esquerda e à direita) e uma nuvem de palavras, que são apresentadas a seguir.

9 Listas de frequência

O Quadro 2 apresenta a lista de frequência, em ordem decrescente, das 20 palavras relevantes mais frequentes das entrevistas face a face com os usuários, considerando o critério de relevância como alguma relação com o tema “autonomia e cuidado de si”.

Quadro 2 – Lista de Frequência das 20 palavras relevantes mais comuns das entrevistas

Palavra	Frequência
eu	2402
fazer	279
comer	152
alimentação	98
diabetes	96
dificuldade(s)	71
difícil	62
exercício(s)	57
saúde	55
caminhada	54
remédio	53
vontade	53
médico	47
doce	42
deus	40
glicose	40
cuidar	39
filha	36
insulina	31
melhorar	31

No Quadro 2, é importante destacar que a frequência da palavra “dificuldade(s)” representa a soma das frequências das palavras “dificuldade” e “dificuldades” e a frequência da palavra “exercício(s)” representa a soma das frequências das palavras “exercício” e “exercícios”.

É possível observar que, em todas as listas de frequência (antes de serem eliminadas palavras que não foram consideradas relevantes), a maior parte das palavras mais frequentes foram palavras gramaticais (artigos, preposições, conjunções), enquanto apenas algumas foram palavras lexicais (verbos, adjetivos, substantivos, pronomes). No entanto, essas palavras gramaticais não foram consideradas importantes para a análise realizada, pois não gerariam resultados relativos à temática analisada.

A frequência elevada de ocorrência nos textos do pronome “eu” pode ser interpretada como uma evidência relativa ao processo de autonomia do cuidado de si. De acordo com o sistema de SUJEITABILIDADE do sistema linguístico do português brasileiro (FIGUEREDO, 2011, p. 194), formas como “nós” e “a gente” são impessoais, diferentemente de “eu”. Assim, a escolha de “eu” pode ser um indicador de percepção de autonomia e responsabilidade perante o cuidar de si.

Para a análise da lista de frequência observou-se também o uso de “alimentação”, devido a sua alta frequência nas transcrições analisadas. Retomando a lista de frequência do Quadro 2 com as palavras relacionadas ao autocuidado e relacionando-as com este tema, pode-se sugerir que o verbo “fazer” e o substantivo “dificuldades” também nos remetem uma ideia de ação, a qual poderá vir acompanhada de dificuldades em se mudar o comportamento, aprofundando no quesito das barreiras para o autocuidado, objeto de estudo de algumas pesquisas nas Ciências da

Saúde (ONWUDIWE et al., 2011). Entretanto, o escopo do estudo de caso deste Capítulo foi demonstrar a técnica de Mineração de Textos e os procedimentos que podem ser adotados, como fazer uma lista de frequência ou uma nuvem de palavras, entre outros resultados obtidos uma vez que o pesquisador pode optar pela automatização da análise e adentrar nas possibilidades de apresentação de resultados mais inovadores.

10 Nuvem de palavras

Uma vez que a nuvem de palavras é a representação gráfica de uma lista de frequência um texto ou corpus, por meio dela é possível visualizar de forma espacial o modo como as palavras se organizam de acordo com sua frequência. Com isso, é possível analisar mais rapidamente as palavras consideradas mais relevantes (neste caso por serem as mais frequentes)

para o tema de estudo, sem recorrer a outros recursos. Assim, as palavras com maior frequência possuem fontes de maior tamanho e estão mais centralizadas que as outras palavras e as cores (quando presentes) ou os diferentes tons de cinza diferenciam também as frequências de ocorrência das palavras. Vale ressaltar que a opção de cores na apresentação gráfica da nuvem de palavras fica a critério do pesquisador.

A Figura 5 apresenta a nuvem de palavras gerada a partir da lista de frequência do Quadro 2. Nota-se também a presença de substantivos, adjetivos e verbos, que expressam os sentimentos dos usuários, sentimentos estes, positivos e negativos, levando-nos a refletir alguns aspectos do cuidar de si e viver com a condição do diabetes.

A Figura 5 mostra, em destaque, diversas palavras associadas a sentimentos e aos aspectos psicossociais de se viver com a condição crônica do diabetes:

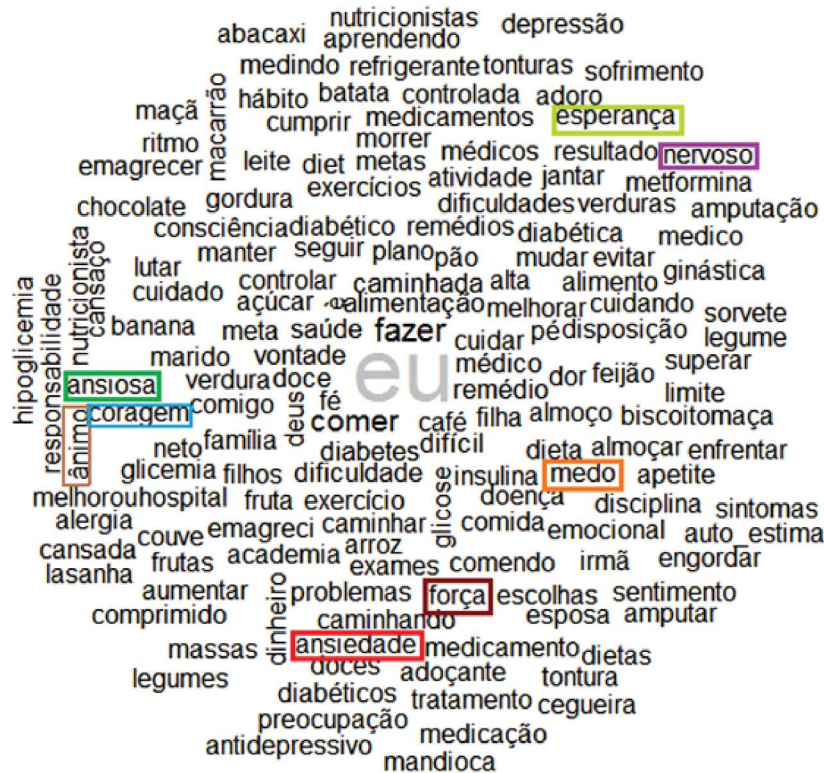


Figura 5 – Nuvem de palavras das entrevistas com destaque para uma seleção de palavras relacionadas a sentimentos e aspectos psicossociais de se viver com a condição do diabetes

É importante ressaltar também que outra possibilidade de análise, embora não utilizada nesse Capítulo, é a seleção das palavras de interesse a partir da lista de frequência completa para a criação de uma nova lista de frequência. Assim, para exemplificar esse ponto, com base na lista de palavras da Figura 5, foi gerada a lista de frequência do Quadro 3, contendo apenas os sentimentos ou os indicadores de sentimentos a serem analisados mais adiante. Como já mencionado na Introdução, esses sentimentos foram selecionados especificamente para a análise deste Capítulo.

Quadro 3 – Lista de frequência das palavras associadas aos sentimentos

Palavra associada a sentimento	Frequência
medo	13
força	11
coragem	8
ansiosa	7
esperança	5
ânimo	3
cansada	3
nervoso	2

No Quadro 3 podem ser observadas as palavras marcadas na nuvem de palavras, as quais são consideradas indicadoras de sentimentos ou estão relacionadas a eles. A partir disso, é possível confirmar a afirmativa de que, para algumas análises – como a realizada neste Capítulo, a escolha de determinadas palavras

Quadro 4 – Linhas de concordância das palavras associadas a sentimentos destacadas na nuvem de palavras

A1	A2	A3	A4	A5	Sentimento	D1	D2	D3	D4	D5
nenhum	tem		que	criar	coragem	e	sair	ter	ânimo	
para	tudo	Eu	não	tenho	ânimo	para	nada	antes	eu	não
sete	porque	eu	ainda	tenho	esperança	de	não	moner	sem	um
Eu	tenho	que	ir	com	fé	que	a	caminhada	vai	controlar
Aí	entra	um	pouco	de	depressão	porque	você	sente	que	não
minha	dificuldade	mesmo	é	a	ansiedade	onde	extravaso	na	comida	
Ficar	cega	ne.	eu	tenho	medo	da	cegueira	e	da	insulina

Como já foi dito, a escolha por gerar linhas de concordância a partir de palavras que indicassem um estado emocional partiu da

de análise não precisa ser baseada necessariamente em sua (alta ou baixa) frequência, visto que sua relação com o tema de estudo (neste caso, o exemplo dos sentimentos em relação a condição do diabetes) pode ser mais importante para o objetivo proposto.

Como evidenciado pela nuvem de palavras, há ocorrência de diversas palavras que denotam sentimentos, a saber, “esperança”, “coragem”, “fé”, “ânimo” e “força”: “sofrimento”, “medo”, “cansaço”, “ansiedade”, “nervoso”, “ansiosa”, “cansada”, “problemas” e “dificuldade”. Essas ocorrências sugerem ser a análise de sentimentos, conforme será utilizada no Capítulo 7, uma perspectiva produtiva para abordagem das entrevistas.

11 Linhas de concordância

O Quadro 4 apresenta exemplos de linhas de concordância com 5 palavras antes e 5 palavras depois da ocorrência da palavra de busca nas entrevistas, utilizando como palavra de busca aquelas associadas a sentimentos, identificadas na nuvem de palavras das transcrições. As colunas da linha de concordância indicam as posições antes e depois do nóculo, que no caso é palavra de busca (ver Quadro 6 mais adiante). Assim como a nuvem, essas linhas de concordância também foram elaboradas utilizando um script do R.

importância em se considerar os sentimentos e aspectos psicológicos e comportamentais do usuário do serviço de Saúde. Sentimentos

expressos pelas palavras “coragem” e “esperança”, no Quadro 4, apareceram em linhas de concordância que sugerem um sentido de motivação para a realização do autocuidado.

Cabe acrescentar ainda que o item “força de vontade”, embora não tenha ocorrido

na nuvem, também serviu de palavra de busca para linhas de concordância. O seu sentido, de acordo com as linhas de concordância, expressa um estado emocional que se associa à motivação discutida acima, por isso a opção de apresentá-lo no Quadro 5 de maneira separada.

Quadro 5 – Linhas de concordância de “força de vontade”

A1	A2	A3	A4	A5	Sentimento	D1	D2	D3	D4	D5
É	que	fala	assim	a	força de vontade	de	continuar	observando	os	passos
Esforçar	sabe,	tem	que	ter	força de vontade	né				

Tomando novamente como base no sistema de POLARIDADE da linguística sistêmico-funcional (HALLIDAY; MATTHIESSEN, 2014, p. 180), a análise das palavras de busca das linhas de concordância apresentadas no Quadro 4 e no Quadro 5 mostra que: i) a oração pode ser positiva ou negativa ii) pode haver ‘negatividade’ na ordem do grupo ou da palavra (MATTHIESSEN, 1995, p. 476-491). Por exemplo, no caso de “eu não tenho ânimo para nada”, retirado do Quadro 4, temos o adjunto “não”, classificando a oração como negativa. Assim, quando a polaridade da linha de concordância analisada for negativa, o sentimento da linha também será.

Dessa forma, ao analisarmos a palavra “ânimo”, indicadora do sentimento ânimo, ainda que “ânimo” sugira uma polaridade positiva, se no exemplo da linha de concordância ela ocorrer em “falta de ânimo”, essa expressão estará associada a negatividade. Se essa palavra fosse analisada somente pelo recurso da nuvem de palavras, poderíamos manter a interpretação de que, no contexto analisado,

a palavra expressa um sentimento positivo e se enquadra em um sentido de motivação para mudança de comportamento. Entretanto, ao observar a linha de concordância, foi possível verificar sua negatividade.

Adicionalmente, a palavra “ansiedade” (ver Quadro 4) vem contextualizada em situações negativas, relacionadas às dificuldades para seguir um padrão alimentar saudável, confirmando o que foi discutido na sessão “Listas de Frequência”, ao início do estudo de caso. A instabilidade emocional caracterizada pela ansiedade acarreta em um padrão alimentar desregular. Autores apontam a relação existente entre o intitulado “Transtorno da compulsão alimentar periódica” (TCAP) e o Diabetes Mellitus tipo 2, sendo que o TCAP é caracterizado por um episódio frequente e incontrolado de ingestão de alimentos (compulsão alimentar). Geralmente, a compulsão ainda ocorre associada ao sobrepeso, sintomas depressivos e qualidade de vida prejudicada (REBELO; VAZ, 2014).

Quadro 6 – Linhas de concordância de palavras indicadoras de sentimentos

A1	A2	A3	A4	A5	Sentimento	D1	D2	D3	D4	D5
you	do	bad	is	a	tired	hurt	ai	you	will	
o	doctor	found	that	these	problems	could	ser	more	emotional	and
o	doctor	found	that	these	problems	could	ser	more	emotional	e
until	then	was	making	only	choice	wrong	né	or	not	se
	eu	acho	que	nas	escolhas	you	tem	um	benefício	muito
eu	estou	dando	conta	a	difficult	está	ai	no	doce	que
you	do	bad	is	a	tired	hurt	ai	you	will	

12 Listas de colocados

Prosseguindo na apresentação de possibilidades de análise de dados a partir da geração de scripts no ambiente R, a Tabela 1 apresenta resultados elaborados com base na lista de colocados considerando as 5 palavras à esquerda da palavra de busca, na qual foram apresentados dados de diferentes palavras

de busca relacionados a sentimentos, assim como os totais de linha e de coluna. No entanto, para sumarizar os dados, apresentamos apenas a frequência de “eu”, verbos conjugados na primeira pessoa do singular (*verbos 1SG*), como “fico” ou “tenho”, e do “não”. O objetivo desta lista é auxiliar a comprovação da discussão do ambiente linguístico ao redor das palavras de estudo.

Tabela 1 – Lista de colocados à esquerda a 5 posições de diferentes palavras de busca relacionadas a sentimentos

Palavra de busca	Frequência (“eu”)	Frequência (verbos 1SG)	Frequência (“não”)	Total
ansiedade	1	2	0	3
ansiosa	1	4	0	5
ânimo	1	2	0	3
coragem	2	4	0	6
esperança	3	5	0	8
medo	6	7	3	16
nervoso	1	1	0	2
Total	15	25	3	43

Com base na Tabela 1, pode-se observar, primeiramente, que no caso de todos os sete itens analisados, houve ocorrência do “eu” e de algum verbo conjugado na primeira pessoa do singular antes da palavra de busca, como exemplo: “eu fico ansioso” ou “eu tenho coragem”.

Isso sugere, conforme discutido na seção “Lista de Frequências”, que há uma relação entre o “eu” e os verbos na primeira pessoa do singular, reafirmando uma condição do usuário considerar as disposições e as sensibilidades para lidar com suas dificuldades ao atuar em sua própria vida: uma vida caracterizada pela sua atitude no cuidado com a saúde, que implica em perseverança, autoconhecimento e aceitação de seus sentimentos.

13 Considerações finais

A contribuição da linguística por meio da Linguística Sistêmico-Funcional neste

Capítulo possibilitou uma interpretação da linguagem utilizada nas entrevistas de forma mais abrangente, apresentando como a linguagem expressava as interações humanas na construção do conhecimento. Foram sugeridas, ainda, novas possibilidades de análise dos dados não estruturados que podem ser realizadas em pesquisas futuras, levando em conta que a análise poderia ir além do item lexical e considerar também outras expressões indicando sentimentos, por meio da análise dos processos mentais e seus fenômenos (os complementos dos verbos a eles relacionados). Por fim, por meio da estatística e uma análise de dados automatizada, foi possível interpretar e ter acesso a uma análise de dados não estruturados com um enfoque mais aprofundado.

Assim, por meio desses recursos computacionais e estatísticos, é possível processar dados não estruturados (no caso deste Capítulo, as transcrições das entrevistas), transformando-os em dados estruturados

(em linhas e colunas) e obter resultados mais relevantes tanto do aspecto qualitativo quanto quantitativo, de acordo com os interesses do pesquisador.

Síntese do capítulo

- A opção pela Mineração de Textos para uma análise de grande volume de dados não estruturados mostra ser um caminho promissor para as metodologias de análise de dados em saúde, uma vez que integra recursos tecnológicos ao processo;
- A Mineração de Textos e a Análise de Conteúdo são técnicas que guardam semelhanças entre si e podem ser utilizadas de maneira complementar;
- O uso de uma teoria linguística subsidia a análise ao permitir interpretar as frequências obtidas, complementando a análise das Ciências da Saúde e da Mineração de Dados e de Textos;
- O uso de recursos tecnológicos pode beneficiar a análise de textos, tanto em termos de rapidez quanto de eficiência, gerando resultados confiáveis que podem até mesmo ser validados por meio do uso da estatística.

Referências

ALMEIDA, M. B. Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares. **Ciência da Informação**, v. 31, n. 2, p. 5–13, 2002.

ARANHA, C.; PASSOS, E. A tecnologia de mineração de textos. **Revista Eletrônica de Sistemas de Informação**, v. 5, n. 2, p. 1-8, 2006.

BARDIN, L. **Análise de Conteúdo**. São Paulo: Edições 70, 2011.

BRANCO, A.; SILVA, J. Evaluating Solutions for the Rapid Development of State-of-the-Art POS taggers for Portuguese. In: LINO, M. T. et al. (Eds.). **Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC2004)**. Paris: ELRA, 2004. p. 507–510. Disponível em: <<http://lxcenter.di.fc.ul.pt/services/pt/LXParserPT.html>>. Acesso em: 1 out. 2015.

CARVALHO FILHO, J. A. **Mineração de textos: análise de sentimento utilizando tweets referentes à copa do mundo**. Monografia (Graduação em Engenharia de Software)—Quixadá: Universidade Federal do Ceará, 2014.

CHAVES, F. A. **Tradução e adaptação cultural do “Behavior Change Protocol” para a língua portuguesa-Brasil**. Dissertação (Mestrado em Enfermagem)—Belo Horizonte: Universidade Federal de Minas Gerais. Escola de Enfermagem, 2014.

DOMINGUES, M. L. C. S. **Mineração de dados utilizando aprendizado não-supervisionado: um estudo de caso para bancos de dados da saúde**. Dissertação (Mestrado em Computação)—Porto Alegre: Universidade Federal do Rio Grande do Sul. Instituto de Informática., 2003.

FEINERER, I.; HORNIK, K.; MEYER, D. Text Mining Infrastructure in R. **Journal of Statistical Software**, v. 25, n. 5, p. 1–54, 2008.

FIGUEREDO, G. P. **Introdução ao perfil metafuncional do português brasileiro: contribuições para os estudos multilíngues**. Tese (Doutorado em Linguística Aplicada)—Belo Horizonte: Universidade Federal de Minas Gerais. Faculdade de Letras, 2011.

FUNNELL, M. M.; ANDERSON, R. M. Empowerment and Self-Management of Diabetes. **Clinical Diabetes**, v. 22, n. 3, p. 123-127, 2004.

- GALVÃO, N. D.; MARIN, H. DE F. Técnica de mineração de dados: uma revisão da literatura. **Acta Paulista de Enfermagem**, v. 22, n. 5, p. 686–690, 2009.
- GOMES, D. D. S. Inteligência Artificial: Conceitos e Aplicações. **Olhar Científico**, v. 1, n. 2, p. 234-246, 2010.
- GTECH.EDU RESEARCH GROUP. **SOBEK Text Mining**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2015.
- HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. **Halliday's introduction to functional grammar**. 4. ed. Milton Park, Abingdon, Oxon: Routledge, 2014.
- HAN, J.; KAMBER, M. **Data mining: concepts and techniques**. 2. ed. Amsterdam: Elsevier, 2006.
- INTEXT. **Intext Mining - Descoberta de Conhecimento em Textos**. Porto Alegre: InText, 2005.
- MACIEL, T. V. et al. Mineração de dados em triagem de risco de saúde. **Revista Brasileira de Computação Aplicada**, v. 7, n. 2, p. 26-40, 2015.
- MATTHIESSEN, C. M. I. M. **Lexicogrammatical cartography: English systems**. Tokyo: International Language Sciences Publishing, 1995.
- ONWUDIWE, N. C. et al. Barriers to self-management of diabetes: a qualitative study among low-income minority diabetics. **Ethnicity & Disease**, v. 21, n. 1, p. 27-32, 2011.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation for Statistical Computing, 2015.
- REBELO, A. B.; VAZ, E. M. Associação de Diabetes Mellitus com transtornos alimentares. **Revista da Sociedade Brasileira de Alimentação e Nutrição - Nutrire**, v. 39, n. 2, p. 252–264, 2014.
- REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. Barueri: Editora Manole Ltda, 2003.
- SILVA, J. et al. Out-of-the-Box Robust Parsing of Portuguese. In: PARDO, T. A. S. et al. (Eds.). **Proceedings of the 9th International Conference on the Computational Processing of Portuguese (PROPOR'10)**. Porto Alegre: Springer Berlin Heidelberg, 2010. p. 75–85. Disponível em <<http://lxcenter.di.fc.ul.pt/services/en/LXServicesSuite.html>>. Acesso em: 1 out. 2015.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to data mining**. 1. ed. Boston: Pearson Addison Wesley, 2005.
- ZAHRA, F. M.; CARVALHO, D. R.; MALUCELLI, A. Poronto: ferramenta para construção semi automática de ontologias em português. **Journal of Health Informatics**, v. 5, n. 2, 2013.

