

Guia ReINVenTA para a Audiodescrição Fílmica

CONTRIBUIÇÕES DA TECNOLOGIA
LINGUÍSTICA PARA A EXPERIÊNCIA
CINEMATOGRÁFICA ACESSÍVEL



G943

Guia ReINVenTA para a audiodescrição fílmica : contribuições da tecnologia linguística para a experiência cinematográfica acessível / Organização: Maucha Andrade Gamonal... [et al.]. – Belo Horizonte : FALE/UFMG, 2026.

83 p. : il., color.

ISBN: 978-85-7758-398-0

Bibliografia: p. 79-82.

1. Tradução e interpretação. 2. Tradução fílmica. 3. Audiodescrição fílmica. I. Gamonal, Maucha Andrade. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD : 418.02

ORGANIZAÇÃO

Maucha Andrade Gamonal
Flávia Affonso Mayer
Frederico Belcavello
Igor Antônio Lourenço Da Silva
Mariane De Carvalho Pinto
Adriana Silvina Pagano
Tiago Timponi Torrent



FICHA TÉCNICA

Projeto Gráfico

Marcelo Viridiano

Realização

ReINVenTA - Research and Innovation Network for Vision and Text Analysis
Universidade Federal de Juiz de Juiz de Fora | FrameNet Brasil
Universidade Federal de Minas Gerais | LETra
Universidade Federal de Uberlândia
Universidade Federal da Paraíba

Financiamento

FAPEMIG – Fundação de Amparo a Pesquisa do Estado de Minas Gerais. (Processo nº CHE-RED-00106-21)
CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico. (Processo nº 420945/2022-9)

SUMÁRIO

1. Introdução	6
2. Fundamentos da Semântica de Frames e da FrameNet Brasil	12
3. Tradução Audiovisual Acessível	22
4. Audiodescrição Fílmica: Princípios e Desafios	26
4.1. Diretrizes e outras normativas	26
4.2. A tensão entre objetividade e interpretação	29
4.3. Semântica de Frames e Audiodescrição	30
5. Audiodescrição Fílmica Baseada em Frames	34
5.1. O que significa audiodescrever com frames semânticos	35
5.2. Como os frames orientam a seleção e organização do conteúdo	44
5.3 Consulta à FrameNet Brasil como apoio à audiodescrição	53
6. Inteligência Artificial na Audiodescrição Fílmica	64
6.1. Potenciais da IA orientada por frames na geração e apoio à AD	65
6.2. Limites, riscos e desafios éticos do uso de IA	73
6.2.1. <i>Vieses algorítmicos</i>	74
6.2.2. <i>Alucinações artificiais</i>	75
6.3 Diretrizes para uso responsável da IA	76



01 INTRODUÇÃO

1. Textos inseridos entre imagens para transmitir diálogos, explicar ações, contextualizar a narrativa ou marcar transições tempo.

Desde os primórdios do cinema, a imagem em movimento nunca se apresentou como completamente autossuficiente. O público sempre contou com algum tipo de orientação para acompanhar a narrativa – seja por meio dos intertítulos¹ do cinema mudo, seja pelas convenções narrativas e técnicas que se consolidaram ao longo do século XX, como a montagem clássica hollywoodiana. Esses recursos funcionam, historicamente, como **guias de interpretação** voltados a um público capaz de acessar diretamente as imagens, os textos escritos em cena e os códigos visuais do cinema.

No presente século, contudo, o compromisso social com acessibilidade e inclusão tem ampliado o entendimento de quem participa da **experiência cinematográfica**. Nesse contexto, a audiodescrição surge como um recurso fundamental para garantir que pessoas com deficiência visual ou com baixa visão também possam acessar e vivenciar obras audiovisuais. Mais do que um recurso técnico, a audiodescrição atua como uma forma de

mediação que permite ao espectador navegar pela complexidade das imagens, preservando o princípio de que a experiência estética deve ser compartilhável.

No cinema, essa mediação é especialmente relevante porque o sentido não se constrói apenas a partir da imagem isolada, mas da articulação entre diferentes **modos semióticos**: enquadramentos, cores, movimentos de câmera, gestos, expressões faciais, iluminação, trilhas sonoras e diálogos. Para o público com deficiência visual ou baixa visão, a audiodescrição precisa funcionar como uma ponte entre esses modos, tornando acessíveis informações que, de outro modo, permaneceriam restritas ao campo visual.

Nesse processo, o audiodescritor assume papel central. Ao descrever ações, cenários, expressões faciais e atmosferas, ele orienta a percepção do espectador sem impor interpretações fechadas, o que requer um equilíbrio entre congruência com o produto audiovisual e abertura interpretativa. Com isso, esse profissional – em colaboração com outros agentes, como consultores com deficiência visual e locutores – contribui para que o filme continue sendo um espaço de fruição estética, não um relatório oral dos acontecimentos em tela.

A prática da audiodescrição também evidencia o **caráter social do cinema**. Filmes ganham vida na interação entre pessoas. O gesto de descrever democratiza o acesso e reafirma a ideia de que o conhecimento estético é um patrimônio comum, construído coletivamente. Em tempos de aceleradas transformações tecnológicas, manter viva essa sensibilidade talvez seja o traço mais clássico e, paradoxalmente, mais moderno da audiodescrição.

Apesar de sua importância, ainda persistem dúvidas recorrentes sobre como audiodescrever e o que priorizar em uma audiodescrição fílmica. Em contextos artísticos, costuma-se afirmar que uma boa audiodescrição deve respeitar o estilo da obra, evitar re-

Mais do que um recurso técnico, a audiodescrição atua como uma forma de mediação que permite ao espectador navegar pela complexidade das imagens, preservando o princípio de que a experiência estética deve ser compartilhável.

dundâncias e nomear com precisão sem interpretar em excesso. Essas orientações são relevantes, mas, muitas vezes, permanecem vagas para quem está em formação ou iniciando sua atuação profissional.

Audiodescritores em início de carreira, bem como profissionais envolvidos na produção audiovisual, frequentemente, sentem falta de diretrizes mais concretas sobre o que selecionar, como organizar o conteúdo e como justificar suas escolhas. Ao mesmo tempo, tratar a audiodescrição apenas como cumprimento de normas pode limitar seu caráter expressivo e narrativo, especialmente no cinema, em que cada obra mobiliza recursos estéticos próprios.

Além disso, o cinema contemporâneo tem explorado linguagens cada vez mais diversas: linguagens híbridas, formatos imersivos, multiplicidade de telas, gestos sem diálogo, cortes abruptos, montagens fragmentadas e estruturas narrativas não lineares. Essas características desafiam modelos de audiodescrição excessivamente padronizados e mostram que não há soluções universais capazes de dar conta de toda a variedade do audiovisual.

Outro aspecto fundamental é reconhecer que o público com deficiência visual ou baixa visão, assim como o público vidente, é diverso em suas formas de fruição cinematográfica. Há quem prefira descrições mais objetivas e concisas; há quem valorize maior densidade expressiva; há ainda quem deseje mais contextualização cultural ou narrativa. Essa diversidade torna improvável a existência de um único modelo ideal de audiodescrição.

Diante disso, práticas como testes de recepção, diálogo com usuários e revisão constante de procedimentos tornam-se essenciais. Ao mesmo tempo, essa pluralidade reforça a importância de compreender a audiodescrição como um gênero criativo, aberto a diferentes escolhas linguísticas e estilísticas, e não apenas como um procedimento técnico de conversão de imagens em palavras.

É nesse cenário que se insere o **Guia ReINVenTa para Audiodescrição Fílmica**. Este material não pretende substituir outros guias ou manuais já existentes, mas oferecer uma forma complementar de pensar a audiodescrição a partir de uma base linguística: a **Semântica de Frames**. Nosso objetivo é mostrar como essa abordagem pode auxiliar audiodescritores — pesquisadores ou não — a tomar decisões mais conscientes sobre o que e como descrever, considerando a construção de sentido no cinema.

O guia foi desenvolvido no âmbito de uma rede de pesquisa financiada pela Fundação de Apoio em Pesquisa do Estado de Minas Gerais (FAPEMIG) e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a qual congrega pesquisadores de diferentes laboratórios e grupos de pesquisas situados em diversas universidades mineiras (Universidade Federal de Juiz de Fora, Universidade Federal de Minas Gerais, Universidade Federal de Uberlândia, PUC-Minas) e outras universidades nacionais (Universidade Federal da Paraíba) e estrangeiras (Case Western Reserve University e Universität Leipzig).

A rede, denominada **ReINVenTA (Research and Innovation Network for Vision and Text Analysis)**, desenvolve pesquisas que vão da iniciação científica ao pós-doutoramento e tem como foco a investigação da multimodalidade, especialmente no contexto do programa de pesquisa lexicográfica e computacional **FrameNet Brasil (FN-Br)**.

Este Guia se apoia nesse percurso de pesquisa e se apresenta como uma base analítica e interpretativa ancorada na Semântica de Frames. Ao longo do material, mostramos como essa perspectiva linguística pode contribuir para a criação de roteiros de audiodescrição fílmica, oferecendo apoio às escolhas descritivas e à construção de sentido na experiência cinematográfica.

Este material não pretende substituir outros guias ou manuais já existentes, mas oferecer uma forma complementar de pensar a audiodescrição a partir de uma base linguística: a Semântica de Frames.

O material está estruturado da seguinte maneira:

A **Seção 2** apresenta os fundamentos da Semântica de Frames e suas investigações no âmbito da FrameNet Brasil.

A **Seção 3** aborda a Tradução Audiovisual Acessível, situando a audiodescrição no campo mais amplo das práticas tradutórias intersemióticas.

A **Seção 4** discute a audiodescrição fílmica, seus princípios e desafios, incluindo um panorama das diretrizes e normativas existentes, além da tensão entre objetividade e interpretação.

A **Seção 5** constitui o núcleo deste guia e apresenta a proposta de uma audiodescrição fílmica baseada em frames, detalhando o que significa audiodescrever a partir dessa abordagem e como os frames orientam a seleção e a organização do conteúdo descritivo.

Por fim, a **Seção 6** discute o uso da Inteligência Artificial na audiodescrição fílmica, explorando seus potenciais, limites, riscos e desafios éticos, bem como diretrizes para um uso responsável dessas tecnologias no apoio à prática da audiodescrição.



02

FUNDAMENTOS DA SEMÂNTICA DE FRAMES E DA FRAMENET BRASIL

2. A teoria da Semântica de Frames tem como base os trabalhos de Fillmore (1982, 1985), adotados em todas as pesquisas desenvolvidas no âmbito da FrameNet Brasil.

A Semântica de Frames² é um programa de pesquisa em semântica empírica que propõe o **frame** como a estrutura central para a compreensão do significado linguístico. Em inglês, frame significa literalmente “**moldura**” ou “**quadro**”, mas, nesse contexto, o termo se refere a uma cena mental ou situação típica que organiza nossa interpretação do que está sendo dito ou mostrado. Em termos simples, frames representam os “cenários de fundo” que ativamos mentalmente para compreender ações, eventos, estados e experiências do cotidiano. Neste guia, sempre que um frame for mencionado, ele será apresentado, na sequência, na forma de um quadro que segue a estrutura básica mostrada no quadro de exemplo.

Nome_do_Frame (Indica, de forma concisa, a cena representada)

Definição do frame. (Apresenta a descrição da cena que é definida pelo frame)

★ **Nome do Elemento de Frame Essencial:** Definição do elemento essencial.








☰ “Um exemplo de como o **elemento essencial** seria usado na AD.”

+ **Nome do Elemento de Frame Complementar:** Definição do elemento complementar.

☰ “Um exemplo de como o **elemento complementar** seria usado na AD.”

Por exemplo, quando alguém diz “**visitamos o centro histórico da cidade com um guia local**”, não pensamos apenas em pessoas andando por um lugar. Ativamos uma cena relacionada à atividade de turismo: há alguém que visita (*Turista*), um local reconhecido por seu valor cultural ou histórico (*Atração*), um espaço específico (*Lugar*) e uma finalidade implícita, como conhecer, aprender ou apreciar aquele lugar. Essa cena corresponde ao frame **Fazer_turismo**, mostrado no Quadro 1, que organiza nossa compreensão do evento mesmo quando nem todos esses elementos são explicitados linguisticamente.

Quando falamos em frames, estamos falando também de **ativação de sentido**. Um frame entra em funcionamento quando uma palavra, uma expressão ou mesmo um gesto em cena nos faz reconhecer uma situação conhecida. Na língua, isso acontece por meio de palavras que acionam essas cenas mentais. Na Semântica de Frames, essas palavras são chamadas de *Unidades Lexicais* — mas, em termos simples, podemos pensar nelas como **palavras-chave que fazem uma situação inteira ganhar sentido**.

Fazer_turismo
<p>Este frame representa situações em que uma pessoa realiza uma atividade turística, visitando uma atração ou lugar reconhecido socialmente por sua relevância histórica, cultural, artística ou simbólica. O objetivo principal dessa experiência costuma ser conhecer, observar e aprender sobre a atração. Em muitos casos, o próprio local oferece informações sobre si mesmo, como placas, guias, exposições, panfletos ou recursos audiovisuais.</p>
<p>★ Turista: É a pessoa que realiza a visita e vivencia a experiência turística.  “Os visitantes caminham lentamente pelo museu.”</p>
<p>★ Atração: É o local, objeto ou espaço socialmente reconhecido como relevante e que motiva a visita turística.  “Ela aprecia a fachada histórica da catedral.”</p>
<p>★ Lugar: Indica a localização geral onde a atividade turística acontece.  “Eles visitam o museu no centro da cidade.”</p>
<p>+ Duração: Indica quanto tempo a atividade turística dura.  “Eles passam a tarde inteira no museu.”</p>
<p>+ Finalidade: Aponta o objetivo do turismo, como lazer, entretenimento ou aprendizado.  “Eles visitam o local para conhecer a história da região.”</p>
<p>+ Maneira: Descreve como a experiência acontece ou como o turista se sente durante a visita.  “Ela observa as exposições com curiosidade.”</p>
<p>+ Tempo: Indica quando a atividade turística ocorre.  “A visita acontece durante a manhã.”</p>

Quadro 1: O frame de Fazer_turismo.

Ao ativar um frame, essas palavras não funcionam sozinhas. Elas se conectam a outras palavras da frase, que passam a ocupar papéis específicos dentro da cena evocada. Esses papéis correspondem aos participantes, objetos ou circunstâncias que normalmente fazem parte daquela situação. É essa rede de relações que permite que o sentido se construa no contexto — e não de forma isolada.

Isso está diretamente ligado a um princípio central da Semântica de Frames: **as palavras não carregam significado fixo por si mesmas**. O que elas fazem é ativar sentidos diferentes conforme a situação em que aparecem. Vamos avaliar dois exemplos com o verbo *ver*:

- (1) Edimburgo é uma das cidades mais lindas que já **vi**.
- (2) Onde você **vê** Singapura daqui a dez anos?









Na primeira frase, *ver* está ligado a uma experiência sensorial, dizemos que evoca o frame **Experiência_de_percepção**, como apresentado no Quadro 2. A cena ativada é a de alguém que percebe algo visualmente, sem que isso tenha sido necessariamente planejado ou intencional.

Já na segunda frase, *ver* não tem relação direta com os olhos, mas com expectativa e projeção de futuro. Aqui, *ver* significa imaginar, prever ou antecipar uma situação que ainda não aconteceu. Por isso, o frame ativado é outro: **Expectativa**, como apresentado no Quadro 3.

Embora a palavra seja a mesma, o sentido muda porque **o frame ativado é outro**. No primeiro caso, trata-se de uma experiência perceptiva; no segundo, de uma expectativa em relação ao futuro. As palavras ao redor — *“cidades, daqui a dez anos”* — ajudam a definir qual cena está em jogo e quais papéis precisam ser preenchidos para que o sentido se complete.

Experiência_de_percepção
<p>Este <i>frame</i> representa situações em que um personagem tem uma experiência sensorial sem agir de forma intencional para perceber. A percepção acontece com o personagem, como quando ele vê algo de repente, ouve um som inesperado, sente um cheiro ou percebe uma sensação no corpo.</p> <p>Diferentemente de situações em que o personagem procura algo com os olhos ou presta atenção de forma deliberada, aqui a percepção ocorre de maneira espontânea ou involuntária.</p>
<p>★ Perceptor_passivo: É o personagem que vivencia a experiência perceptiva, sem necessariamente buscá-la de forma consciente ou intencional.</p> <p>💬 “Ela percebe um movimento.”</p>
<p>★ Fenômeno: É aquilo que é percebido pelo personagem. Pode ser uma pessoa, um objeto, um som, uma mudança de luz ou qualquer estímulo sensorial relevante.</p> <p>💬 “Ela percebe um movimento atrás da porta.”</p>
<p>★ Parte_do_corpo (quando relevante): Indica a parte do corpo pela qual a percepção ocorre, especialmente em sensações físicas ou de toque. Nem sempre precisa ser mencionada, pois muitas percepções já indicam implicitamente o sentido envolvido.</p> <p>💬 “Ele sente o frio nas mãos.”</p>
<p>+ Direção: Indica para onde a atenção perceptiva do personagem se volta.</p> <p>💬 “Ela percebe algo à sua esquerda.”</p>
<p>+ Lugar: Indica o espaço geral onde a percepção acontece.</p> <p>💬 “Ele ouve um barulho no corredor.”</p>
<p>+ Grau: Indica a intensidade da percepção.</p> <p>💬 “Ele sente levemente o chão vibrar.”</p>

Quadro 2: O frame de Experiência_de_percepção.

