



Análise semântica de audiodescrição em curta-metragem: uma abordagem multimodal a partir da Semântica de Frames

Semantic analysis of audio description in short films: a multimodal approach based on Frame Semantics

Letícia Grossi DORNELAS*

Maucha Andrade GAMONAL**

Adriana Silvina PAGANO***

RESUMO: Este trabalho apresenta um estudo exploratório de anotação semântica multimodal com base na Semântica de *Frames* (Fillmore, 1982) e nos Estudos da Tradução Audiovisual. Por meio de um corpus anotado semanticamente pela metodologia da FrameNet Brasil para a multimodalidade (Belcavello *et al.*, 2022), avaliam-se as relações semânticas dos *frames* evocados na transcrição da audiodescrição e nas imagens dinâmicas (cenas em vídeo) do curta-metragem brasileiro *Eu não quero voltar sozinho*, produzido em 2010 pela Lacuna Filmes. Para tanto, organizam-se os *frames* a partir das categorias raízes (*root types*), que são agrupamentos de *frames* estruturalmente semelhantes (Ruppenhofer *et al.*, 2016), levando em consideração os critérios adotados na criação de uma audiodescrição, conforme o Guia para Produções Audiovisuais Acessíveis (Naves *et al.*, 2016). Os resultados demonstram que os elementos dos modos comunicativos atuam simultaneamente na evocação e invocação de *frames* semânticos, contribuindo para a construção do sentido em uma obra audiovisual, que, por essência, é de natureza multimodal.

Palavras-chave: Estudos da Tradução Audiovisual. Audiodescrição. Semântica de *Frames*. FrameNet Brasil. Anotação semântica multimodal.

ABSTRACT: This work presents an exploratory study of multimodal semantic annotation based on Frame Semantics (Fillmore, 1982) and Audiovisual Translation Studies. Through a corpus semantically annotated using the FrameNet Brasil methodology for multimodality (Belcavello *et al.*, 2022), the semantic relations of the frames evoked in the transcription of the audio description and the dynamic images (video scenes) of the Brazilian short film *Eu não quero voltar sozinho*, produced in 2010 by Lacuna Filmes, are analyzed. For this purpose, frames are organized based on root types, which are groupings of structurally similar frames (Ruppenhofer *et al.*, 2016), considering the criteria adopted in the creation of an audio

* Mestre em Linguística Aplicada pela Universidade Federal de Minas Gerais. leticiagdornelas@gmail.com

** Doutora em Estudos Linguísticos pela Universidade Federal de Juiz de Fora e residente de pós-doutorado na Universidade Federal de Minas Gerais. mgamonal@ufmg.br

*** Doutora em Estudos Literários pela Universidade Federal de Minas Gerais e professora titular da Universidade Federal de Minas Gerais. apagano@ufmg.br

description, according to the Guide for Accessible Audiovisual Productions (Brazil, 2016). The results demonstrate that the elements of the communicative modes simultaneously act in the evocation and invocation of frames, contributing to the meaning in audiovisual work, which by nature is multimodal.

Keywords: Audiovisual Translation Studies. Audio description. Frame Semantics. FrameNet Brasil. Multimodal semantic annotation.

Artigo recebido em: 12.09.2024

Artigo aprovado em: 27.11.2024

1 Introdução

A Tradução Audiovisual Acessível (Tava) com todas as suas modalidades é um subcampo disciplinar dos Estudos da Tradução. Neste trabalho, atua-se com o modo Audiodescrição de Filmes (ADF), relacionando-o à teoria linguística da Semântica de *Frames* (Fillmore, 1982; 1985). O objetivo é realizar uma análise multimodal exploratória dos *frames* evocados no texto transcrito da ADF e nas imagens dinâmicas (cenas) do curta-metragem (filme de curta duração) *Eu não quero voltar sozinho*, lançado em 2010 pela Lacuna Filmes e dirigido por Daniel Ribeiro. A ADF usada neste trabalho foi criada pelo Tramad (Tradução, Mídia e Audiodescrição), grupo de pesquisa da UFBA.