Leituras suplementares

- AQUINO, J. A. DE. **R para cientistas sociais**. Ilhéus: Editus, 2014.
- BARR, A.; FEIGENBAUM, E. A. (EDS.). **The handbook of artificial intelligence**. Los Altos: William Kaufmann, 1981.
- WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data mining: practical machine learning tools and techniques**. 3. ed. Burlington: Morgan Kaufmann, 2011.
- WONG-RIEGER, D.; RIEGER, F. P. Health Coaching in Diabetes: Empowering Patients to Self-Manage. **Canadian Journal of Diabetes**, v. 37, n. 1, p. 41-44, 2013.

GLOSSÁRIO

CORPORA: conjunto de corpus utilizados que servirão como base para algum estudo.

CORPUS: conjunto de dados de origem linguística, ou seja, provenientes do uso oral ou escrito da língua, que podem ser apresentados por meio de transcrições ou tratados de forma que possam ser processados pelo computador.

LISTA DE COLOCADOS: lista das palavras mais frequentes à esquerda ou à direita da palavra escolhida para análise. É um dos resultados obtidos pela análise de dados utilizando uma ferramenta computacional, como o ambiente R.

LISTA DE FREQUENCIA: apresentação das palavras mais frequentes no corpus que pode ser filtrada para exibir apenas determinada quantidade de palavras de acordo com a escolha do pesquisador. É um dos resultados obtidos pela análise de dados utilizando uma ferramenta computacional, como o ambiente R.

LINHAS DE CONCORDÂNCIA: linhas que indicam quais são as palavras antes e após a ocorrência do nóculo (palavra de busca). Por exemplo, se a palavra “alimentação” for escolhida como o nóculo, temos acesso a todos os trechos do corpus que contém “alimentação”, assim como o que está antes e após tal palavra. É um dos resultados obtidos pela análise de dados utilizando uma ferramenta computacional, como o ambiente R.

PALAVRA GRAMATICAL: classes de palavras que, para o cientista da computação, não

possuem significado por si só. São os artigos, preposições, pronomes e as conjunções.

PALAVRA LEXICAL: classes de palavras que, para o cientista da computação, possuem significado por si só. São os verbos, adjetivos, substantivos. Esse conceito pode se estender, inclusive, a estruturas linguísticas formadas por essas palavras lexicais.

NUVEM DE PALAVRAS: representação gráfica das palavras mais frequentes em um texto ou corpus. O tamanho de uma palavra na nuvem está relacionado à frequência com que essa palavra aparece no texto.

SCRIPT: na computação, o script se refere ao conjunto de instruções dados a um programa (software). No caso deste capítulo, script é um conjunto de linhas que geram um comando ao ambiente R, para que ele execute alguma ação.

TRANSCRIÇÃO: ato de transcrever de maneira criteriosa dados provenientes de falas, muitas vezes considerando determinadas expressões de interesse de acordo com o objetivo do estudo a ser feito.

TOKENS: palavras ou sequência de caracteres entre espaços.

TOKENIZAÇÃO: divisão de um texto em tokens.

ANÁLISE DE DADOS NÃO ESTRUTURADOS COLETADOS DE GRAVAÇÕES DE ÁUDIO E VÍDEO

Arthur de Melo Sá
Rodrigo Araújo e Castro
Sumaya Giarola Cecílio
Wanderley Gonçalves dos Santos Junior

Objetivos do capítulo

- Apresentar uma proposta de metodologia de análise de dados coletados de gravações de áudio e vídeo;
- Destacar as vantagens de se trabalhar com softwares de anotação de gravações.

1 Introdução

Nós usamos a língua para interagir uns com os outros, para estabelecer nossas relações interpessoais e manter a ordem social que as subjaz. Ao empregarmos a língua desta forma, nós também interpretamos e representamos o mundo para nós mesmos e uns para os outros. Portanto, a língua é uma parte natural do processo de estar vivo; ela é usada para “armazenar” a nossa experiência ao longo da vida, tanto da nossa vida individual, quanto da nossa vida coletiva. Entre outras coisas, a língua é ferramenta para representar o conhecimento, ou, em termos da própria língua, é ferramenta para construir significado

(MATTHIESSEN; HALLIDAY, 2009, p. 41).¹

¹ Tradução de Figueredo (2011) de: “We use language to interact with one another — to construct and maintain our interpersonal relations and the social order that lies behind them; and in doing so we interpret and represent the world for one another and for ourselves. Language is a natural part of the process of living; it is also used to ‘store’ the experience built up in the course of that process, both personal and collective. It is (among other things) a tool for representing knowledge — or, to look at this in terms of language itself, for constructing meaning.”

Retomando o conceito de linguagem apresentado no Capítulo 2, a citação acima de Matthiessen e Halliday (2009) ressalta nossa perspectiva de análise de que é por meio da linguagem que nos organizamos em uma sociedade. Isso nos permite, não só compartilhar da mesma cultura, como também compreender o que o outro fala. É justamente por existir a linguagem que podemos escrever e ler textos como o deste Capítulo. E seguindo a conceptualização da nossa teoria linguística, – a Linguística Sistêmico-Funcional – é também por meio da linguagem que as relações (e a ordem social a elas subjacente) entre o pesquisador e o profissional da área da Saúde são estabelecidas, assim como as relações entre o profissional da área da Saúde e o usuário do serviço de Saúde. Seu papel fundamental nas relações entre o usuário do serviço de Saúde, o profissional e o pesquisador reforçam, assim, a necessidade de se contemplar a linguagem como um objeto de estudo nas Ciências da Saúde, conforme reiterado em capítulos anteriores.

No Capítulo 2, vimos que, na revisão feita por Candlin e Candlin (2003), a linguagem é enfocada tanto em pesquisas das Ciências da Saúde como da Linguística Aplicada, verificando-se um número crescente de pesquisas nas Ciências da Saúde. Isso se reflete na forma como, nos últimos anos, a linguagem vem sendo investigada por diferentes pesquisadores, sobretudo no Brasil, tais como Vieira (2011), Santos e Torres (2012), David, Torres e Reis (2012), Cecílio *et al.* (2013), Chaves (2014) e Barroso (2015).

A título de exemplo, a pesquisa de Cecílio *et al.* (2013) teve como objetivo conhecer os sentimentos de companheiros de mulheres com histórico de câncer de mama, sentimentos esses expressos por meio da língua. Nesse estudo, os autores coletaram dados de gravações de áudio de entrevistas, as quais foram transcritas e depois analisadas. As análises apontaram que os sentimentos mencionados pelos entrevistados eram de “preocupação, tristeza, choque, angústia, transtorno, tensão e medo da perda da parceira” (CECÍLIO *et al.*, 2013,

p. 26). Além disso, evidenciaram que, após a confirmação do diagnóstico do câncer de mama, os parceiros apresentaram uma tendência em “desenvolver maior exigência sobre suas atitudes diante da companheira e da família” (CECÍLIO *et al.*, 2013, p. 26), pois eles sentiam-se na obrigação de “manter a imagem da força, de apoio e de superação da doença, ocultando, por vezes, seus reais sentimentos” (CECÍLIO *et al.*, 2013, p. 26). A identificação desses sentimentos dos parceiros das doentes só foi possível graças à investigação de suas falas.

Assim como a pesquisa de Cecílio *et al.* (2013), o estudo da linguagem nas interações no âmbito do serviço de Saúde permite compreender aspectos da relação entre os interlocutores (profissionais e usuários) que de outra forma não poderiam ser identificados. Sendo assim, a linguagem constitui-se como um importante objeto de estudo para o empoderamento do pesquisador, cujas pesquisas poderão vir a oferecer subsídios para o empoderamento do profissional da área da Saúde e, em virtude da sua prática profissional, propiciar o empoderamento do usuário do serviço de Saúde, como exposto no Capítulo 1.

Como vimos no Capítulo 3, dados coletados de textos escritos ou transcrições da fala podem ser abordados como dados não estruturados, passíveis de serem examinados por meio da Mineração de Textos, ou se forem dotados de estruturação prévia, como dados estruturados, pela Mineração de Dados. Neste Capítulo, gravações de vídeo de entrevistas são importadas num *software* para serem anotadas com categorias definidas pelo pesquisador. As categorias anotadas geram dados estruturados que podem ser abordados pelos métodos descritos no Capítulo 3 e que serão analisados pelas técnicas descritas no Capítulo 8. Trata-se de uma alternativa ao tratamento de entrevistas gravadas que são transcritas manualmente pelo pesquisador, examinado no Capítulo 5, uma vez que a anotação pode ser feita sem necessidade de se transcrever as falas.

A próxima seção apresenta alguns referenciais teóricos que justificam a utilização de ferramentas computacionais. Em seguida, é apresentada a metodologia objeto de descrição deste Capítulo, bem como um breve estudo de caso que a exemplifica.

2 Dados não estruturados

Como já exposto nos Capítulos 3 e 6, dados não estruturados podem ser definidos como dados que não estão armazenados em uma estrutura, como na forma de linhas e colunas em uma planilha de dados. Esse tipo de dado não é passível de ser usado em uma análise computacional sem preparação prévia, necessária, até mesmo, para a realização de procedimentos como a busca por determinada palavra, caso se trate de texto. Dessa forma, para a obtenção de resultados quantitativos, é necessário que sejam preparados previamente, seja pela *tokenização* (a separação de cada palavra) para ser feita a contagem ou pela geração de dados estruturados, dispostos, especificamente, em uma planilha de dados, que seria o equivalente a gerar um novo conjunto de dados (estruturados) a partir dos iniciais (não estruturados).

Uma forma de gerar dados não estruturados é, no caso de texto falado, fazer a transcrição do arquivo de áudio ou vídeo, para em seguida prepará-los para a análise e realização de outros procedimentos, como a geração de listas de frequências ou nuvens de palavras (exploradas no Capítulo 6).

Entretanto, um *software* de anotação de entrevistas gravadas possibilita a anotação de segmentos da gravação e dispensa a transcrição do texto falado, lembrando que a decisão de se transcrever os dados ou não irá depender do objeto de estudo do pesquisador.

Uma vez feita a anotação e gerados dados estruturados, pode-se, por exemplo, contar as

frequências das categorias anotadas e gerar tabelas/gráficos que podem ser interpretados ou apresentados como resultados dos estudos.

Nas Ciências da Saúde, muitos estudos são realizados por meio de anotações manuais, como Onwudiwe *et al* (2011), Cecílio *et al* (2013), Santos *et al* (2014), Peres *et al* (2014), Oliveira *et al* (2015), Gvozd *et al* (2014) e Nery *et al* (2015). Todas essas pesquisas foram baseadas na análise manual, envolvendo leituras exaustivas (i.e., até que o pesquisador não considere possível extrair mais dados) de transcrições de entrevistas gravadas, e na criação de categorias, contagem de suas frequências e retirada de exemplos, a fim de gerar resultados e conclusões. Dentre esses estudos, Cecílio *et al* (2013), Peres *et al* (2014), Oliveira *et al* (2015), Gvozd *et al* (2014) fizeram a anotação manual utilizando a metodologia da Análise de Conteúdo de Bardin (2011), apresentada no Capítulo 6.

Todavia, alguns estudos realizados nas Ciências da Saúde (ZINKEN; CRADOCK; SKINNER, 2008) e também em outras áreas como a Linguística (CRASBORN *et al.*, 2006; CRASBORN; SLOETJES, 2008; DREUW; NEY, 2008) e a Engenharia (ROS; BARONI; DEMIRIS, 2014), fizeram uso do *software* de anotação ELAN (MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS, 2015). Todos eles apontam as vantagens da anotação das gravações em vídeo em comparação às metodologias baseadas na realização de transcrições manuais.

Analogamente às pesquisas acima mencionadas, neste Capítulo apresentamos um estudo de caso utilizando uma abordagem semiautomatizada fazendo uso do *software* ELAN. O estudo visa ilustrar os benefícios de se utilizar o *software*, que dispensa a transcrição, reduz a participação humana no processo de análise e permite ao pesquisador delegar mais funções à máquina, que se destaca pela

velocidade e consistência ao executar certas funções (REZENDE, 2003, p. 14). O *software* permite, ainda, caso for necessária ou desejável a transcrição, que ela seja feita de forma mais integrada, uma vez que se pode ouvir o som e digitar a transcrição no ambiente do *software*, bem como armazenar a mesma junto à anotação. Ainda, o *software* permite que o pesquisador insira uma tradução ou legendagem da transcrição para uma outra língua, caso seja condizente aos objetivos da pesquisa. Por meio do *software* ELAN é também possível associar as anotações às categorias em um mesmo momento, otimizando o processo de análise dos dados. Nos casos em que a filmagem também será analisada (e não somente o recurso de voz), há a possibilidade de se analisar a comunicação não verbal, potencializando a contabilização dos sentimentos dos respondentes.

Na próxima seção detalhamos a metodologia de análise utilizada.

3 Uma proposta de metodologia

A metodologia apresentada aqui utiliza o *software* ELAN (MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS, 2015) para o tratamento e análise dos dados. Nosso estudo enfocou gravações realizadas com usuários do sistema de Saúde – mais especificamente, uma gravação de vídeo de uma entrevista realizada com uma usuária com condição crônica do diabetes. A entrevista foi escolhida como tipo de texto objeto deste estudo devido ao fato de ser uma instância linguística caracterizada pela interação dialógica, face-a-face, registrada em tempo real por meio de filmagem em vídeo.

É importante ressaltar que a entrevista filmada e anotada aqui foi gerada com base num roteiro de coleta de dados semiestruturado, o qual consiste em perguntas baseadas nos objetivos da pesquisa, mas também possibilita

a realização de outras perguntas pertinentes ao tema pesquisado, dando ao respondente mais oportunidades de verbalizar suas dúvidas e questionamentos. Quando o recurso da filmagem é planejado, o pesquisador pode prestar mais atenção a detalhes que podem contribuir para melhores resultados de análise. Como exemplo, caso a coleta de dados ocorra por meio de um grupo focal, um recurso de coleta de dados em grupo, é preciso que exista mais de um equipamento de filmagem e de microfones com boa qualidade, os quais permitam a captação de imagem e som de todos os envolvidos.

Neste capítulo recomendamos planejar a filmagem antes de sua importação no ELAN, ainda que seja possível utilizar filmagens que já tenham sido feitas anteriormente, ou mesmo gravações realizadas por terceiros.

Para realizar a filmagem, pode-se utilizar mais de uma câmera, geralmente não mais que quatro, sendo o número flexível de acordo com o número de entrevistados. A filmagem pode ser realizada tanto no domicílio do entrevistado quanto em uma sala de um centro de Saúde, podendo esta última ser a opção escolhida quando o entrevistado não possui dificuldades de locomoção. Um bom ângulo para a realização da filmagem é aquele em que o entrevistado é colocado de frente para a câmera, ficando o entrevistador de costas ou ao lado da câmera. Para ambientes maiores, como uma sala de reuniões com mais de dois participantes, é recomendado, além de uma câmera, um gravador de som conectado a ela a fim de capturar o áudio de todo o ambiente com maior qualidade. Nesse caso, depois de terminada a entrevista, será necessário sincronizar a imagem e o áudio captado pelo gravador de som. Por fim, para a escolha do equipamento e do ambiente em que ocorre a gravação da mídia, devemos levar em conta tanto a natureza do estudo (e.g. se é necessário maior ou

menor qualidade de áudio/vídeo para analisar o fenômeno estudado) quanto o objetivo e natureza da pesquisa: exploratória – proporciona maior familiaridade com o problema de pesquisa; descritiva – visa descrever o fenômeno que se quer estudar; explicativa – identifica os fatores que contribuem para a ocorrência do fenômeno (GIL, 2008). Se a filmagem for planejada antes de sua importação no ELAN, aconselhamos (i) utilizar uma resolução não muito alta para filmar, porém com qualidade razoável (por exemplo 800x600 pixels); e (ii) obter uma filmagem com duração que não exceda 30 minutos. Se o vídeo já tiver sido gravado anteriormente e não atender aos requisitos acima, aconselhamos editar o vídeo, alterando seu formato e sua resolução, e fazer o recorte de um trecho deste, utilizando algum *software* de edição de vídeos.²

No estudo apresentado neste Capítulo, utilizamos um trecho de 2 minutos com a resolução 640x480 pixels. É importante esclarecer que os critérios adotados aqui quanto à preparação do material não são baseados nas atribuições do *software* de anotação; é possível trabalhar com arquivos maiores, isto é, de alta resolução e de longa duração, como um filme comercial. Sugerimos utilizar arquivos menores com o objetivo de melhorar o desempenho do processamento de dados por parte do computador.

Todas as anotações feitas com o ELAN, incluindo transcrições da fala, são armazenadas em um arquivo com extensão própria (.eaf). A estrutura de anotação, isto é, o sistema hierárquico de etiquetas criadas e que serão anotadas, também é armazenado em formato próprio com extensão .etf. Ambos os arquivos são armazenados de forma separada do arquivo de áudio (normalmente em .mp3) ou vídeo (que pode estar em .mp4, .mts, .avi, etc.) importados no ELAN.

² Um *software* livre de edição de vídeos sugerido é o Lightworks, disponível em <<https://www.lwks.com/>>.

Ressaltamos aqui a importância de o arquivo de áudio ou vídeo a ser importado estar em um formato “legível” para o computador. Muitas vezes, câmeras e gravadores de som armazenam os dados em formatos proprietários de uma marca ou sistema operacional (como o .mts, para vídeo, e o .aac, para áudio). Alguns desses formatos não são suportados pelo sistema operacional do computador, sendo necessária a instalação de *codecs*³ para sua execução. O ELAN somente conseguirá lidar com esses arquivos se o sistema operacional conseguir executá-los. Portanto, caso o equipamento armazene a gravação de áudio ou vídeo em um formato proprietário, é necessário instalar o *codec* específico para esse formato para trabalhar com o arquivo no ELAN. No entanto, se o computador já é capaz de executar o arquivo, isso significa que ele já possui um *codec* instalado.

Ademais, o ELAN oferece a possibilidade de integração com outros *softwares*, o que permite que os dados trabalhados em um determinado projeto possam ser exportados para outros programas de manipulação de dados, como o R (R CORE TEAM, 2015), editores de planilhas eletrônicas, editores de textos, etc., nos quais os resultados podem ser armazenados e manipulados em formatos diversos, como .txt, .xls, .htm, entre outros. É importante ressaltar que a opção de exportação é apenas uma sugestão, já que dentro do próprio *software* há diversos recursos, como busca por palavras, criação de tabelas de frequência, ferramentas de manipulação de áudio e vídeo, dentre outros.

Além da exportação dos seus dados para outras plataformas, o ELAN também permite

³ Codec é uma sigla para Compressor-Descompressor. É um software responsável por codificar e decodificar arquivos em um determinado formato. Dessa forma, o equipamento de gravação comprime (isto é, codifica) a gravação em um arquivo (e.g. .mp3, .mp4, .avi, etc.), e o computador descomprime (isto é, decodifica) o arquivo para reproduzi-lo.

a importação de dados de outros programas de transcrição, anotação e até mesmo de legendagem para sua plataforma, o que possibilita ao usuário associar o seu projeto com formatos de arquivos de outros *softwares*.