Expectativa
<p>Este frame representa situações em que um personagem demonstra, por meio de comportamentos observáveis, que acredita que algo acontecerá no futuro. A expectativa não é descrita como um pensamento interno, mas inferida a partir de ações, expressões, gestos ou da organização da cena.</p> <p>Na AD, esse frame é utilizado com cautela, quando a expectativa do personagem é claramente perceptível a partir de sinais visuais ou do contexto narrativo. Os elementos do frame funcionam principalmente como categorias de análise da cena: nos exemplos de AD, eles não correspondem, necessariamente, a informações explicitamente verbalizadas, sendo muitas vezes inferíveis a partir do que é descrito.</p>
<p>★ Pensador: É o personagem que demonstra expectativa em relação a um evento futuro, por meio de ações ou expressões visíveis.  “Ela olha repetidamente para o relógio.”</p>
<p>★ Fenômeno: É o evento futuro ao qual a expectativa se relaciona, quando esse evento pode ser inferido a partir da cena.  “Ela olha para a porta fechada.”</p>
<p>★ Tópico: É a entidade ou situação que concentra a expectativa do personagem.  “Ele segura o celular, atento à tela.”</p>
<p>+ Evidência: Indica o sinal visual ou contextual que sustenta a expectativa.  “O céu escurece, e o vento aumenta.”</p>
<p>+ Lugar: Indica o local associado ao evento esperado.  “Ele aguarda na plataforma vazia.”</p>
<p>+ Maneira: Indica como o personagem manifesta a expectativa.  “Ele observa a estrada com o corpo inclinado para frente.”</p>
<p>+ Tempo: Indica o momento em que a expectativa se manifesta na cena.  “No fim da tarde, ela permanece à janela.”</p>
<p>+ Tempo do fenômeno: Indica quando o evento esperado parece estar prestes a acontecer, conforme o contexto visual.  “O painel anuncia a chegada do trem.”</p>

Quadro 3: O frame de Expectativa.

O audiodescritor, ao observar uma cena, pode se perguntar: que tipo de situação está sendo construída aqui? Que experiência está sendo proposta ao espectador?

3. Para mais informações sobre a rede FrameNet Brasil, recomendamos acessar o site <https://www2.ufjf.br/framenetbr/> Acesso em 20 de jan. de 2026.

No cinema, esse funcionamento não depende apenas das palavras. Muitas vezes, são as imagens que ativam os frames. Um plano aberto de uma cidade, um personagem observando a paisagem, um silêncio prolongado acompanhado de um olhar atento: tudo isso pode evocar uma experiência perceptiva, uma expectativa, uma lembrança ou um desejo — mesmo sem diálogo.

É aqui que a Semântica de Frames começa a dialogar diretamente com a audiodescrição. O audiodescritor, ao observar uma cena, pode se perguntar: **que tipo de situação está sendo construída aqui?** Que experiência está sendo proposta ao espectador? Trata-se de perceber algo, esperar algo, decidir algo, lembrar algo?

A FrameNet Brasil³ é o projeto que transforma essa teoria em uma base organizada de dados. Ela reúne descrições de frames e mostra como diferentes palavras e expressões costumam ativá-los em contextos reais de uso. Cada frame é descrito a partir de seus **elementos** — isto é, das partes que normalmente compõem aquela cena —, e essas descrições são construídas a partir da análise de exemplos reais.

Mais recentemente, esse trabalho foi ampliado para além do texto verbal. Na abordagem multimodal da FrameNet Brasil, parte-se da ideia de que imagens e vídeos também evocam frames, assim como as palavras. Um elemento visual pode ativar uma cena inteira ou completar informações que o áudio não explicita.

Por exemplo, a frase “*Edimburgo é uma das cidades mais lindas que já vi*” é proferida pelo apresentador Pedro Andrade no programa Pedro Pelo Mundo, exibido pelo canal televisivo GNT. Considerada isoladamente, a frase não informa quem é essa pessoa nem exatamente o que está sendo visto. Essas informações, no entanto, são contextualizadas na Figura 1.



Figura 1: Cena da série Pedro pelo Mundo, em que o apresentador caminha com outra pessoa por uma rua da cidade de Edimburgo. Captura de tela do software de anotação semântica WebTool⁴.

4. Audiodescrição referente à Figura 1: Pessoas caminham por uma rua de Edimburgo ao entardecer, com prédios ao lado e o sol baixo ao fundo.

Por meio da imagem, é possível compreender quem está vivendo a experiência de ver (o apresentador), o que está sendo visto (parte da cidade, representada pelo espaço urbano enquadrado pela câmera) e em que contexto essa experiência acontece (enquanto ele caminha com outra pessoa por uma rua da cidade).

Essa complementação visual torna explícitos elementos fundamentais da cena evocada pelo verbo ver. Enquanto a frase ativa uma experiência perceptiva, a imagem preenche

5. Para aprofundamento na abordagem multimodal da FrameNet Brasil, sugerimos Torrent et al. (2022), Belcavello et al. (2024), Viridiano et al. (2024), Abreu e Matos (2025) e Gamonal et al. (2025b).

papéis que permanecem implícitos no texto verbal. **Na anotação semântica multimodal⁵**, o personagem que caminha pela rua corresponde ao **Elemento de Frame** *Perceptor_passivo*, isto é, aquele que vivencia a percepção, e o espaço urbano mostrado corresponde ao Elemento de Frame *Fenômeno*, aquilo que é percebido.

É justamente esse tipo de articulação entre linguagem e imagem que a anotação procura tornar visível. A proposta da FrameNet Brasil parte da ideia de que imagens e vídeos também evocam frames, de modo semelhante às palavras. Pessoas em cena, objetos, espaços, gestos ou movimentos podem ativar uma situação inteira ou complementar informações que o áudio não explicita. Texto, som e imagem, assim, trabalham juntos na construção do sentido, oferecendo diferentes perspectivas sobre a mesma cena.

Os frames são especialmente úteis porque permitem representar situações recorrentes do nosso cotidiano, como:

- ações e eventos (comprar algo, viajar, caminhar pela cidade);
- tipos de pessoas ou objetos (papéis sociais, artefatos, personagens);
- estados mentais e emocionais (pensar, desejar, desconfiar);
- e situações tipicamente cinematográficas, em que o sentido emerge da relação entre enquadramento, movimento, som e narrativa.

A anotação semântica multimodal, nesse processo, é fundamental. É por meio dela que se registram, de forma explícita, quais frames estão sendo ativados e como seus elementos aparecem no texto, no áudio e na imagem. Essa prática sustenta tanto a pesquisa linguística quanto aplicações computacionais, mas também ajuda a tornar visível algo que, no cinema, costuma operar de maneira implícita.

Uma forma simples de imaginar tudo isso é pensar na Semântica de Frames como um **mapa mental**. Sempre que uma palavra é dita ou uma cena é mostrada, esse mapa se acende, trazendo consigo pessoas, lugares, objetos e ações típicas daquela situação. A anotação multimodal sobrepõe esse mapa ao filme, conectando linguagem, imagem e som na construção do sentido.

Para a audiodescrição, essa perspectiva é especialmente valiosa. Ela oferece uma base para compreender o que está acontecendo em uma cena antes mesmo de decidir como descrever. Em vez de apenas converter imagens em palavras, o audiodescritor passa a atuar como alguém que reconhece cenas, ativa frames e escolhe, conscientemente, quais elementos são mais relevantes para que o espectador tenha acesso à experiência cinematográfica como um todo.



Uma forma simples de imaginar tudo isso é pensar na Semântica de Frames como um mapa mental. Sempre que uma palavra é dita ou uma cena é mostrada, esse mapa se acende, trazendo consigo pessoas, lugares, objetos e ações típicas daquela situação.

TRADUÇÃO AUDIOVISUAL ACESSÍVEL

6. Pagano, Teixeira e Mayer (2020).

Para inserir a Semântica de Frames como teoria de base para a análise e a produção de audiodescrições, é necessário compreender, antes, o escopo mais amplo da **Tradução Audiovisual Acessível**⁶ (TAVa). A TAVa é a subárea dos Estudos da Tradução dedicada às práticas de mediação de conteúdos multimodais para públicos que, por razões sensoriais, linguísticas ou cognitivas, encontram barreiras no acesso pleno ao produto audiovisual.

Esse campo consolidou-se historicamente a partir do reconhecimento de que o cinema, a televisão e, mais recentemente, as plataformas digitais dependem da articulação de

múltiplos modos de significação — imagem, som, fala, gesto, música e silêncio — e, por isso, exigem estratégias tradutórias específicas para que a experiência narrativa e estética possa ser efetivamente compartilhada. Entre as modalidades mais difundidas da TAVA estão a audiodescrição, a legendagem para surdos e ensurdecidos (LSE), a interpretação em Língua de Sinais, a dublagem acessível, além de adaptações textuais ou visuais voltadas a públicos neurodivergentes. Todas essas práticas reafirmam que a acessibilidade não constitui uma concessão periférica, mas uma dimensão estruturante do processo tradutório em contextos multimodais.

Nesse cenário, a audiodescrição se destaca como uma forma paradigmática de **tradução intersemiótica**. Seu propósito é transpor, de maneira criteriosa e esteticamente informada, elementos visuais e cinéticos para um código verbal sonoro que permita ao ouvinte construir um modelo mental da cena. Não se trata de uma simples “leitura” da imagem, mas de um processo tradutório que envolve interpretação e escolhas conscientes por parte do audiodescritor, incluindo decisões lexicais, sintáticas e retóricas que buscam equilibrar fidelidade ao produto audiovisual, clareza comunicativa e coerência narrativa.

Ao descrever ações, eventos, gestos, ambientes, expressões faciais, movimentos de câmera e atmosferas, a audiodescrição reorganiza o sistema semiótico original para outro canal perceptivo. O objetivo é possibilitar que as relações narrativas, afetivas e culturais centrais da obra sejam acessíveis por meio do som, de modo convergente com a experiência proposta ao público vidente. Por essa razão, a audiodescrição insere-se na mesma tradição tradutória que compreende a linguagem como prática social de construção de sentido orientada por um material-fonte.

Quando analisamos a TAVA sob a perspectiva da Semântica de Frames, observamos que, tanto na audiodescrição quanto nas demais modalidades acessíveis, o tradutor — aqui entendido em sentido amplo, incluindo audiodescritores, legendadores e outros profis-

Ao descrever ações, eventos, gestos, ambientes, expressões faciais, movimentos de câmera e atmosferas, a audiodescrição reorganiza o sistema semiótico original para outro canal perceptivo. O objetivo é possibilitar que as relações narrativas, afetivas e culturais centrais da obra sejam acessíveis por meio do som, de modo convergente com a experiência proposta ao público vidente.

sionais da acessibilidade — ativa e reconstrói, dentro de restrições de tempo e espaço, os frames mais salientes do produto audiovisual. Esse processo envolve decidir o que deve ser explicitado verbalmente, o que pode ser inferido pelo usuário e como organizar essas informações de modo compatível com a dinâmica da obra.

Descrever que “Fulano sorri de lado, desconfiado”, por exemplo, não consiste apenas em enumerar um detalhe visual. Trata-se de evocar um conjunto de conhecimentos e cenas socialmente compartilhadas que permitem ao espectador com deficiência visual ou com baixa visão interpretar intenções, emoções e dinâmicas sociais em articulação com as pistas sonoras já presentes no filme. De modo semelhante, uma legenda para surdos e ensurdecidos, ao marcar pausas, ruídos significativos ou entonações específicas, também opera sobre frames, oferecendo informações contextuais que possibilitam a compreensão da situação comunicativa original.

Dessa forma, as práticas de TAVa podem ser compreendidas como instâncias privilegiadas de materialização da Semântica de Frames. A cada decisão tradutória, o profissional seleciona quais elementos do ou dos frames ativados são indispensáveis para que o usuário possa construir interpretações compatíveis com aquelas produzidas pelo público geral, a partir do modo sensorial de que dispõe.

Nesse sentido, o presente guia parte do pressuposto de que a Semântica de Frames dialoga de maneira especialmente produtiva com a TAVa, oferecendo contribuições relevantes para a busca de coerência e convergência semântica entre as produções acessíveis e os produtos audiovisuais que lhes servem de base. Ao explicitar as estruturas de sentido mobilizadas pelas obras cinematográficas, essa abordagem fornece subsídios teóricos e analíticos para decisões mais conscientes na prática da audiodescrição.

Na próxima seção, discutiremos os princípios e desafios da audiodescrição fílmica, abordando as tensões e escolhas que antecedem a elaboração de um roteiro de audiodescrição e que fundamentam as propostas apresentadas neste guia.





AUDIODESCRIÇÃO FÍLMICA: PRINCÍPIOS E DESAFIOS

Pensar de maneira prática a audiodescrição baseada em frames implica conhecer mais de perto o universo do cinema e do audiovisual no Brasil, bem como o modo como a audiodescrição vem sendo concebida, regulamentada e praticada nesse contexto. Nesta seção, discutimos como a audiodescrição é abordada na legislação brasileira e apresentamos princípios, desafios e procedimentos que orientam a produção da audiodescrição fílmica.

4.1. Diretrizes e outras normativas

A audiodescrição fílmica constitui um componente essencial das políticas públicas de acessibilidade comunicacional no Brasil. Ao longo deste Guia, utilizaremos o termo **audiodescrição** (AD) para nos referirmos a essa prática. Ela figura como critério avaliativo em projetos submetidos a leis de incentivo, bem como em produtos e eventos audiovisuais financiados com recursos públicos. No país, a audiodescrição é regulamentada por um conjunto de normativas, tendo como principal referência o

Guia para Produções Audiovisuais Acessíveis⁷, publicado pelo Ministério da Cultura e complementado por documentos oriundos de pesquisas acadêmicas e de políticas de acessibilidade voltadas para emissoras de televisão e plataformas de streaming.

Nesse sentido, a **Lei nº 13.146/2015**, conhecida como Lei Brasileira de Inclusão (LBI), assegura o direito à comunicação acessível e define a audiodescrição como recurso de tecnologia assistiva. A LBI tem como objetivo garantir e promover, em condições de igualdade, o acesso das pessoas com deficiência aos seus direitos e liberdades fundamentais, visando à sua **inclusão social**⁸. Já o **Decreto Federal nº 5.296/2004** regulamenta as Leis nº 10.048/2000 e nº 10.098/2000, estabelecendo prioridade de atendimento às pessoas com deficiência e fixando normas gerais e critérios básicos para a promoção da acessibilidade de pessoas com deficiência ou com mobilidade reduzida.

Outra normativa relevante é a da **Associação Brasileira de Normas Técnicas** (ABNT), que, por meio da NBR 15290, aborda a acessibilidade na comunicação televisiva, estabelecendo critérios que também se aplicam à narrativa fílmica. A **Agência Nacional do Cinema** (ANCINE), por sua vez, dispõe de três Instruções Normativas que tratam especificamente da audiodescrição.

A **Instrução Normativa nº 116/2014** determina que todos os projetos audiovisuais financiados com recursos públicos federais geridos pela ANCINE devem prever, em seus orçamentos, serviços de acessibilidade, incluindo a audiodescrição. A **Instrução Normativa nº 128/2016** estabelece que as salas de exibição comercial devem dispor de tecnologia assistiva que possibilite ao público a fruição da audiodescrição sempre que solicitada, além de definir que os distribuidores devem fornecer às salas de cinema cópias das obras com os recursos de acessibilidade. Por fim, a **Instrução Normativa nº 165/2022** prevê penalidades para os casos de descumprimento da disponibilização desses recursos quando demandados.

7. Naves et al. (2016).

8. A discussão sobre acessibilidade, deficiência e inclusão no âmbito da Semântica de Frames e da rede FrameNet Brasil é aprofundada em Rabelo, Gamonal e Pagano (2025).

O *Guia para Produções Audiovisuais Acessíveis*, elaborado pelo Ministério da Cultura e pela Secretaria do Audiovisual, aborda **questões técnicas, linguísticas e tradutórias** do processo de produção da audiodescrição. No que se refere ao tempo de inserção das descrições, o documento recomenda que elas sejam posicionadas, preferencialmente, nos intervalos entre os diálogos, de modo a não interferir nos efeitos musicais e sonoros da obra.

O Guia orienta que a **narração da AD** seja fluida e não monótona, acompanhando de forma sutil a atmosfera emocional da cena, sem competir com o produto audiovisual. A AD deve integrar-se de modo orgânico à experiência fílmica, evitando a sobreposição a diálogos e sons relevantes para a compreensão do enredo. Excepcionalmente, quando cenas inteiras são ocupadas por sons significativos, a audiodescrição pode cobrir trechos de menor relevância sonora, desde que isso seja indispensável para a compreensão visual da narrativa.

No que diz respeito ao **uso da linguagem**, o Guia recomenda descrições diretas, simples e sucintas, com variação lexical e adequação à poética e à estética da obra audiovisual. O uso de adjetivos e advérbios é considerado relevante, pois contribui para tornar ações, cenas, personagens e ambientes mais compreensíveis ao espectador. Esses recursos devem expressar estados de humor e emoções que estejam explicitamente presentes no produto audiovisual, evitando inferências subjetivas por parte do audiodescritor. Quanto aos verbos, recomenda-se o emprego de formas específicas e precisas, que indiquem claramente a realização das ações, bem como o uso do presente do indicativo, que favorece a fluidez e a simultaneidade entre descrição e ação.