A partir desse material, as anotações semânticas feitas para as sentenças que compõem a audiodescrição e para os objetos visuais audiodescritos são avaliadas, buscando os *frames* evocados nesses dois modos comunicativos. As anotações semânticas de texto seguem a metodologia da FrameNet Brasil (FN-Br) (Ruppenhofer *et al.*, 2016; Torrent; Ellsworth, 2013), e as anotações de imagens, a proposta de anotação multimodal em Belcavello *et al.* (2022) e Belcavello (2023). Em seguida, verificam-se os resultados obtidos de acordo com os critérios de estruturação de audiodescrições disponíveis no Guia para Produções Audiovisuais Acessíveis (Naves *et al.*, 2016).

A pesquisa é estruturada por meio das anotações dos *frames* da transcrição da audiodescrição e das imagens dinâmicas, após isso, faz-se o levantamento dos *frames* evocados em cada modalidade a partir do agrupamento de *frames* dispostos pela FN-

Br no recurso Lutma (Torrent *et al.*, 2022) e, por fim, exploram-se os critérios adotados na elaboração de uma audiodescrição de modo a identificar os tipos de *frames* semânticos atuantes na experiência do público usuário.

Dos resultados deste trabalho, tem-se um corpus anotado semanticamente no *dataset* multimodal da ReINVenTA (Research and Innovation Network for Vision and Text Analysis), rede de pesquisa em processamento semântico computacional que alia a análise semântica de objetos multimodais a *frames* semânticos. Dentre os objetivos da rede, estão previstas a constituição de um *gold standard dataset* de objetos multimodais anotados semanticamente e validado psicolinguisticamente (Viridiano *et al.*, 2022; Belcavello, 2023) e o desenvolvimento de algoritmos de inteligência artificial para rotulação semântica automática de objetos multimodais. No que tange ao domínio da Tava, este trabalho, assim como outras pesquisas (De Souza *et al.*, 2022), evidenciou a pertinência da Semântica de *Frames* tanto nas reflexões quanto na prática da elaboração de audiodescrição. Essa teoria, como destacado por Czulo (2024) na enciclopédia da Associação Ibérica de Estudos de Tradução e Interpretação, atua tanto na análise quanto no método de tradução, tendo o enfoque no estudo empírico das relações entre unidades linguísticas e conceitos mentais.

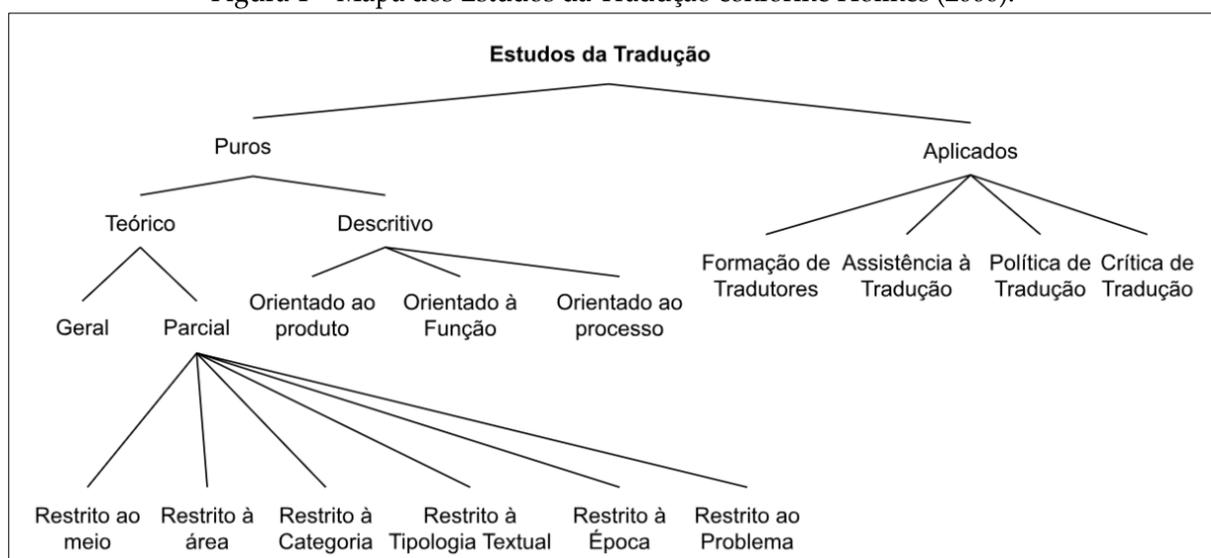
Na organização deste artigo, além desta introdução, discutem-se os Estudos da Tradução, a Semântica de *Frames* e os princípios da multimodalidade. Em seguida, o desenvolvimento do trabalho é composto pela apresentação da FrameNet Brasil, bem como os procedimentos para a anotação de texto e de imagem dinâmica e, por último, as diretrizes para elaboração de audiodescrição. Com isso feito, exibem-se o levantamento das anotações e a sua relação com as diretrizes. Por fim, são expostas as considerações finais, os agradecimentos e as referências bibliográficas.