Para realizar a anotação de categorias, é necessário definir as mesmas e isto se faz com base em uma teoria que delimite categorias passíveis de anotação. Para este Capítulo, utilizamos categorias de análise de sentimentos baseadas em Plutchik (1980), que permitem caracterizar o sentimento expresso pelo entrevistado em suas falas e gestos. Sendo assim, para a análise de sentimentos, neste Capítulo trabalhamos com os 8 sentimentos humanos básicos definidos em Plutchik (1980): alegria (*joy*); tristeza (*sadness*); confiança (*trust*); aversão (*disgust*); medo (*fear*); raiva (*anger*); surpresa (*surprise*); expectativa (*anticipation*). O objetivo foi tentar identificar, por meio da frequência das anotações, a prevalência de algum sentimento específico que permitisse traçar um perfil do usuário e este pudesse ser analisado em relação com outros dados desse mesmo usuário. Como a atribuição de uma categoria a um trecho da fala é feita manualmente pelo pesquisador e, no caso de sentimentos, com base na sua ponderação, sugere-se sempre a utilização de um mínimo de 2 anotadores independentes para que os resultados de suas anotações possam ser comparados e a consistência na atribuição das categorias verificada. Mais detalhes são fornecidos na seção seguinte.

4 Estudo de caso

Como dito, o estudo de caso deste Capítulo consistiu na anotação dos dois primeiros minutos de uma das entrevistas conduzida por uma pesquisadora da Escola de Enfermagem da UFMG. A entrevistada era uma usuária do serviço de Saúde com a condição crônica do Diabetes Mellitus tipo 2.

Essa entrevista faz parte do banco de dados do projeto Empoder@ - *Protótipo conceitual e metodológico para avaliação de intervenções orientadas ao autocuidado em diabetes*, desenvolvido em um trabalho conjunto de pesquisadores das Ciências da Saúde, da Estatística e da Linguística Aplicada.

A equipe do projeto Empoder@ procura acompanhar de perto a saúde dos usuários do serviço de Saúde com a condição crônica do diabetes por meio de visitas domiciliares, ligação telefônica por meio de software de comunicação pela Internet e interlocução coletiva, conforme apresentado no Capítulo 1. Entre os profissionais que integram o projeto, estão enfermeiros e nutricionistas, que desenvolvem estratégias sistematizadas de educação em saúde aos usuários pertencentes aos Centros de Saúde colaboradores do projeto.

Além disso, um dos principais objetivos do projeto é o estabelecimento de vínculos entre profissionais e usuários do sistema de Saúde, para uma melhor compreensão das vivências dos usuários a fim de alcançar melhores resultados no tratamento das condições crônicas. Logo, a realização de entrevistas favorece um contato mais próximo com o usuário do sistema de Saúde, permitindo que este compartilhe suas vivências com sua condição e que o pesquisador e/ou profissional da Saúde compartilhem de seu conhecimento técnico sobre os cuidados com a condição crônica do diabetes.

Neste Capítulo, o trecho da entrevista foi analisado com o objetivo de mostrar como pode ser feita a anotação das falas do entrevistado por meio do *software* ELAN. Como acima dito, foram anotadas 8 categorias referentes aos 8 sentimentos humanos básicos definidos em Plutchik (1980).

A seguir são apresentados os procedimentos detalhados para a anotação no ELAN conforme os objetivos deste estudo de caso. Cumpre esclarecer que os procedimentos descritos não caracterizam um tutorial para a

utilização do ELAN. Mais informações a esse respeito podem ser encontradas em Auer *et al* (2010) e Brugman e Russel (2004).

O arquivo de vídeo utilizado no nosso estudo encontrava-se no formato *.mts*. Após a conversão para *.mp4*, o vídeo foi importado no ELAN e foram definidas as “trilhas”, que são camadas criadas dentro do *software* reservadas à segmentação e anotação do áudio ou vídeo. A título de comparação, essas trilhas são informações acrescentadas ao conteúdo como é o caso das legendas em filmes; ao legendarmos um filme, estamos criando uma camada adicional com informações que antes não estavam presentes. Essas informações podem ser a tradução da língua falada no vídeo (e.g. inglês) para a língua do público alvo (e.g. português), a descrição do áudio para portadores de deficiência auditiva, ou, no caso de longas metragens, comentários do diretor. Em todos os casos, criamos uma camada com informações adicionais, sendo possível até criamos múltiplas camadas e deixar que o telespectador escolha qual(is) deseja em determinado momento. De forma semelhante, no ELAN as trilhas permitem adicionar informações diversas e também definir qual(is) tipos de informação serão trabalhadas num determinado momento.

Após as “trilhas”, foram criados os “tipos linguísticos”, necessários para poder segmentar trechos de interesse e anotá-los. Um tipo linguístico diz respeito a um critério para selecionar o que será anotado, sendo definido segundo o interesse de pesquisa. Como exemplo, no estudo de caso deste Capítulo, investigamos as falas do entrevistado; portanto, não era de interesse anotar as falas do entrevistador nem os momentos de silêncio; assim foi criado um tipo linguístico chamado “interesse”, com o objetivo de anotar somente esses trechos em que

o entrevistado falava. Além disso, para anotar os sentimentos expressos pelo entrevistado, foi necessário definir um outro tipo linguístico que permitisse anotá-los segundo os trechos de interesse definidos anteriormente. Em suma, os tipos linguísticos indicam o tipo de anotação a ser inserido em uma trilha.

Para que se possa definir os tipos linguísticos, o ELAN permite quatro tipos de segmentação e anotação. Esses quatro tipos de anotação recebem o nome de “estereótipos”, sendo eles: (i) “*time subdivision*”, que permite segmentar o trecho de interesse em trechos menores, não permitindo intervalos vazios entre esses trechos menores; (ii) “*included in*”, que é semelhante ao primeiro estereótipo, mas permitindo intervalos vazios entre os trechos menores; (iii) “*symbolic association*”, que permite a associação de uma categoria qualquer a um trecho de interesse completo; e (iv) “*symbolic subdivision*”, que é semelhante ao anterior mas permite a associação de categorias a partes de um trecho de interesse.

Como já mencionado, nosso estudo visou anotar apenas as falas do entrevistado e optamos por fazer a transcrição do texto falado para escrito dessas falas exclusivamente. Adicionalmente, fizemos a tradução das falas transcritas para a língua a inglesa. Assim, os tipos linguísticos usados para a anotação dos dois primeiros minutos da entrevista foram: (i) “Interesse”, com o estereótipo “*included in*”; (ii) “texto”, com o estereótipo “*symbolic association*”; e (iii) “Sentimentos”, também com o estereótipo “*symbolic association*”. Neste último tipo, optamos pela opção de inserir as categorias de anotação na forma de vocabulário controlado possibilitando somente a anotação das 8 categorias correspondentes aos 8 sentimentos de Plutchik (2001) – ver Figura 1.



Figura 1 – Captura de tela do software ELAN mostrando a criação de tipos linguísticos

Esses tipos linguísticos foram criados a fim de obtermos uma etiqueta para cada trecho de fala. É importante reiterar que foi feita a transcrição apenas dos trechos de interesse de acordo com o objetivo do nosso estudo. Isso permitiu exemplificar os resultados deste estudo de caso – conforme o Quadro 2. Além

disso, a tradução das transcrições foi feita unicamente com o objetivo de demonstrar as possibilidades de trabalho com o ELAN. A Figura 2 mostra uma tela capturada, na qual vemos, na parte inferior, a estruturação de anotação criada.

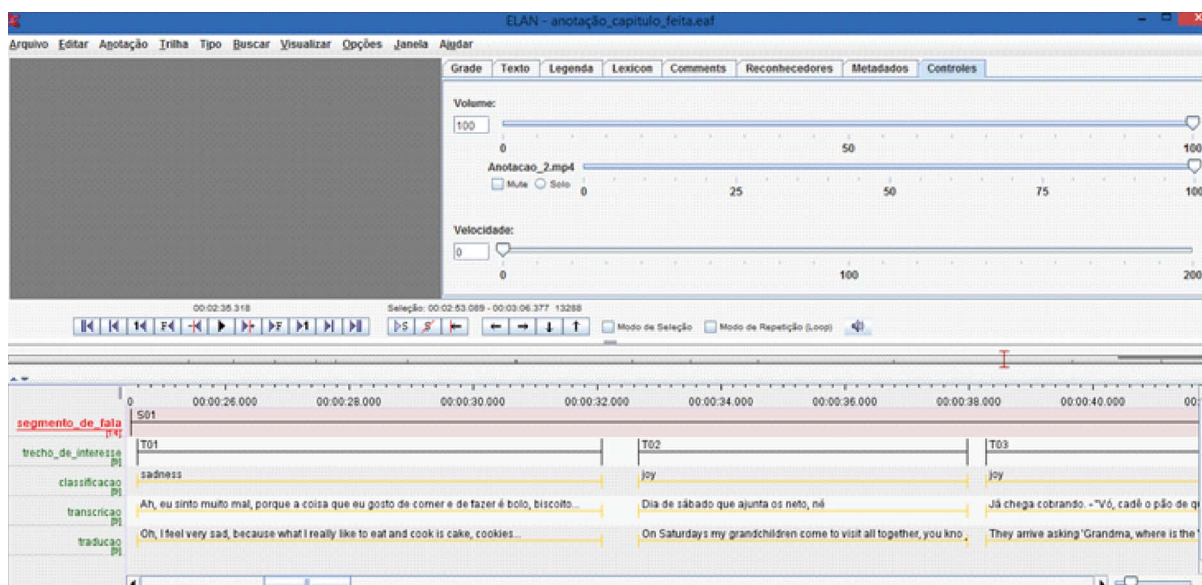


Figura 2 – Captura de tela do software ELAN com a estrutura de anotação utilizada

Com os tipos linguísticos criados, foram, então, desenvolvidas as trilhas a serem utilizadas na anotação: (i) trilha “segmento_de_fala”, que visa indicar no vídeo quais são

os trechos que efetivamente contém a fala do sujeito, excluindo assim momentos de pausa e também momentos com a fala do entrevistador; (ii) “Trecho_de_interesse”, que permite a

seleção de trechos relevantes e que serão posteriormente classificados; (iii) “classificacao”, com a anotação do sentimento identificado no “Trecho_de_interesse”; (iv) “Transcrição”; contendo a transcrição do trecho de interesse; e (v) “tradução”, com a tradução do trecho de interesse para o inglês.

A trilha “segmento_de_fala” foi associada ao tipo linguístico “default”, a trilha “trecho_de_interesse” ao tipo linguístico “interesse”; a trilha “classificacao” ao tipo “Sentimentos”; e as trilhas “transcricao” e “traducao”

ao tipo “texto”, cuja característica (texto livre, sem a seleção de categorias fixas) se adequa a ambas as trilhas, conforme mostrado na Figura 3. Essas trilhas foram organizadas em uma hierarquia da seguinte forma: “segmento_de_fala” é a trilha mãe de “trecho_de_interesse”, que por sua vez é a trilha mãe de “classificacao”, “transcricao” e “traducao”. A Figura 2 mostra o processo de criação no ELAN das trilhas de acordo com os tipos linguísticos criados anteriormente.



Figura 3 – Captura de tela do software ELAN mostrando a criação de trilhas

Além disso, a Figura 4 apresenta as relações de hierarquia entre as trilhas e os tipos linguísticos.

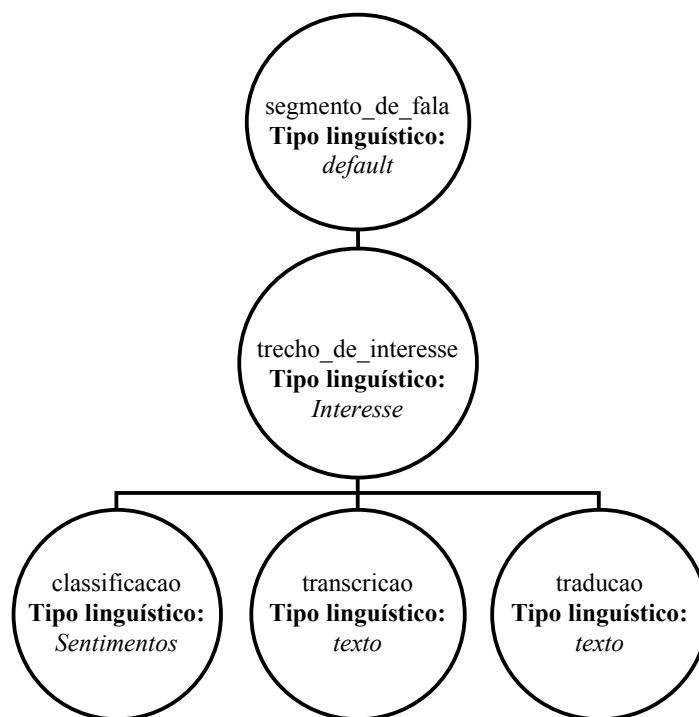


Figura 4 – Relações hierárquicas entre as trilhas e os tipos linguísticos

Em seguida, foram marcados na trilha “segmento_de_fala” os intervalos de tempo com as falas da entrevistada por meio do modo de segmentação (*segmentation mode*) e, para melhor identificá-los, esses segmentos receberam as etiquetas (*tags*) de S01 (segmento 01), S02 (segmento 02), e assim sucessivamente. Os trechos de interesse foram então identificados com as etiquetas T01, T02, T03, etc. e também foram selecionados por meio do modo de segmentação. Na trilha “classificacao” havia sido criado um vocabulário controlado contendo as categorias correspondentes aos 8 sentimentos de Plutchick (1980), de forma que o anotador simplesmente escolhe uma delas e não necessita digitá-las sempre que for fazer uma anotação. Por fim, esses trechos de interesse foram transcritos na trilha “transcricao” (ver Figura 2, na qual o vídeo foi obscurecido para preservar as identidades do entrevistado e do entrevistador).

Terminado o processo de anotação, o projeto foi salvo em um arquivo no formato *.eaf* e o modelo de anotação no formato *.etf*. As anotações foram exportadas em um arquivo no formato *.txt* tabulado (separado por tabulações), o qual pode ser aberto por qualquer editor de planilhas ou por um editor de textos.

A seguir, são apresentados os resultados do estudo de caso gerados a partir dos procedimentos descritos acima.

5 Resultados do estudo de caso

Um dos resultados que podem ser extraídos diretamente do ELAN e que pode ser utilizado como base para gerar outros resultados é a lista de frequência das categorias de sentimentos anotados na entrevista. Essa lista é gerada como um arquivo de texto tabulado, como pode ser observado no Quadro 1.

Quadro 1 – Lista de frequência de categorias anotadas na trilha de “classificacao”

Categoria	Frequência
alegria	4
tristeza	4
aversão	1
confiança	0
medo	0
raiva	0
surpresa	0
expectativa	0

Nesta lista de frequência, podemos observar ocorrências de três categorias anotadas na trilha “classificacao” de tipo linguístico

“Sentimentos”, utilizadas no estudo de caso e ordenadas por frequência: *alegria*, *tristeza* e *aversão*, sendo que as demais (*confiança*, *medo*, *raiva* e *surpresa*) não apresentaram ocorrência. Com base nessa lista de frequência, podem ser feitos, por exemplo, gráficos para representar os dados de forma distinta e enfatizar diferentes aspectos da análise.

Por exemplo, com base na lista de frequência (incluindo os itens de frequência zero) do Quadro 1, é possível desenhar o Gráfico 1: um gráfico de barras representando visualmente a mesma informação disposta no Quadro 1 mas de forma percentual.

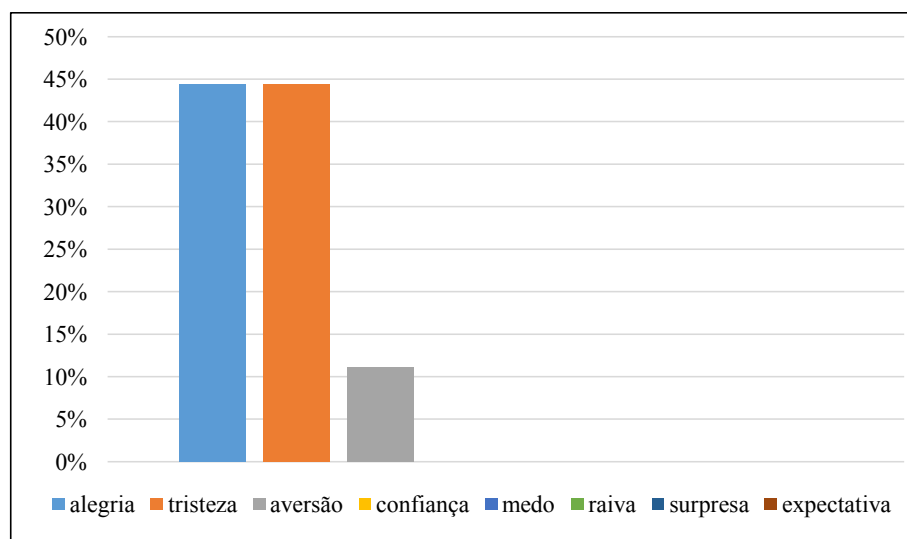


Gráfico 1 – Ocorrência percentual das categorias anotadas nas entrevistas

No Gráfico 1, que permite uma maior comparabilidade entre a ocorrência das categorias estudadas, verifica-se que, do total de 8 categorias utilizadas, apenas *alegria*, *tristeza* e *aversão* foram anotadas no trecho da entrevista estudado. De acordo com esse gráfico, pode-se observar que as duas primeiras categorias (*alegria* e *tristeza*) ocorreram com a mesma frequência (44,44%) e a terceira (*aversão*) com uma frequência menor (11,11%). Esta igualdade das duas primeiras em relação à frequência

pode ser justificada pelo fato de que sempre que a entrevistada falava sobre uma vivência prazerosa (e.g. como ela adorava cozinhar para os netos e a família), em seguida ela descrevia as limitações que sua condição crônica lhe impunha, sentimento ao qual foi atribuído a categoria *tristeza*. Portanto, agrupando a categoria *tristeza* e *aversão*, têm-se como resultado 55,44% de sentimentos negativos, revelando uma predominância destes em relação aos sentimentos positivos.

Podemos também extrair do ELAN, como ilustrado no Quadro 2, um arquivo de texto tabulado, no qual podemos visualizar a classificação dos sentimentos, assim como o segmento de fala, o trecho de interesse e a

transcrição da fala do entrevistado em que se observa um sentimento. Com isso, é possível relacionar a classificação atribuída a um trecho de interesse e extrair exemplos a partir das transcrições.

Quadro 2 – Dados das trilhas exportadas do ELAN

Trilhas			Transcrição
segmento_ de_fala	trecho_de_ interesse	classificacao	
S01	T01	<i>tristeza</i>	“Ah, eu sinto muito mal, porque a coisa que eu gosto de comer e de fazer é bolo, biscoito...”
S01	T02	<i>alegria</i>	“Dia de sábado que ajunta os neto, né”
S01	T03	<i>alegria</i>	“Já chega cobrando: “Vó, cadê o pão de queijo?” Aí eu não resisto...”
S01	T04	<i>aversão</i>	“300, 400, aí tem que tomar aquela insulina glandular...”
S01	T05	<i>alegria</i>	“Baixar mesmo não abaixa não, mas...”
S02	T06	<i>alegria</i>	“A coisa mais gostosa nesse mundo é poder comer gostoso”
S03	T07	<i>tristeza</i>	“Igual estou te falando, eu sinto tristeza de ter essa doença, porque a melhor...”
S03	T08	<i>tristeza</i>	“Isso não dá, isso não é nada bom”
S04	T09	<i>tristeza</i>	“E dá aquela crise de suadeira, durante a noite, tem dia que eu acordo assim molhadinha de suor, eu tenho que levantar e tomar um pouco de...”

O Quadro 2 apresenta transcrições que representam os três sentimentos observados. Por exemplo, no segmento 1 (S01), trecho 1 (T01), a entrevistada expressa-se negativamente em relação a sua condição crônica de diabetes no que diz respeito à impossibilidade de poder comer as coisas que ela considera mais saborosas. Já no segmento 1 (S01), trecho 3 (T03), a entrevistada afirma que os seus netos pedem insistentemente que ela cozinhe para eles, o que leva ao consumo de alimentos fora da dieta indicada pelo profissional da Saúde. Nesse momento, ela expressa *alegria* ao descrever o fato de cozinhar para os netos e, logo após, fala da tristeza que sente por ter que seguir uma dieta prescrita. Como apontado acima, os resultados dos dados analisados por meio da anotação de categorias podem ser examinados em relação aos dados demográficos e

clínicos do usuário para assim poder ser traçado um perfil do usuário, que pode oferecer indicações sobre o percurso mais ou menos bem-sucedido no cuidado de si.

