

# Um Estudo sobre Descrição de Imagens em Mídias Sociais Online na Perspectiva de Pessoas com Cegueira Congênita

Carolina Sacramento<sup>1</sup>, Simone Bacellar Leal Ferreira<sup>2</sup>, Sara Remedios<sup>2</sup>

<sup>1</sup>Casa de Oswaldo Cruz – Fundação Oswaldo Cruz – Rio de Janeiro – RJ – Brasil

<sup>2</sup>Departamento de Informática Aplicada – Universidade Federal do Estado do Rio de Janeiro (UNIRIO) – Rio de Janeiro – RJ – Brasil

carolina.sacramento@fiocruz.br, simone@uniriotec.br,  
sara.lobato@edu.unirio.br

**Abstract.** *A great challenge for screen reader users when using online social media systems is interacting with thousands of images posted daily, as screen readers are not yet able to interpret them. Despite some existing initiatives to describe images, when considering people who have never experienced the sense of sight (congenitally blind or without visual memory), we questioned whether there are particularities and differences in the way they describe the world, represented by photographs when compared to sighted people. This paper presents the results of description exercises, with the participation of congenitally blind and sighted people, to identify the particularities and differences in the descriptions made by them and to verify if a set of questions to assist the description process proposed in the literature contemplates the elements used in the description made by someone with congenital blindness.*

**Resumo.** *Um grande desafio dos usuários de leitores de tela no uso das mídias sociais online é a interação com as milhares de imagens postadas diariamente, já que os leitores ainda não são capazes de interpretá-las. Apesar de algumas iniciativas existentes para descrever imagens, ao considerar pessoas que nunca experimentaram o sentido da visão (cegos congênitos ou sem memória visual), questiona-se se existem particularidades e diferenças na forma em que eles descrevem o mundo, representado por fotografias, em comparação com videntes. Neste artigo apresenta-se os resultados de exercícios de descrição, com participação de cegos congênitos e videntes, no intuito de identificar as particularidades e diferenças nas descrições feitas e verificar se um conjunto de questões para auxiliar a descrição proposta na literatura contempla os elementos utilizados na descrição feita por quem tem cegueira congênita.*

## 1. Introdução

Muitos conteúdos que exigem o sentido da visão, como fotografias, animações, gráficos, vídeos, entre outros, são compartilhados diariamente nas mídias sociais [Internet Live Stats 2022]. Para interagir nessas mídias, pessoas com deficiência visual (cegas ou em casos específicos de baixa visão) utilizam leitores de tela, tecnologia que captura o código relacionado à informação exibida na tela e a disponibiliza em forma de som, por meio dos sintetizadores de voz [Ferreira et al. 2006]. No entanto, quando se trata de conteúdo visual é necessário que sejam fornecidas alternativas a este conteúdo [W3C, 2018].

Apesar da maioria das imagens digitais serem inacessíveis para usuários de leitor de telas [Morris et al. 2018], diferentes abordagens vêm sendo adotadas, desde a descrição manual, feita por uma pessoa que enxerga até descrições automáticas, providas por algoritmos [Gleason et al. 2019], como no Facebook, que fornece uma descrição textual superficial automaticamente para todas as fotografias [Sacramento et al. 2020].

Porém, mesmo em situações em que alternativas ao conteúdo visual estão sempre disponíveis, nem sempre essas alternativas são consideradas confiáveis [Wu et al. 2017]. Dentre os fatores que impactam na produção de alternativas destacam-se a dificuldade e subjetividade da ação de descrever, os diversos tipos de conteúdo visuais existentes [Sacramento et al. 2020] e as diferenças sensoriais que norteiam a percepção desses conteúdos por pessoas com e sem deficiência visual, principalmente quando se trata de pessoas cegas de nascença ou que ficaram cegas nos primeiros anos da infância (cegos congênitos ou sem memória visual), já que este público apresenta modelos mentais diferentes das pessoas videntes (que enxergam) [Nunes e Lomônaco 2008].