2 Estudos da Tradução

De acordo com Holmes (1988), os Estudos da Tradução se dividem entre puros e aplicados. Os puros são classificados em teóricos (geral, que descreve as teorias, e

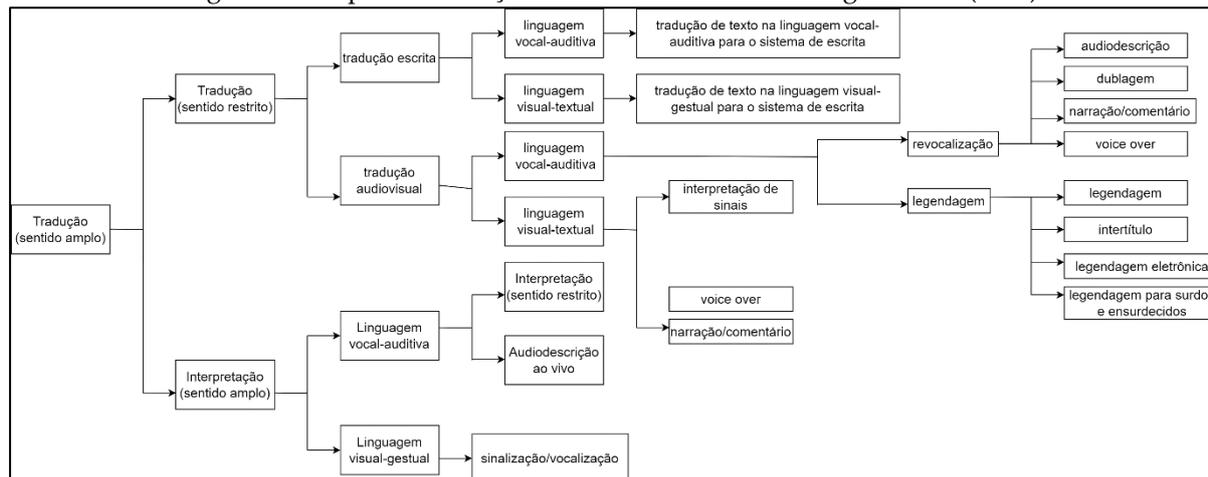
parcial, que descreve os parâmetros) e descritivos (orientado ao produto, orientado ao processo e orientado à função). Já os estudos aplicados atuam com a formação de tradutores, auxílio à tradução, política da tradução e críticas à tradução. Esta pesquisa se afilia aos estudos descritivos orientados ao produto uma vez que a ADF é o produto final de uma tradução e também está aliada aos estudos aplicados de materiais de assistência à tradução, visto que se trabalha com ferramentas que analisam a ADF com o intuito de, posteriormente, auxiliar na prática tradutória. A Figura 1 mostra o mapa dos Estudos da Tradução em que essas divisões estão ilustradas:

Figura 1 – Mapa dos Estudos da Tradução conforme Holmes (2000).