Em relação à **apresentação de personagens**, o Guia orienta que se priorizem características relevantes para a trama, seguindo, de modo geral, a seguinte sequência: gênero, faixa etária, etnia, cor da pele, estatura, olhos, cabelos e características marcantes. A

nomeação dos personagens deve ocorrer conforme sua identificação na narrativa; até que isso aconteça, recomenda-se que sejam referidos por características físicas ou funcionais.

Mudanças de cenário, ambiente ou tempo devem ser explicitadas assim que ocorrem. Expressões como “é dia”, “fim de tarde” ou “de madrugada” são importantes para marcar a passagem do tempo diegético, isto é, o tempo que transcorre dentro da história narrada. Da mesma forma, os ambientes devem ser descritos a partir dos elementos mais relevantes para a compreensão da obra. Também se recomenda indicar quantos personagens estão em cena e quem são eles. Quanto ao **figurino**, o Guia sugere iniciar pelas peças maiores e pela parte superior do corpo, avançando depois para peças menores e acessórios, evitando descrições excessivas e repetitivas ao longo das cenas. Importante destacar que não é necessário descrever o figurino de todos os personagens em todas as cenas, para evitar o excesso de informação.

Por fim, **elementos visuais verbais** — como créditos, títulos, textos em cena, legendas e intertítulos — devem ser lidos pelo audiodescritor sempre que forem relevantes para a compreensão da obra. O Guia também recomenda explicitar a **origem de sons** cuja fonte não seja imediatamente inferível, como o som de gotas caindo, que pode corresponder, por exemplo, a uma goteira e não a uma torneira aberta.

4.2. A tensão entre objetividade e interpretação

A audiodescrição audiovisual lida com um desafio recorrente: equilibrar a oferta de informações necessárias para que o público possa interpretar a cena e a interpretação inevitável feita pelo próprio audiodescritor. Esse desafio se dá pelo fato de a audiodescrição audiovisual abranger elementos simbólicos, estilísticos ou ambíguos, frequentemente descritos como um **impasse entre objetividade e subjetividade na audiodescrição**.

Neste guia, defendemos que esse impasse é apenas aparente. Mais do que buscar uma objetividade absoluta, o que se impõe é a necessidade de uma formação sólida do audiodescritor. No campo da tradução — e a audiodescrição se insere nesse campo — a objetividade total é inalcançável, pois toda tradução passa, necessariamente, por escolhas humanas. **Descrever é sempre selecionar, e toda seleção envolve interpretação.**

No campo da tradução — e a audiodescrição se insere nesse campo — a objetividade total é inalcançável, pois toda tradução passa, necessariamente, por escolhas humanas. Descrever é sempre selecionar, e toda seleção envolve interpretação.

O desafio, portanto, não é eliminar a interpretação, mas torná-la responsável e justificada. Isso implica identificar o que é mais relevante para a compreensão da narrativa, reconhecer metáforas visuais, atmosferas e signos culturais, e **fazer escolhas** que garantam inteligibilidade sem acrescentar sentidos que não estejam sustentados pela obra.

Nesse sentido, não se trata de defender uma audiodescrição “neutra”, mas **uma prática cuidadosa**, capaz de lidar com as sutilezas da imagem sem recorrer a julgamentos de valor ou inferências indevidas. Emoções, intenções ou estados psicológicos só devem ser mencionados quando estiverem claramente expressos por elementos visuais ou narrativos do filme.

Para equilibrar essas dimensões, audiodescritores podem recorrer a estratégias como protocolos de decisão, listas de prioridades narrativas, glossários de personagens, cenários e objetos, além do diálogo constante com consultores com deficiência visual. Essas práticas contribuem para maior consistência, transparência e qualidade da audiodescrição.

Por fim, uma possibilidade especialmente produtiva — e que orienta o Guia ReINVenTA — é o uso da Semântica de Frames como ferramenta analítica para fundamentar escolhas na audiodescrição fílmica. Ao oferecer critérios para reconhecer situações, participantes e relações relevantes em cada cena, essa abordagem ajuda o audiodescritor a lidar de forma mais consciente e sistemática com a inevitável dimensão interpretativa do seu trabalho.

4.3. Semântica de Frames e Audiodescrição

Ao contrário do que muitas vezes se afirma no campo da audiodescrição, nem tudo o que aparece na tela deve — ou mesmo pode — ser audiodescrito. Isso ocorre, por um lado,

porque nem todas as informações visuais são relevantes para a compreensão da cena e, por outro, porque, no audiovisual, a audiodescrição precisa ser inserida majoritariamente nos intervalos entre diálogos e sons significativos, que costumam ser breves.

Nesse contexto, a prática da AD se aproxima da **noção de usabilidade** proposta por Louise Fryer⁹, autora do livro *An Introduction to Audio Description: A Practical Guide* (ainda não traduzido para o português). Entendida aqui como um critério mais realista para orientar escolhas descritivas, a usabilidade desloca o foco do “descrever tudo” para o “descrever o que é funcional para a experiência do espectador”.

A audiodescrição passa, assim, a atuar como uma camada integrada à trilha audiovisual, contribuindo para a construção de sentido da cena fílmica e substituindo parcialmente — e de forma situada — a experiência visual. Como discutido anteriormente, um frame organiza um **conjunto de elementos semânticos inter-relacionados**, tais como participantes, objetos, ações, papéis e relações. Na AD, esse modelo contribui diretamente para a identificação do que é essencial em cada cena, dentro de um determinado contexto conceitual, orientando a seleção dos elementos nucleares que sustentam o sentido narrativo.

De forma prática, a aplicação da Semântica de Frames pode auxiliar o audiodescritor em diferentes níveis. Em frames de ação ou de entidade, por exemplo, ajuda a determinar quem age, o que é feito, sobre quem ou o quê, e com qual propósito. Em frames de emoção, contribui para identificar situações em que sentimentos são explicitamente representados por gestos ou expressões faciais, auxiliando na distinção entre descrição e inferência indevida. Já em frames espaciais, orienta a organização de elementos como local, objetos salientes e relações de disposição, favorecendo a manutenção da coerência espacial ao longo da narrativa fílmica.

9. Fryer (2016).

Um frame organiza um conjunto de elementos semânticos inter-relacionados, tais como participantes, objetos, ações, papéis e relações. Na AD, esse modelo contribui diretamente para a identificação do que é essencial em cada cena, dentro de um determinado contexto conceitual, orientando a seleção dos elementos nucleares que sustentam o sentido narrativo.

Como se pode observar, o uso da Semântica de Frames favorece um maior **alinhamen-
to entre a descrição e o sentido narrativo da obra**. Além disso, contribui para reduzir a subjetividade arbitrária, oferecendo critérios analíticos mais claros para a tomada de decisões, e facilita o processo interpretativo por parte da audiência da audiodescrição.

A partir desse enquadramento, torna-se possível avançar da fundamentação teórica para a prática da audiodescrição. Se, por um lado, a Semântica de Frames oferece uma forma de compreender como o sentido é construído no audiovisual, por outro, ela também fornece critérios operacionais para orientar escolhas descritivas concretas. Na seção seguinte, apresentamos como essa perspectiva pode ser aplicada à audiodescrição fílmica, detalhando o que significa audiodescrever com frames semânticos e de que maneira eles podem orientar a seleção e a organização do conteúdo descritivo.





AUDIODESCRIÇÃO FÍLMICA BASEADA EM FRAMES

10. Para mais informações sobre a aplicação da Semântica de Frames à audiodescrição, sugerimos Gamonal et al. (2025a); Gamonal et al. (2025b); Dornelas, Gamonal e Pagano (2022), e De Souza, Gamonal e Pagano (2022).

Dando continuidade ao enquadramento teórico apresentado anteriormente, esta seção concentra-se na aplicação prática da Semântica de Frames à audiodescrição fílmica¹⁰. Parte-se da compreensão de que os frames não apenas explicam como o sentido é construído no audiovisual, mas também funcionam como critérios operacionais para orientar escolhas descritivas concretas.

Assim, discutimos o que significa audiodescrever com frames semânticos e como essa perspectiva pode orientar a seleção e a organização do conteúdo da audiodescrição, deslocando o foco da enumeração de elementos visuais para a identificação das situações interpretativas centrais mobilizadas pela cena.

5.1. O que significa audiodescrever com frames semânticos

O processo de audiodescrição fílmica envolve **compartilhamento de significados entre sistemas semióticos distintos** — da modalidade visual para a verbal —, o que caracteriza sua **natureza intersemiótica**. No cinema, a audiodescrição não apenas verbaliza o que é visto, mas participa ativamente da **construção de sentido da cena**, articulando-se com elementos sonoros, narrativos e estilísticos da obra.

Sob a perspectiva da Semântica de Frames, a audiodescrição pode ser compreendida como uma prática que ativa, por meio da linguagem, frames semânticos — estruturas de conhecimento que organizam nossa compreensão de ações, estados, emoções, ambientes, relações espaciais, figurinos, caracterização de personagens e efeitos estéticos. Assim como a imagem e o som mobilizam frames no espectador vidente, a audiodescrição busca evocar, verbalmente, aquilo que permita ao espectador com deficiência visual ou baixa visão reconstruir mentalmente a experiência fílmica.

A Figura 2¹¹ apresenta uma imagem do curta-metragem *A árvore do dinheiro*, animação brasileira inspirada na literatura de cordel, integrante da mostra Alumiar – I Mostra de Curtas Pernambucanos Acessíveis On-line (Fundaj). A obra conta com recursos de audiodescrição, Libras e legenda para surdos e ensurdecidos (LSE). O curta foi produzido por Marcos Buccini, Diego Credidio e Luciana de Mari. O roteiro e a narração da audiodescrição foram elaborados por Liliane Tavares, com consultoria de Michelle Alheiros.

A imagem é acompanhada pela seguinte audiodescrição:

(3) Agora, na praça diante da igreja, ele entrega uma flor vermelha para ela e se olham, enamorados.

Assim como a imagem e o som mobilizam frames no espectador vidente, a audiodescrição busca evocar, verbalmente, aquilo que permita ao espectador com deficiência visual ou baixa visão reconstruir mentalmente a experiência fílmica.





11. Todas as imagens apresentadas nesta seção foram extraídas de curtas-metragens previamente anotados semanticamente e integram o dataset Audition, desenvolvido no âmbito do projeto ReINVenTA, da FrameNet Brasil, dedicado à anotação semântica multimodal de obras audiovisuais acessíveis (Gamonal et al., 2025b).

Figura 2: Cena do curta-metragem A árvore do dinheiro. Captura de tela do software de anotação semântica Webtool¹².






12. Audiodescrição referente à Figura 2: Ilustração em preto e branco de um homem e uma mulher em uma praça em frente à igreja; a mulher segura uma flor vermelha.

Este exemplo ilustra como uma única sentença de audiodescrição pode ativar simultaneamente diferentes frames semânticos, organizando a interpretação da cena de forma integrada. A expressão “na praça diante da igreja” ativa os frames **Locais_por_uso** (Quadro 4) e **Proximidade_não_graduável** (Quadro 5), situando a ação em um espaço socialmente marcado e relevante para a narrativa.

Locais_por_uso
<p>Este frame representa lugares definidos principalmente por sua função social ou uso recorrente, e não apenas por suas características físicas. O Local é reconhecido culturalmente por aquilo que se faz nele (por exemplo, sala de aula, cozinha, hospital, ponto de ônibus), o que permite ao espectador ativar rapidamente expectativas espaciais, funcionais e narrativas.</p> <p>Na AD, <i>Locais_por_uso</i> é mobilizado quando a identificação funcional do espaço é suficiente — e mais eficaz — do que uma descrição detalhada de sua aparência. O frame permite economia descritiva, pois o uso do local aciona inferencialmente objetos, disposições espaciais e tipos de ação associados a ele.</p>
<p>★ Local: É o lugar definido por seu uso socialmente reconhecido.  “Ele entra na cozinha.”</p>
<p>★ Uso: Refere-se à atividade típica associada ao Local. (<i>frequentemente implícito</i>)  “Ela espera na clínica (médica).”</p>
<p>+ Descritor: Indica uma característica do Local que não está diretamente ligada ao uso, mas contribui para a construção da cena.  “Eles caminham pela fazenda silenciosa”</p>
<p>+ Localização_relativa: Indica a relação espacial do Local com outro espaço.  “O mercado ao lado da praça está aberto.”</p>

Quadro 4: O frame de **Locais_por_uso**.







Quadro 5: O frame de **Proximidade_não_graduável**.

Proximidade_não_graduável
<p>Uma Figura é caracterizada como próxima de um Fundo, de acordo com uma escala de proximidade definida contextualmente, sem gradação (isto é, não se trata de “mais perto” ou “menos perto”, mas de uma relação categórica de proximidade). Essa relação pode ser inferida a partir de entidades físicas, eventos ou configurações espaciais representadas linguisticamente ou visualmente.</p>
<p>★ Figura: É a entidade (ou evento) cuja posição é descrita em relação a um Fundo.  “O homem está ao lado do carro.”</p>
<p>★ Fundo: É a entidade de referência que serve de base para localizar a Figura..  “A criança permanece perto da mãe.”</p>
<p>+ Tempo: Indica o período durante o qual a relação de proximidade se mantém.  “Durante a conversa, ela permanece próxima da janela.”</p>

O verbo “*entrega*” ativa o frame **Dar** (Quadro 6), estruturando a relação entre os participantes (*ele, ela* e *a flor*), enquanto “*uma flor vermelha*” funciona como elemento simbólico que reforça valores culturalmente associados ao afeto.

Por fim, “*se olham, enamorados*” aciona frames ligados à **Percepção_ativa** (Quadro 7) e à **Emoção_direcionada** (Quadro 8), permitindo que o espectador infira o vínculo afetivo entre os personagens sem que a emoção precise ser explicitada de forma analítica.






Quadro 6: O frame de **Dar**.

Dar	
<p>Este frame representa situações em que um personagem transfere intencionalmente a posse de um objeto para outro personagem. A ação é iniciada pelo Doador, que detém o controle ou a posse do Tema no início da cena e faz com que esse objeto passe a ser possuído pelo Recipiente ao final da ação.</p>	
★	<p>Doador: É o personagem que inicia a ação e que possui o Tema antes da transferência.  “O homem entrega o documento.”</p>
★	<p>Tema: É o objeto que muda de possuidor durante a ação.  “Ele entrega o bilhete.”</p>
★	<p>Recipiente: É a entidade que passa a possuir o Tema após a transferência.  “Ele entrega o bilhete ao colega.”</p>
+	<p>Lugar: É o espaço onde a ação de dar ocorre.  “Na sala de aula, ele entrega o livro.”</p>
+	<p>Maneira: Refere-se a como o Tema é transferido, podendo revelar aspectos expressivos da ação.  “Ela lhe dá o presente com cuidado.”</p>
+	<p>Meio: Indica o modo ou ação intermediária pela qual o Tema é transferido.  “Ele passa o envelope por cima da mesa.”</p>

Quadro 7: O frame de **Percepção_ativa**.

Percepção_ativa
<p>Este frame representa situações em que um Perceptor_agentivo direciona intencionalmente sua atenção para um Fenômeno, com o objetivo de ter uma experiência perceptiva. Diferentemente de percepções passivas, aqui há uma ação voluntária e orientada, como olhar, observar, escutar atentamente ou examinar.</p> <p>Na AD, este frame é utilizado quando a ação perceptiva do personagem é visualmente relevante para a progressão narrativa, para a construção de foco atencional da cena ou para indicar relações entre personagens, objetos e eventos.</p>
<p>★ Perceptor_agentivo: É o personagem que realiza ativamente a ação perceptiva, direcionando sua atenção.</p> <p>💬 “Ela observa a rua pela janela.”</p>
<p>★ Fenômeno: É a entidade ou evento para o qual a atenção do Perceptor_agentivo é dirigida.</p> <p>💬 “O homem examina o mapa.”</p>
<p>† Lugar: Indica o espaço geral onde ocorre o ato perceptivo.</p> <p>💬 “Na sala escura, ele observa a tela.”</p>
<p>† Maneira: Indica como a percepção é realizada, podendo sugerir intensidade ou cuidado.</p> <p>💬 “Ele observa atentamente o movimento.”</p>
<p>† Finalidade (quando inferível): Indica o objetivo visualmente sugerido da ação perceptiva.</p> <p>💬 “Ele examina a fechadura para ver se está trancada.”</p>
<p>† Duração: Indica o período durante o qual a relação de proximidade se mantém.</p> <p>💬 “Ela olha o quadro por alguns segundos.”</p>
<p>† Meio: Indica o instrumento ou método utilizado para perceber o Fenômeno.</p> <p>💬 “Ela observa a cena com binóculos.”</p>
<p>† Tempo: Indica o momento em que a percepção ocorre na progressão da cena.</p> <p>💬 “À noite, ele observa o céu.”</p>

Quadro 8: O frame de **Emoção_direcionada**.