Logo, a partir dos exemplos do Quadro 2, podemos observar certa regularidade na ocorrência dos sentimentos e de que forma eles acontecem no ambiente linguístico retratado pelas transcrições.

6 Conclusão

Este Capítulo teve como objetivo apresentar uma proposta de análise de dados não estruturados na forma de gravações de áudio e vídeo por meio do *software* ELAN, que permite sua análise, sem necessidade de se realizar a etapa de transcrição.

Primeiramente, discutimos a importância de se estudar a linguagem nas Ciências da Saúde. Em seguida, delimitamos um breve referencial teórico que nos permitiu definir o que são dados não estruturados e como estudá-los. Na sequência, apresentamos a proposta de metodologia, que é o foco deste Capítulo. E, por fim, exemplificamos essa proposta com base em um estudo de caso, realizado com dados do projeto Empoder@, mostrando alguns resultados que podem ser obtidos por meio da análise de gravações de áudio e vídeo.

Portanto, a abordagem metodológica proposta neste Capítulo mostra que as inovações tecnológicas têm muito a contribuir com o pesquisador da área das Ciências da Saúde, possibilitando que ele tenha uma visão mais abrangente da interação entre o usuário do sistema de Saúde e o profissional da área da Saúde. Dessa forma, esta metodologia pode

contribuir para estudos mais aprofundados da relação que os usuários têm com o profissional, com o pesquisador, com as eventuais condições crônicas (no caso deste Capítulo, o diabetes), com seus familiares e consigo mesmo. Nesse sentido, este Capítulo contribui à abordagem deste livro, que busca apresentar uma proposta para auxiliar o empoderamento do pesquisador, do profissional da área da Saúde e do usuário do sistema de Saúde.

Em suma, a interdisciplinaridade entre a Linguística, a Estatística e a Enfermagem permite alcançar novos patamares nas pesquisas realizadas no âmbito das ciências da saúde. Esperamos, assim, que outras abordagens, metodologias e ferramentas possam ser desenvolvidas para a geração de resultados ainda mais pertinentes para os usuários, profissionais e pesquisadores das Ciências da Saúde.

Síntese do capítulo

- O estudo da língua é muito importante para o empoderamento do usuário do serviço de Saúde, do profissional e do pesquisador.
- Dados não estruturados podem ser estudados de diferentes maneiras e, a partir da proposta apresentada neste Capítulo, vimos que é possível estudar gravações de áudio e vídeo usando ferramentas computacionais.
- Uma maior interação entre o ser humano e a máquina leva a uma investigação mais ampla das relações entre usuário, profissional e pesquisador.

Referências

AUER, E. et al. **ELAN as Flexible Annotation Framework for Sound and Image Processing Detectors** (N. Calzolari et al., Eds.) Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). **Anais...** Valletta, Malta: European Language Resources Association (ELRA), 2010

BARDIN, L. **Análise de Conteúdo**. São Paulo: Edições 70, 2011.

BARROSO, R. A. A. **Multidisciplinaridade e inovação metodológica na dinâmica de atuação do comitê de juízes na adaptação de instrumentos: Diabetes Empowerment Scale-Short Form (DES-SF)**. Dissertação (Mestrado em Enfermagem)—Belo Horizonte: Universidade Federal de Minas Gerais. Escola de Enfermagem, 2015.

BRUGMAN, H.; RUSSEL, A. **Annotating Multi-media/Multi-modal resources with ELAN** (M. T. Lino et al., Eds.) Proceedings of

- the 4th International Conference on Language Resources and Language Evaluation (LREC 2004). Paris: European Language Resources Association (ELRA), 2004
- CANDLIN, C. N.; CANDLIN, S. Health Care Communication: a problematic site for applied linguistics research. **Annual Review of Applied Linguistics**, v. 23, 2003.
- CECÍLIO, S. G. et al. A visão do companheiro da mulher com histórico câncer de mama. **Revista Mineira de Enfermagem**, v. 17, n. 1, p. 24–32, 2013.
- CHAVES, F. A. **Tradução e adaptação cultural do “Behavior Change Protocol” para a língua portuguesa-Brasil**. Dissertação (Mestrado em Enfermagem)—Belo Horizonte: Universidade Federal de Minas Gerais. Escola de Enfermagem, 2014.
- CRASBORN, O. et al. Combining video and numeric data in the analysis of sign languages within the ELAN annotation software. In: VETOORI, C. (Ed.). **Proceedings of the 2nd Workshop on the Representation and Processing of Sign languages: Lexicographic matters and didactic scenarios**. Paris: ELRA, 2006. p. 82–87.
- CRASBORN, O.; SLOETJES, H. Enhanced ELAN functionality for sign language corpora. In: **Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora**. Hamburg: LREC, 2008. p. 39–43.
- DAVID, G. F.; TORRES, H. DE C.; REIS, I. A. Atitudes dos profissionais de saúde nas práticas educativas em diabetes mellitus na atenção primária. **Ciência, Cuidado e Saúde**, v. 11, n. 4, p. 758–766, 2012.
- OLIVEIRA, J. L. C. et al. Qualidade do cuidado: concepções de graduandos de enfermagem. **Reme: Revista Mineira de Enfermagem**, v. 19, n. 1, p. 29–35, 2015.
- DREUW, P.; NEY, H. Towards Automatic Sign Language Annotation for the ELAN Tool. In: **Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora**. Hamburg: LREC, 2008.
- FIGUEREDO, G. P. **Introdução ao perfil metafuncional do português brasileiro: contribuições para os estudos multilíngues**. Tese (Doutorado em Linguística Aplicada)—Belo Horizonte: Universidade Federal de Minas Gerais. Faculdade de Letras, 2011.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. São Paulo: Atlas, 2008.
- GVOZD, R.; SAKAI, A. M.; HADDAD, M. DO C. L. Sentimentos e perspectivas de trabalhadores de instituição universitária pública frente à aposentadoria. **REME: Revista Mineira de Enfermagem**, v. 19, n. 1, p. 79–83, 2015.
- MATTHIESSEN, C. M. I. M.; HALLIDAY, M. A. K. **Systemic Functional Grammar: A First Step Into the Theory**. Beijing: Higher Education Press, 2009.
- MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS. **ELAN Linguistic Annotator**. Nijmegen: The Language Archive, 2015.
- NERY, I. S. et al. Abordagem da sexualidade no diálogo entre pais e adolescentes. **Acta Paulista de Enfermagem**, v. 28, n. 3, p. 287–292, 2015.
- ONWUDIWE, N. C. et al. Barriers to self-management of diabetes: a qualitative study among low-income minority diabetics. **Ethnicity & Disease**, v. 21, n. 1, p. 27–32, 2011.

- PERES, R. R. et al. Percepções de trabalhadores e estudantes atuantes em um pronto-socorro, sobre meio ambiente e saúde. **Reme: Revista Mineira de Enfermagem**, v. 18, n. 1, 2014.
- PLUTCHIK, R. A general psychoevolutionary theory of emotion. In: PLUTCHIK, R.; KELLERMAN, H. (Eds.). . **Emotion: Theory, research, and experience: Vol. 1. Theories of emotion**. New York: Academic press, 1980. p. 3–33.
- PLUTCHIK, R. The Nature of Emotions. **American Scientist**, v. 89, n. 4, p. 344, 2001.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation for Statistical Computing, 2015.
- REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. Barueri: Editora Manole Ltda, 2003.
- ROS, R.; BARONI, I.; DEMIRIS, Y. Adaptive human-robot interaction in sensorimotor task instruction: From human to robot dance tutors. **Robotics and Autonomous Systems**, v. 62, p. 707–720, 2014.
- SANTOS, A. C. C. et al. Percepção dos usuários de um centro de saúde acerca de sua participação no grupo de ginástica chinesa – Lian Gong: uma análise compreensiva. **Reme: Revista Mineira de Enfermagem**, v. 18, n. 1, 2014.
- SANTOS, L.; TORRES, H. DE C. Práticas Educativas em Diabetes Mellitus: compreendendo as competências dos profissionais de saúde. **Texto e Contexto - Enfermagem**, v. 21, n. 3, p. 574–580, 2012.
- VIEIRA, G. DE L. C. **Avaliação da educação em grupos operativos com usuários diabéticos tipo 2 em Unidades Básicas de Saúde - Belo Horizonte - MG**. Dissertação (Mestrado em Enfermagem)—Belo Horizonte: Universidade Federal de Minas Gerais. Escola de Enfermagem, 2011.
- ZINKEN, K. M.; CRADOCK, S.; SKINNER, T. C. Analysis System for Self-Efficacy Training (ASSET). Assessing treatment fidelity of self-management interventions. **Patient Education and Counseling**, v. 72, n. 2, p. 186–193, 2008.

ANÁLISE ESTATÍSTICA DESCRITIVA DE DADOS ESTRUTURADOS

Edna Afonso Reis
Ilka Afonso Reis

Objetivos do capítulo

- Fornecer as principais ferramentas para a análise estatística descritiva de dados estruturados;
- Conceituar os tipos de variáveis nos estudos estatísticos;
- Apresentar as tabelas e gráficos para a descrição da distribuição dos valores de uma variável;
- Introduzir as principais medidas de síntese numérica;
- Apresentar as tabelas e gráficos para o estudo da relação entre duas variáveis.

1 Introdução

Como visto no Capítulo 3, dados são a base para a geração de informação e, por consequência, do conhecimento. Sejam os dados estruturados ou não, sua coleta deve ser bem planejada e executada (Capítulo 4) e seu pré-processamento, se necessário, deve ser bem feito (Capítulos 5, 6 e 7) para que a análise desses dados possa revelar informações válidas.

Neste capítulo, vamos tratar da primeira análise pela qual deve passar todo conjunto de dados estruturados: a análise descritiva e exploratória.

Olhar para a extensa listagem de dados coletados não permite obter praticamente nenhuma conclusão, especialmente para grandes conjuntos de dados, com muitas características sendo investigadas. A Análise Estatística Descritiva é a fase inicial deste processo de estudo dos dados coletados.

Utilizamos métodos descritivos para organizar, resumir e descrever os aspectos importantes das características observadas. As ferramentas descritivas são os muitos tipos de tabelas e gráficos e também medidas de síntese como porcentagens e médias.

A descrição dos dados também tem como objetivo identificar anomalias, até mesmo resultantes do registro incorreto de valores, por exemplo, um dado de peso corporal igual a zero, e dados atípicos, aqueles que não seguem a tendência geral do restante do conjunto como um indivíduo com um dado de altura igual a 2,50m.

Neste Capítulo, vamos ilustrar as definições e ferramentas da Estatística Descritiva utilizando o conjunto de dados dos *indivíduos com diagnóstico de Diabetes Mellitus* (que chamaremos de *indivíduos DM*), apresentado no Capítulo 3.

2 Tipo de variáveis

Variável é a característica de interesse que é *medida* ou *observada* em cada indivíduo. Como o nome diz, seus valores *variam* de indivíduo para indivíduo. As variáveis podem ter valores numéricos ou não numéricos, definindo o que chamamos de variáveis *quantitativas* ou *qualitativas*, respectivamente.

As *variáveis quantitativas* são as características que podem ser medidas em uma escala quantitativa, ou seja, apresentam valores numéricos que fazem sentido. Podem ser *contínuas* ou *discretas*.

- **Contínuas:** Assumem valores em uma escala contínua, para as quais valores não inteiros (com casas decimais) fazem sentido. Usualmente devem ser medidas por meio de algum instrumento, como balança, régua, etc. Exemplos: peso, em gramas; altura, em centímetros; circunferência abdominal, em centímetros;

nível sanguíneo de glicohemoglobina, em porcentagem.

- **Discretas:** assumem um número finito ou infinito contável de valores, como as contagens. Exemplos: número de filhos, número de cigarros fumados por dia, idade em anos completos.

As *variáveis qualitativas* (ou *categóricas*) são as características que não possuem valores quantitativos, mas, ao contrário, são definidas por várias categorias, ou seja, representam uma classificação dos indivíduos. Podem ser *nominais* ou *ordinais*.

- **Nominais:** não existe (e não faz sentido existir) uma ordenação entre as categorias. Exemplos: estado do indivíduo (doente/sadio), sexo (feminino/masculino), hábito de fumar (fumante/não fumante); estado civil (solteiro, casado, separado, viúvo).
- **Ordinais:** existe uma ordenação natural entre as categorias. Exemplos: escolaridade (analfabeto, fundamental, médio, superior), estágio da doença (inicial, intermediário, terminal), mês de observação (janeiro, fevereiro, ... , dezembro).

Uma variável originalmente quantitativa pode ser coletada de forma qualitativa. Por exemplo, a variável *idade*, medida em anos completos, é quantitativa (discreta). No entanto, se for informada apenas a faixa etária (0 a 5 anos, 6 a 10 anos, etc.), ela passa a ser classificada como qualitativa ordinal.

Um ponto importante é que nem sempre uma variável representada por números é quantitativa, por exemplo, o número do telefone de uma pessoa, o número da sua casa ou o número de sua identidade. Às vezes, o sexo do indivíduo é registrado na planilha de dados como 1, se masculino, e 2, se feminino, por exemplo. Entretanto, isto não significa que a variável sexo passou a ser quantitativa.

Como visto no Capítulo 4, outra maneira de classificar as variáveis é quanto ao nível

de mensuração: nominal, ordinal, intervalar e razão. Os níveis de mensuração nominal e ordinal coincidem com a classificação aqui apresentada para as variáveis qualitativas. No entanto, uma variável quantitativa contínua, por exemplo, pode ser medida tanto no nível intervalar (ex.: temperatura corporal, para o qual o valor zero não indica a ausência de calor) quanto no nível razão (ex. altura de um indivíduo, para o qual o valor zero indica a ausência de altura).

A classificação das variáveis usando o seu nível de mensuração não é incomum nos textos sobre análise estatística e também é utilizada por alguns softwares para tratamento estatístico de dados. No entanto, para os propósitos deste Capítulo, utilizaremos a classificação primeiramente apresentada, ou seja, trabalharemos com variáveis nominais, ordinais, discretas e contínuas.

3 Tabelas e gráficos para o estudo da distribuição de frequências de uma variável

Os dados gerados por tipos diferentes de variável são de naturezas diferentes e, portanto, devem receber tratamentos diferentes. Nesta seção, vamos estudar separadamente as ferramentas descritivas – tabelas e gráficos – mais adequados para cada tipo de dado.

Iniciaremos esta apresentação com os dados de natureza qualitativa, que são os mais fáceis de tratar do ponto de vista da análise descritiva.

3.1 Variáveis qualitativas – nominais e ordinais

No exemplo do estudo com indivíduos DM, uma das duas variáveis qualitativas

presentes é o *estado civil*. Para organizar os dados provenientes de uma variável qualitativa, é usual fazer uma *tabela de frequências*, como a Tabela 1, na qual estão apresentadas as frequências com que ocorre cada uma das categorias de estado civil nos indivíduos DM observados. Cada categoria é representada em uma linha da tabela. Há uma coluna com as *contagens* de indivíduos DM em cada categoria (*frequência absoluta*) e outra com os *percentuais* que essas contagens representam no total de indivíduos DM (*frequência relativa*).

Como a variável estado civil é *qualitativa nominal*, ou seja, não há uma ordem entre as categorias, a ordem das linhas da tabela pode ser qualquer uma. É comum a disposição das linhas pela ordem decrescente das frequências das classes.

Tabela 1 – Distribuição de frequências dos indivíduos DM segundo estado civil

Estado Civil	Frequência Absoluta	Frequência Relativa
Casado	40	45%
Separado	28	31%
Solteiro	12	13%
Viúvo	10	11%
Total	90	100%

Quando a variável tabelada é qualitativa ordinal, as linhas da tabela de frequências devem ser dispostas na ordem existente para as categorias. A Tabela 2 mostra a distribuição de frequências dos indivíduos DM segundo a variável qualitativa ordinal *escolaridade*. Nesse caso, podemos acrescentar mais duas colunas, as *frequências acumuladas* (*absoluta e relativa*), que mostram a frequência de indivíduos DM observados até aquela categoria de escolaridade. Por exemplo, 49% dos indivíduos DM cursaram até o ensino fundamental.

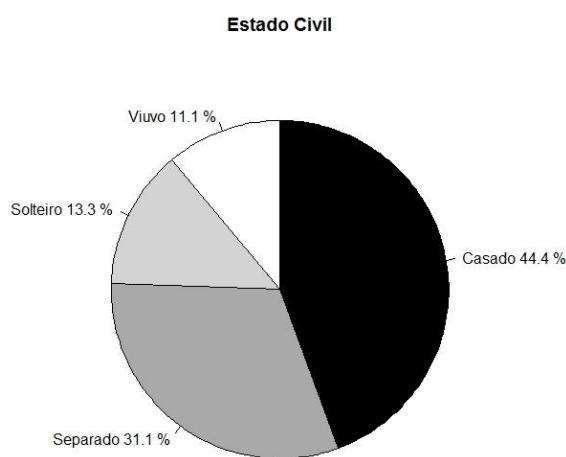
Tabela 2 – Distribuição de frequências dos indivíduos DM segundo escolaridade

Escolaridade	Frequências Simples		Frequências Acumuladas	
	Absoluta	Relativa	Absoluta	Relativa
Analfabeto	15	17%	15	17%
Ensino Fundamental	29	32%	44	49%
Ensino Médio	25	28%	69	77%
Ensino Superior	21	23%	90	100%
Total	90	100%	-	-

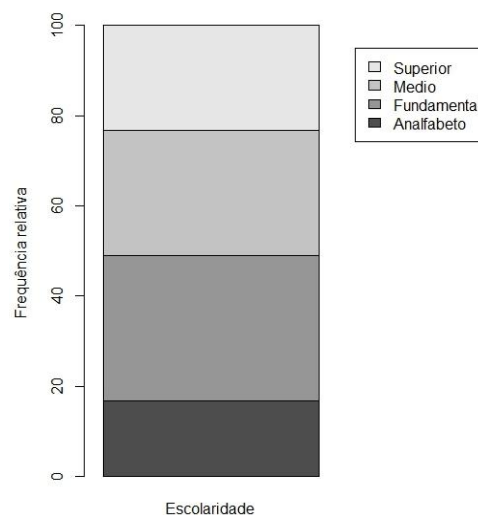
A visualização da distribuição de frequências de uma variável fica mais fácil se fizermos um gráfico a partir da tabela de frequências. Para as variáveis qualitativas, abordaremos dois tipos de gráficos: de *setores* e de *barras*.

O *gráfico de setores*, mais conhecido como gráfico de pizza ou torta, é construído dividindo-se um círculo (pizza) em setores (fatias), um para cada categoria da variável qualitativa, que serão proporcionais (em área) à frequência daquela categoria. A Figura 1(a) mostra um gráfico de setores para a variável estado civil, construído a partir da Tabela 1.

Uma alternativa ao gráfico de setores é o *gráfico de barras*, como o da Figura 1(b), construído para a variável *escolaridade*. Ao invés de dividirmos um círculo, dividimos uma barra. Note que, em ambos os gráficos, as frequências relativas das categorias devem somar 100%. Aliás, esta é a ideia dos gráficos: mostrar como se dá a divisão (distribuição) do total de elementos (100%) em partes (fatias). As vantagens da representação gráfica das distribuições de frequências ficam ainda mais evidentes quando há a necessidade de comparar vários grupos com relação a variáveis que possuem muitas categorias, como veremos mais adiante.



(a) Gráfico de setores para estado civil



(b) Gráfico de barra segmentada para escolaridade

Figura 1 – Representação gráfica da distribuição de frequências dos indivíduos DM, segundo estado civil e escolaridade

3.2 Variáveis quantitativas discretas

Quando estamos trabalhando com uma *variável discreta que assume poucos valores*, podemos dar a ela o mesmo tratamento dado às variáveis qualitativas ordinais, assumindo que cada valor é uma classe e que existe uma ordem natural nessas classes.

A Tabela 3 apresenta a distribuição de frequências dos indivíduos DM segundo seu número de filhos, que, nesse caso, assumiu apenas oito valores distintos. Podemos perceber que os indivíduos DM têm, em geral, dois ou três filhos (50%) e que cerca de três quartos deles (76%) deles têm até três filhos.