Dentre as pesquisas existentes sobre este tópico [Jandrey et al. 2021], destaca-se a de Salisbury e colegas [Salisbury et al 2017]. O trabalho combinou intervenção humana (abordagem *crowdsourcing*) com descrições automáticas, para identificar os requisitos necessários a criação de textos alternativos de qualidade e, conseqüentemente, gerar um conjunto de perguntas para orientar as descrições. Algumas dessas questões são apresentadas na Tabela 1.

**Tabela 1 Algumas das questões propostas por Salisbury e colegas (2017) e seus respectivos códigos de identificação**

Cód.	Questão
Q1	Quem são os principais sujeitos da imagem (pessoas, animais, objetos notáveis etc.)? Descreva suas características físicas (características notáveis, roupas, poses, posições relativas etc.)
Q2	Onde está este conjunto? Descreva a localização e as características proeminentes do plano de fundo.
Q3	O que os sujeitos da imagem estão fazendo? Descreva suas ações e suas intenções.
Q4	Que emoção esta imagem evoca? Ou quais são as emoções dos sujeitos presentes na imagem?
Q5	Descreva quaisquer aspectos notáveis do estilo visual da imagem.
Q6	Esta é uma imagem famosa ou conhecida?

Fonte: Salisbury et al. (2017). Tradução: dos autores.

No entanto, a pesquisa de Salisbury et al. (2017) propõe descrições sob o ponto de vista das pessoas videntes. Ao considerar a perspectiva dos cegos congênitos, que não experimentaram o sentido visual e possuem formas diferentes de lidar com elementos do mundo quando comparado com pessoas que enxergam ou já enxergaram, questionou-se se poderiam existir particularidades e diferenças na forma em que eles descrevem tais elementos, representado por fotografias, no contexto das mídias sociais.

Neste artigo apresenta-se os resultados de uma pesquisa qualitativa, caracterizada por exercícios de descrição, com participação de cegos congênitos e videntes, no intuito de identificar as particularidades, similaridades e diferenças na forma em que ambos criam descrições que poderiam ser utilizadas como alternativas ao conteúdo visual em

mídias sociais online. Adicionalmente, os resultados obtidos com os cegos congênitos foram comparados com o conjunto de questões de Salisbury e colegas [Salisbury et al. 2017] e geraram uma adaptação das questões para o público em estudo.

## **2. Método de Pesquisa**

O exercício de descrição foi realizado remotamente e contou com a participação de vinte e dois voluntários, onze cegos congênitos e onze videntes, com idade média de 35 anos em ambos os perfis. A pesquisa foi submetida e aprovada por um Comitê de Ética em Pesquisa (CAAE: 17056719.1.0000.5285). O foco do exercício foi a investigação de como alguém que nunca enxergou descreveria elementos com base nas suas experiências e percepção. As seguintes etapas foram empreendidas:

1. Definição dos elementos a serem descritos pelos participantes (detalhado na Subseção 3.1);
2. Preparação para a coleta de dados, que contemplou elaboração do roteiro do exercício (apresentado na Subseção 3.2);
3. Recrutamento dos participantes;
4. Coleta de dados, realizada de forma remota via chamadas de áudio, incluindo a realização de testes piloto;
5. Preparação para análise de dados, contemplando a transcrição dos exercícios;
6. Análise dos dados obtidos nas respostas, com a técnica de Análise de Conteúdo (apresentado na Subseção 3.3);
7. Validação dos atributos criados por um segundo pesquisador, que realizou um novo processo de codificação com base nos atributos definidos;
8. Comparação dos resultados entre os perfis, utilizando Teste Exato de Fisher, escolhido por ser um teste estatístico não-paramétrico de associação, baseado no Teste Qui-Quadrado, que permite verificar a relação entre duas variáveis qualitativas nominais independentes em situações em que o número de amostras é pequeno [Guimarães 2016] (exemplificado na Subseção 3.4);
9. Relação dos atributos obtidos no exercício dos cegos congênitos com as questões criadas por Salisbury et al. (2017), com uso da técnica de diagramas de afinidades, gerando uma adaptação das questões.

Dentre as limitações do método destacou-se a necessidade de adaptação do exercício para o contexto remoto, já que a coleta de dados ocorreu entre os meses de fevereiro e março de 2021, quando ainda existiam restrições impostas pela pandemia de covid-19 no Brasil. Inicialmente estava previsto um exercício de exploração efetiva de sentidos, que foi substituída pela exploração de sentidos baseada em lembranças e experiências prévias dos participantes.