Emoção_direcionada
<p>Este frame representa situações em que um Experienciador manifesta uma resposta emocional orientada a um Estímulo ou a um Tópico, podendo essa emoção ser descrita como um Estado mais duradouro.</p> <p>Na AD, a ativação desse frame acontece quando há forte ancoragem visual ou narrativa que legitime o uso de um traço emocional convencionalizado.</p>
<p>★ Experienciador: É a pessoa ou outro ser vivo que vivencia a emoção.  “O menino permanece parado, satisfeito.”</p>
<p>★ Estado: É o que nomeia a experiência emocional relativamente estável do Experienciador. Na AD, é mencionado quando há consenso interpretativo visual suficiente para sustentar sua inferência.  “Ela observa a cena, emocionada.”</p>
<p>★ Estímulo / Tópico: O Estímulo é a entidade ou acontecimento específico que evoca a emoção; o Tópico designa o domínio geral ao qual essa emoção se relaciona. Na AD, esses elementos frequentemente permanecem implícitos, sendo recuperáveis pelo contexto narrativo.  “Ela demonstra surpresa ao encontrar o diário.”</p>
<p>+ Grau: Indica a intensidade da emoção, quando visualmente marcada.  “Ele parece profundamente emocionado.”</p>
<p>+ Maneira: Refere-se à forma como a emoção se manifesta corporal ou expressivamente.  “Ela reage contida, mas satisfeita.”</p>

A combinação desses frames mostra como a AD seleciona elementos da cena para construir espaço, ação e expressividade, orientando a compreensão narrativa e estética do momento.

Audiodescrição e Frames Semânticos

A audiodescrição é uma modalidade de tradução audiovisual acessível, de natureza intersemiótica, que se baseia na ativação linguística de frames semânticos para reconstruir, em palavras, os sentidos produzidos pela experiência audiovisual.

O que significa audiodescrever com frames

Audiodescrever com frames significa identificar quais situações a cena mobiliza, selecionar os elementos de frame essenciais para a compreensão e escolher palavras e expressões que os evoquem de modo claro, econômico e coerente com a dinâmica multimodal da obra.

Essa compreensão dialoga diretamente com o que já é apontado pelo Guia para Produções Audiovisuais Acessíveis, segundo o qual a audiodescrição consiste em uma locução adicional roteirizada que não se limita à enumeração de elementos visuais, mas reconstrói, pela linguagem, os sentidos ativados pela experiência cinematográfica. A abordagem baseada em frames aprofunda esse princípio ao explicitar a **estrutura cognitivo-semântica** subjacente às escolhas do audiodescritor, oferecendo critérios para selecionar, entre múltiplos aspectos da cena, aqueles que são essenciais tanto para a compreensão narrativa quanto para a experiência estética.

A Figura 3 apresenta *bounding boxes* (caixas delimitadoras), isto é, marcações visuais que indicam explicitamente os elementos da cena destacados pela audiodescrição, como os personagens, a rosa, a igreja e a praça. No âmbito do projeto ReINVenTA, a anotação semântica multimodal também desempenha um papel estratégico para o aprendizado de máquina: ao associar *bounding boxes* e rótulos semânticos a frames e a seus Elementos de Frame, a cena audiovisual é transformada em dados estruturados. Isso torna explícita a relação entre imagem e linguagem verbal, possibilitando tanto a análise linguística e cognitiva quanto o treinamento, a avaliação e o aperfeiçoamento de modelos computacionais capazes de integrar imagem, linguagem verbal e som. Assim, a anotação constitui a base para o desenvolvimento de tecnologias assistivas e de ferramentas de apoio à audiodescrição.

Home / Annotation / Dynamic / Árvore_do_dinheiro_animação

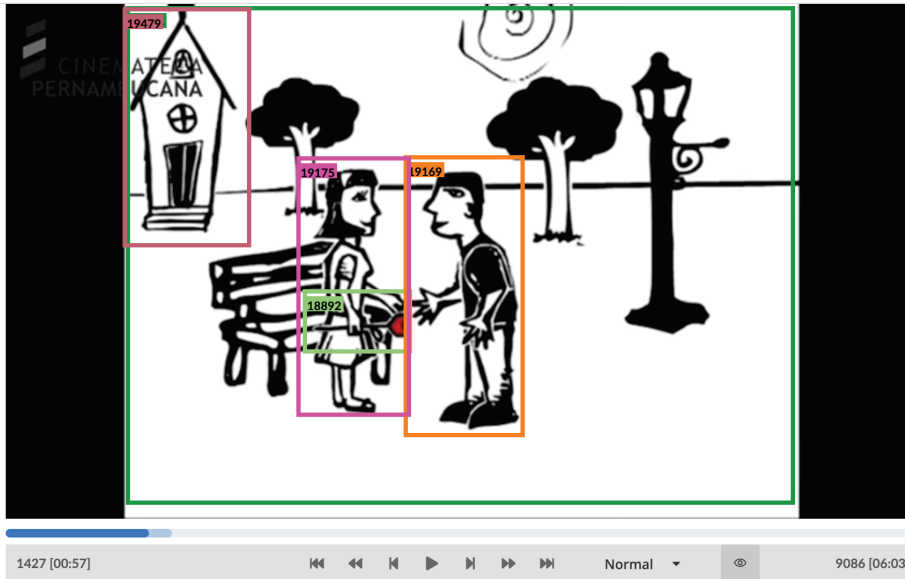


Figura 3: Cena do curta-metragem A árvore do dinheiro com caixas delimitadoras. Captura de tela do software de anotação semântica Webtool¹³.

Dessa forma, uma AD orientada por frames oferece ao audiodescritor critérios analíticos para escolher conscientemente palavras e expressões que ativem sentidos específicos, permitindo ao usuário construir mentalmente ações, intenções e emoções. Como recurso de acessibilidade, a AD mantém sua função social fundamental; como prática discursiva, assume também um papel narrativo integrado ao filme, contribuindo para a dinâmica multimodal da obra.

13. Audiodescrição referente à Figura 3: Ilustração em preto e branco de um homem e uma mulher em uma praça em frente à igreja; a mulher segura uma flor vermelha. Há quatro *bounding boxes* coloridas delimitando, respectivamente, o homem, a mulher, a praça e a igreja ao fundo.

5.2. Como os frames orientam a seleção e organização do conteúdo.

Como discutido na seção anterior, audiodescrever com frames semânticos implica compreender a audiodescrição como uma prática orientada pela ativação linguística de estruturas de conhecimento compartilhadas. A análise de frames permite ao audiodescritor decidir não apenas o que deve ser dito na audiodescrição, mas também como esse conteúdo deve ser organizado de modo coerente com a narrativa e com a experiência estética da obra fílmica. Em vez de partir de listas de elementos visuais ou de categorias gramaticais pré-definidas, o audiodescritor passa a observar quais frames são ativados pela cena e a identificar quais elementos desses frames são essenciais para a compreensão e para a construção interpretativa do espectador.

Esse deslocamento teórico-metodológico reorganiza a prática da audiodescrição em três etapas analíticas interdependentes: (a) a identificação do **frame dominante** da cena; (b) a seleção dos **elementos de frame essenciais**; e (c) a integração desses frames à **dinâmica multimodal da obra cinematográfica**.

a) Identificação do frame dominante da cena.

Cada cena ou segmento de um filme pode ativar simultaneamente diversos frames, incluindo aqueles associados a situações dinâmicas (*Quadro 6*), a estados físicos ou mentais (*Quadro 8*), a relações espaciais (*Quadro 5*) ou a processos perceptivos (*Quadro 7*).

O frame dominante é aquele que expressa o núcleo interpretativo da cena, organizando o que acontece nela e orientando a leitura narrativa proposta pelo filme. Identificá-lo evita que a audiodescrição se disperse em detalhes periféricos e auxilia o audiodescritor a priorizar os elementos essenciais para a construção de sentido. A Figura 4 tem como AD uma sentença do curta-metragem Cinema Glória, produzido por Fernando Spencer e Félix Filho. A audiodescrição, o roteiro e a narração são de Liliana Tavares e consultoria de Michelle Alheiros.

(4) O barbeiro usa uma navalha.

Neste ponto, a audiodescrição não precisa retomar o ambiente da feira nem as ações periféricas ao redor do barbeiro. A partir do som, do áudio original e da sequência anterior da AD¹⁴, o público já construiu mentalmente o cenário de uma feira movimentada, em um espaço aberto, com pessoas circulando. Desse modo, o audiodescritor pode concentrar-se apenas nos elementos visuais centrais da cena — o barbeiro, o cliente e a navalha — garantindo fluidez narrativa e economia informacional.

Home / Annotation / Dynamic / Cinema_Glória_documentário



14. Um pátio ensolarado onde há uma feira movimentada. / Várias pessoas circulam. / Barracas com lona colorida. / Alguns carros parados ao redor da feira. / Muitos guarda-sóis. / Uma grande árvore sombreia a praça. / Um homem de pé lendo jornal. / A imagem desliza para a esquerda. / Pessoas ao redor. / Carros estacionados com porta aberta. / Um espaço grande com lixo no chão. / E vários homens por perto. / Embaixo da sombra da árvore, um barbeiro atende um cliente. / O homem está com a cabeça recostada na cadeira de barbeiro. / Um babador branco sobre a camisa e o rosto com espuma.

Figura 4: Cena do curta-metragem Cinema Glória. Captura de tela do software de anotação semântica Webtool¹⁵.

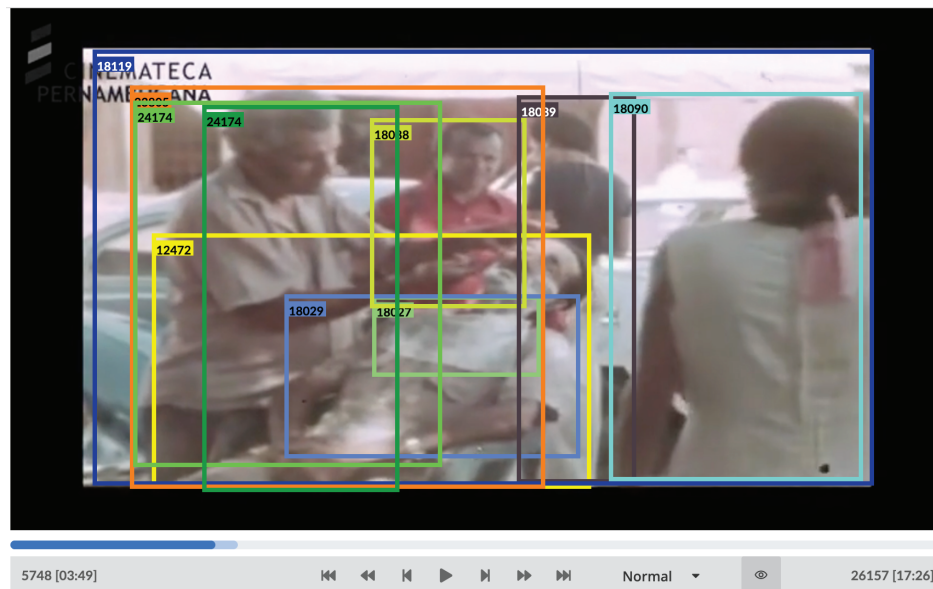
15. Audiodescrição referente à Figura 4: Cena de um barbeiro fazendo a barba de um homem sentado, com rosto coberto por espuma. Ao fundo, carros antigos e outras pessoas, que observam e passam pelo local.

A Figura 5 apresenta a mesma cena com as marcações da anotação semântica multimodal, nas quais diferentes *bounding boxes* delimitam os elementos visuais envolvidos em múltiplas ações simultâneas (diferentes pessoas, babador, espuma, cadeira, navalha etc). Observa-se o barbeiro em atividade, manipulando uma navalha com uma das mãos enquanto, com a outra, encosta no rosto do cliente, que está recostado na cadeira e tem o rosto coberto por espuma branca. Além disso, outros eventos ocorrem no mesmo espaço: um homem observa a cena e uma mulher passa ao lado deles, evidenciando a coexistência de ações centrais e periféricas no ambiente representado.

Figura 5: Cena do curta-metragem Cinema Glória com *bounding boxes*. Captura de tela do software de anotação semântica Webtool¹⁶.

16. Audiodescrição referente à Figura 5: Cena de um barbeiro fazendo a barba de um homem sentado, com rosto coberto por espuma. Ao fundo, carros antigos e outras pessoas, que observam e passam pelo local. Bounding boxes coloridas delimitam diferentes pessoas e objetos, como babador, espuma, cadeira e navalha.

Home / Annotation / Dynamic / Cinema_Glória_documentário



Nesse contexto, o frame dominante funciona como o eixo organizador da audiodescrição, orientando:

- quais ações, estados ou emoções devem ser descritos;
- quais participantes precisam ser explicitados;
- quais elementos de caracterização são relevantes naquele momento;
- quais pistas visuais ou sonoras merecem destaque;
- e qual é o foco interpretativo da cena na progressão narrativa.

A escolha do frame dominante constitui, portanto, o primeiro passo para a produção de uma audiodescrição coerente, precisa e sensível à experiência estética da obra.

b) Seleção dos elementos de frame essenciais.

A seleção dos elementos que compõem a audiodescrição não deve seguir listas fixas de aspectos visuais, mas a lógica dos frames ativados pela cena. Em cada momento do filme, diferentes frames estruturam a interpretação — como **Partes_do_Corpo** (Quadro 9), **Percepção_ativa** (Quadro 7) ou **Chegar** (Quadro 10) — definindo o tipo de situação representada e o que nela se torna relevante para o espectador.

Cada frame é constituído por Elementos de Frame, entendidos como **funções semânticas que organizam a cena**, tais como participantes (*Pessoa, Tema*), características (*Atributo, Tipo*), configurações espaciais (*Lugar, Direção*), modos de realização (*Maneira*), entre outros componentes que qualificam o clima expressivo e o efeito estético das imagens. Esses elementos só adquirem sentido pleno dentro da situação representada e, por isso, devem ser selecionados a partir do frame dominante previamente identificado.

Quadro 9: O frame de **Partes_do_corpo**.

Partes_do_corpo
<p>Este frame representa situações em que uma parte do corpo de um personagem é colocada em foco visual, seja de forma isolada ou em relação ao corpo como um todo. Na AD, este frame é mobilizado quando a menção a uma parte específica do corpo contribui para a compreensão da ação, da postura, da expressividade corporal ou do estado físico do personagem.</p> <p>A descrição de partes do corpo não tem como objetivo listar elementos anatômicos, mas destacar visualmente aquilo que organiza a interpretação da cena, frequentemente em articulação com frames de ação, percepção ou emoção.</p>
<p>★ Parte_do_corpo: É a parte do corpo que se torna relevante na cena por estar em evidência visual, em movimento ou envolvida em uma ação.</p> <p>☰ “As mãos tremem.”</p>
<p>★ Possuidor: É o personagem ao qual a parte do corpo pertence.</p> <p>☰ “Os braços dele encostam na grade.”</p>
<p>+ Descritor: Indica uma característica perceptível da parte do corpo, como forma, estado, aparência ou condição momentânea.</p> <p>☰ “O joelho machucado está inchado.”</p>
<p>+ Localização_orientacional: Descreve a posição da parte do corpo em relação à orientação do corpo do personagem.</p> <p>☰ “A mão direita está sobre a mesa.”</p>
<p>+ Sub-região: Especifica uma região interna ou delimitada da parte do corpo mencionada.</p> <p>☰ “A ponta dos dedos aparece molhada.”</p>

Chegar
<p>Este frame representa situações em que um personagem ou entidade se desloca até um ponto final do movimento, isto é, alcança um Alvo, ainda que esse ponto não seja explicitamente nomeado na descrição.</p> <p>Na AD, o evento de chegada é geralmente ativado pela própria ação que designa o movimento, que pressupõe a existência de um ponto de término, permitindo que o espectador reconstrua a situação a partir de pistas espaciais e contextuais.</p>
<p>★ Tema: É o personagem ou entidade que se desloca e realiza o movimento de chegada, identificado como participante central do evento.</p> <p>☺ “Ele chega ao portão.”</p>
<p>★ Alvo: É o ponto final do movimento, isto é, o local onde o Tema passa a estar após o deslocamento. Na audiodescrição, o Alvo pode ser explicitado ou permanecer implícito, sendo inferido a partir do contexto.</p> <p>☺ “Ela entra no quarto.”</p>
<p>+ Lugar: Indica o espaço mais amplo em que o evento de chegada ocorre, contribuindo para a organização espacial da cena.</p> <p>☺ “Ele chega ao portão da escola.”</p>
<p>+ Fonte: Indica o ponto de início do movimento, quando esse dado é visualmente relevante para a compreensão da cena.</p> <p>☺ “Ela vem do corredor e entra na sala.”</p>
<p>+ Trajectoria: Refere-se ao percurso realizado pelo Tema até o ponto de chegada, geralmente expresso como via ou caminho intermediário.</p> <p>☺ “Ele atravessa o pátio até a entrada.”</p>
<p>+ Maneira: Descreve como o movimento de chegada é realizado, destacando propriedades como velocidade, hesitação ou fluidez.</p> <p>☺ “Ela chega correndo.”</p>
<p>+ Meio: Indica o recurso ou condição através do qual o Tema se desloca até o Alvo.</p> <p>☺ “Ele chega de bicicleta.”</p>
<p>+ Modo de transporte: Especifica se o deslocamento ocorre pelo próprio corpo do Tema ou por meio de um veículo.</p> <p>☺ “O ônibus chega ao ponto.”</p>
<p>+ Tempo: Indica o momento em que a chegada ocorre na progressão da narrativa.</p> <p>☺ “Ao amanhecer, ela chega à praia.”</p>

Quadro 10: O frame de **Chegar**.