Fonte: elaborada pelas autoras com base em Holmes (2000).

Este estudo também atuou no campo da TAVA na modalidade de AD. Por isso, a seguir, a Figura 2 mostra o mapa da TAVA que inclui a ADF. De acordo com a figura, a tradução no sentido amplo se divide em tradução no sentido restrito e interpretação no sentido amplo. Na categoria de tradução no sentido restrito, há as modalidades de tradução escrita e tradução audiovisual. Fazem parte da Tava as modalidades da linguagem visual-gestual e linguagem vocal-auditiva. Essa última se ramifica em legendagem e revocalização. No subcampo de revocalização, estão classificadas a audiodescrição, a dublagem, o *voice-over* e a narração/comentário. Esta pesquisa investiga os estudos de revocalização focados na audiodescrição.

Figura 2 – Mapa da Tradução Audiovisual conforme Pagano *et al.* (2020).

Fonte: Pagano *et al.* (2020, p. 76)¹.

A ADF tem como finalidade oferecer acesso a um material audiovisual às pessoas com deficiência visual. Essa funcionalidade é inserida por meio de uma narração nos intervalos entre os diálogos, compartilhando informações que completam o sentido da cena como ações, expressões faciais, figurinos, cenários. Essa categoria trabalha com os meios de produção e reprodução, por exemplo, diálogo de signos, leitura, metacriação, entre outros. Para ilustrar melhor esta modalidade, Mayer (2012) descreve seus principais objetivos e sua forma de aplicação:

Tendo como um de seus principais objetivos fornecer informação adicional ao público – deficiente visual ou não –, a audiodescrição disponibiliza por meio do som detalhes visuais importantes como figurinos, indicação de tempo e espaço, movimentações e ações dos personagens. Para um melhor resultado, essas narrações extras devem ser inseridas nos intervalos dos diálogos e dos ruídos importantes, de modo a não se sobrepor aos efeitos musicais e sonoros já existentes – extremamente importantes para construção da cena (Mayer, 2012, p. 23).

Nesse sentido, a audiodescrição não é feita apenas de descrições, e sim narrações com o intuito de tornar acessíveis as informações mais importantes, possibilitando o acesso do público usuário a um determinado material audiovisual.

¹ Imagem traduzida pelas autoras.

Em 2016, foi desenvolvido e publicado, pela Secretaria do Audiovisual do Ministério da Cultura, o Guia para Produções Audiovisuais Acessíveis. No guia, são abordados critérios para elaboração do roteiro de AD. O material será abordado posteriormente.

A fim de aprofundar o entendimento das teorias utilizadas neste estudo, em seguida, a Semântica de *Frames* é apresentada.

3 Semântica de Frames

A Semântica de *Frames* (SF), teoria na qual o foco de estudo é a linguagem e a experiência, descreve a língua a partir do uso num viés descritivo juntamente à semântica empírica. O objetivo do fundador, Charles Fillmore, é encontrar as razões que fundamentam as categorizações humanas. Para tanto, ele leva em consideração o uso linguístico e o contexto utilizado pelos falantes. Para Fillmore:

com o termo semântica de *frames*, eu tenho em mente um programa de pesquisa em linguística empírica e uma metodologia descritiva para apresentar os resultados de tal pesquisa. [...] Pelo termo *frame*, eu tenho em mente qualquer sistema de conceitos relacionado de tal forma que, para entender qualquer um deles, você tem de entender toda a estrutura na qual ele se encaixa; quando um dos conceitos em dada estrutura é introduzido dentro de um texto ou de uma conversa, todos os outros são automaticamente disponibilizados (Fillmore, 1982, p. 111).

Este autor defende que a definição das palavras está diretamente ligada à compreensão das instituições sociais ou da estrutura de experiência que elas pressupõem, isto é, uma palavra para a SF representa uma categoria de experiência do falante. Assim, os *frames*, estão ligados à experiência humana e possuem uma porção significativa de cultura (Fillmore, 1982, p. 118).