Tabela 3 – Distribuição de frequências dos indivíduos DM segundo número de filhos

Número de filhos	Frequência Absoluta		Frequência Acumulada	
	Simples	(Relativa)	Absoluta	(Relativa)
0	9	(10%)	9	(10%)
1	14	(16%)	23	(26%)
2	22	(24%)	45	(50%)
3	23	(26%)	68	(76%)
4	10	(11%)	78	(87%)
5	10	(11%)	88	(98%)
6	1	(1%)	89	(99%)
8	1	(1%)	90	(100%)
Total	90	(100%)	-	-

Quando trabalhamos com uma *variável discreta que pode assumir um grande número de valores distintos* como, por exemplo, a idade dos indivíduos (31 valores distintos), a

construção da tabela de frequências e de gráficos considerando cada valor como uma categoria fica inviável. A solução é agrupar os valores em classes, como mostra a Tabela 4.

Tabela 4 – Distribuição de frequências dos indivíduos DM segundo idade

Idade (anos completos)	Frequência Absoluta		Frequência Acumulada	
	Simples	(Relativa)	Absoluta	(Relativa)
45 a 49	4	(4%)	4	(4%)
50 a 54	11	(12%)	15	(16%)
55 a 59	15	(17%)	30	(33%)
60 a 64	26	(29%)	56	(62%)
65 a 69	19	(21%)	75	(83%)
70 a 74	9	(10%)	84	(93%)
75 a 80	6	(7%)	90	(100%)
Total	90	(100%)	-	-

O *gráfico de barra verticais* da Figura 2(a) mostra que os indivíduos estudados têm, principalmente, de dois a três filhos, embora haja uma grande variabilidade nessa variável.

Na Figura 2(b), podemos perceber que os valores de idade desses indivíduos se distribuem de maneira aproximadamente simétrica em torno da faixa de 60 a 64 anos.

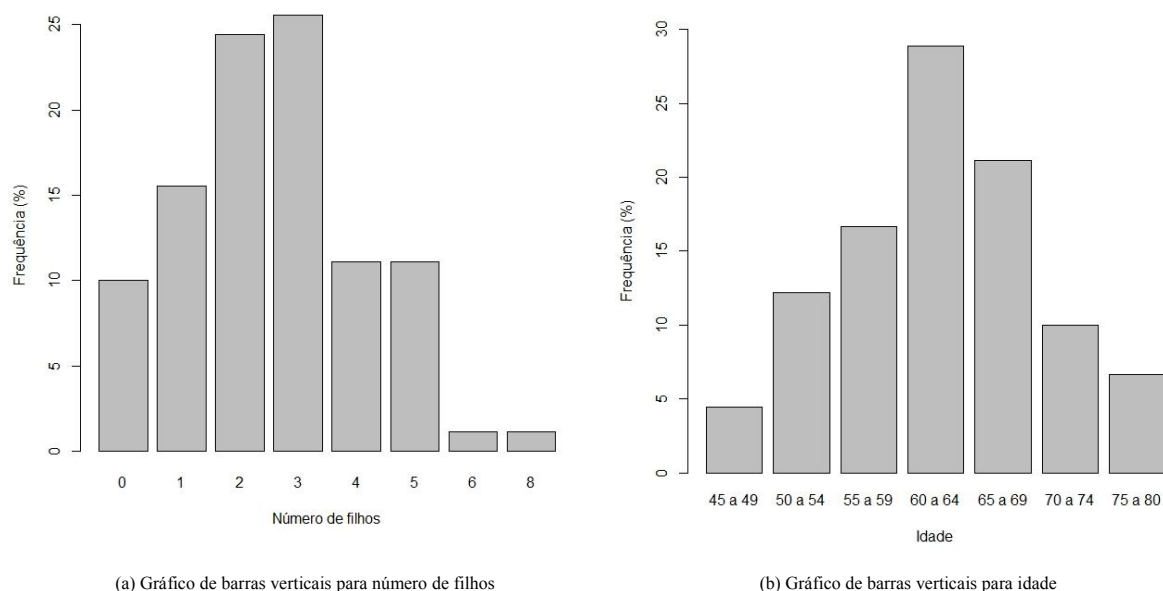


Figura 2 – Representação gráfica da distribuição de frequências dos indivíduos DM, segundo número de filhos e idade

3.3 Variáveis quantitativas contínuas

Quando a variável em estudo é do tipo contínua, que assume muitos (teoricamente, infinitos) valores distintos, o agrupamento

dos dados em classes será sempre necessário na construção das tabelas de frequências. A Tabela 5 apresenta a distribuição de frequências para o peso dos indivíduos DM do sexo masculino.

Tabela 5 – Distribuição de frequências dos indivíduos DM do sexo masculino segundo faixa de peso

Faixa de Peso (kg)	Frequência Absoluta		Frequência Acumulada	
	Simples	(Relativa)	Absoluta	(Relativa)
65 - 70	1	(2,4%)	1	(2,4%)
70 - 75	8	(19,0%)	9	(21,4%)
75 - 80	11	(26,2%)	20	(47,6%)
80 - 85	17	(40,5%)	37	(88,1%)
85 - 90	5	(11,9%)	42	(100%)
Total	42	(100%)	-	-

Os limites das classes são representados de modo diferente daquele usado nas tabelas para variáveis discretas: o limite superior de uma classe é igual ao limite inferior da classe seguinte. Mas, afinal, onde ele está incluído? O símbolo |- resolve essa questão. Na primeira classe (65 |- 70), por exemplo, estão incluídos todos os indivíduos com peso de 65,00 a 69,99 kg. Os indivíduos que pesarem exatos 70 kg serão incluídos na classe seguinte (70 |- 75).

A representação gráfica mais adequada para a distribuição de frequências de uma variável contínua é feita por meio de um gráfico chamado *histograma*, que nada mais é do que o gráfico de barras verticais, porém construído com as barras unidas devido ao caráter contínuo dos valores da variável.

A Figura 3 mostra o histograma do peso dos indivíduos DM, separadamente por sexo. Notamos que os indivíduos do sexo feminino

são, em geral, mais leves, pesando principalmente entre 65 e 75 kg, enquanto os indivíduos do sexo masculino pesam, em geral, entre 70 e 90 kg. Observamos, também, que o peso dos indivíduos do sexo feminino é mais homogêneo (valores mais próximos entre si) do que o

peso dos indivíduos do sexo masculino. Para esta comparação, os dois histogramas foram feitos com escala idêntica nos eixos de “peso” e de “frequência relativa”. Quando comparamos grupos de tamanhos diferentes, devemos sempre utilizar a frequência relativa.

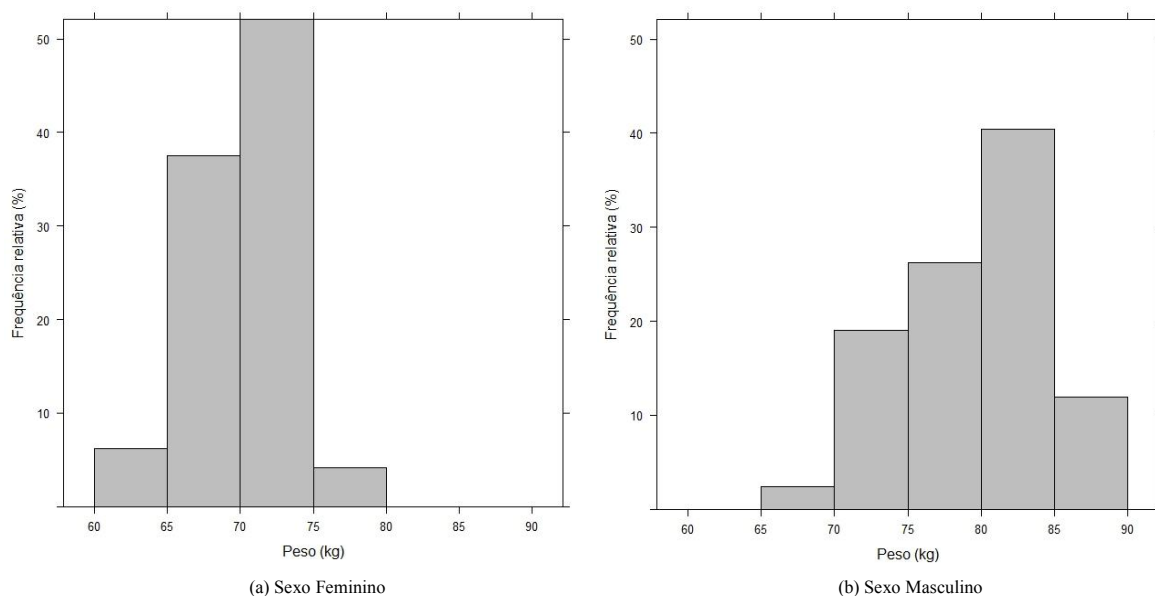


Figura 3 – Histograma do peso dos indivíduos DM segundo o sexo do indivíduo

4 Medidas de síntese numérica para variáveis quantitativas

Nesta seção, abordaremos as medidas de síntese numérica, que são usadas para resumir os dados de uma variável do tipo quantitativa. Serão discutidas as medidas de tendência central, variabilidade e ainda as medidas de posição.

4.1 Medidas de tendência central

A *tendência central* da distribuição de frequências de uma variável quantitativa é caracterizada pelo *valor típico* dessa variável. Essa é uma maneira de resumir os dados, pois escolheremos um valor para representar todos os outros. Esse resumo gera informação a partir dos dados, como discutido no Capítulo 3.

Assim, poderíamos perguntar, por exemplo, qual é a idade típica dos indivíduos DM do sexo feminino e compará-la com a idade típica dos indivíduos do sexo masculino. Fazer essa comparação utilizando medidas-resumo é bem mais sensato do que comparar os dois conjuntos de dados valor a valor, o que seria inviável.

Mas, como identificar o valor típico de um conjunto de dados? Vamos apresentar duas das medidas que podem ser utilizadas para descrever a tendência central de um conjunto de dados: a *média* e a *mediana*.

A *média aritmética simples* (que chamaremos apenas de *média*) é a medida de tendência central mais conhecida e usada para o resumo de dados. Essa popularidade pode ser devida à facilidade de cálculo e à ideia simples que ela nos sugere. De fato, se quisermos um

valor que represente, por exemplo, a altura dos indivíduos DM, por que não pegar as 90 medidas de altura na amostra, somar os valores e dividir esse “bolo” igualmente entre os 90 indivíduos? Essa é a ideia da *média*, que denotaremos por \bar{x} (leia-se “xis barra”):

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

onde $x_1, x_2, x_3, \dots, x_n$, são os valores da variável altura para cada um dos $n=90$ indivíduos da amostra.

Se, por exemplo, temos apenas os $n=6$ valores {3 ; 4,5 ; 5,5 ; 2,5 ; 1,3 ; 6 } de uma variável em uma amostra, a média destes valores será dada por

$$\bar{x} = \frac{3 + 4,5 + 5,5 + 2,5 + 1,3 + 6}{6} = \frac{22,8}{6} = 3,8.$$

Se esses seis valores representassem, por exemplo, a renda (em salários-mínimos) dos funcionários de um escritório, o salário típico nesse escritório seria de 3,8 salários-mínimos. Como estamos representando o valor típico pela média aritmética, podemos falar em *salário médio*. Um outro escritório, que apresentasse uma média de 4,2 salários-mínimos, teria funcionários com uma renda típica maior.

Outra medida do valor típico de uma variável quantitativa é dada pela mediana. A *mediana* de um conjunto de dados é definida como sendo o “valor do meio” desse conjunto de dados, dispostos em ordem crescente, deixando metade dos valores *acima* dela e metade dos valores *abaixo* dela.

Como calcular a mediana? Basta seguir sua definição. Vejamos:

<ul style="list-style-type: none"> • n é ímpar: 	<p>Existe apenas um “valor do meio”, que é a mediana. Seja o conjunto de dados (2 ; 3,3 ; 2,5 ; 5,6 ; 5 ; 4,3 ; 3,2). Ordenando os valores, temos (2 ; 2,5 ; 3,2 ; 3,3 ; 4,3 ; 5 ; 5,6). O valor do meio é o 3,3 . A mediana é o valor 3,3.</p>
<ul style="list-style-type: none"> • n é par: 	<p>Existem dois “valores do meio”. A mediana é média aritmética simples deles. Seja o conjunto de dados (3 ; 4,5 ; 5,5 ; 2,5 ; 1,3 ; 6). Ordenando os valores, temos (1,3 ; 2,5 ; 3 ; 4,5 ; 5,5 ; 6). Os valores do meio são 3 e 4,5. A mediana é $(3 + 4,5)/2 = 3,75$.</p>

Como medida de tendência central, a mediana é mais intuitiva do que a média, pois representa, de fato, o centro (meio) do conjunto de valores ordenados. Entretanto, a média é a medida-resumo muito mais utilizada. Existem várias razões para essa popularidade da média. Entre elas, estão a facilidade de tratamento estatístico e algumas propriedades interessantes que a média apresenta quando precisamos usá-la para estender os resultados da amostra para a população de onde ela veio (inferência estatística).

No entanto, a média é uma medida muito influenciada pela presença de valores extremos

em um conjunto de dados (valores muito grandes ou muito pequenos em relação aos demais). Como a média usa os valores de cada observação em seu cálculo, esses valores extremos “puxam” o valor da média em direção a si. Isto desloca a representação do centro, que já não será tão central como deveria ser. A mediana, por sua vez, não é tão influenciada por valores extremos, pois o que utilizamos para calculá-la é a ordem dos elementos e não diretamente seus valores. Assim, se um elemento do conjunto de dados tem o seu valor alterado (um erro, por exemplo), mas sua ordem continua a mesma, a mediana não sofre influência nenhuma.

No exemplo dos salários dos seis funcionários de uma empresa, se o valor do maior salário fosse alterado de 6 para 24 salários-mínimos, a média subiria de 3,8 para 6,8 salários mínimos e não seria uma medida-resumo adequada para representar os salários pagos nessa empresa. A mediana, entretanto, não seria afetada por este valor discrepante e permaneceria em 3,75 salários mínimos.

4.2 Medidas de variabilidade

As medidas de tendência central conseguem resumir, em um único número, o valor que é “típico” no conjunto de dados. Mas, será que, somente com essas medidas, conseguimos descrever adequadamente o que ocorre em um conjunto de dados?

Vejamos um exemplo: quando pesamos algo em uma balança, esperamos que ela nos dê o verdadeiro peso daquilo que estamos pesando. No entanto, se fizermos várias medições do peso de um mesmo objeto em uma mesma balança,

teremos *diferentes* valores para o peso desse objeto. Ou seja, existe *variabilidade* nas medições de peso fornecidas pela balança. Nesse caso, quanto menor a variabilidade desses valores, mais precisa é a balança (considerando que a média das medidas de peso coincida com o seu valor real).

Observe a Figura 4, na qual estão representadas as distribuições de frequências das medições do peso de uma esfera de 1000 g, feitas por duas balanças (A e B). As duas balanças registram o mesmo *peso médio de 1000 g* (média dos pesos de todas as medições feitas). Isto é, as duas balanças tipicamente acertam o verdadeiro peso da esfera. Porém, pela Figura 4, podemos notar que

- As medições da balança A *variam pouco* em torno de 1000 g: entre cerca de 950 g e 1050 g (uma “dispersão” de 100 g);
- As medições da balança B *variam muito* em torno de 1000 g: entre 900 g e 1100 g (uma “dispersão” de 200 g).

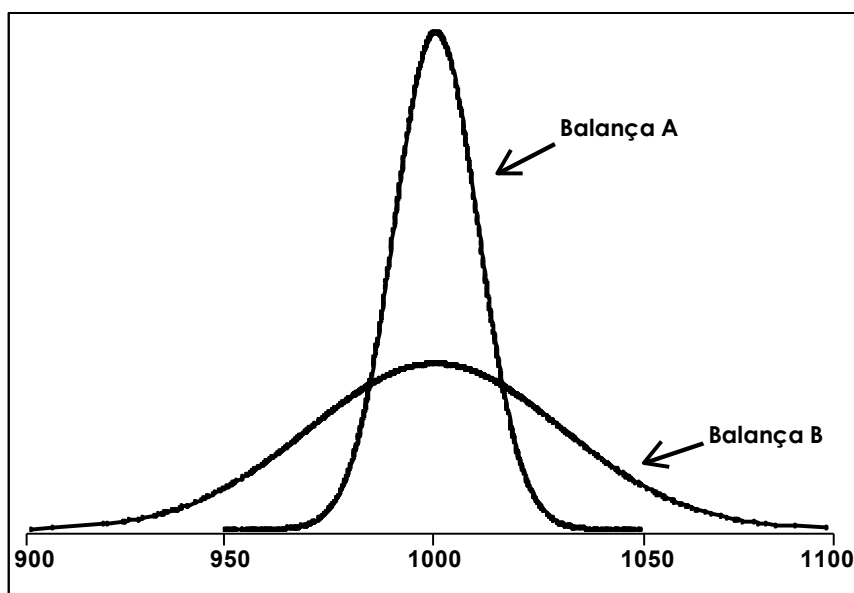


Figura 4 – Representação esquemática da distribuição de frequências das medições repetidas do peso da esfera de 1000 gramas feitas nas balanças A e B

Fonte: Reproduzida de Reis e Reis (2002) com autorização dos autores

Dois conjuntos de dados podem ter a mesma medida de centro (valor típico), porém com uma dispersão diferente em torno desse valor. Desse modo, além de uma medida que nos diga qual é o valor “típico” do conjunto de dados, precisamos de uma medida do *grau de dispersão (variabilidade)* dos dados em torno do valor típico. O objetivo das medidas de variabilidade é quantificar esse grau de dispersão.

A medida de variabilidade mais simples é a chamada *amplitude total (AT)*, que é a diferença entre os valores máximo e o mínimo observados da variável. No exemplo das balanças, as medições de peso da esfera apresentaram $AT = 1100 - 900 = 200$ gramas na balança A e $AT = 1050 - 950 = 100$ gramas na balança

B, mostrando que a variabilidade das medições de peso foi maior na balança A.

Embora seja uma medida simples de variabilidade, a amplitude total é um tanto grosseira, pois depende somente de dois valores do conjunto de dados (máximo e mínimo), não captando o que ocorre com os outros valores. Vejamos o exemplo mostrado na Figura 5, no qual uma mesma variável foi medida em dois conjuntos de dados. A variabilidade é claramente maior no segundo conjunto de dados. No entanto, se compararmos a dispersão dos valores usando apenas a amplitude total (igual a 50 em ambos casos), concluiríamos que os dois conjuntos de dados são igualmente dispersos nesta variável.

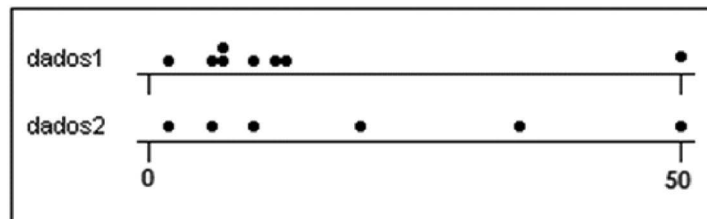


Figura 5 – Dois conjuntos de dados com mesma amplitude total

Fonte: Reproduzida de Reis e Reis (2002) com autorização dos autores

Uma boa medida de dispersão deve considerar todos os valores do conjunto de dados e resumir o grau de dispersão desses valores em torno do valor típico. Considerando a média como a medida de tendência central, podemos pensar em medir a dispersão (desvio) de cada valor do conjunto de dados em relação a ela. A medida mais simples de desvio entre duas quantidades é a diferença entre elas. Assim, para cada valor X_i , teremos o seu desvio em relação à média \bar{X} , representado por $(X_i - \bar{X})$.

Por exemplo, no conjunto de dados $\{1; 1; 2; 3; 4; 4; 5; 6; 7; 7\}$, relativo ao número de filhos de 10 mulheres, temos que a média é $\bar{X} = 4$ filhos. A coluna 1 da Tabela 6 mostra esses 10 valores e a coluna 2 mostra o desvio de cada um deles até a média. Como temos

um desvio para cada elemento, poderíamos pensar em resumi-los em um desvio típico, a exemplo do que fizemos com a média. Porém, quando somarmos esses desvios para o cálculo do desvio médio, a soma será *sempre* zero, como mostrado na última linha da coluna 2 da Tabela 6. Isto ocorre com qualquer conjunto de dados, pois os desvios negativos sempre compensarão os positivos. No entanto, os sinais dos desvios não são importantes para nossa medida de dispersão, já que estamos interessados na quantidade de dispersão e não na direção dela. Portanto, eliminaremos os sinais elevando os desvios ao quadrado, como mostrado na coluna 3 da Tabela 6. A soma desses desvios ao quadrado pode ser, então, dividida entre os participantes do “bolo”. Na verdade, por razões absolutamente teóricas e que

fogem ao escopo deste Capítulo, dividiremos essa soma pelo total de participantes menos 1 (n-1). Assim, usando a notação definida anteriormente, teremos

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Para os dados da Tabela 6, teremos $46 / (10-1) = 5,11$. Esse valor pode ser visto como uma *quase* média dos desvios ao quadrado e é chamado de *variância*.