Tal ajuste tornou o estudo mais subjetivo, pois apesar da escolha por situações e elementos de conhecimento coletivo, existem variações entre os elementos, mesmo os mais simples. Além disso, lembranças e experiências individuais prévias também podem ser variadas, não apenas entre os diferentes perfis (cegos e videntes), mas também entre os diversos sujeitos de um mesmo perfil.

### 3. Exercícios de Descrição

#### 3.1. Definição dos elementos a descrever

Optou-se por escolher elementos que representassem fotografias presentes em mídias sociais, que pudessem ser criadas e descritas a partir de uma experiência real de interação das pessoas com deficiência, seja pela exploração de sentidos ou por experiências vivenciadas. Assim, definiu-se que seriam solicitadas as seguintes descrições:

- de um autorretrato (selfie de rosto);
- de uma situação vivenciada específica (ida à uma praia);
- dos atributos utilizados na descrição de uma pessoa e de um local desconhecidos;
- de quatro elementos concretos.

O trabalho de Nunes & Lomônaco (2008), que investigou o desenvolvimento de conceitos por crianças cegas congênitas, foi utilizado como referência na definição dos elementos concretos por apresentar uma estruturação de conceitos baseada nos sentidos. Os autores organizaram os conceitos dois grupos: concretos e abstratos. Os concretos foram divididos em tateáveis e não tateáveis, que por sua vez foram subdivididos em amplamente manuseáveis e de manuseio restrito (tateáveis) e cognoscíveis e não cognoscíveis pelos sentidos dos cegos (não tateáveis), como ilustra a Figura 1.

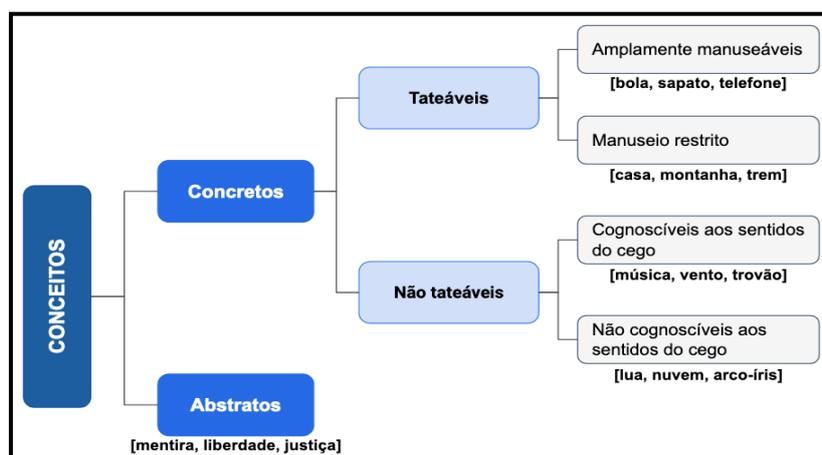


Figura 1 Estrutura de conceitos de Nunes e Lomônaco (2008), com exemplos para cada tipo de conceito.

Após diversas seções para validar a escolha dos elementos concretos, com participação de pessoas com e sem deficiência visual, foram estabelecidos os seguintes elementos para o exercício: laranja e escova de dentes (elementos tateáveis amplamente manuseáveis), edifício (elemento tateável, porém manuseio restrito) e fumaça de cigarro (elemento não tateável, mas cognoscível pelos sentidos de uma pessoa cega).

#### 3.2. Condução do exercício

A maioria dos voluntários foi recrutada a partir do interesse registrado em pesquisa anterior [Sacramento et al. 2020], onde foram questionados se gostariam de participar de novos estudos. Para validar o roteiro do exercício foram realizados estudos-piloto com um cego congênito e um vidente.

Antes de iniciar o exercício, os participantes foram avisados que a chamada de áudio seria gravada, e fez-se a leitura do Termo de Consentimento Livre e Esclarecido (enviado previamente aos participantes, via *WhatsApp* ou e-mail, em formato PDF legível pelo leitor de telas). Ao final da leitura, os voluntários foram orientados a declarar seu nome completo e que aceitavam participar do estudo, de acordo com os termos mencionados. Na sequência, o roteiro do exercício, apresentado na Tabela 2, foi aplicado.