Frames evocados e invocados foram duas distinções feitas para mostrar a natureza deles na interação humana:

Frames interpretativos podem ser introduzidos no processo de compreensão de um texto sendo invocados pelo intérprete ou sendo evocados pelo texto. Um *frame* é invocado quando o intérprete, no processo de construção de sentido de dado segmento textual, é capaz de fazer uma interpretação situando o conteúdo textual em um padrão que é conhecido independentemente do texto. Um *frame* é evocado pelo texto se alguma forma ou padrão linguístico é convencionalmente associado com o *frame* em questão (Fillmore, 1985, p. 232).

Em Gamonal (2013), a autora apresenta essa distinção por meio dos seguintes exemplos:

- (1) Não há nada mais prazeroso que chegar a um lugar que ninguém te conhece. Você acorda bem cedo, vai à praia e assiste ao pôr do sol, deixando todas as obrigações de lado por uns dias.
- (2) Sempre que chegar a uma cidade turística, tome cuidado com a bolsa e demais pertences. Os assaltantes reconhecem os visitantes de longe.

Na frase (1), há um exemplo de *frame* invocado. Tem-se um conjunto de informações que sinalizam uma atividade turística, sem que haja qualquer material linguístico específico desse domínio. Já na frase (2), **cidade turística** e **visitantes** atuam como materiais linguísticos que especificamente evocam o contexto turístico.

Na audiodescrição em análise neste trabalho, esse fenômeno foi percebido. A sentença (3) é um exemplo do curta-metragem. Não há evocação do *frame* *Subpartes_de_prédio*², mas a integração dos modos comunicativos invoca tal *frame*, sugerindo se tratar de uma sala de aula.

- (3) A professora sentada diante de sua mesa olha a turma.

A [professora^{Pessoas_por_vocação}] [sentada^{Movimento_corporal}] [diante^{Proximidade_não_graduável}] de sua [mesa^{Móveis}] [olha^{Percepção_ativa}] a [turma^{Agregado}].

² A definição e demais informações do *frame* *Subpartes_de_prédios* está disponível em: <https://webtool.frame.net.br/report/frame/169>. Acesso em: 10 de nov. de 2024.

De acordo com Torrent *et al.* (2022), as informações linguísticas visuais expressas nas cenas se tornam fonte de informações utilizadas para construir um significado. Por isso a importância de vincular os fundamentos da multimodalidade à teoria exposta nesta seção. Considera-se que, assim como o texto evoca/invoca *frames*, as informações visuais também o fazem na construção de sentido. Diante disso, apresentam-se a seguir princípios da multimodalidade.

4 Princípios da Multimodalidade

A anotação multimodal presente nesta pesquisa é estruturada com a transcrição do texto da ADF e com as imagens dinâmicas do curta-metragem. Para Bateman *et al.* (2017), os textos participam do gênero, isto é, os textos não pertencem ao gênero. Para esses autores, texto e gênero não podem ser separados. “Qualquer texto deve ser visto como existente ao gênero” (Bateman *et al.*, 2017, p. 129). Para Lemke (1999) *apud* Bateman *et al.* (2017):

Construímos gêneros interpretando certos tipos de padrões semânticos no que consideramos ser textos distintos, e dizemos que tais textos pertencem ao mesmo gênero. Textos cogenéricos são intertextos privilegiados para a interpretação uns dos outros” (Lemke 1999 *apud* Bateman *et al.*, 2017, p. 129, tradução nossa).

Swales (1990) *apud* Bateman *et al.* (2017) define gênero como:

Um gênero é uma classe de eventos comunicativos que compartilham um propósito comunicativo reconhecível, que exibem uma estrutura esquemática que suporta a realização desse propósito e que mostram semelhanças em forma, estilo, conteúdo, estrutura e público-alvo (Swales 1990 *apud* Bateman *et al.* 2017, p. 130, tradução nossa).