Na audiodescrição orientada por frames semânticos, portanto, os Elementos de Frame não funcionam como rótulos a serem preenchidos no texto, mas como estruturas interpretativas subjacentes à seleção e à organização do conteúdo descritivo.

Selecionar o conteúdo da audiodescrição significa, assim, identificar quais elementos são essenciais para que o usuário possa construir mentalmente a ação, o espaço, as relações entre os participantes e a expressividade da cena, preservando seus sentidos narrativos e estéticos. Em vez de partir de categorias gramaticais, a AD se ancora nas funções semânticas ativadas pela imagem e no papel que elas desempenham na organização da experiência cinematográfica.

Essa seleção pode ser orientada por perguntas como:

- Quais elementos são indispensáveis para que o espectador compreenda a situação?
- Quais elementos comprometem a interpretação se forem omitidos?
- Quais contribuem para a qualidade expressiva da cena, produzindo suspense, humor etc?

Ao privilegiar os elementos essenciais, a audiodescrição evita transformar-se em uma simples listagem visual e passa a focalizar aquilo que sustenta a estrutura interpretativa da cena. Em muitos casos, a escolha de um único elemento já permite que o espectador construa mentalmente toda a situação representada.

Em diversas cenas, pequenos detalhes visuais funcionam como **pistas interpretativas** capazes de ativar significados mais amplos. Gestos, expressões faciais, objetos em evidência ou variações de iluminação podem assumir função central na audiodescrição, pois contribuem para a construção do clima da cena, das intenções dos personagens e da progressão narrativa. Essas pistas não correspondem, necessariamente, à explicitação direta e convencionalizada de um Elemento de Frame, mas são suficientes para ativar estruturas de conhecimento mais amplas, permitindo que o espectador construa a situação representada a partir de informações parciais, porém relevantes.

Na audiodescrição orientada por frames semânticos, portanto, os Elementos de Frame

não funcionam como rótulos a serem preenchidos no texto, mas como **estruturas interpretativas subjacentes** à seleção e à organização do conteúdo descritivo. Em muitos casos, a AD não nomeia um Elemento de Frame específico, mas oferece indícios visuais ou contextuais compatíveis com ele, de modo a preservar a capacidade inferencial do espectador.

Essa lógica pode ser observada no exemplo apresentado a seguir. A Figura 6 apresenta uma cena do curta-metragem *Eu Não Quero Voltar Sozinho*, dirigido por Daniel Ribeiro e produzido pela Lacuna Filmes. A audiodescrição foi realizada pelo Grupo Tramad, com roteiro de Sandra Rosa, revisão de Iracema Vilarunga e narração de Adriana Urpia. O curta retrata, de forma sensível, o desenvolvimento do afeto entre dois adolescentes. A AD da cena é:

(5) Ele abaixa a cabeça e dá um sorrisinho.

Home / Annotation / Dynamic / Eu_nao_quero_voltar_sozinho_AD_v01



Figura 6: Cena do curta-metragem *Eu não quero voltar sozinho*. Captura de tela do software de anotação semântica Webtool¹⁷.

17. Audiodescrição referente à Figura 6: Um rapaz de pele clara e cabelos cacheados olha para baixo com um leve sorriso. À direita, outro rapaz, de costas e desfocado.

Essa descrição é inserida imediatamente antes de um beijo romântico entre os personagens. A escolha descritiva não explicita diretamente o estado emocional vivido pelo personagem, mas oferece pistas interpretativas para que o espectador com deficiência visual ou baixa visão construa inferencialmente sentidos associados à timidez, ao constrangimento ou à emoção contida. Do ponto de vista da Semântica de Frames, a descrição ativa inferencialmente o frame **Emoção_direcionada** (Quadro 8) por meio da referência a ações corporais sutis, evitando a nomeação explícita da emoção.

Essa característica é particularmente relevante em frames associados a processos perceptivos ou emocionais, nos quais a descrição direta de estados internos seria inadequada ou eticamente problemática. Nos Quadros 3 e 7, os frames **Expectativa** e **Percepção_ativa** ilustram como a audiodescrição opera a partir de comportamentos observáveis, gestos, objetos em foco ou configurações espaciais que sustentam a ativação do frame, sem determinar de forma explícita o conteúdo exato de cada elemento.

c) Integração dos frames com a dinâmica multimodal da obra cinematográfica.

A audiodescrição orientada por frames semânticos requer que o audiodescritor considere não apenas o que descrever, mas também como cada escolha linguística se articula com a dinâmica multimodal da obra cinematográfica.

No cinema, o sentido é construído pela interação de múltiplos recursos semióticos — como iluminação, gestos, trilha sonora, enquadramento, silêncio e movimento de câmera — que atuam conjuntamente na ativação de frames. Por essa razão, o roteiro de audiodescrição deve levar em conta de que modo esses recursos já estão operando na construção da interpretação da cena.

Integrar frames a essa dinâmica implica:

I. Conceber a cena como uma configuração multimodal de frames semânticos.

Uma mesma sequência pode ativar simultaneamente diversas situações, como **Emoção_direcionada** (Quadro 8) e **Chegar** (Quadro 9). O audiodescritor deve observar quais desses frames estão sendo acionados pelos recursos cinematográficos e decidir como reforçá-los ou introduzi-los verbalmente, respeitando a hierarquia interpretativa proposta pela cena.

II. Descrever apenas o necessário para completar a composição multimodal da obra.

A AD deve evitar redundâncias e preservar a capacidade inferencial do espectador diante dos elementos já percebidos por meio do som, do silêncio ou da organização visual da cena. Desse modo, sua função é complementar os recursos multimodais existentes, priorizando elementos de frame essenciais para a interpretação e para a experiência estética.

III. Ajustar o ritmo, o tempo e a extensão da AD ao fluxo cinematográfico.

O frame dominante de uma cena pode se modificar rapidamente em função de mudanças nos gestos, no enquadramento, na trilha sonora ou na entrada de novos personagens e objetos em cena. A AD deve acompanhar essa dinâmica, descrevendo apenas o que se torna relevante no momento em que assume valor narrativo.

5.3 Consulta à FrameNet Brasil como apoio à audiodescrição

Ao longo do processo de audiodescrição, a consulta à FrameNet Brasil funciona como **ferramenta de apoio ao trabalho do audiodescritor**. Ela não substitui a análise da cena nem define automaticamente o que deve ser descrito, mas auxilia na tomada de decisões quando há dúvidas sobre como caracterizar uma ação, um gesto ou uma situação representada no audiovisual.

Em termos práticos, a FrameNet Brasil pode ser mobilizada quando o audiodescritor precisa confirmar qual situação semântica está sendo acionada por uma ação observada, identificar quais elementos são centrais para a compreensão da cena, comparar escolhas lexicais possíveis ou manter consistência descritiva ao longo de uma obra.

Dito de outra forma

Integrar *frames* à multimodalidade implica compreender que a audiodescrição não é um texto paralelo, mas uma camada narrativa que dialoga diretamente com os demais modos comunicativos da obra cinematográfica. Sua função é reforçar os elementos visuais essenciais para manter a coerência interpretativa da cena.

O objetivo da AD não é verbalizar tudo o que aparece na tela, mas possibilitar que o espectador acesse os *frames* que a obra mobiliza, acompanhando a ordem, a intensidade e a combinação multimodal pela qual a própria cena os apresenta.

Em termos práticos, a FrameNet Brasil pode ser mobilizada quando o audiodescritor precisa confirmar qual situação semântica está sendo acionada por uma ação observada, identificar quais elementos são centrais para a compreensão da cena, comparar escolhas lexicais possíveis ou manter consistência descritiva ao longo de uma obra.

Figura 7: Cena do episódio 7 de Pedro pelo Mundo, mostrando Pedro, que observa um homem fumar um charuto. Captura de tela do software de anotação semântica Webtool¹⁸.

18. Audiodescrição referente à Figura 7: Na rua, o apresentador Pedro conversa com um senhor de terno bege e chapéu de abas largas. O senhor fuma um charuto.

Situação inicial de análise

Para ilustrar esse percurso de consulta, tomamos como ponto de partida uma imagem de Pedro pelo Mundo, programa televisivo que não oferece audiodescrição. A imagem selecionada serve como estímulo visual para simular o processo de tomada de decisão do audiodescritor diante da necessidade de construir uma AD para a cena.

Home / Annotation / Dynamic / PPM_NLG_ep_07



O objetivo não é propor uma audiodescrição finalizada, mas **reconstituir o raciocínio analítico** que antecede a escrita da descrição, evidenciando como a observação da imagem pode orientar a identificação de ações relevantes, a escolha de itens lexicais e a consulta a frames semânticos na base da FrameNet Brasil.

A Figura 7 apresenta uma interação entre duas pessoas que estão em uma rua de Cuba. Uma das pessoas é o apresentador Pedro Andrade, a outra pessoa é um homem mais velho, vestido formalmente. O homem segura um charuto aceso próximo à boca, enquanto Pedro, que veste roupas casuais, mantém o corpo levemente inclinado em sua direção, com a mão apoiada na cintura, acompanhando a ação com o olhar.

Nesse momento da sequência, não há fala associada à imagem, mas uma música cubana de fundo. A construção de sentido se dá, principalmente, pelos elementos visuais: o direcionamento do olhar de Pedro, a ação corporal do homem ao fumar o charuto e a relação espacial estabelecida entre os dois personagens.

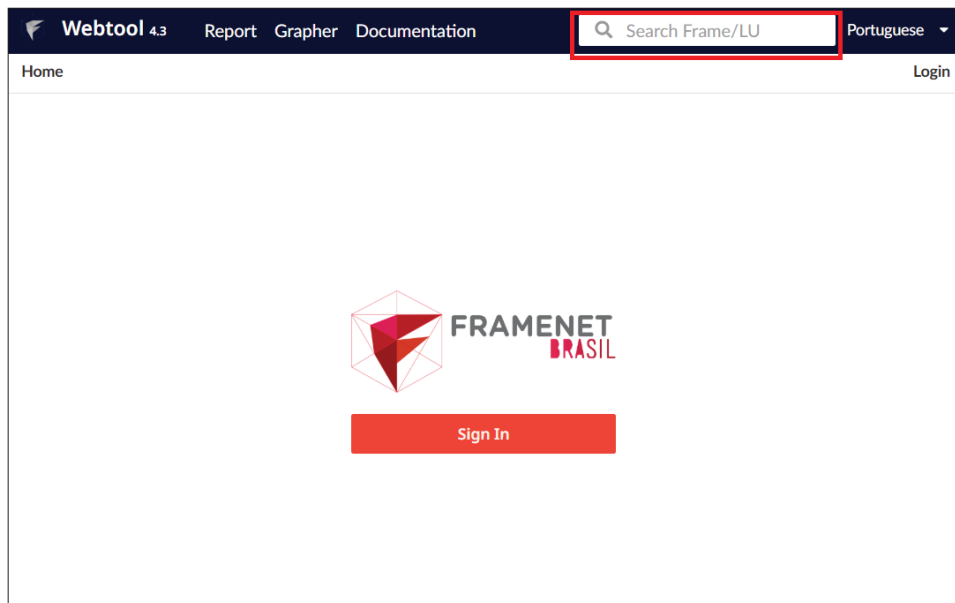
É a partir dessa observação inicial que se define a estratégia de consulta à FrameNet Brasil. Em vez de buscar diretamente estados emocionais ou intenções inferidas, o audiodescritor identifica **ações e relações perceptíveis na imagem**, que podem ser tomadas como ponto de entrada para a busca por itens lexicais na base. Essas observações orientam a escolha inicial dos itens lexicais a serem consultados, dando início ao percurso de navegação na FrameNet Brasil descrito a seguir.

Passo 1 – Acesso à base da FrameNet Brasil

Com base na ação observada na imagem, o audiodescritor acessa a base oficial da FrameNet Brasil para iniciar a consulta, <https://webtool.frame.net.br/>. A interface permite buscas tanto por itens lexicais do português brasileiro quanto diretamente por frames semânticos, oferecendo diferentes caminhos de exploração conforme a necessidade analítica.

A Figura 8 mostra a página inicial da FrameNet Brasil com destaque para a interface pública de busca disponível na WebTool.

Figura 8: Página inicial da FrameNet Brasil, área de busca. Captura de tela do software de anotação semântica Webtool¹⁹.



19. Audiodescrição referente à Figura 8: Logotipo Framenet Brasil e logo abaixo o botão Sign In, para logar no software Webtool.

Passo 2 – Busca por itens lexicais relacionados à ação observada

Partindo da ação visualmente identificável, o audiodescritor realiza a busca por um item lexical que poderia organizar a descrição da cena. O objetivo dessa etapa não é definir imediatamente o texto da AD, mas verificar **quais frames estão associados às opções lexicais disponíveis** e que tipo de situação elas tendem a ativar.

A Figura 9 mostra o campo de busca preenchido pelo verbo *fumar* e o resultado da pesquisa. Dentre as possibilidades apresentadas, o frame **Ingerir_substância** parece ser o de nosso interesse considerando as informações da Figura 7.

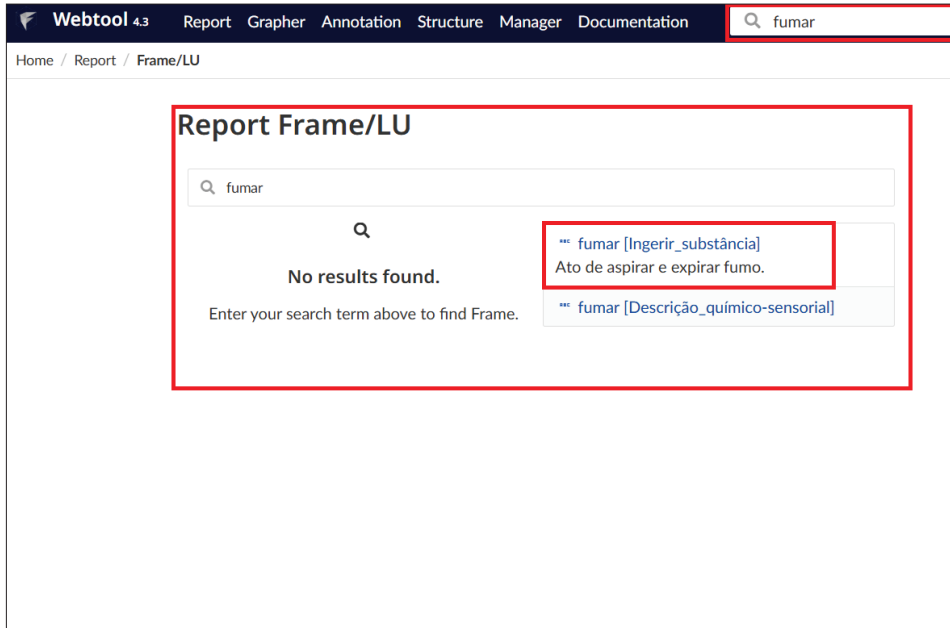


Figura 9: Campo de busca com a Unidade Lexical *fumar* e lista de resultados. Captura de tela do software de anotação semântica Webtool²⁰.

20. Audiodescrição referente à Figura 9: Captura de tela do software de anotação semântica Webtool. Campo de busca com a Unidade Lexical *fumar* e lista com resultado incluindo **Ingerir_substância**, ato de aspirar e expirar fumo; e **Descrição_químico_sensorial**.

Passo 3 – Acesso à página do frame e a identificação dos Elementos de Frame em articulação com a imagem

Após seleccionar o item lexical com o sentido específico, ou seja, decidir qual é a Unidade Lexical, o usuário é direcionado à página do frame correspondente. Encontram-se, assim, informações para a análise semântica da cena, como a definição do frame, seus Elementos

Figura 10: Página do frame de **Ingerir_substância**, com a definição e a lista de Elementos de Frame essenciais. Captura de tela do software de anotação semântica Webtool²¹.

21. Audiodescrição referente à Figura 10: Captura de tela do software de anotação semântica Webtool. Página do frame de **Ingerir_substância**, com a definição um ingestor leva uma substância para dentro de seu corpo, e a indicação que o Dispositivo_de_fornecimento pode aparecer como objeto direto. A lista de Elementos de Frame essenciais inclui o Dispositivo_de_fornecimento como meio pelo qual a substância é ingerida; o Ingestor como quem ingere a substância; e a Substância como a droga ou outra substância que é ingerida pelo ingestor.

Ingerir_substância
 Nesse frame, um **Ingestor** leva uma **Substância** para dentro de seu corpo. O **Dispositivo_de_fornecimento** pode aparecer como um objeto direto.

Frame Elements

Core Frame Elements

- Dispositivo_de_fornecimento**
 O **Dispositivo_de_fornecimento** é o meio pelo qual a substância é ingerida.
 CoreSet: Substância
- Ingestor**
 O **Ingestor** ingere a **Substância**.
 sentient
- Substância**
 A **Substância** é a droga ou outra substância que é ingerida pelo **Ingestor**.

de Frame (EFs) e exemplos de uso. A Figura 10 apresenta o frame **Ingerir_substância** com sua definição geral e a lista de seus elementos de frames essenciais. As demais informações associadas ao frame podem ser acessadas por meio da navegação direta na base.