Tabela 6 – Exemplo de cálculo do desvio-padrão

Coluna 1	Coluna 2	Coluna 3
X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	-3	9
1	-3	9
2	-2	4
3	-1	1
4	0	0
4	0	0
5	1	1
6	2	4
7	3	9
7	3	9
Soma	-	46

A variância seria nossa medida de variabilidade se não fosse o fato de que ela está

expressa em uma unidade diferente da unidade dos dados. De fato, ao elevarmos os desvios ao quadrado, elevamos também as unidades de medida em que eles estão expressos. No caso dos dados da Tabela 6, expressos em número de filhos, a variância vale 5,11 “filhos ao quadrado”, algo que não faz nenhum sentido. Para eliminar esse problema, extraímos a raiz quadrada da variância e, finalmente, temos a nossa medida de variabilidade, que chamaremos *desvio-padrão* (DP ou *s*).

$$DP = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

O desvio-padrão, como o nome já diz, representa o desvio típico (padrão) dos dados em relação à média, que foi escolhida como medida de tendência central. No exemplo, temos que o desvio-padrão vale 2,26 filhos. Isto significa que a distância típica (padrão) de cada mãe até o número médio de filhos (4 filhos) é de 2,26 filhos. O desvio-padrão nos permite distinguir numericamente conjuntos de dados de mesmo tamanho, mesma média, mas que são visivelmente diferentes, como mostra a ilustração da Figura 6: quanto maior a variabilidade dos valores em torno da mesma média, maior é o desvio-padrão.

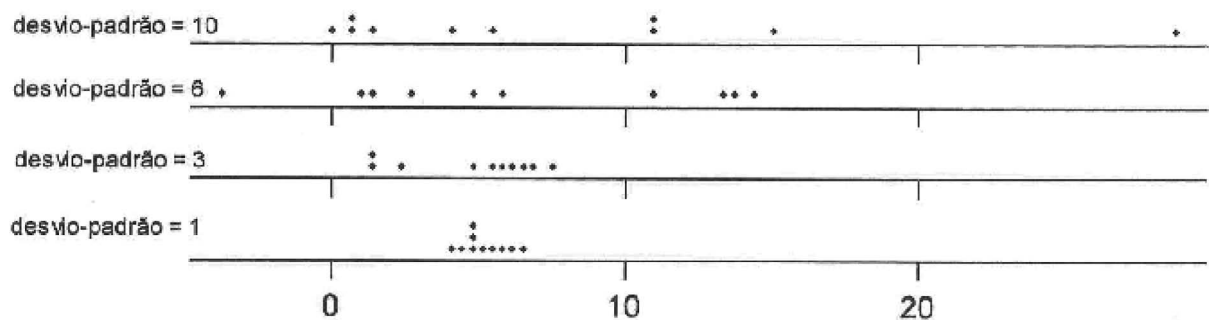


Figura 6 – Conjuntos de dados de mesmo tamanho e com a mesma média, mas com variabilidades diferentes

Fonte: Reproduzida de Reis e Reis (2002) com autorização dos autores

Ao analisarmos o grau de dispersão de um conjunto de dados, poderemos nos deparar com uma questão do tipo: um desvio-padrão de 10 unidades é pequeno ou grande? Se estivermos trabalhando com um conjunto de dados cuja média é 1000, um desvio típico de 10 unidades em torno dessa média significa pouca dispersão. Entretanto, se a média for igual a 100, um desvio típico de 10 unidades em torno dessa média significa muita dispersão. Assim, antes de responder se um desvio-padrão de 10 unidades é grande ou pequeno, devemos avaliar sua magnitude em relação à média: no primeiro caso, o desvio-padrão corresponde a 1% da média; no segundo caso, o desvio-padrão corresponde a 10% da média.

A essa razão entre o desvio-padrão e a média damos o nome de *coeficiente de variação*¹:

$$CV = \frac{\text{desvio padrão}}{\text{média}}$$

Quanto menor o coeficiente de variação de um conjunto de dados, menor é a sua variabilidade. O Coeficiente de Variação expressa o quanto da escala de medida, representada pela média, é ocupada pelo desvio-padrão. O Coeficiente de Variação é uma medida adimensional, isto é, não depende da unidade de medida. Essa característica nos permite usá-lo para comparar a variabilidade de variáveis medidas em unidades diferentes, o que não é possível usando a amplitude total nem desvio-padrão. Na Tabela 7, apresentamos o cálculo do coeficiente de variação para quatro variáveis quantitativas da amostra de indivíduos DM, mostrando que eles são mais homogêneos nas variáveis altura e peso (menores CV) e mais heterogêneos em idade e circunferência abdominal (maiores CV).

¹ Em algumas áreas, o Coeficiente de Variação é chamado de Desvio-Padrão Relativo (DPR).

Tabela 7 – Média, desvio-padrão (DP) e coeficiente de variação (CV) dos indivíduos DM nas variáveis idade, altura, peso e circunferência abdominal

Variável	Média	DP	CV
Idade	62,5 anos	7,5 anos	12%
Altura	164,8 cm	6,1 cm	4%
Peso	74,3 kg	6,6 kg	9%
Circunferência Abdominal	96,7 cm	10,2 cm	10%

Quanto menor o coeficiente de variação, mais homogênea é a variável naquela população. A definição do que pode ser considerado pouco ou muito homogêneo segundo o coeficiente de variação varia de acordo com a área de estudo da qual provêm os dados. Em geral, um coeficiente de variação menor de que 0,25 indica uma variável homogênea. Em populações para as quais já se espera uma variabilidade maior entre os indivíduos, essas faixas de homogeneidade devem ser redefinidas. Depois de alguma experiência analisando dados de uma população, consegue-se definir as faixas de homogeneidade apropriadas.

4.3 Medidas de posição

Quando falamos de *posição* ou *colocação* de um indivíduo em uma corrida ou em um teste como o vestibular, frequentemente nos referimos ao seu *posto*, como 1º, 2º, 3º, 29º ou último lugar. Mas, para sabermos se uma dada colocação é ou não um bom resultado, precisamos informar quantos indivíduos estão sendo ordenados. As duas medidas de posição que veremos aqui, os *percentis* e os *escores padronizados*, solucionam este e outros problemas de posicionamento (*ranking*). Estas duas medidas de posição podem ser usadas para comparar a posição do indivíduo em relação ao seu grupo considerando uma mesma variável ou variáveis diferentes.

4.3.1 Percentil

A posição de um indivíduo no conjunto de dados é mostrada pelo *percentil* contando-se (em porcentagem) quantos indivíduos do conjunto têm valores *menores ou iguais* a valor deste indivíduo. Desse modo, definimos:

O percentil de ordem K (onde K é qualquer valor entre 0 e 100), denotado por P_k , é o valor tal que K% dos valores do conjunto de dados são menores ou iguais a ele.

Por exemplo, na amostra dos indivíduos DM, o percentil de ordem 10 da variável altura é $P_{10} = 157,3$ cm. Isso significa que 10% dos indivíduos medem até 157,3 cm; na variável peso, temos que $P_{95} = 85,7$ kg, ou seja, 95% dos indivíduos pesam 85,7 kg ou menos (o que equivale a dizer que 5% dos indivíduos pesam mais que 85,7 kg).

Os percentis de ordem 10, 20, 30, ... 90 dividem o conjunto de dados em dez partes com mesmo número de observações (10%) e são chamados de decis (1º decil, 2º decil, 3º decil, ..., 9º decil). Os percentis de ordem 25, 50 e 75 dividem o conjunto de dados em quatro partes com o mesmo número de observações (25%). Assim, esses três percentis recebem o nome de *quartis*, *primeiro quartil* (Q_1), *segundo quartil* (Q_2) e *terceiro quartil* (Q_3), respectivamente. O segundo quartil é a nossa conhecida mediana.

No exemplo dos indivíduos DM, o indivíduo ID1 tem 158,9 cm de altura, pesa 69,3 kg e tem circunferência abdominal igual a 77,9 cm, que correspondem, respectivamente, aos percentis de ordem 15, 20 e 2. Ou seja, este indivíduo está entre os 15% mais baixos, entre os 20% mais leves e apenas 2% dos indivíduos têm circunferência abdominal menores ou iguais à dele.

4.3.2 Escore padronizado

O *escore padronizado* mostra a posição do indivíduo em relação ao valor típico da variável (dado pela média) de todos os indivíduos, considerando também a variabilidade de tais medidas (dada pelo desvio-padrão):

$$\text{Escore Padronizado} = \frac{\text{escore original} - \text{média}}{\text{desviopadrão}}$$

onde média e desvio-padrão são calculados para cada variável usando os escores originais de todos indivíduos naquela variável. Assim, o escore padronizado mede a distância do escore original à média em *número de desvios-padrão*, o que o torna uma medida comparável entre variáveis diferentes.

No exemplo indivíduos DM, o indivíduo ID1 tem escore padronizado de $(158,9 - 164,8)/6,1 = -0,97$ na variável *altura*, e de $(69,3 - 74,3)/6,6 = -0,76$ na variável *peso*. Desse modo, esse indivíduo está mais distante do grupo na variável *altura* (0,97 desvios-padrões abaixo da média) do que na variável *peso* (0,76 desvios-padrões abaixo da média).

Como medida de posição, a vantagem dos escores padronizados em relação aos percentis é que os escores padronizados de um indivíduo em várias variáveis podem ser combinados em um único valor (por soma ou média, por exemplo) para gerar uma posição *global* do indivíduo no conjunto de dados. Além disso, o escore padronizado leva em conta explicitamente a variabilidade do grupo, enquanto o percentil não o faz.

4.4 O gráfico box-plot

O *box-plot* (gráfico de caixa) é um gráfico constituído por uma caixa atravessada por uma linha e é construído usando-se um eixo com uma escala de valores, como mostra a Figura 7. O fundo da caixa é marcado na escala

de valores na altura do primeiro quartil (Q1). O topo da caixa é marcado na altura do terceiro quartil (Q3). Uma linha é traçada dentro da caixa na altura da mediana, que não precisa estar necessariamente no meio da caixa.

Como sabemos, entre o primeiro e o terceiro quartis, temos 50% dos dados. Podemos pensar, então, que essa caixa contém metade dos dados do conjunto.

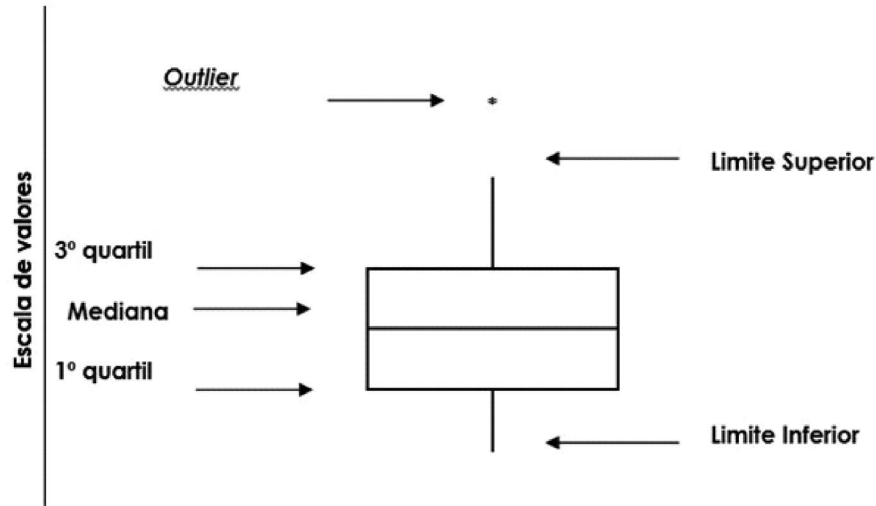


Figura 7 – Representação esquemática do box-plot.

Fonte: Reproduzida de Reis e Reis (2002) com autorização dos autores

A altura da caixa é dada por $(Q3-Q1)$, que é denominada distância interquartílica (DQ). Finalmente, para representar os 25% dos valores que são inferiores ao Q1, uma linha é traçada a partir da base da caixa (Q1) e estendida até, no máximo, 1,5 vezes o valor de DQ (linha inferior). Do mesmo modo, para representar os 25% dos valores que são superiores ao Q3, uma linha é traçada a partir da tampa da caixa (Q3) e estendida até, no máximo, 1,5 vezes o valor de DQ (linha superior). Os valores do conjunto de dados que forem menores do que a linha inferior ou maiores do que a linha superior são marcados com um asterisco. Esses são os valores identificados como discrepantes.

A Figura 8 mostra os *box-plots* da variável altura dos indivíduos DM, construídos separadamente para os três grupos de intervenção (A, B e C). Note que, apenas no grupo B, há um valor discrepante de altura.

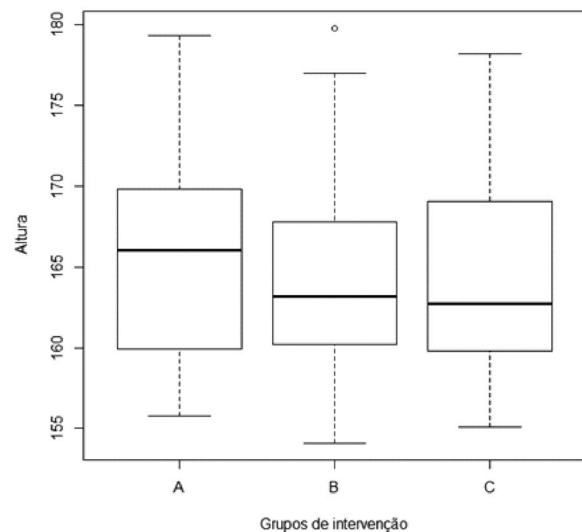


Figura 8 – *Box-plots* para a variável altura dos indivíduos DM, segundo grupo de intervenção

Podemos representar vários *box-plots* em uma mesma figura e comparar a tendência central (mediana) e a variabilidade (pela DQ: quanto maior, mais dispersão) de dois ou mais conjuntos de dados na mesma variável. Na Figura 8, vemos que os indivíduos do grupo A

são tipicamente mais altos (maior mediana) e têm uma variabilidade maior (maior DQ).

O *box-plot* foi inicialmente proposto para a detecção de valores discrepantes (*outliers*), que são aqueles valores muito diferentes do restante do conjunto de dados. Valores discrepantes podem ser indicativo de erro no registro dos dados e devem ser investigados antes de prosseguirmos com a análise.

Para finalizar essa seção, é importante destacar que as medidas de síntese numérica que acabamos de ver só fazem sentido para variáveis do tipo *quantitativas*. Mesmo que uma variável qualitativa (categórica) esteja *rotulada* com números na planilha de dados, o cálculo de medidas de síntese numérica não terá nenhum sentido.

5 Descrevendo a relação entre duas variáveis

Até agora, apresentamos as ferramentas estatísticas para estudo descritivo de cada variável de estudo *individualmente*. Nesta seção, vamos mostrar como estudar duas variáveis *simultaneamente*, verificando se existe ou não uma relação entre elas.

5.1 Duas variáveis qualitativas

Na Tabela 1, mostramos a distribuição de frequências de todos 90 indivíduos DM

segundo uma única variável, seu *estado civil*, e, na Tabela 2, sua *escolaridade*. Essa amostra pode ainda ser classificada segundo as duas variáveis simultaneamente, *estado civil* e *escolaridade*, gerando a Tabela 8(a).

Uma tabela como a Tabela 8(a) é chamada de *Tabela de Contingência*, na qual as frequências correspondem à classificação dos indivíduos simultaneamente em duas variáveis qualitativas: uma classificação está nas linhas da tabela e a outra está nas colunas. Observe que, se considerarmos somente a última linha da Tabela 8(b), teremos uma tabela igual à Tabela 1 e, se considerarmos somente a última coluna da Tabela 8(c), teremos uma tabela igual à Tabela 2.

A Tabela 8(b) mostra a distribuição de frequências da variável estado civil em cada uma das categorias da variável escolaridade (percentuais calculados usando os totais de linha). Com ela, podemos concluir, por exemplo, que a frequência de pessoas solteiras é menor entre as pessoas analfabetas (7%) do que entre as pessoas com ensino fundamental.

A Tabela 8(c) mostra a distribuição de frequências da variável escolaridade em cada uma das categorias da variável estado civil (percentuais calculados usando os totais de coluna). Com ela, podemos concluir, por exemplo, que a frequência de pessoas com curso superior é maior entre as pessoas viúvas (30%) do que entre as pessoas casadas (20%).

Tabela 8 - Distribuição de frequências dos indivíduos DM segundo escolaridade e estado civil

(a) Frequências absolutas					
Escolaridade	Estado Civil				Total
	Casado	Separado	Solteiro	Viúvo	
Analfabeto	7	5	1	2	15
Fundamental	11	10	5	3	29
Médio	14	6	3	2	25
Superior	8	7	3	3	21
Total	40	28	12	10	90

(b) Frequências relativas aos totais das linhas					
Escolaridade	Estado Civil				Total
	Casado	Separado	Solteiro	Viúvo	
Analfabeto	47%	33%	7%	13%	100%
Fundamental	38%	35%	17%	10%	100%
Médio	56%	24%	12%	8%	100%
Superior	38%	34%	14%	14%	100%
Total	44,5%	31,1%	13,3%	11,1%	100%

(c) Frequências relativas aos totais das colunas					
Escolaridade	Estado Civil				Total
	Casado	Separado	Solteiro	Viúvo	
Analfabeto	17,5%	18%	8%	20%	17%
Fundamental	27,5%	36%	42%	30%	32%
Médio	35%	21%	25%	20%	28%
Superior	20%	25%	25%	30%	23%
Total	100%	100%	100%	100%	100%

*Alguns arredondamentos foram feitos para que a soma totalizasse 100%.

Os dados organizados em uma tabela de contingência podem ser representados graficamente por meio dos gráficos de barras segmentadas, como os que são apresentados na Figura 9. Cada categoria de uma variável é representada por uma barra. As barras são segmentadas de acordo com a frequência relativa das categorias da outra variável. Como mencionado anteriormente, o gráfico de barra segmentada é uma alternativa ao

gráfico de setores (“pizza”). Entretanto, quando queremos comparar várias categorias de uma variável em relação à distribuição das categorias de outra variável, o gráfico de barras segmentadas é mais indicado, pois permite que várias barras sejam colocadas lado a lado na mesma figura. Se optássemos pelo gráfico de setores, seriam necessários quatro gráficos para fazer a representação que está na Figura 9(a), por exemplo.

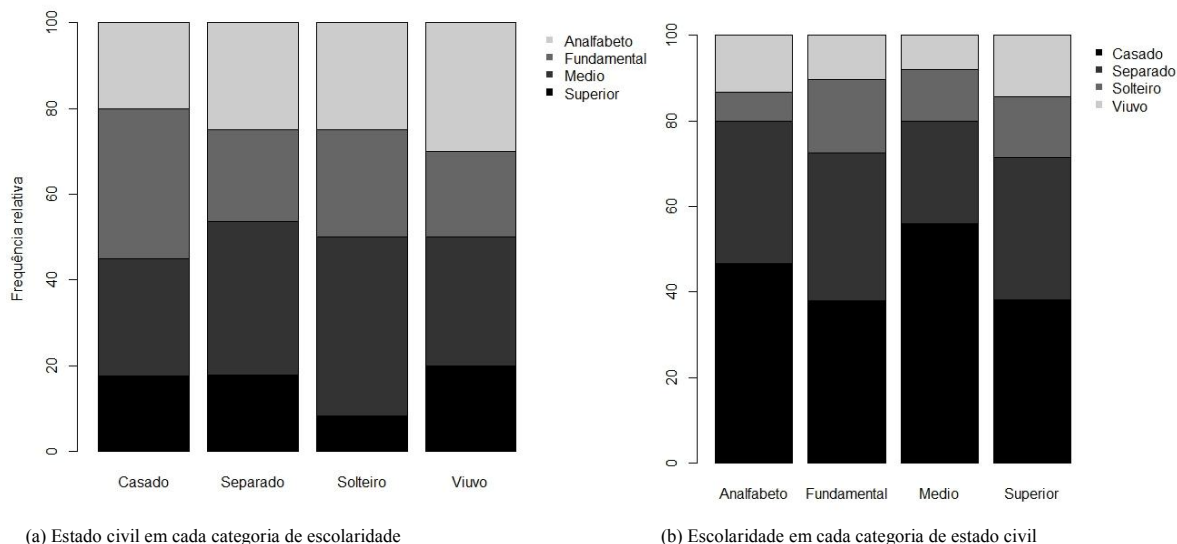


Figura 9 – Gráfico de barras segmentadas da distribuição de frequências dos indivíduos DM segundo (a) e (b)

5.2 Duas variáveis quantitativas

Para visualizar a relação entre as duas variáveis quantitativas, utilizamos o *diagrama de dispersão*, que é um gráfico onde pontos no espaço cartesiano XY são usados para representar simultaneamente os valores de duas variáveis quantitativas medidas em cada indivíduo do conjunto de dados.