**Tabela 2 Roteiro do exercício de descrição: perguntas feitas aos participantes para cada elemento definido**

<b>Elemento</b>	<b>Questões feitas aos voluntários</b>
Selfie	Se você tivesse que compartilhar uma selfie, uma foto sua, do seu rosto, como você descreveria?
Praia	Você já foi à praia em algum momento da sua vida? Lembra da experiência? Essa praia tinha muitas pessoas ou era uma praia deserta? Se você tivesse que compartilhar uma foto dessa praia (cheia ou deserta) com base na sua experiência, como você descreveria essa foto?
Lugar desconhecido	Demos o exemplo da praia anteriormente, um ambiente que apesar de não ser igual em todos os lugares, por exemplo, no Rio de Janeiro a praia é de um jeito, no Nordeste é de outro, ela tem elementos bem característicos (como areia, onda, vento, água) mas quando você conhece algum lugar diferente, como você costuma registrar a lembrança desse lugar? Se tivesse que descrever um lugar, que atributos utilizaria?
Pessoa desconhecida	E com relação às pessoas, como você costuma registrar as pessoas, para lembrar delas posteriormente? Se você tivesse que descrever uma pessoa, que atributos utilizaria?
Elementos concretos	E os elementos do mundo, representado pelos objetos, por criações do homem ou fenômenos da natureza da natureza por exemplo: <ul style="list-style-type: none"> <li>• Se tivesse que descrever uma laranja?</li> <li>• E uma escova de dentes?</li> <li>• E um edifício?</li> <li>• E fumaça de cigarro?</li> </ul>

Fonte: dos autores

### 3.3. Análise de dados

Na Análise de Conteúdo empreendida adotou-se técnica de codificação indutiva, que permitiu a síntese de um conjunto de textos (respostas dos voluntários) em categorias de conteúdo (atributos). Optou por criar categorias a partir das respostas fornecidas (codificação indutiva), ao invés de utilizar uma codificação já existente.

Na análise, a pesquisadora principal categorizou as descrições em atributos, num processo iterativo, no intuito de verificar possíveis falhas e contradições nas atribuições feitas. Posteriormente, uma pesquisadora-colaboradora categorizou as descrições com os atributos previamente definidos em duas fases. Na primeira utilizou apenas os títulos dos atributos para realizar a categorização e na segunda, feita após reuniões de consenso com a pesquisadora principal, contou com suporte de uma tabela de definição de atributos.

Na primeira fase, a maioria dos pontos de divergências entre as categorizações feitas pelas pesquisadoras estava associada à falta de entendimento da pesquisadora-colaboradora em relação ao atributo, uma vez as atribuições eram feitas apenas com base nos títulos. Quando a descrição do atributo foi fornecida (segunda fase), o número de pontos divergentes reduziu significativamente.

A Tabela 3 apresenta exemplos de descrições feitas por voluntários para selfie e fumaça de cigarro, e os atributos gerados.

**Tabela 3 Exemplos de descrição feita para selfie e fumaça de cigarro, em ambos os perfis de voluntário, e atributos definidos na Análise de Conteúdo**

Elemento	Perfil do voluntário	Descrição	Atributos definidos
Selfie	Cego congênito	“Rosto de um homem aparentando ter seus 38 anos, utilizando pouca barba, pele clara e cabelos curtos”	Parte do corpo retratada Gênero Aspecto cronológico Característica física Aspecto visual
Selfie	Vidente	“Mulher descendente de espanhóis, italianos e indígenas. Brasileira. Cabelo castanho, comprido, na altura dos seios, com algumas mechas claras. Olhos castanhos.”	Gênero Etnia e Raça Característica física Aspecto visual
Fumaça de cigarro	Cego congênito	“Pelo cheiro. Pela coisa de cheirar bem forte. Que incomoda ainda mais quem não fuma, que é o meu caso”.	Experiência sensorial não visual: olfativa Opinião, preferência e sentimento
Fumaça de cigarro	Vidente	“É branco, translúcido. Dá uma sensação de leveza... porque assim, eu acho que de leveza por estar no ar, por ser alguma coisa que é um pouco mais fluida assim de se espalhar, com facilidade, mas ao mesmo tempo não passa uma sensação positiva por ser do cigarro, né? Daí você tem a impressão de que é algo que não faz bem...”	Aspecto visual Comportamento Opinião, preferência e sentimento Danos à saúde