Van Leeuwen (2005 *apud* Hiippala, 2014) aponta que os aspectos multimodais são características do gênero multimodal. Para Bateman *et al.* (2017), é necessário identificar as estruturas que compõem este gênero. Dessa forma, conforme Hiippala

(2014), por ser visto como um fenômeno estratificado, as escolhas semióticas devem ser identificadas uma vez que contribuem diretamente para as estruturas de gênero em diversos estratos.

Bateman *et al.* (2017) propõem que o modo semiótico seja composto por três estratos: o material, os recursos semióticos e a semântica do discurso. Hiippala (2014) aponta que o segundo estrato, recursos semióticos, permite diversas escolhas semióticas que juntas se compõem às estruturas do gênero. Além disso, este autor descreve que a semântica do discurso orienta a interpretação contextual de tais recursos.

Para o gênero multimodal, conforme Hiippala (2014), são as estruturas de composição do material como o *layout*, o tamanho das fontes, a posição e o ângulo da imagem, as cores, o espaço entre o texto e a imagem. Todos estes aspectos visuais-verbais são resultados de escolhas semióticas. Esses conceitos contribuem para a construção de um contexto multimodal. Em um material multimodal, de acordo com Bateman *et al.* (2017), todas as facetas semióticas contribuem para a construção do significado como o texto, o discurso, o gênero, as imagens, os movimentos, o *layout*.

Assim, conforme destacam Santos e Pimenta (2014), o campo da multimodalidade leva em consideração os diversos modos e meios de significação às questões socioculturais para explorar a produção de significados.

A seguir, estão descritos os materiais e métodos utilizados para anotação e análise da obra audiovisual.

5 FrameNet Brasil

A FrameNet³ é um projeto que surgiu sob liderança de Fillmore, em Berkeley, Califórnia, e que está em desenvolvimento desde 1997. A metodologia é direcionada pela Semântica de *Frames*, e as análises, amparadas por meio de *corpora*. Os resultados

³ Rede semântica FrameNet. Disponível em: <https://framenet.icsi.berkeley.edu>. Acesso em: 10 de nov. 2024.

são disponibilizados em forma de recurso lexical on-line, utilizado por pesquisadores no mundo todo de diferentes áreas de atuação.

A plataforma conta com um banco de dados de mais de mil frames acompanhados de anotações de sentenças que servem como evidências para análises linguísticas e fonte de pesquisa em diversas áreas, como Linguística de *Corpus*, Estudos da Tradução e Linguística Computacional. De acordo com Fillmore (2010), os objetivos são fornecer descrições sintáticas e semânticas do léxico da língua inglesa e reunir informações relevantes sobre a construção do sentido lexical.

Com o crescente desenvolvimento, segundo Torrent e Ellsworth (2013), a FrameNet passou a ser desenvolvida em outros idiomas, dentre eles o português brasileiro por meio da FrameNet Brasil⁴ (FN-Br), que foi fundada na Universidade Federal de Juiz de Fora, em Minas Gerais, constituindo-se como um laboratório de Linguística Computacional a fim de desenvolver soluções para a Compreensão Automática de Linguagem Natural. A FN-Br, de acordo com os princípios do projeto original, constrói um banco de dados de *frames* semânticos para o português brasileiro.

Para um melhor entendimento da metodologia da FrameNet e da FN-Br em torno do léxico, algumas nomenclaturas utilizadas no procedimento de anotação de texto tornam-se relevantes de serem apresentadas e explicadas, são elas: **formas de palavras, lexemas, lemas e unidade lexical**.

-**Formas de palavras:** formas flexionadas das palavras, por exemplo: **amigo, amiga, amigos** etc.

-**Lexemas:** palavras da mesma classe morfológica. Na FrameNet, prevalece o masculino singular, seguindo o exemplo anterior, o lexema é: **amigo**.

-**Lemas:** compostos por um lexema (monolexêmico) ou mais (multilexêmico), sendo marcado por uma classe de palavra, por exemplo, o lema **pé de moleque** constitui três lexemas (1) **pé**, (2) **de**, (3) **moleque**, isto é um lema multilexêmico.