A leitura dessa página permite compreender **quais participantes, ações, espaços e características são considerados estruturais naquela situação**, oferecendo um repertório conceitual para orientar a seleção do conteúdo da AD. No caso de **Ingerir_Substância**, os elementos de frame **Ingestor**, **Dispositivo_de_fornecimento** e **Substância** são considerados essenciais. Cabe destacar que, dependendo da cena, o dispositivo e a substância podem aparecer de forma alternada ou com diferentes graus de saliência visual.

A análise dos Elementos de Frame não se encerra, contudo, na leitura da página da FrameNet Brasil. Ela se completa na articulação desses elementos com a imagem audiovisual que motivou a consulta. Para tornar esse processo analítico mais observável, a Figura 11 apresenta a imagem selecionada do programa *Pedro pelo Mundo* com os elementos visuais destacados por meio de caixas delimitadoras. O recurso não faz parte do campo de busca aberto da FrameNet Brasil, pois atua no suporte de anotação semântica multimodal.

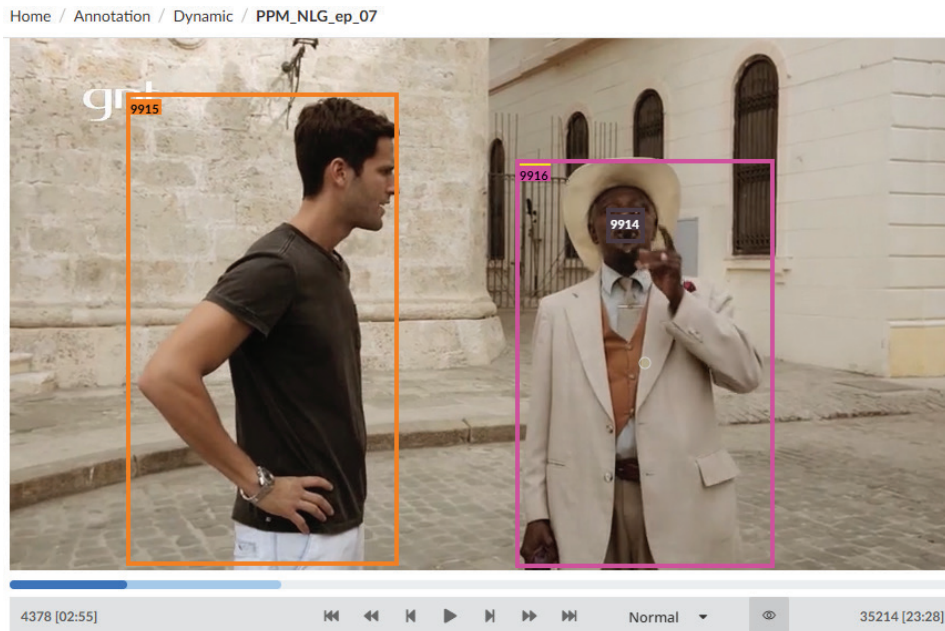
Nessa figura, há a correspondência entre os Elementos de Frame e componentes da imagem: o *Ingestor*, representado pelo homem em primeiro plano; a *Substância*, materializada no charuto; e o *Dispositivo_de_fornecimento*, igualmente associado ao charuto enquanto meio da ação. Esse mapeamento visual evidencia como os Elementos de Frame auxiliam o audiodescritor a identificar quais informações são semanticamente centrais e visualmente acessíveis na cena.

Ao explicitar essa relação entre frame e imagem, o objetivo não é transformar os Elementos de Frame em um checklist a ser descrito, mas mostrar como eles organizam o raciocínio descritivo, apoiando decisões sobre o que é relevante verbalizar e o que pode ser deixado à inferência do espectador.

Ao explicitar essa relação entre frame e imagem, o objetivo não é transformar os Elementos de Frame em um checklist a ser descrito, mas mostrar como eles organizam o raciocínio descritivo, apoiando decisões sobre o que é relevante verbalizar e o que pode ser deixado à inferência do espectador.

Figura 11: Elementos de Frame essenciais demarcamos por caixas delimitadoras. Captura de tela do software de anotação semântica Webtool²².

22. Audiodescrição referente à Figura 11: Na rua, o apresentador Pedro conversa com um senhor de terno bege e chapéu de abas largas. O senhor fuma um charuto. *Bounding boxes* coloridas delimitam Pedro e o senhor com o charuto.



No mesmo recorte analisado, identificamos ainda que Pedro observa o homem enquanto este fuma o charuto. Essa ação pode ser associada ao frame **Percepção_ativa**. A Figura 7 ilustra essa relação visual entre os personagens, enquanto o Quadro 7, apresentado anteriormente, mostra o frame.

Neste ponto, convidamos o(a) leitor(a) a realizar a busca por essa ação diretamente na base, de modo a acompanhar o percurso de consulta e reconhecer como diferentes frames podem ser acionados a partir de uma mesma imagem.

Passo 4 – Pesquisa em frames relacionados

A Figura 12 apresenta a navegação pela rede de frames, evidenciando relações de herança e de herdado por (em inglês, Inherits from e Inherited by). A exploração de frames relacionados possibilita que o audiodescritor compreenda melhor como diferentes frames podem se articular em uma mesma cena, favorecendo descrições alinhadas à progressão narrativa do audiovisual.

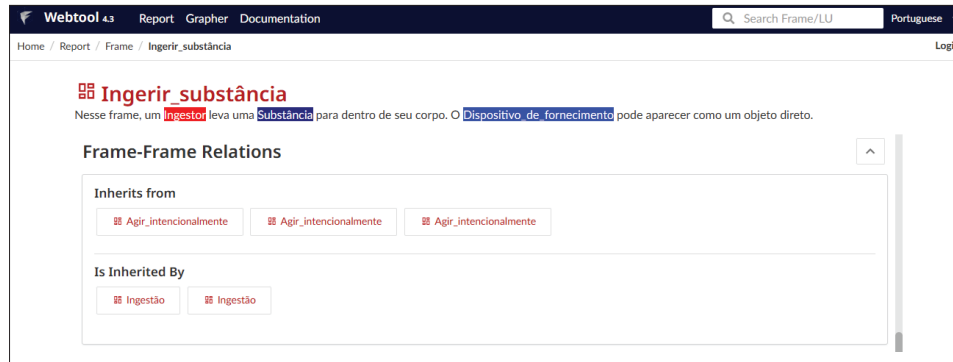


Figura 12: Frames relacionados ao de **Ingerir_Substância**. Captura de tela do software de anotação semântica Webtool²³.

23. Audiodescrição referente à Figura 12: Captura de tela do software de anotação semântica Webtool. Página do frame de Ingerir_substância, com a definição um ingestor leva uma substância para dentro de seu corpo, e a indicação que o Dispositivo_de_fornecimento pode aparecer como objeto direto. A lista de relações de frame inclui frame herdado de Agir_intencionalmente, e frame herdado por Ingestão.

Passo 5 – Pesquisa em itens lexicais e entidades visuais associadas

A Figura 13 apresenta as Unidades Lexicais e as Unidades Visuais associadas ao frame **Ingerir_substância**, conforme disponibilizadas na interface da Webtool da FrameNet Brasil.

Figura 13: Unidades Lexicais e Unidades Visuais associadas ao frame de **Ingerir_substância**. Captura de tela do software de anotação semântica Webtool²³.

A listagem evidencia como um mesmo frame pode ser ativado por diferentes itens lexicais do português brasileiro, bem como por entidades visuais previamente anotadas na base.

The screenshot shows the Webtool interface for the frame 'Ingerir_substância'. The main text reads: 'Nesse frame, um **Ingestor** leva uma **Substância** para dentro de seu corpo. O **Dispositivo de fornecimento** pode aparecer como um objeto direto.'

Lexical Units

aspirar consumir fumar mascar

Visual Units

Visual Unit	Doc
Corpus_oficina_SEM_sentença_1	Doc
Corpus_oficina_SEM_sentença_1	Doc
Corpus_oficina_SEM_sentença_1	Doc
Corpus_oficina_SEM_sentença_2	Doc
Corpus_oficina_SEM_sentença_2	Doc
Corpus_oficina_SEM_sentença_2	Doc
Corpus_oficina_SEM_sentença_2	Doc
Corpus_oficina_SEM_sentença_2	Doc
Corpus_oficina_SEM_sentença_2	Doc
Corpus_oficina_SEM_sentença_2	Doc
Corpus_oficina_SEM_sentença_2	Doc

Legend: ■ Ingestor ■ Substância

The image shows a man in a blue shirt holding a cigarette. A red bounding box highlights the man, and a blue bounding box highlights the cigarette.

24. Audiodescrição referente à Figura 13: Captura de tela do software de anotação semântica Webtool. Página do frame de Ingerir_substância, com a definição um ingestor leva uma substância para dentro de seu corpo, e a indicação que o Dispositivo de fornecimento pode aparecer como objeto direto. As unidades lexicais são aspirar, consumir, fumar, mascar. À direita, a imagem de um homem fumando, com *bounding boxes* coloridas delimitando o homem como ingestor e o cigarro como substância.

A observação dessas associações permite ao audiodescritor comparar alternativas lexicais possíveis para descrever a ação observada na imagem, avaliando quais escolhas lexicais se alinham melhor ao tipo de situação representada e ao efeito narrativo desejado. Além

disso, a presença de Unidades Visuais associadas ao frame reforça a relação entre padrões recorrentes de representação imagética e a organização semântica proposta pela rede, o que oferece subsídios adicionais para a tomada de decisão na elaboração da AD.

Em síntese

A consulta à FrameNet Brasil, quando ancorada em uma imagem concreta, funciona como um instrumento de apoio, análise e organização da AD. O percurso apresentado evidencia que o uso da base não parte da teoria em direção à imagem, mas da necessidade prática de audiodescrever uma cena construída pela articulação entre áudio, gestos, ações, enquadramentos etc. Ao avaliar frames e itens lexicais, o audiodescritor pode refinar suas escolhas linguísticas, manter a coerência semântica da descrição e alinhar a AD à dinâmica multimodal pela qual o sentido é construído na obra audiovisual.

INTELIGÊNCIA ARTIFICIAL NA AUDIODESCRIÇÃO FÍLMICA

Quando integrada a uma abordagem baseada na Semântica de Frames, a IA pode auxiliar na preparação de rascunhos iniciais e na organização de informações visuais recorrentes, funcionando como suporte ao trabalho interpretativo humano.

O uso de sistemas de inteligência artificial (IA) na audiodescrição tem se ampliado nos últimos anos, impulsionado pelo avanço de modelos capazes de processar simultaneamente texto, imagem e som. No contexto deste guia, concebemos que a IA pode ser útil como ferramenta de apoio a etapas específicas do processo, sobretudo na identificação preliminar de ações, participantes, objetos, espaços e relações visuais que podem ativar frames semânticos relevantes.

Quando integrada a uma abordagem baseada na Semântica de Frames, a IA pode auxiliar na preparação de rascunhos iniciais e na organização de informações visuais recorrentes, funcionando como suporte ao trabalho interpretativo humano. No entanto, como esses sistemas operam de forma probabilística e dependem dos dados e dos critérios de treinamento, sua atuação exige mediação especializada para evitar ativações indevidas de frames, inferências não sustentadas pela imagem e distorções narrativas.

Para isso, o texto se estrutura em três eixos centrais. Primeiro, discutem-se os potenciais da IA na geração e no apoio à audiodescrição, com ênfase em seu uso articulado à Semântica de Frames. Em seguida, examinam-se os limites e os desafios éticos associados à utilização dessas tecnologias na produção de AD, especialmente no que diz respeito à qualidade, à responsabilidade autoral e à adequação ao público. Por fim, apresentam-se diretrizes para o uso responsável da IA, reunindo recomendações práticas voltadas a profissionais e equipes que pretendem incorporar essas ferramentas aos seus fluxos de trabalho, sem perder de vista o papel central da curadoria humana.

6.1 Potenciais da IA orientada por frames na geração e apoio à AD

O uso de inteligência artificial na audiodescrição tem se consolidado, principalmente, como um **recurso de apoio ao trabalho do audiodescritor**²⁵. Na prática profissional, ferramentas baseadas em IA podem contribuir para agilizar etapas específicas do processo, como a elaboração de rascunhos iniciais, a verificação de consistência interna do texto e a comparação entre diferentes versões de uma mesma descrição. Esse apoio tende a reduzir tarefas repetitivas e a liberar tempo para que o audiodescritor se concentre nas decisões interpretativas, narrativas e estilísticas, que permanecem sob responsabilidade humana.

Algumas ferramentas e sistemas experimentais adotam um **modelo colaborativo**, no qual a IA gera uma primeira versão da audiodescrição e o profissional realiza revisões sucessivas. Nesse tipo de fluxo, como o proposto pelo sistema **DescribePro**²⁶, a IA atua como produtora de uma versão preliminar, enquanto o audiodescritor ajusta estilo, precisão descritiva, terminologia e coerência narrativa. A possibilidade de gerar múltiplas versões de um mesmo trecho permite comparar alternativas e apoiar escolhas mais conscientes sobre o que deve integrar o roteiro final, reforçando o papel da curadoria humana no processo.

25. Campos (2019); Cheema et al. (2025); Ribeiro e Andrade (2025).

26. Para mais informações, pesquisar em Cheema et al. (2025).

27. Campos (2019).

28. Ribeiro e Andrade (2025).

29. Jacques et al. (2025).

30. Para detalhamento das pesquisas, sugerimos Ribeiro e Andrade (2025).

31. Pinto (2025).

No contexto brasileiro, o uso de IA na audiodescrição vem sendo explorado sobretudo em iniciativas ligadas à acessibilidade educacional²⁷, audiovisual²⁸ e cultural²⁹. De modo geral, essas experiências investigam como modelos multimodais podem auxiliar na análise de imagens e na preparação de descrições iniciais, funcionando como ferramentas de apoio ao trabalho do audiodescritor, e não como produtos finais prontos para uso.

No âmbito da audiodescrição fílmica, aplicações como o **Be My Eyes** — que combina modelos multimodais e suporte humano — e **sistemas de IA generativa**, como o ChatGPT, demonstram potencial para acelerar as etapas iniciais da AD. Esses recursos costumam apresentar ganhos importantes em termos de rapidez e cobertura geral das informações visuais, sendo particularmente úteis na produção de rascunhos preliminares³⁰. No entanto, a qualidade final da audiodescrição depende diretamente da **clareza das instruções fornecidas ao sistema** e, sobretudo, da **revisão especializada do audiodescritor**, responsável por assegurar coerência narrativa, consistência terminológica, adequação ao gênero da obra e respeito ao ritmo cinematográfico.

Uma das formas de tornar essas instruções mais claras e estruturadas consiste no **uso de anotações semânticas** que organizem os eventos e os participantes presentes em cada cena da produção audiovisual antes da geração do texto. Nesse sentido, a anotação multimodal de eventos, desenvolvida no âmbito da FrameNet Brasil com base na Semântica de Frames, tem sido investigada quanto à sua contribuição para a geração de roteiros de audiodescrição por sistemas de inteligência artificial, atuando como copiloto para audiodescritores humanos³¹. A ideia é que esse tipo de anotação, ao mapear semanticamente os eventos e seus participantes na obra audiovisual, favorece a criação de roteiros mais precisos, rápidos e adequados ao contexto narrativo.

Para verificar isso, os episódios 1 e 7 da série de viagens *Pedro Pelo Mundo*, da GNT, receberam anotações semânticas segundo a metodologia da FrameNet Brasil para anotação

multimodal de eventos, que identifica ações e participantes relevantes em cada cena. Com base nessas anotações, sistemas de IA auxiliaram na geração de roteiros preliminares de audiodescrição, tanto com quanto sem o apoio das informações baseadas em frames. Esses roteiros foram analisados qualitativamente, considerando sua aproximação aos critérios do Guia para Produções Audiovisuais Acessíveis³², e também quantitativamente, por meio de medidas de similaridade de cosseno entre as descrições produzidas. A comparação entre as versões revelou resultados promissores para o uso da Semântica de Frames, com descrições mais específicas e alinhadas à dinâmica das cenas.

Do ponto de vista quantitativo, as anotações de frames funcionaram como um modulador do desempenho³³ do modelo: quando a versão apoiada nas anotações semânticas apresentou desempenho inferior, as diferenças negativas tenderam a ser pequenas; por outro lado, quando apresentou desempenho superior, os ganhos foram mais expressivos. Além disso, a versão com metadados alcançou níveis mais altos de alinhamento com o conteúdo visual com maior frequência e, mesmo quando não gerou melhorias imediatas, não comprometeu a qualidade das descrições.