Fonte: dos autores

### 3.4 Comparação dos resultados entre os perfis

A comparação foi feita com o Teste Exato de Fisher, que utiliza tabelas de contingências 2x2 para comparar dois grupos independentes (ex.: cegos congênitos e videntes), em circunstâncias em que um dos valores esperados da tabela é menor do que cinco [Guimarães 2016]. Para ilustrar a aplicação do Teste toma-se como exemplo a indicação de “aspectos visuais” na descrição de “praia”. Três de onze cegos utilizaram referências visuais para descrever o item em questão, em contraste com sete de onze videntes. A tabela de contingência desta associação é apresentada na Tabela 4.

**Tabela 4 Tabela de contingência montada na aplicação do Teste de Fisher em relação ao atributo “aspectos visuais” na descrição de “praia”**

Perfil de participante	Indicou aspectos visuais	Não indicou aspectos visuais
Cegos congênitos	3	8
Videntes	7	4

Cada categoria identificada nos diferentes itens descritos (selfie, praia etc.) foi comparada entre os perfis em termos de quantidade de vezes que foram ou não citadas. Para o teste, utilizou-se o intervalo de confiança de 95% (nível de significância de 5%, p-value de 0.05) e estabeleceu-se as seguintes hipóteses para cada atributo:

H0: Não há diferença entre cegos congênitos e videntes no uso do atributo X.

H1: Há diferença entre cegos congênitos e videntes no uso do atributo X.

Onde X é o atributo em análise.

No exemplo apresentado na Tabela 4, o teste resultou um p-value de 0.198, não sendo possível rejeitar H<sub>0</sub>, ou seja, indicando a inexistência de diferença significativa para este atributo entre os perfis na descrição de “praia”.

#### **4. Síntese dos resultados**

A análise comparativa entre os atributos gerados para cada descrição demonstrou que poucas foram as diferenças significativas entre os perfis. Na descrição do autorretrato, por exemplo, percebeu-se que os cegos tenderam a descrever da mesma forma que recebem descrições nas mídias sociais, usando trechos presentes em descrições automáticas: “Foto pode conter...”. O mesmo aconteceu na descrição da praia. Neste caso, os aspectos da natureza e arquitetura tiveram destaque, apesar dos videntes apresentarem descrições mais detalhadas.

Já na descrição de um local novo/desconhecido, os videntes continuaram usando atributos relacionados a itens da natureza, enquanto os cegos, por não terem o local registrado em memória e nem conseguirem interagir de imediato com os elementos do local (pois foram incentivados a refletir sobre os atributos que usariam para descrever um local desconhecido), fizeram descrições mais próxima da realidade de interação deles com o mundo. O mesmo em relação a descrição de uma pessoa desconhecida.

Nos demais elementos, o único que apresentou diferença significativa nas descrições entre cegos e videntes foi fumaça de cigarro. Os cegos tenderam a mencionar o cheiro, comparar com outros elementos e dar sua opinião sobre o cheiro ruim da fumaça. Já os videntes tenderam a utilizar referências visuais na descrição. Cabe ressaltar que, tanto os participantes com deficiência quanto os videntes consideraram a descrição deste elemento uma tarefa difícil e abstrata.

Ao relacionar dos atributos fornecidos por cegos com as questões de Salibury e colegas (2017) percebeu-se que a maioria dos atributos obtidos exercício estava contemplada nas questões estabelecidas (Tabela 1). Ainda assim, foram propostos ajustes nas questões para que pudessem incorporar aspectos vivenciados por este público:

- I. Na Q1, indicar paisagem como possível foco da imagem
- II. Na Q1, acrescentar os termos cores, composição, formato e itens da natureza/arquitetura nas orientações de descrição;
- III. Na Q2, acrescentar a frase incluindo a presença de pessoas e características do tempo, nas orientações de descrição;
- IV. Incluir nova orientação, direcionada a possíveis experiências sensoriais não visuais ou climáticas (ex.: frio, quente) vivenciadas quando o registro fotográfico foi feito;
- V. Incluir nova orientação, direcionada a descrição de elementos incomuns, com indicações sobre utilidade e informações de uso e comparações com elementos similares.