⁴ Rede semântica FrameNet Brasil. Disponível em: <https://webtool.frame.net.br>. Acesso em: 10 de nov. 2024.

-Unidade lexical (LU, do inglês): uma associação de um lema a um *frame* e pode ser entendida como a junção de forma e significado, por exemplo, o lema *visitar.v*, que evoca significados diferentes dependendo do *frame*. Quando esse lema está associado ao *frame* *Fazer_turismo*⁵ significa conhecer uma nova cidade e/ou um ponto turístico. Já ao associá-lo ao *frame* *Visitar* evoca diversos significados, como rever uma pessoa (*Os sobrinhos **visitaram**^{VISITAR} a avó*) e realizar uma fiscalização (*Os policiais **visitaram**^{VISITAR} os bares de São Paulo*). Isso porque conforme a definição “*um Agente combina a localização com uma Entidade para satisfazer alguma Finalidade. Geralmente, a Finalidade é social ou de entretenimento*” (FrameNet Brasil, 2024).

A FrameNet Brasil desenvolveu um recurso de criação de *frames*, o Lutma⁶. Este projeto teve o interesse de propiciar a colaboração entre laboratórios e pesquisadores de diferentes países com o objetivo de facilitar o compartilhamento de dados e pesquisas sobre a metodologia da FrameNet para o público não especializado. De acordo com Torrent *et al.* (2022), o Lutma possui dois diferenciais: (1) durante a criação dos *frames* e das LUs deve ser indicada a língua/cultura a que pertencem, (2) os usuários devem ser pessoas interessadas na FrameNet e ter pouca ou nenhuma experiência na criação de *frames*. Para os autores, estes propósitos contribuem em construir um recurso linguístico baseado em cognição que seja acessível para não especialistas.

No Lutma, os *frames* são categorizados pelos tipos de raízes (*Root Type*) a que pertencem. Além desta nomenclatura, também se utiliza *frames* de topo (*Top Level*). As categorias raízes abordadas no Lutma, segundo expostas na plataforma, são Entidade, Evento, Estado, Relação e Atributo. Nesta pesquisa, o estudo dos *frames* segue as categorias utilizadas no Lutma. As definições apresentadas para cada categoria, conforme está demonstrado na interface do recurso, são:

⁵ Seguindo convenções estabelecidas, o nome do *frame* é registrado em fonte Courier, e os Elementos de *Frame* em letras maiúsculas.

⁶ Ferramenta para criação de *frames* semânticos por não especialistas. Disponível em: <https://lutma.frame.net.br>. Acesso em: 10 de nov. 2024.

- (i) Entidade: um conceito, uma coisa, um ser vivo.
- (ii) Evento: envolve pelo menos um participante e ocorre em um local e hora.
- (iii) Estado: uma situação estável.
- (iv) Relação: uma relação entre duas ou mais entidades, eventos ou estados.
- (v) Atributo: um atributo com um valor é aplicado a uma entidade.

Os *frames* que não se enquadram em nenhuma das opções acima são inseridos no conjunto Indefinido. A diferença entre as categorias e o conjunto é que os *frames* ligados às categorias mantêm relação de herança, isto é, aqueles categorizados como Entidade herdam do *frame* Entidade, e isso ocorre nas outras categorias. Já o conjunto Indefinido engloba os que não têm relação de herança com as categorias pré-determinadas.

A fim de expandir os estudos com *frame* para além do texto verbal, a FN-Br desenvolve a anotação semântica de materiais multimodais. Por exemplo, em um filme no qual se apresentam audiodescrição, cenário e diálogos, considera-se que haja diversos objetos modais atuando na evocação de *frames*.

A Webtool, plataforma de anotação multimodal da FN-Br, trabalha com anotação de imagens dinâmicas e estáticas. Por meio deste *software* de anotação, vinculam-se as anotações de texto e imagem/vídeo. Assim, é possível analisar um mesmo material através do texto e dos demais modos comunicativos.