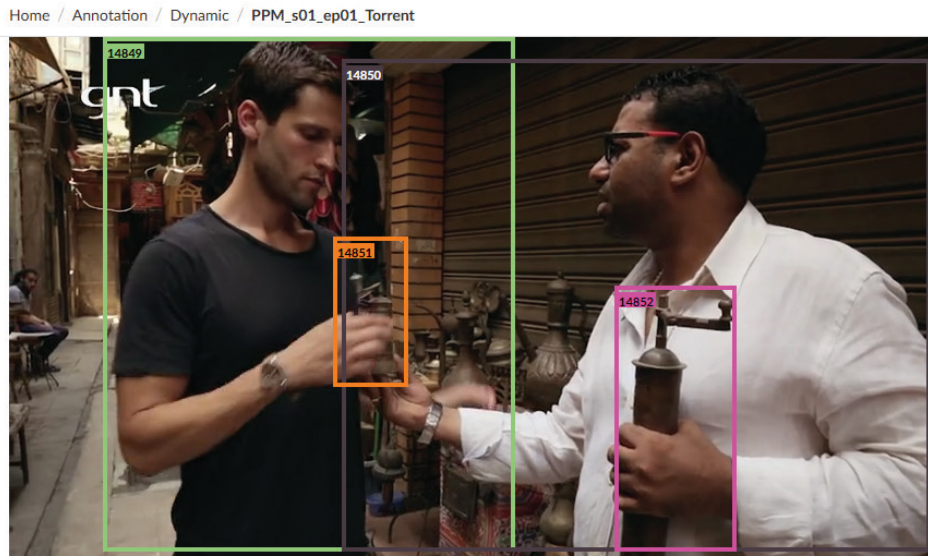
No plano qualitativo, as anotações também influenciaram as escolhas lexicais realizadas pelo modelo. Em um dos trechos analisados, por exemplo, ilustrado pela Figura 14, tanto a versão sem anotações quanto a versão com metadados registram Pedro examinando uma mercadoria no mercado, descrevendo a cena da seguinte forma, respectivamente:

- (6) O apresentador examina um objeto de metal.
- (7) Pedro examina um moedor.

32. Naves et al. (2016).

33. No contexto deste trabalho, desempenho refere-se ao grau de proximidade entre as descrições geradas pela IA e o conteúdo visual da cena, avaliado por indicadores como a similaridade de cosseno, que varia de 0 a 1, sendo 1 o alinhamento ideal. Para mais detalhes sobre os resultados, ver Pinto (2025).

Figura 14: Cena do episódio 1 de *Pedro pelo Mundo*, mostrando Pedro examinando uma mercadoria no mercado. Captura de tela do software de anotação semântica Webtool³⁴.



34. Audiodescrição referente à Figura 14: O apresentador Pedro analisa um moedor pequeno nas mãos de um vendedor de camisa branca e óculos, que também segura um moedor maior. Bounding boxes coloridas delimitam Pedro, o vendedor e cada um dos dois moedores.

Na versão sem anotações, no entanto, a descrição se mantém mais geral, referindo-se a “um objeto de metal”. Já a versão com anotações especifica o item como “um moedor”, seguindo a anotação semântica realizada para o objeto na cena, que o associa à mercadoria no frame **Comércio_Vender** (Quadro 10) e à Unidade Lexical “moedor” no frame **Utensílios** (Quadro 11).

Esse contraste mostra como as anotações orientam escolhas lexicais mais precisas e contextualizadas: nesse caso, o modelo seleciona uma denominação mais informativa para

Quadro 11: O frame de **Comércio_vender**.

Comércio_vender	
<p>Este frame representa situações de transação comercial em que um Vendedor transfere Mercadorias a um Comprador em troca de Dinheiro, assumindo a perspectiva daquele que detém inicialmente as mercadorias.</p> <p>Na AD, este frame é ativado quando a cena apresenta visualmente uma situação de venda — como a entrega de um produto, a interação em um balcão, o manuseio de dinheiro ou a organização de mercadorias.</p>	
★	<p>Vendedor: É o personagem que possui as mercadorias e realiza a ação de vendê-las.</p> <p>☰ “O atendente vende o produto ao cliente.”</p>
★	<p>Mercadorias: São os bens ou serviços que estão sendo transferidos na transação comercial.</p> <p>☰ “Ele vende pão diariamente.”</p>
★	<p>Comprador: É o personagem que recebe as mercadorias em troca de dinheiro.</p> <p>☰ “O atendente vende o produto ao cliente.”</p>
+	<p>Dinheiro: É o valor dado em troca das mercadorias, quando visualmente salientado.</p> <p>☰ Na placa, está escrito que o padeiro vende o pão por vinte reais.</p>
+	<p>Lugar: Indica o local onde a transação ocorre.</p> <p>☰ “Na feira de rua, o homem anuncia e vende as frutas.”</p>
+	<p>Meio: Indica o meio pelo qual a transação é realizada.</p> <p>☰ “A venda foi feita por cartão de crédito.”</p>

o objeto, evitando a formulação genérica da versão sem anotações. Assim, as anotações funcionam como um guia para a escolha lexical, reduzindo generalizações e aproximando a descrição da organização semântica da cena. Para o espectador, isso proporciona uma compreensão mais clara da narrativa, aproximando a descrição daquilo que realmente ocorre na tela.

Quadro 12: O frame de **Utensílios**.

Utensílios	
<p>Este frame descreve materiais criados para um uso específico, especialmente associados a práticas domésticas. Um utensílio é identificado por suas propriedades físicas e funcionais observáveis na cena.</p> <p>Na AD, o frame é ativado quando a presença, a manipulação ou a disposição do objeto é relevante para a compreensão da ação ou do espaço, evitando inferências sobre finalidade não sustentadas visualmente.</p>	
★	<p>Utensílio: Objeto criado para um uso específico, identificável por sua forma e função aparentes.</p> <p>☺ “Uma panela está sobre o fogão.”</p>
+	<p>Tipo: Indicação do subtipo do utensílio, com base em categorias reconhecíveis.</p> <p>☺ “Uma faca de cozinha.”</p>
+	<p>Material: Indicação do material que compõe o utensílio, quando visível.</p> <p>☺ “Uma tigela de vidro aparece molhada sobre a mesa.”</p>
+	<p>Partes_constituíntes: Partes salientes que compõem o utensílio e são relevantes para a cena.</p> <p>☺ “A tampa da panela está apoiada ao lado da pia.”</p>

Ainda nessa mesma cena, a versão com anotações aproveitou uma nova janela de silêncio para registrar outra ação, mostrada na sentença (8):

(8) Pedro se afasta do vendedor.

Com a inserção dessa informação e a especificação do item como “moedor”, fica mais fácil visualizar o que realmente aparece na tela e entender o papel de cada participante. Vemos Pedro interagindo com o vendedor dentro de um contexto de compra e venda, ativando o frame **Comércio_vender** (Quadro 10). Com isso, a ação deixa de ser apenas a análise de um objeto isolado e passa a ser compreendida como parte de uma relação social, enriquecendo a descrição.

Em outro segmento do vídeo, o contraste entre as versões se manifesta não apenas nas escolhas lexicais, mas também na organização do evento descrito. A cena ilustrada pela Figura 15 faz parte de uma sequência que apresenta diferentes espaços de um mercado, como cafés e corredores, e a versão sem e a versão com anotações, nessa ordem, resumem o que aparece na tela, mas enquadram a situação de forma distinta:

- (8) Pessoas sentadas em cafés na rua. Placa “Fechado para oração” pendurada.
- (9) Homens jogam gamão em café, perto de uma mesquita.

Home / Annotation / Dynamic / PPM_s01_ep01_Torrent

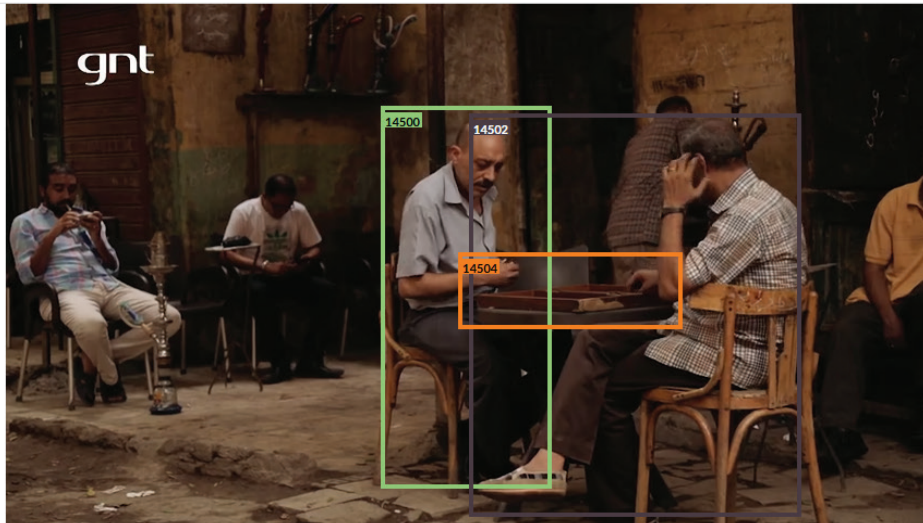


Figura 15: Cena do episódio 1 de *Pedro pelo Mundo*, mostrando homens jogando gamão em um café. Captura de tela do software de anotação semântica Webtool³⁵.

35. Audiodescrição referente à Figura 15: Em um café ao ar livre, dois homens jogam gamão em uma mesa. Ao fundo, outras pessoas estão sentadas e um homem fuma narguilé. Bounding boxes coloridas delimitam os dois homens que jogam gamão e o tabuleiro do jogo.

Na versão sem anotações, a descrição se limita a enumerar os elementos visíveis na cena e indicar o estado das pessoas, apresentando uma situação estática. O foco recai sobre o fato de estarem sentadas e sobre a configuração espacial, sem explicitar qualquer relação de ação entre os participantes. A versão com anotações, por outro lado, organiza a cena em torno da atividade realizada, destacando que os homens jogam gamão. Isso acontece porque as anotações identificam os participantes como agentes da ação dentro do frame **Atividade** (Quadro 12), associando-os diretamente ao jogo. Desse modo, os participantes deixam de ser elementos passivos no espaço e passam a ser percebidos como envol-

Quadro 13: O frame de **Atividade**.

Atividade
<p>Este frame descreve situações em que um personagem se envolve de forma intencional em uma atividade contínua, permanecendo nesse estado por um determinado intervalo de tempo. A Atividade pode ter início, desenvolvimento, término ou ser suspensa. Na AD, ele atua para organizar a progressão visível das ações, sem recorrer à atribuição de estados mentais ou motivações internas.</p>
<p>★ Agente: Aquele que realiza a atividade. 💬 “O homem joga cartas.”</p>
<p>★ Atividade (<i>costuma ser expresso pela própria ação</i>): A ação realizada pelo agente. 💬 “O homem joga cartas.”</p>
<p>+ Duração: Quantidade de tempo durante a qual a atividade se estende. 💬 “O homem joga cartas durante toda a manhã.”</p>
<p>+ Lugar: Local onde a atividade ocorre 💬 “O homem joga cartas na sala.”</p>
<p>+ Maneira: Forma visível como a atividade é realizada, incluindo ritmo, intensidade ou características observáveis da ação 💬 “O homem joga rapidamente as cartas”</p>

vidos em uma ação concreta, o que resulta em uma descrição mais dinâmica, que evidencia a relação entre os participantes e o evento central da cena.

Dessa forma, a Semântica de Frames se mostra uma base promissora para a AD mediada por IA. De modo geral, ao longo das cenas analisadas, a organização semântica dos eventos contribui para escolhas lexicais mais precisas, promove descrições mais dinâmicas e alinhadas à narrativa e permite que cada ação seja enquadrada corretamente dentro do contexto, aproximando os rascunhos gerados das diretrizes da prática profissional.

Considerando esses potenciais, torna-se igualmente importante discutir os limites, os riscos e os desafios éticos associados ao uso de sistemas generativos na audiodescrição, especialmente quando aplicados a conteúdos complexos como produções cinematográficas.

6.2 Limites, riscos e desafios éticos do uso de IA

A adoção de sistemas de inteligência artificial na produção de audiodescrição costuma ser acompanhada pela ideia de que tecnologias automatizadas seriam mais neutras ou objetivas do que descritores humanos. Essa percepção, embora recorrente, desconsidera como esses sistemas são desenvolvidos e treinados e pode gerar uma compreensão equivocada sobre seu funcionamento e seus impactos no campo da acessibilidade.

Modelos de IA são construídos a partir de **decisões humanas** e treinados com **grandes volumes de dados**, cujos critérios de seleção e organização, em especial no caso de modelos proprietários, carecem de transparência. Como consequência, esses sistemas incorporam padrões e valores presentes nos dados de treinamento, o que impede que sejam considerados neutros³⁶. No contexto da audiodescrição, essa falsa neutralidade é particularmente problemática, pois a AD atua como mediação direta entre o conteúdo visual e o usuário com deficiência visual ou com baixa visão e deve ser feita de forma cuidadosa.

36. Para mais discussão sobre ética e uso de novas tecnologias, sugerimos Mitchell et al. (2025).

Omissões, distorções ou inferências inadequadas na descrição afetam diretamente a compreensão da cena e a experiência do espectador. Sistemas automatizados podem, por exemplo, apagar traços identitários relevantes, reforçar estereótipos ou inserir informações que não estão presentes no vídeo. Também podem atribuir intenções, estados emocionais ou características que não são visíveis ou confirmáveis a partir das imagens. Esses comportamentos comprometem a confiabilidade da audiodescrição e tornam indispensável a discussão sobre limites, riscos e desafios éticos no uso da IA nesse campo. Reconhecer que a inteligência artificial também não é neutra é, portanto, um passo fundamental para compreender suas limitações e para orientar práticas responsáveis de uso na audiodescrição.

6.2.1. Vieses algorítmicos

Os vieses algorítmicos são padrões de distorção que aparecem quando sistemas de inteligência artificial são treinados com dados que refletem desigualdades históricas, lacunas de representatividade ou formas estereotipadas de retratar determinados grupos. Como esses modelos se baseiam em grandes volumes de exemplos para identificar regularidades, acabam reproduzindo tendências presentes nesses conjuntos, as quais podem resultar de escolhas de curadoria, de falhas de representatividade nos dados de treinamento ou, até mesmo, de formas historicamente enviesadas de retratar determinados grupos.

Na prática, esses vieses influenciam a forma como pessoas, objetos e situações são reconhecidos e descritos pelos sistemas. Muitas vezes, essas distorções passam despercebidas, pois são apresentadas com aparência de precisão técnica ou objetividade. Na audiodescrição, esse efeito é especialmente sensível: ao serem apresentadas como descrição objetiva da cena, essas saídas adquirem peso de verdade para o usuário, reforçando estereótipos e representações excludentes.

Em descrições automáticas de imagens³⁷, por exemplo, é comum observar o **apagamento de identidades**, a **falta de sensibilidade ao contexto cultural**, a **reprodução de estereótipos de gênero** ou a **invisibilização de aspectos relacionados à deficiência**. Mesmo quando tecnicamente corretas em termos de objetos e ações, essas descrições podem falhar em representar de forma justa e contextualizada os sujeitos retratados.

Assim, o uso de sistemas automatizados na audiodescrição envolve riscos que vão além de falhas técnicas. Trata-se também de uma questão ética, ligada à forma como corpos, identidades e experiências são representados na mediação do conteúdo visual. Reconhecer a existência de vieses algorítmicos é essencial para que o audiodescritor exerça um papel ativo de supervisão, revisão e curadoria, de modo a buscar descrições mais precisas, inclusivas e sensíveis à diversidade de corpos, culturas e identidades e aproveitar o potencial das tecnologias de IA sem comprometer a equidade ou a experiência do usuário.

6.2.2. Alucinações artificiais

Além dos vieses, outro limite relevante dos sistemas de IA aplicados à AD são as chamadas alucinações artificiais³⁸. Esse termo designa a geração de informações incorretas, imprecisas ou não verificáveis, produzidas quando o sistema não dispõe de dados suficientes sobre o conteúdo descrito.

Na audiodescrição automática, as alucinações se manifestam, por exemplo, na descrição de elementos inexistentes na cena, na criação de ações que não ocorrem no vídeo ou na atribuição de emoções, intenções e características que não podem ser confirmadas visualmente. Esse comportamento está relacionado ao funcionamento estrutural dos modelos gerativos, que tendem a preencher lacunas para manter a fluidez e a coerência do texto, mesmo em situações de incerteza³⁹.

37. Para esse tema, sugerimos Rosetti, Borges e Souza (2025).

38. Para mais informações, pesquisar Kalai et al. (2025).

39. A pesquisa de Kalai et al. (2025) mostra por que esses modelos de língua alucinam.

Para a audiodescrição, esse tipo de saída representa um risco significativo. Ao inventar detalhes ou extrapolar informações, a IA compromete a fidelidade narrativa e pode induzir o espectador a formar uma compreensão equivocada da cena. Isso inclui, por exemplo, a atribuição indevida de idade, etnia ou estado emocional a personagens, a descrição de objetos inexistentes ou a criação de eventos que não se desenvolvem no filme. Esses desvios afetam a coerência temporal e narrativa da obra e prejudicam a experiência do usuário que depende da descrição.

Diante desses riscos, o uso de IA na audiodescrição deve ser sempre acompanhado de **supervisão humana** e de **diretrizes específicas para curadoria do que é gerado**. É igualmente importante fornecer ao sistema entradas bem delimitadas e contextualizadas, reduzindo ambiguidades que favorecem a produção de informações imprecisas.

Embora as tecnologias gerativas ofereçam ganhos importantes em produtividade e possam apoiar etapas preliminares do processo descritivo, seu uso sem verificação crítica tende a introduzir ruídos na experiência audiovisual. Reconhecer as limitações estruturais dos modelos e sua propensão a produzir informações não verificáveis é, portanto, indispensável para evitar que a audiodescrição automatizada amplifique equívocos, lacunas interpretativas ou representações imprecisas.