#### **5. Conclusão**

Este artigo apresentou uma pesquisa qualitativa, caracterizada por exercícios de descrição, cujo objetivo foi investigar como pessoas com cegueira congênita e videntes criam descrições de elementos do mundo real, representados por fotografias em mídias sociais online, no intuito de identificar as particularidades, similaridades e diferenças existentes na forma em que estas pessoas criam descrições. Os resultados obtidos com os

cegos congênitos foram comparados com as questões de Salisbury et al (2017), gerando uma adaptação das questões para este público.

A comparação entre perfis resultou em pouca diferença significativa na descrição feita por cegos congênitos e videntes, com destaque apenas para descrição de local ou pessoa nova/desconhecida e de fumaça de cigarro, onde os cegos congênitos tenderam a fazer descrições mais próxima da realidade de interação deles com o mundo. Como contribuição, este trabalho apresenta uma adaptação das questões propostas por Salisbury et al. (2017), onde foram incorporados aspectos vivenciados por cegos congênitos.

Um desdobramento da pesquisa foi a realização de uma primeira avaliação das questões adaptadas com a participação de cinco pessoas videntes (produtores de conteúdo visual em mídias sociais), para verificar se as questões, de fato, dão suporte na tarefa de descrição e dois cegos congênitos para avaliar a satisfação em relação às descrições criadas. Apesar de não ser o objetivo deste artigo apresentar os resultados desta avaliação, é importante ressaltar uma de suas conclusões: as questões não devem substituir a descrição livre, mas servirem de complemento para ajudar na composição de descrições.

## Referências

- Ferreira, S. B. L., Chauvel, M. A., Ferreira, M. G. A. L. (2006) “e-Acessibilidade: tornando visível o invisível”. In: Encontro da ANPAD, Salvador.
- Gleason, C. et al. (2020) "Twitter A11y: a browser extension to make Twitter images accessible". In: Conference on Human Factors in Computing Systems, CHI20, p. 1-20
- Guimarães, P. R. B. (2016) “Estatística não paramétrica”. Material didático, [https://docs.ufpr.br/~prbg/public\\_html/ce050/aluno\\_2016\\_np.pdf](https://docs.ufpr.br/~prbg/public_html/ce050/aluno_2016_np.pdf)
- Internet Live Stats. (2022) “In one second, each and every second, there are...”, <https://www.internetlivestats.com/one-second/>
- Jandrey, A. et al. (2021) “Image Descriptions' Limitations for People with Visual Impairments: Where Are We and Where Are We Going?": In: Brazilian Symposium on Human Factors in Computing Systems, IHC21, p. 1–11
- Nunes, S., Lomônaco, J. F. B. (2008) “Desenvolvimento de conceitos em cegos congênitos: caminhos de aquisição do conhecimento”. In: Psicologia Escolar e Educacional, v. 12, n. 1, p. 119–138.
- Morris, M. R. et al. (2018) "Rich representations of visual content for screen reader users". In: Conference on Human Factors in Computing Systems, CHI18, p. 1–11
- Sacramento, C. et al. (2020) "#PraCegoVer: investigating the description of visual content in brazilian online social media". In: Brazilian Symposium on Human Factors in Computing Systems, IHC2020, p. 1-10
- Salisbury et al. (2017) "Toward scalable social alt text: conversational crowdsourcing as a tool for refining vision-to-language technology for the blind". In: Conference on Human Computation and Crowdsourcing, HCOMP2017, p. 147–156
- W3C. (2018) “Web content accessibility guidelines (WCAG) 2.1”., <https://www.w3.org/TR/WCAG21/>
- WU, S. et al. (2014) "Visually impaired users on an online social network". In: Conference on Human Factors in Computing Systems, CHI14, p. 3133–3142