A Figura 10 mostra um esquema do desenho do diagrama de dispersão. Nesse exemplo, foram medidos os valores de duas variáveis quantitativas, X e Y, em quatro indivíduos. O eixo horizontal do gráfico representa a variável X e o eixo vertical representa a variável Y.

Indivíduos	Variável X	Variável Y
A	2	3
B	4	3
C	4	5
D	8	7

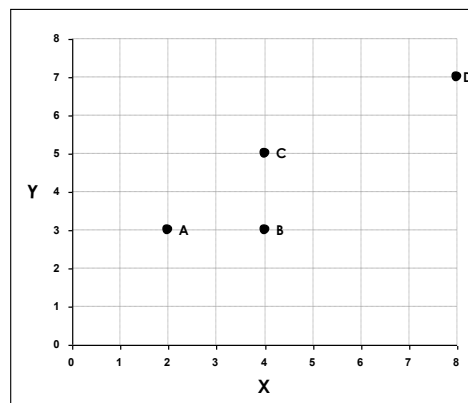


Figura 10 – Ilustração do diagrama de dispersão com valores de quatro indivíduos nas variáveis fictícias X e Y

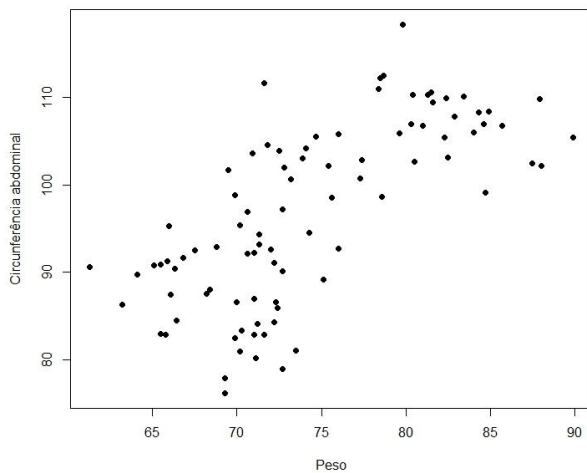
Fonte: Reproduzida de Reis e Reis (2002) com autorização dos autores

Na Figura 11(a), é apresentado o diagrama de dispersão das variáveis peso versus circunferência abdominal para os indivíduos DM. Verificamos que, quanto maior o peso

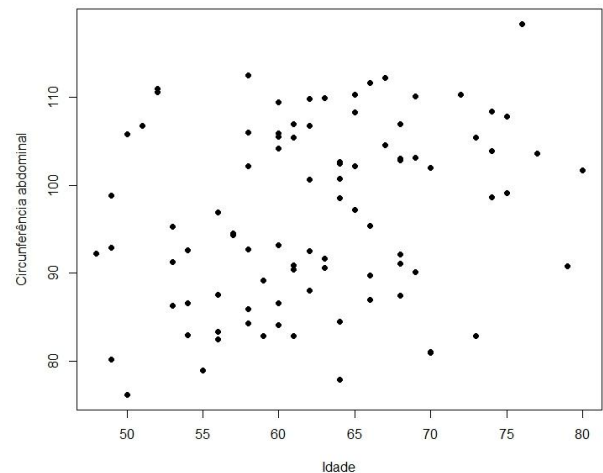
do indivíduo, maior tende a ser sua medida de circunferência abdominal: os indivíduos mais leves, com peso abaixo de 75 kg, têm circunferência abdominal medindo basicamente entre

80 e 105 cm, enquanto os indivíduos mais pesados (peso acima de 75 kg) estão em uma faixa mais elevada da circunferência abdominal (entre 90 e 110 cm, no geral). Como os pontos estão dispostos como se “tentassem seguir” uma reta crescente, dizemos que parece haver uma relação linear crescente (ou positiva) entre essas duas variáveis.

Por outro lado, quando construímos um diagrama de dispersão com as variáveis idade versus circunferência abdominal (Figura 11(b)), não vemos relação alguma entre elas: os valores de circunferência abdominal se distribuem basicamente entre 80 e 110 cm em toda a extensão dos valores de idade.



(a) Variáveis peso e circunferência abdominal



(b) Variáveis idade e circunferência abdominal

Figura 11 – Diagrama de dispersão para variáveis no exemplo dos indivíduos DM

O diagrama de dispersão também é muito útil para comparar o efeito de dois tratamentos no mesmo indivíduo e para verificar o efeito tipo antes/depois de um tratamento. Um exemplo é mostrado na Figura 12, onde cada ponto no diagrama de dispersão corresponde ao valor da glicohemoglobina de um indivíduo DM, medida antes (eixo horizontal) e depois da intervenção (eixo vertical). A linha marcada no diagrama corresponde à situação onde a glicohemoglobina não se alterou depois do indivíduo passar pela intervenção. Veja que a maioria dos pontos estão abaixo desta linha, ou seja, para a maioria dos indivíduos, houve uma diminuição do valor da glicohemoglobina depois da intervenção. Grande parte desses pontos está bem distante da linha, mostrando que a redução no valor da glicohemoglobina depois da intervenção foi grande.

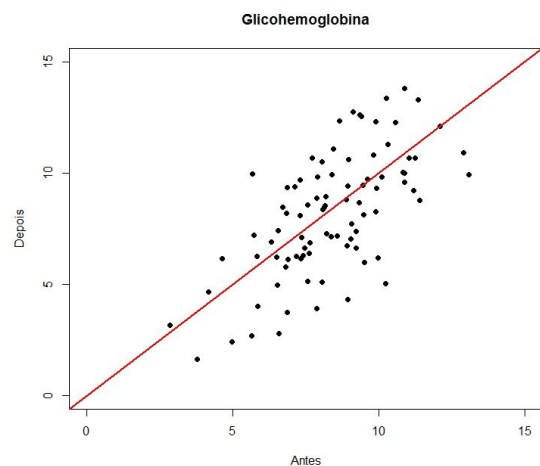


Figura 12 – Diagrama de dispersão da glicohemoglobina nos indivíduos DM, medida antes e depois da intervenção

6 Considerações finais

A análise descritiva deve ser a primeira etapa de qualquer análise de dados, precedendo

a etapa de Inferência Estatística, na qual os resultados obtidos na amostra são estendidos para a população de interesse.

Neste Capítulo, foram abordadas algumas das técnicas de Estatística Descritiva existentes. Com as ferramentas aqui apresentadas, é possível fazer uma análise descritiva e exploratória dos tipos de dados mais comuns na área das Ciências da Saúde. Todos os gráficos mostrados neste Capítulo foram criados no ambiente de programação estatística R, que pode ser obtido gratuitamente no website <https://cran.r-project.org/>. O R é uma ferramenta poderosa para análise estatística de

dados, característica que, aliada à sua gratuidade, tem atraído a atenção da comunidade acadêmica, de pesquisadores de instituições públicas e até mesmo de instituições privadas. Os comandos utilizados para gerar os gráficos aqui apresentados estão listados ao final deste Capítulo. Sugerimos o trabalho de Ribeiro *et al.* (2012) como texto introdutório sobre a utilização do ambiente R para análise estatística de dados.

Para maiores detalhes e outros exemplos sobre análise estatística de dados, sugerimos o texto de Siqueira e Tibúrcio (2011).

Síntese do capítulo

- As características investigadas podem ser classificadas como variáveis *qualitativas (nominais ou ordinais)* ou *quantitativas (discretas ou contínuas)*;
- *Tabelas e gráficos* são ferramentas muito úteis para organizar e resumir dados, sendo os gráficos particularmente úteis para a visualização dos dados;
- A *média* e a *mediana* são utilizadas para identificar o valor típico de um conjunto de dados quantitativos;
- A *variabilidade (dispersão)* dos valores da variável quantitativa pode ser medida pelo *desvio-padrão* e pelo *coeficiente de variação*;
- O *percentil de ordem p* é o valor da variável que deixa abaixo dele *p%* dos valores na amostra;
- O *box-plot* é um gráfico que representa a distribuição dos valores de uma variável por meio da mediana, do primeiro e terceiro quartil. É um gráfico útil para detecção de valores atípicos em um conjunto de dados (valores muito baixos ou muito altos em comparação com os demais);
- Uma *tabela de contingência* organiza os dados segundo duas variáveis qualitativas simultaneamente e é útil no estudo da relação entre elas;
- O *diagrama de dispersão* permite a visualização da relação entre duas variáveis quantitativas.

Referências

REIS, E. A.; REIS, I. A. **Análise Descritiva de Dados: Síntese Numérica**. Belo Horizonte: Universidade Federal de Minas Gerais, 2002.

RIBEIRO, A. J. F. et al. **Bioestatística Básica Usando o Ambiente Computacional R**. Belo

Horizonte: Universidade Federal de Minas Gerais, 2012.

SIQUEIRA, A. L.; TIBÚRCIO, J. D. **Estatística na área de saúde: conceitos, metodologia, aplicações e prática computacional**. Belo Horizonte: COOPMED, 2011.

Comandos do R utilizados neste Capítulo

```
#####  
# Comandos R para gráficos e tabelas do CAPITULO 08  
#####  
  
# Importando o arquivo de dados:  
diabetes = read.table("dados_diabetes_CAP08.txt", header=T)  
  
# Vendo o nome das colunas de diabetes:  
names(diabetes)  
  
# Fixando o conjunto de dados diabetes:  
# Este comando deve ser usado sempre que um conjunto de dados  
# for importado para o R  
attach(diabetes)  
  
#####  
# Tabelas de Distribuição de Frequências:  
#####  
  
# Tabela com as frequências absolutas:  
table(escolaridade)  
  
# Tabela com as frequências relativas:  
prop.table(table(escolaridade))  
  
#####  
# Figura 1(a): Gráfico de setores do estado civil:  
#####  
  
# Calculando porcentagens  
prop = round(100*prop.table(table(estadocivil)),1)  
  
# Criando vetor de rotulos, funcao paste()  
rotulos = paste(names(prop),prop,"%")  
  
pie(table(estadocivil), main="Estado Civil", labels=rotulos,  
clockwise=T,col=c("black","darkgrey","lightgrey","white"))  
  
#####  
# Figura 1(b): Gráfico de barra segmentada da escolaridade:  
#####  
  
aux = rep("Escolaridade",dim(diabetes)[1])  
tab = table(escolaridade,aux)  
  
#Para que a legenda fique no lugar correto
```

```

tab<-cbind(tab,rep(0,nrow(tab)))
barplot(100*prop.table(tab,margin=2),legend=names(table(escolaridade)),
        ylab="Frequência relativa")

#####
# Figura 2(a): Gráfico de barras verticais para numero de filhos:
#####

barplot(prop.table(table(filhos))*100, ylab="Frequência (%)",
        xlab="Número de filhos")

#####
# Figura 2(b): Gráfico de barras verticais para faixa etária:
#####

tab = hist(idade,breaks=seq(45,80,by=5),right=F,plot=F)$counts

names(tab) = c("45 a 49","50 a 54","55 a 59","60 a 64","65 a
              69","70 a 74","75 a 80")

barplot(100*tab/sum(tab), ylab="Frequência (%)", xlab="Idade",
        ylim=c(0,30))

#####
# Figura 3 (a e b): Histograma de frequencias relativas para peso
#                segundo sexo:
#####

breaks = seq(60,90,by=5)
tabF = hist(peso[sexo=="F"], right=F, breaks=breaks, plot=F)$counts
tabF = 100*tabF/sum(tabF)
tabM = hist(peso[sexo=="M"], right=F, breaks=breaks, plot=F)$counts
tabM = 100*tabM/sum(tabM)
limy = range(c(tabF,tabM))
library(lattice)
histogram(peso[sexo=="F"], right=F, breaks=breaks,
          ylim=limy, xlab="Peso (kg)",
          ylab="Frequência relativa (%)", main="", col="grey")
histogram(peso[sexo=="M"], right=F, breaks=breaks, ylim=limy,
          xlab="Peso (kg)",
          ylab="Frequência relativa (%)", main="", col="grey")
rm(breaks,tabF,tabM,limy)
#####
# Medidas Síntese:
#####

# Variável idade: média, Desvio-padrão, CV, (mínimo, máximo), aptitude total:
mean(idade)
sqrt(var(idade))
sqrt(var(idade))/mean(idade)
range(idade)
max(idade)-min(idade)

```



```

# Variável idade: mediana, Q1, Q3, P5, P95:
median(idade)
quantile(idade, 0.25)
quantile(idade, 0.75)
quantile(idade, 0.05)
quantile(idade, 0.95)

# Mostra simultaneamente (mínimo, Q1, mediana, Q3, máximo):
summary(idade)

#####
# Figura 8: Box-plot de altura segundo grupo de intervenção:
#####

boxplot(altura~grupo, ylab="Altura", xlab="Grupos de intervenção")

#####
# Tabelas de Contingência:
#####

# Frequências absolutas:
tab = table(escolaridade,estadocivil)
tab

# Totais nas marginais:
margin.table(tab,1) # retorna o total das linhas da tabela
margin.table(tab,2) # retorna o total das colunas da tabela
margin.table(tab) # soma todas as caselas

# Tabela de frequências relativas:
prop.table(tab) # percentuais somando 1 no total geral
prop.table(tab,margin=1) # percentuais somando 1 nas linhas
prop.table(tab,margin=2) # percentuais somando 1 nas colunas

#####
# Figura 9: Gráficos de barras segmentadas:
#####

# (a) Estado civil dentro de escolaridade:
tab = table(estadocivil,escolaridade)
tab<-cbind(tab,rep(0,nrow(tab)))
cores<-c("grey0","grey20","grey40","grey80")
barplot(100*prop.table(tab,margin=2), ylab="Frequência relativa",
        col=cores)
legend(5,100,names(table(estadocivil)),pch=15, col=cores,bty="n")

# (b) Escolaridade dentro de estado civil :
tab = table(escolaridade,estadocivil)
tab<-cbind(tab,rep(0,nrow(tab)))
barplot(100*prop.table(tab,margin=2), ylab="Frequência relativa",
        col=cores)
legend(5,100,names(table(escolaridade)),pch=15, col=cores[rev(1:4)],bty="n")

```

```
#####
#
# Figura 11: Diagrama de dispersão de peso (idade) e
#   circunferência abdominal:
#####
#

plot(peso, circ.abdom, xlab="Peso",
      ylab="Circunferência abdominal", pch=20, cex=1.5)

plot(idade, circ.abdom, xlab="Idade",
      ylab="Circunferência abdominal", pch=20, cex=1.5)

#####
# Figura 12: Diagrama de dispersão de glicohemoglobina antes e
#   depois da intervenção:
#####

plot(ghb1,ghb2, xlab="Antes", ylab="Depois",
      main="Glicohemoglobina", xlim=c(0,15), ylim=c(0,15),
      pch=20, cex=1.5)
abline(0,1, col="red", lwd=2)
```


BOAS PRÁTICAS PARA A PESQUISA NAS CIÊNCIAS DA SAÚDE

Adriana S. Pagano
Heloisa de Carvalho Torres
Ilka Afonso Reis
Arthur de Melo Sá
Edna Afonso Reis
Fernanda Figueredo Chaves
Júlia Santos Nunes Rodrigues
Kícila Ferreguetti
Rodrigo Araújo e Castro
Sumaya Giarola Cecílio
Wanderley Gonçalves dos Santos Junior

Seguem nossas dicas para você se tornar um pesquisador experto:

- Conduza a pesquisa com conhecimentos e habilidades que possibilitarão a obtenção de resultados efetivos.
- As relações interpessoais, experiências, e prática do dia-a-dia são importantes no desenvolvimento de sua expertise como pesquisador.
- Faça um plano de formação pessoal para desenvolver suas habilidades de leitura, escrita e apresentação de trabalhos em eventos.
- Desenvolva uma leitura crítica e aprendizado de métodos e técnicas. Dialogue e colabore ativamente com seus colegas pesquisadores, estejam eles mais ou menos avançados na carreira acadêmica.
- Procure engajar-se em todas as atividades de pesquisa, desde o planejamento de um estudo, a coleta de dados, sua análise e interpretação dos resultados.

- Gerencie suas atividades acadêmicas para atender seu plano de metas.
- Desvincilhe-se da ideia de que a linguagem é um mero veículo de transmissão ou um conjunto de regras. Lembre: a linguagem é um sistema de recursos para gerar significados e é ela que nos permite construir e representar nossa realidade.
- Observe bem como a linguagem varia em cada tipo de texto e procure sistematizar suas observações para apreender à medida que toma contato com um repertório crescente de textos.
- Não subestime decisões importantes na hora de produzir e analisar textos: pergunte-se sempre que atividade a linguagem está desempenhando, como os interlocutores negociam significados e que tipo de texto você tem em mãos.
- Ao preparar uma planilha para a coleta de dados, lembre-se de que
 - cada linha da planilha deve conter os dados de apenas um sujeito;
 - cada coluna da planilha deve conter os dados de apenas uma característica;
 - cada sujeito deve possuir um identificador único, que pode ser composto por números (ex: 003) ou por letras e números (ex: IAR003). A coluna com os identificadores geralmente é a primeira coluna da planilha de dados.
- O registro de dados categóricos cujas categorias permitam flexão de gênero (ex.: casado/casada, ativo/ativa, enfermeiro/enfermeira, dentre outras) deve ser feito usando-se o gênero masculino (ex: casado, ativo, enfermeiro). O dado sobre o sexo do sujeito deve ser registrado em outra coluna.
- Dados numéricos que geralmente são apresentados em conjunto (ex: pressão sistólica e pressão diastólica, 120/85 mmHg) devem ser registrados em colunas separadas (ex: na coluna “pressão sistólica”, registrar o valor 120; na coluna “pressão diastólica”, registrar o valor 85).
- Dados numéricos devem ser registrados sem a unidade de medida utilizada e apenas números devem aparecer na coluna destinada a eles. A unidade de medida pode ser registrada no nome da coluna (ex: peso em Kg, altura em cm, número de filhos).
- Ao registrar os dados de uma questão que permite respostas múltiplas, use uma coluna para cada opção de resposta disponível. Para cada uma dessas colunas, registre SIM se o sujeito escolheu a opção de resposta representada naquela coluna. Caso contrário, registre NAO.
- Certifique-se de que
 - o nome das colunas ou das categorias não possua caracteres especiais (acentos, cedilhas, símbolos). Por exemplo, no lugar de usar “NÃO”, use “NAO”; no lugar de usar “criança”, use “crianca”);
 - o separador das casas decimais nos números não inteiros (ex: 34.5 ou 34,5) esteja padronizado em toda planilha de dados. Use vírgula OU ponto.

Se for trabalhar com questionários:

- Antes de escolher o tipo de questionário da sua pesquisa, seu meio de divulgação e o tipo de preenchimento, relembre as vantagens e desvantagens de cada tipo de questionário
- No caso de questionários eletrônicos, certifique-se de que a plataforma online que será usada na criação do questionário atende aos objetivos da pesquisa.

- Leia as versões do questionário como se fosse um respondente. No caso de questões estruturadas, certifique-se de que todas as possíveis respostas estão contempladas nas opções disponíveis para o respondente.
- Nunca dispense o pré-teste da versão final do questionário, mesmo que ele tiver que ser feito com poucas pessoas.
- Lembre-se de que um questionário extenso demais pode cansar o respondente e levá-lo a dar respostas vagas, no caso de questões não estruturadas, ou a escolher aleatoriamente uma opção de resposta, no caso de questões estruturadas. Assim, antes de incluir mais uma questão, reflita sobre sua utilidade para o que está sendo estudado.