6.3 Diretrizes para uso responsável da IA

A partir dos potenciais ganhos e dos riscos discutidos anteriormente, torna-se essencial definir diretrizes para o uso responsável de sistemas de inteligência artificial na audiodescrição. Essas orientações buscam integrar ferramentas gerativas ao trabalho de forma ética e transparente, sem comprometer a precisão nem a experiência do público com deficiência visual e baixa visão. As recomendações a seguir sintetizam cuidados fundamentais e indicam caminhos para o uso seguro e consistente da IA nesse contexto.

I. Uso da IA como apoio ao trabalho do audiodescritor

No estágio atual de desenvolvimento, a IA deve ser utilizada prioritariamente como ferramenta de apoio ao trabalho humano, e não como substituta do audiodescritor⁴⁰. O uso isolado de sistemas automatizados, sem supervisão especializada, é problemático devido à ocorrência de alucinações artificiais, limitações de compreensão contextual e dificuldades na interpre-

tação narrativa. As decisões envolvidas na AD exigem sensibilidade interpretativa, julgamento situacional e repertório sociocultural, competências que permanecem dependentes da atuação humana.

40. Retomamos aqui os estudos de Ribeiro e Andrade (2025).

II. Necessidade de curadoria humana especializada

Toda saída gerada por sistemas de IA deve ser submetida à revisão criteriosa de audiodescritores qualificados. A curadoria humana é indispensável para corrigir informações imprecisas, eliminar inferências indevidas, ajustar o estilo e garantir a coerência temporal e narrativa com a obra audiovisual. Além disso, a revisão atua como proteção ética, reduzindo o risco de reprodução de estereótipos, apagamentos identitários ou representações enviesadas que podem estar presentes nas saídas automatizadas.

III. Fornecimento de informações completas e contextualizadas aos modelos

O uso de IA na produção de AD requer a oferta de informações claras, completas e bem delimitadas. Sempre que possível, devem ser fornecidos ao sistema dados sobre o contexto narrativo, a sequência temporal das cenas, os personagens, os diálogos e os elementos visuais relevantes. Quanto mais contextualizada for a entrada, menor tende a ser a incidência de descrições imprecisas ou inventadas, aumentando a confiabilidade dos rascunhos gerados.

IV. Transparência no uso da IA

A adoção de tecnologias generativas no processo de audiodescrição deve ser transparente para todos os envolvidos — instituições produtoras, audiodescritores, revisores e público. Isso inclui registrar a participação da IA no fluxo de trabalho, indicando em que etapas foi utilizada e quais escolhas foram automatizadas ou validadas por humanos. A transparência permite auditoria ética e colabora para o desenvolvimento de boas práticas no campo.

V. Atualização contínua e formação profissional

A integração responsável da IA à audiodescrição exige formação contínua. Audiodescritores, revisores e equipes técnicas devem ser capacitados para compreender o funcionamento das tecnologias gerativas, reconhecer limitações técnicas e éticas, identificar alucinações e vieses e decidir quando e como a IA deve ser utilizada. A formação contínua fortalece a autonomia do profissional e evita usos acríticos ou inadequados da tecnologia.

VI. Salvaguarda da privacidade e dos direitos autorais

O uso de IA em contextos audiovisuais envolve circulação de imagens, cenas e roteiros que podem ser protegidos por direitos autorais. Recomenda-se verificar termos de uso, garantir consentimento adequado e adotar práticas de segurança para evitar exposição indevida de materiais. Ao estabelecer diretrizes para o uso responsável da IA na AD, não se trata apenas de incorporar novas tecnologias ao processo, mas de assegurar que elas atuem como aliadas na promoção da acessibilidade. O uso ético e cuidadoso da IA deve fortalecer o trabalho do audiodescritor, ampliar a produção de conteúdo acessível e garantir que o público com deficiência visual e baixa visão receba informações precisas, coerentes e sensíveis ao contexto artístico e comunicativo da obra.

Em síntese, as diretrizes apresentadas reforçam que o uso de IA na audiodescrição deve ser sempre orientado por princípios de responsabilidade, ética e transparência. Ao atuar como ferramenta de apoio ao trabalho do audiodescritor, a IA pode potencializar a produção de roteiros, desde que acompanhada de curadoria humana, contextualização adequada das entradas, formação profissional contínua e respeito à privacidade e aos direitos autorais. Seguir essas recomendações possibilita integrar a tecnologia de forma segura e eficiente, apoiando a produção de audiodescrições e ampliando o acesso do público com deficiência visual ao conteúdo audiovisual.



REFERÊNCIAS

ABREU, Helen de Andrade; MATOS, Ely Edison da Silva. A FrameNet Brasil Approach to Annotation of Pragmatic Frames Evoked by Turn Organization Gestures. *Revista Caligrama: Revista de Estudos Românicos*, [S. l.], v. 30, n. 1, p. 94–109, 2025. DOI: 10.35699/2317-2096.2025.57543.

BELCAVELLO, Frederico; TORRENT, Tiago Timponi; MATOS, Ely E.; PAGANO, Adriana S.; GAMONAL, Maucha; SIGILIANO, Natalia; DUTRA, Livia Vicente; ABREU, Helen de Andrade; SAMAGAIO, Mairon; CARVALHO, Mariane; CAMPOS, Franciany; AZALIM, Gabrielly; MAZZEI, Bruna; OLIVEIRA, Mateus Fonseca de; LUZ, Ana Carolina Locasso; RUIZ, Lúvia Padua; BELLEI, Júlia; PESTANA, Amanda; COSTA, Josiane; RABELO, Iasmin; SILVA, Anna Beatriz; ROZA, Raquel; SOUZA, Mariana; OLIVEIRA, Igor. Frame2: a FrameNet-based multimodal dataset for tackling text-image interactions in video. In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Torino, Itália: ELRA; ICCL, 2024. p. 7429–7437. Disponível em: <https://aclanthology.org/2024.lrec-coling.744/>. Acesso em: 9 jan. 2026.

CAMPOS, Virginia. *Sistema de geração automática de audiodescrição a partir de análise de conteúdo de vídeo*. 2019. 92 f. Tese (Doutorado em Engenharia Elétrica e de Computação) — Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, Natal, 2019. Disponível em: <https://repositorio.ufrn.br/jsui/handle/123456789/28616>. Acesso em: 06 nov. 2025.

CHEEMA, Maryam S.; ELAHIMANESH, Sina; MARTIN, Samuel; FAZLI, Pooyan; SEIFI, Hasti. DescribePro: collaborative audio description with human-AI interaction. In: *Proceedings of the 27th international acm sigaccess conference on computers and accessibility (assets '25)*. New York: Association for Computing Machinery, 2025. Artigo 86, p. 1–19. DOI: 10.1145/3663547.3746320.

DE SOUZA, Débora S.; PAGANO, Adriana Silvina; GAMONAL, Maucha Andrade. A audiodescrição sob a perspectiva da Semântica de Frames: um estudo exploratório. *Revista Gatilho*, Juiz de Fora, v. 23, n.1, p. 101-125, 2022. DOI: 10.34019/1808-9461.2022.v23.38564.

DORNELAS, Letícia; GAMONAL, Maucha Andrade; PAGANO, Adriana Silvina. Análise semântica de audiodescrição em curta-metragem: uma abordagem multimodal a partir da Semântica de Frames. *Revista Domínios de Linguagem*, Uberlândia, v. 18, p. e1866, 2024. DOI: 10.14393/DLv18a2024-66. Disponível em: <https://seer.ufu.br/index.php/dominiosdelinguagem/article/view/75272>. Acesso em: 26 jan. 2026.

FILLMORE, Charles J. Frame semantics. In: THE LINGUISTIC Society of Korea (org.). *Linguistics in the Morning Calm*. Seul: Hanshin Publishing Co., 1982. p. 111-137.

FILLMORE, Charles J. Frames and the semantics of understanding. *Quaderni di Semantica*, v. 6, n. 2, p. 222-254, 1985.

FRYER, Louise. An introduction to audio description: A practical guide. London: Routledge, 2016.

GAMONAL, Maucha Andrade; PAGANO, Adriana Silvina; TORRENT, Tiago Timponi; MATOS, Ely Edison. Frame-Based Semantic Representation and Similarity Analysis in Audio Description Scripts. In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA (STIL), 16., 2025, Fortaleza, CE. *Anais do XVI Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*. Porto Alegre: Sociedade Brasileira de Computação, 2025a. p. 534-541. DOI: 10.5753/stil.2025.37854.

GAMONAL, Maucha Andrade; TORRENT, Tiago Timponi; MATOS, Ely Edison; PAGANO, Adriana S.; BELCAVELLO, Frederico; MAYER, Flavia Affonso; LORENZI, Arthur; SIGILIANO, Natalia S.; ABREU, Helen de Andrade; DUTRA, Livia Vicente; VIRIDIANO, Marcelo; CONEGLIAN, Andre; HERBST, Victor A. S.; CAMPOS, Franciany O.; BROWN, Kenneth; RUIZ, Livia Padua; BONOTO, Lisandra Carvalho; PEREIRA, Luiz Fernando; NAVARRO, Yulla Liquer. Audition: a frame-annotated multimodal dataset for accessible audiovisual content. In: *Proceedings of the 21st Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-21), IWCS 2025*. Düsseldorf, Germany: Tilburg University, 2025b.

JACQUES, Edson Gabriel; SACRAMENTO, Carolina; GOUVEIA, Yndiana; SILVA, Wagner Nascimento; BARROS, Ygor Santos; FERREIRA, Simone Bacellar Leal. Preservação da memória com acessibilidade digital: um plugin para descrição de imagens com IA generativa. TRILHA DE INDÚSTRIA E INOVAÇÃO EM SISTEMAS DE INFORMAÇÃO – SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO (SBSI), 21., 2025, Recife, PE. *Anais Estendidos do XXI Simpósio Brasileiro de Sistemas de Informação*. Porto Alegre: Sociedade Brasileira de Computação, 2025. p. 157–161. Disponível em: https://sol.sbc.org.br/index.php/sbsi_estendido/article/view/34596. Acesso em: 6 nov. 2025.

KALAI, Adam; NACHUM, Ofir; VEMPALA, Santosh; ZHANG, Edwin. *Why language models hallucinate*. 2025. Disponível em: <https://arxiv.org/abs/2509.04664>. Acesso em: 6 nov. 2025.

NAVES, Sylvia Bahiense; MAUCH, Carla; ALVES, Soraya Ferreira; ARAÚJO, Vera Lúcia Santiago. (org.). *Guia para Produções Audiovisuais Acessíveis*. Brasília: Secretaria do Audiovisual do Ministério da Cultura, 2016. Disponível em: <https://www.gov.br/culturaviva/pt-br/biblioteca-cultura-viva/documentos-e-publicacoes/documentos/minc-guia-para-producoes-audiovisuais-acessiveis-com-audiodescricao-das-imagens-2016.pdf>. Acesso em: 6 nov. 2025.

MITCHELL, Margaret; ATTANASIO, Giuseppe; BALDINI, Ioana; CLINCIU, Miruna; CLIVE, Jordan; DELOBELLE, Pieter; DEY, Manan; HAMILTON, Sil; DILL, Timm; DOUGHMAN, Jad; DUTT, Ritam; GHOSH, Avijit; FORDE, Jessica Zosa; HOLTERMANN, Carolin; KAFFEE, Lucie-Aimée; LAUD, Tanmay; LAUSCHER, Anne; LOPEZ-DAVILA, Roberto L.; MASOUD, Maraim; NANGIA, Nikita; OVALLE, Anaelia; PISTILLI, Giada; RADEV, Dragomir; SAVOLDI, Beatrice; RAHEJA, Vipul; QIN, Jeremy; PLOEGER, Esther; SUBRAMONIAN, Arjun; DHOLE, Kaustubh; SUN, Kaiser; DJANIBEKOV, Amirbek; MANSUROV, Jonibek; YIN, Kayo; CUEVA, Emilio Villa; MUKHERJEE, Sagnik; HUANG, Jerry; SHEN, Xudong; GALA, Jay; AL-ALI, Hamdan; DJANIBEKOV, Tair; MUKHITULY, Nurdautlet; NIE, Shangrui; SHARMA, Shanya; STANCZAK, Karolina; SZCZECZLA, Eliza; TORRENT, Tiago Timponi; TUNUGUNTLA, Deepak; VIRIDIANO, Marcelo; VAN DER WAL, Oskar; YAKEFU, Adina; NÉVÉOL, Aurélie; ZHANG, Mike; ZINK, Sydney; TALAT, Zeerak. SHADES: Towards a multilingual assessment of stereotypes in large language models. *In: Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies*. v. 1, p. 11995–12041, 2025. Disponível em: <https://aclanthology.org/2025.naacl-long.600/>. Acesso em: 26 jan. 2026.

PAGANO, Adriana S.; TEIXEIRA, André L. R.; MAYER, Flávia A. Accessible Audiovisual Translation. In: JI, Meng; LAVIOSA, Sara (ed.). *The Oxford Handbook of Translation and Social Practices*. Oxford: Oxford University Press, 2020. Cap. 4, p. 66-82.

RABELO, Iasmin; GAMONAL, Maucha; PAGANO, Adriana. Disability and assistive technology: modeling accessibility domain frames on FrameNet Brasil. *Cadernos de Estudos Linguísticos*, Campinas, SP, v. 67, n. 00, p. e025032, 2025. DOI: <https://doi.org/10.20396/cel.v67i00.8679802>.

RIBEIRO, Aline; ANDRADE, Luiz Antônio. IA e inclusão: potencialidades e riscos da automação da audiodescrição. *Revista Políticas Públicas & Cidades*, v. 14, n. 7, 2025. Disponível em: <https://journalppc.com/RPPC/article/view/2478>. Acesso em: 06 nov. 2025.

ROSSETTI, Julia; BORGES, Luciana; SOUZA, Patricia. Descrição Automática de Imagens por Ia: Uma Leitura Ética e Representativa. In: SIMPÓSIO BRASILEIRO SOBRE FATORES HUMANOS EM SISTEMAS COMPUTACIONAIS (IHC), 24, 2025, Belo Horizonte/MG. *Anais do XXIV Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*. Porto Alegre: Sociedade Brasileira de Computação, 2025. p. 1608-1628. Disponível em: <https://sol.sbc.org.br/index.php/ihc/article/view/37724>. Acesso em: 07 nov. 2025.

TORRENT, Tiago Timponi; MATOS, Ely Edison da Silva; BELCAVELLO, Frederico; VIRIDIANO, Marcelo; GAMONAL, Maucha Andrade; COSTA, Alexandre Diniz da; MARIM, Mateus Coutinho. Representing context in FrameNet: a multidimensional, multimodal approach. *Frontiers in Psychology*, v. 13, 2022. DOI: <https://doi.org/10.3389/fpsyg.2022.838441>

VIRIDIANO, Marcelo; LORENZI, Arthur; TORRENT, Tiago Timponi; MATOS, Ely E.; PAGANO, Adriana S.; SIGILIANO, Natalia Sathler; GAMONAL, Maucha; ABREU, Helen de Andrade; DUTRA, Livia Vicente; SAMAGAI, Mairon; CARVALHO, Mariane; CAMPOS, Franciany; AZALIM, Gabrielly; MAZZEI, Bruna; OLIVEIRA, Mateus Fonseca de; LUZ, Ana Carolina; RUIZ, Livia Padua; BELLEI, Julia; PESTANA, Amanda; COSTA, Josiane; RABELO, Iasmin; SILVA, Anna Beatriz; ROZA, Raquel; MOTA, Mariana Souza; OLIVEIRA, Igor; FREITAS, Marcio Henrique Pelegrino de. Framed Multi30K: a frame-based multimodal-multilingual dataset. In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Torino, Italia: ELRA; ICCL, 2024. p. 7438–7449.

QUEM SOMOS

Maucha Andrade Gamonal é doutora em Linguística pela UFJF, com período de doutorado sanduíche na University of California, Berkeley. É pesquisadora na FrameNet Brasil na área de Linguística Cognitiva e Computacional.

Flávia Affonso Mayer é doutora em Linguística pela PUC Minas, com estágio na Case Western Reserve University. É professora do Departamento de Comunicação da UFPB. Coordena o Observatório da Linguagem e Inclusão.

Frederico Belcavello é doutor em Linguística pela UFJF, com estágios na Case Western Reserve University. É professor do UniAcademia e pesquisador da FrameNet Brasil.

Igor Antônio Lourenço da Silva é doutor em Estudos Linguísticos pela UFMG, com estágio na Universidade do Sarre. É professor da UFU onde pesquisa tradução, inteligência artificial e rastreamento ocular.

Mariane de Carvalho Pinto é professora da rede municipal de Juiz de Fora e mestre em Linguística pela UFJF. Desenvolve pesquisas em Semântica de Frames, multimodalidade, IA e audiodescrição.

Adriana Silvina Pagano é professora da UFMG. Doutora em Letras pela UFMG, com estágio na University of Nottingham. Desenvolve pesquisas sobre modelagem sistêmico-funcional da linguagem, abordagens quantitativas da tradução e Processamento de Língua Natural.

Tiago Timponi Torrent é doutor em Linguística pela UFRJ, com estágio na University of California, Berkeley. É professor da UFJF e visitante na Universidade de Gotemburgo. Lidera a FrameNet Brasil e desenvolve recursos, datasets multimodais e modelos de IA para o português brasileiro.