Se optar por transcrever um texto falado:

- Certifique-se dos objetivos da pesquisa que está sendo desenvolvida para decidir se o processo de transcrição beneficiará o estudo.
- Não se esqueça de desenvolver uma metodologia de transcrição antes de começar o processo de transcrição e repassá-la para todos os transcritores.
- Procure antecipar possíveis problemas de transcrição e análise para que eles possam ser solucionados e sua solução incluída na metodologia.
- Lembre-se que os computadores e os *softwares* concordanciadores são ferramentas que auxiliam o processo de transcrição e de análise.
- Priorize os recursos de anotação e análise que não alteram os arquivos mestres das transcrições.

Para a análise de seus dados:

- Lembre-se que para obter resultados relevantes em pesquisas que realizam mineração de dados, a etapa de

pré-processamento do material (preparação dos dados), ainda que realizada de maneira automatizada, exige a sua participação. Você deve ter muita atenção e dedicação. Atente-se para as características do software escolhido por você para o trabalho, realizando cursos que discutam a sua aplicabilidade e possibilidades de tratamento dos dados.

- Nunca esqueça que a análise dos dados por meio da mineração de dados exige em primeiro lugar, conhecer bem seus dados e definir o resultado a que se espera chegar, pois dessa forma você conseguirá preparar os dados de maneira adequada para análise, delimitar um referencial teórico coerente, uma metodologia clara e efetiva para se obter conclusões relevantes e garantir que o seu estudo seja replicável.

Se você for trabalhar com gravações de áudio e de vídeo:

- Você pode analisar diretamente gravações de áudio ou vídeo por meio de ferramentas computacionais, eliminando a necessidade de transcrição destes. O ELAN é uma boa ferramenta para anotação dessas gravações.
- Para analisar vídeos com o ELAN, é necessário que os vídeos não tenham resolução muito alta ou duração maior que 30 minutos (comprometendo a velocidade de processamento) nem tenham resolução muito baixa e duração muito curta (comprometendo a análise).
- Se você tiver uma única câmera para filmar a entrevista, certifique-se de que ela esteja focalizada no entrevistado, capturando sua fala, feições e gestos.
- É possível utilizar um gravador de áudio para capturar o som na gravação da entrevista, o que pode melhorar significativamente a qualidade do som.

- Caso esteja utilizando um gravador de som, você pode ligá-lo diretamente à câmera (que se encarregará de processar áudio e vídeo) ou gravar o áudio em um arquivo separado (podendo analisá-los separadamente ou juntá-los num único arquivo).
- Lembre-se que seu computador deve conseguir executar o arquivo de áudio ou vídeo para que o ELAN também possa trabalhar com eles.
- Se seu computador não executar os arquivos de áudio ou vídeo gravados, providencie um codec apropriado para sua execução (e.g. para executar um arquivo de áudio do tipo flac, o computador deve ter instalado um *codec* para este tipo de arquivo).
- Além da anotação do áudio/vídeo, o ELAN também possibilita a manipulação desses dados estatisticamente por meio de suas ferramentas. Lembre-se de explorá-las tanto quanto puder.
- Se preferir, você pode usar os dados anotados no ELAN em outros programas. Para isso, você deve exportar os dados em um formato que permita a leitura no programa de destino. Caso vá trabalhar com o R ou com um editor de planilhas, sugerimos exportar os dados no formato *.txt* com dados formatados com tabulações.
- Lembre-se de utilizar um esquema de anotação simples, mas que permita armazenar diferentes informações com uma única etiqueta ou esquema de codificação (e.g. a etiqueta S01-T12, que significa “Sujeito 01, trecho de interesse 12”).
- Tenha sempre a análise fundamentada em uma teoria que busca interpretar os dados que você está trabalhando: se os dados são linguísticos, baseie-se em uma teoria linguística; se são de ordem psicológica, em uma teoria da psicologia; se de ordem social, em uma teoria das ciências sociais; etc.

Na hora de fazer o tratamento estatístico dos dados:

- Lembre-se de classificar todas as variáveis do seu banco de dados antes de começar a análise descritiva. Isto te ajudará a escolher os métodos adequados para analisar os dados de suas variáveis.
- Com respeito aos dados de variáveis quantitativas, certifique-se de que não haja valores discrepantes (faça um box-plot, por exemplo). Se forem encontrados valores discrepantes, verifique se eles são frutos de erros de registros ou são dados de indivíduos que não deveriam participar do estudo, por exemplo. Valores discrepantes devem ser tratados com cuidado, pois eles influenciam medidas de síntese bastante utilizadas, como média e desvio-padrão.
- Ao construir gráficos de barras (ou histogramas) que serão comparados entre si, certifique-se de que os eixos verticais de todos os gráficos estejam na mesma escala, assim como as categorias representadas nos eixos horizontais também devem ser as mesmas em todos os gráficos.
- Ao comparar a variabilidade de variáveis diferentes ou da mesma variável em grupos diferentes, lembre-se de usar o coeficiente de variação.
- Sobre a representação gráfica da relação entre duas variáveis, lembre-se de que
 - Diagramas de dispersão são gráficos utilizados para representar visualmente a relação entre os dados de duas variáveis *quantitativas*;
 - Gráficos de barras segmentadas podem ser utilizados para representar visualmente as informações contidas em uma tabela de contingência, que, por sua vez, organiza os dados de duas variáveis *qualitativas*;
 - Uma sequência de box-plots pode ser utilizada para representar visualmente os dados de uma variável *quantitativa* nas diversas categorias de uma variável *qualitativa*.

SOBRE OS AUTORES

ADRIANA S. PAGANO é Professora Titular de Linguística Aplicada da Faculdade de Letras da Universidade Federal de Minas Gerais (UFMG). É Doutora pela UFMG e fez seu pós-doutoramento na Universidade de Massachusetts, Amherst, Estados Unidos, e na Universidade do Sarre, Saarbruecken, Alemanha. Atua no Programa de Pós-Graduação em Estudos Linguísticos da UFMG e é pesquisadora do Laboratório Experimental de Tradução (LETRA) da UFMG. É bolsista de produtividade em pesquisa do CNPq e pesquisadora do Programa de Pesquisadores Mineiros (PPM) da FAPEMIG. É coautora de *Traduzir com Autonomia* (São Paulo: Contexto, 2000), *Competência em tradução: cognição e discurso* (Belo Horizonte: Editora da UFMG, 2005) e organizadora de *Metodologias de pesquisa em tradução* (Belo Horizonte: FALE, 2001). Possui artigos e capítulos em diversas publicações nacionais e estrangeiras (apagano@ufmg.br).

ARTHUR DE MELO SÁ é mestrando pelo Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais, onde obteve seu título de bacharel em Letras (habilitação em Inglês com ênfase em Estudos da Tradução). É pesquisador do Laboratório Experimental de Tradução (LETRA) da UFMG e participa de projetos de pesquisa sobre a modelagem sistêmico-funcional da produção textual multilíngue (arthurdemelosa@gmail.com).

EDNA AFONSO REIS é bacharel, mestre e doutorada em Estatística. É professora do Departamento de Estatística da Universidade Federal de Minas Gerais desde 1999, pesquisadora do Laboratório de Bioestatística/DEST/UFMG e membro do corpo docente do Programa de Pós-Graduação em Medicamentos e Assistência Farmacêutica da UFMG (edna@est.ufmg.br).

FERNANDA FIGUEREDO CHAVES é mestranda do Programa de Pós-Graduação em Enfermagem da Universidade Federal de Minas Gerais. É enfermeira, especialista em Saúde Coletiva com ênfase em Saúde da Família pela UFMG. É pesquisadora do Núcleo de Gestão, Educação e Avaliação em Saúde da Universidade Federal de Minas Gerais e participa de projetos de pesquisa sobre Tradução, Adaptação Cultural e Validação de instrumentos em doenças crônicas (fernandafchaves@ufmg.br).

HELOISA DE CARVALHO TORRES é Professora Associada da Escola de Enfermagem da Universidade Federal de Minas Gerais (UFMG). É Doutora pela Escola Nacional de Saúde Pública da Fundação Oswaldo Cruz (FIOCRZ). Realizou seu doutorado Sanduíche no *Institut de Recherche en Santé Publique de l'Université de Montréal*, Canadá, e o pós-doutoramento na Escola Nacional de Saúde Pública (FIOCRZ), Rio de Janeiro. Atua no Programa de Pós-Graduação da Escola de Enfermagem da UFMG e é coordenadora do Núcleo de Gestão, Educação e Avaliação em Saúde da UFMG. É autora de estudos de avaliação de programas de empoderamento em diabetes tipo 2 na atenção primária e de elaboração, adaptação e validação de instrumentos em saúde. Possui publicações em diabetes. Atualmente, é residente pós-doutoral do Programa de Pós-Graduação em Estudos Linguísticos da UFMG sob a supervisão da professora Adriana S. Pagano (heloisa@enf.ufmg.com).

ILKA AFONSO REIS é Professora Adjunta do Departamento de Estatística da Universidade Federal de Minas Gerais (UFMG), onde obteve os títulos de bacharel e de mestre em Estatística. É doutora em Sensoriamento Remoto pelo Instituto Nacional de Pesquisas Espaciais (INPE). Está vinculada ao Laboratório de Bioestatística/DEST/UFMG e participa de projetos de pesquisa na área de ciências da saúde em parceria com pesquisadores da UFMG e também do Centro de Pesquisas René Rachou/FioCRUZ. Atualmente, é residente pós-doutoral no Programa de Pós-graduação em Estudos Linguísticos da UFMG sob a supervisão da professora Adriana S. Pagano (ilka@ufmg.br).

JÚLIA SANTOS NUNES RODRIGUES é graduanda do curso de Letras – Bacharelado com Ênfase em Estudos da Tradução – Inglês, da Faculdade de Letras da Universidade Federal de Minas Gerais. É pesquisadora do Laboratório Experimental de Tradução dessa Faculdade e participa como bolsista de iniciação científica (PIBIC/CNPq) do projeto Empoder@ (juliasnrodrigues@ufmg.br).

KÍCILA FERREGUETTI é doutoranda do Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais, onde obteve seus títulos de bacharel em Letras (habilitação em Italiano com ênfase em Estudos da Tradução) e Mestre em Linguística Aplicada. É pesquisadora do Laboratório Experimental de Tradução (LETRA) da UFMG e participa de projetos de pesquisa sobre a modelagem sistêmico-funcional da produção textual multilíngue (kicilaferregueti@let.grad.ufmg.br).

RODRIGO ARAÚJO E CASTRO é mestrando do Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais, onde obteve seu título de bacharel em Letras (habilitação em Inglês com ênfase em Estudos da Tradução). É pesquisador do Laboratório Experimental de Tradução (LETRA) da UFMG e participa de projetos de pesquisa sobre a modelagem sistêmico-funcional da produção textual multilíngue e sobre a abordagem interdisciplinar nas práticas educativas em doenças crônicas: tradução e adequação cultural de instrumentos (rodrigoacastro@ufmg.br).

SUMAYA GIAROLA CECILIO é mestranda do Programa de Pós-Graduação em Enfermagem da Universidade Federal de Minas Gerais. É bacharel em Enfermagem pela Universidade

Federal de São João del Rei e é especialista em Auditoria em Saúde pela Faculdade Internacional de Curitiba (Facinter). É pesquisadora do Núcleo de Pesquisa em Gestão, Educação e Avaliação em Saúde da Universidade Federal de Minas Gerais e participa de pesquisas sobre inovações metodológicas para a construção e tradução de instrumentos em saúde, métodos qualitativos de análise de dados, abordagem do empoderamento em condições crônicas, bem como estratégias educativas em saúde (sumayagc@ufmg.com).

WANDERLEY GONÇALVES DOS SANTOS JUNIOR é graduando do curso de Letras – Bacharelado com Ênfase em Estudos da Tradução – Inglês, da Faculdade de Letras da Universidade Federal de Minas Gerais. É pesquisador do Laboratório Experimental de Tradução dessa Faculdade (wanderleygsjr@gmail.com).

ÍNDICE REMISSIVO

A

Adaptação cultural	4, 8, 16, 19, 32, 33, 35, 56, 57, 62, 75, 95, 112, 143
Adequação cultural	15, 32, 33, 34, 35, 56, 57, 58, 59, 60, 144
Amplitude total	124, 126
Análise de conteúdo	79, 82, 84, 85, 86, 87, 89, 95, 101, 111
Análise de sentimentos	27, 92, 104
Áudio	viii, 7, 22, 25, 34, 35, 42, 44, 69, 80, 84, 99, 100, 101, 102, 103, 105, 110, 111, 141, 142
Automatização da análise	78, 87, 91
Autonomia	vii, viii, 1, 2, 7, 8, 10, 11, 12, 14, 15, 16, 54, 61, 89, 90, 143
Artigo acadêmico	23, 28, 29, 30
Aspectos psicossociais	2, 91
Atitudes	vii, 2, 7, 8, 10, 11, 13, 33, 75, 77, 80, 100, 112
Avaliação	4, 6, 10, 11, 14, 21, 24, 28, 32, 33, 35, 51, 53, 55, 56, 57, 61, 62, 75, 81, 86, 87, 104, 113, 143, 144, 145

C

Capacitação	1, 2, 5, 8, 17
Cartilha	27, 28, 29, 30
Coefficiente de variação	126, 133
Coleta de dados	vii, 3, 7, 15, 22, 41, 45, 49, 50, 55, 62, 63, 65, 102, 140

Competência	viii, 7, 9, 11, 15, 17, 113, 143
Contexto de cultura	19, 21, 22, 25, 35
Contexto de situação	19, 21, 22, 23, 24, 27, 29, 33, 35, 41, 81
Corresponsabilização	1, 10, 14, 80
Corpora	78, 97, 112
Corpus	78, 86, 88, 89, 91, 97
Cuidado em saúde	vii, 1, 2, 7, 10, 12, 13, 14, 15, 19, 87

D

Dados	
Como coletar	45
Dados estruturados	viii, 40, 41, 42, 44, 45, 47, 80, 81, 84, 88, 91, 100, 101, 115
Dados não estruturados	viii, 40, 41, 42, 44, 45, 46, 47, 80, 81, 87, 94, 95, 99, 100, 101, 110, 111
Estruturados	vii, 22, 38, 40, 41, 42, 43, 44, 45, 46, 47, 68, 79, 80, 81, 84, 87, 88, 94, 95, 99, 100, 101, 110, 111, 115
Não-estruturados	vii, 22, 38, 40, 41, 42, 44, 45, 46, 47, 68, 80, 81, 84, 87, 88, 94, 95, 99, 100, 101, 110, 111
Planilha de	42, 43, 44, 45, 101, 116, 129, 140
Decil	127
Desvio-padrão	125, 126, 127, 133, 135, 142
Diagrama de dispersão	131, 132, 133, 137
Distância interquartílica	128

E

ELAN	44, 45, 69, 78, 101, 102, 103, 104, 105, 106, 107, 108, 110, 111, 112, 141, 142
Empoderamento	vii, viii, 1, 2, 3, 4, 5, 6, 7, 8, 10, 14, 15, 33, 68, 80, 100, 111, 144, 145
Escore padronizado	127
Estatística	vii, 5, 6, 15, 20, 32, 38, 39, 42, 46, 52, 56, 57, 61, 63, 79, 80, 82, 88, 94, 95, 104, 111, 115, 116, 117, 122, 129, 133, 143, 144
Etiquetas (<i>tags</i>)	108

F

Filmagem	102, 103
Frequência	
Frequência absoluta	117, 119, 120
Frequência relativa	117, 121, 130, 135, 136

G

Gráfico	24, 25, 29, 35, 41, 81, 83, 101, 109, 110, 115, 116, 117, 118, 119, 120, 127, 130, 131, 134, 135, 136, 142
de barras	109, 118, 120, 130, 131, 135
de box-plot	127, 128, 129, 133, 136, 142
de setores	118, 130, 134
Gravação	25, 34, 35, 101, 102, 103, 141

H

Histograma	120, 121, 135, 142
------------	--------------------

I

Informação	2, 20, 32, 37, 38, 42, 43, 45, 46, 47, 72, 80, 81, 82, 95, 105, 109, 115, 121
Interlocução coletiva	11, 12, 80, 104

L

Ligação telefônica por meio de <i>software</i> de comunicação pela Internet	104
Língua escrita	25
Língua falada	105
Linguística sistêmico-funcional	20, 22, 33, 87, 93, 94, 100
Linhas de concordância	42, 44, 45, 70, 81, 87, 88, 89, 90, 92, 93, 97

Lista de colocados	87, 89, 94, 97
Lista de frequência	83, 87, 89, 90, 91, 92, 94, 108, 109

M

Média	11, 13, 116, 121, 122, 123, 124, 125, 126, 127, 133, 135, 142
Mediana	121, 122, 123, 127, 128, 129, 133, 136
Metodologia de transcrição	68, 141
Metodologia semiautomática	79
Mineração de dados	34, 44, 79, 80, 81, 82, 83, 95, 96, 100, 141
Mineração de textos	7, 34, 52, 69, 72, 79, 80, 81, 82, 84, 86, 87, 89, 91, 95, 100

N

Nuvem de palavras	83, 87, 90, 91, 92, 93, 97
-------------------	----------------------------

P

Padronização	4, 61, 69, 74, 76
Palavra gramatical	97
Palavra lexical	97
Percentil	127, 133
Pesquisador novato	1, 3, 4, 5, 6, 7, 14, 15, 20, 27, 28, 32, 35
Planilha de dados	42, 43, 44, 45, 101, 116, 129, 140
Práticas educativas em saúde	4, 15
Pré-processamento de dados	46
Profissional da área da saúde	vii, viii, 1, 2, 3, 4, 7, 8, 9, 10, 13, 15, 31, 46, 55, 75, 80, 100, 111

Q

Quartil	127, 128, 133
Questionário	3, 4, 15, 25, 27, 29, 30, 33, 34, 35, 41, 45, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 65, 67, 68, 140, 141

S

<i>Script</i>	88, 89, 90, 94, 97
<i>Software</i>	11, 13, 14, 55, 54, 69, 70, 72, 73, 74, 75, 76, 78, 83, 84, 86, 88, 95, 97, 99, 100, 101, 102, 103, 104, 105, 106, 107, 110, 112, 117, 141
<i>Software de comunicação pela Internet</i>	11, 14, 104
<i>Software concordanciador</i>	69

T

Tabela de contingência	39, 43, 129, 130, 133, 142
Tabela de frequências	117, 118, 119, 136
Tipo de texto	21, 22, 23, 25, 30, 33, 35, 41, 65, 68, 69, 102, 140
Tipos de questionário	50, 53
Tokenização	83, 97, 101
Tokens	83, 89, 97
Tradução	2, 6, 8, 16, 19, 20, 32, 33, 35, 56, 57, 62, 75, 95, 102, 105, 106, 107, 112, 143, 144, 145
Transcrição	viii, 21, 22, 25, 27, 34, 35, 43, 65, 67, 68, 69, 70, 72, 73, 74, 75, 77, 81, 88, 97, 101, 102, 103, 105, 106, 107, 110, 141

V

Variável	23, 24, 26, 27, 28, 41, 81, 116, 117, 118, 119, 120, 121, 122, 124, 126, 127, 128, 129, 130, 131, 133, 135, 136, 142
Variável qualitativa	117, 118, 129, 142
nominal	51, 52, 83, 117
ordinal	51, 52, 116, 117
Variável quantitativa	117, 121, 133, 142
Variável contínua	120
Variável discreta	119
Vídeo	vii, 7, 22, 34, 35, 42, 44, 62, 69, 80, 84, 99, 100, 101, 102, 103, 104, 105, 106, 108, 110, 111, 142
Visita domiciliar	11, 16, 80

Este volume está direcionado aos pesquisadores das Ciências da Saúde e parte do pressuposto de que o empoderamento do pesquisador conduz à obtenção de resultados mais esclarecedores que poderão informar as práticas educativas do profissional da área da Saúde, corresponsável pela autonomia do usuário do serviço de Saúde. Interdisciplinaridade e inovação tecnológica são os dois pilares desta proposta pela qual os autores convidam seus leitores a se engajarem nas suas pesquisas com os subsídios teóricos e recursos metodológicos oferecidos.



ISBN: 978-85-7758-264-8



9 788577 582648