Para a tarefa de anotação multimodal, adota-se uma abordagem perspectivista. Cabitza *et al.* (2021) apontam que uma anotação no viés perspectivista é uma interação entre o anotador e o esquema de anotação em que ele expressa seu julgamento sobre o fenômeno anotado. Estes autores apontam duas metodologias nesse viés: o perspectivismo fraco e o perspectivismo forte.

O perspectivismo forte, segundo os autores, tem o objetivo de coletar múltiplos dados sobre um fenômeno específico. Já o perspectivismo fraco não tem o intuito de avaliar um único fenômeno, por isso visa a coletar o maior número de avaliadores e anotações. Nas tarefas de anotação desempenhadas neste estágio da pesquisa, adotou-

se o viés perspectivista fraco como metodologia de anotação. É importante ressaltar que essa metodologia também adotada pelas demais pesquisas multimodais na FN-Br, em que o anotador expressa o próprio julgamento de análise, é desenvolvida mediante orientações vindas dos estudos sobre os *frames*. Isto é, para o anotador realizar uma anotação ele deve ter obtido formação tanto sobre a abordagem da SF e sobre a metodologia de anotação que inclui compreender a definição do *frame*, o comportamento dos Elementos de *Frames* e o procedimento de anotação, no caso, semântica. Desse modo, o anotador poderá imprimir a própria perspectiva na anotação.

6 Anotação de texto e imagem dinâmica

As anotações de texto⁷ são feitas essencialmente em três camadas⁸. Os Elementos de *Frame* (FE, do inglês, *Frame Element*) são os primeiros pontos a serem analisados, ou seja, a primeira camada na qual buscamos informações semânticas. A segunda camada é a análise da Função Gramatical (GF, do inglês, *Gramatical Function*), e a terceira é o Tipo Sintagmático (PT, do inglês, *Phrase Type*), em ambas, investigam-se as informações sintáticas.

Na Webtool, ao acessar a constituição de um *frame*, encontra-se sua definição a partir dos elementos que o compõem. Conforme demonstrado na Figura 3, os participantes do *frame* são nomeados como Elementos de *Frames* (FE, do inglês, *Frame Element*) e já estão definidos na plataforma. Os FEs são divididos em nucleares quando são essenciais na composição do *frame* e não nucleares quando não são essenciais.

⁷ Todas os exemplos de anotações textuais nesta seção advêm da transcrição da audiodescrição do curta “Eu não quero voltar sozinho” e foram anotados semanticamente por meio da Webtool.

⁸ Para saber mais sobre as camadas de anotações da FrameNet, sugere-se o artigo *Behind The Labels: Criteria For Defining Analytical Categories In Framenet Brasil* (Ellsworth, M.; Torrent, T., 2013)

Figura 3 – O frame Comércio Comprar.

Comércio_comprar @Lexical @Business #155 Commerce_buy [en]

Definition
Este frame descreve uma transação comercial básica, envolvendo um **Comprador** e um **Vendedor** que trocam **Dinheiro** e **Mercadorias** assumindo-se a perspectiva do comprador. As palavras variam individualmente quanto aos padrões de realização dos elementos de frame. Por exemplo, a valência típica do verbo COMPRAR é: O COMPRADOR compra MERCADORIAS do VENDEDOR por DINHEIRO.

Frame Elements

Core

- Comprador** O **Comprador** quer **Mercadorias** e oferece **Dinheiro** a um **Vendedor** em troca.
- Mercadorias** As **Mercadorias** referem-se a qualquer coisa (incluindo tempo de trabalho, por exemplo), que é trocada por **Dinheiro** em uma transação.

Non-Core

- Dinheiro** O **Dinheiro** refere-se ao que é dado em troca por **Mercadorias** em uma transação.
- Finalidade** A finalidade pela qual um ato intencional é realizado.
- Lugar** Onde o evento ocorre.
- Maneira** Qualquer descrição de um evento de comprar que não é dada por EFs mais específicos, incluindo efeitos secundários (silenciosamente, em voz alta), e descrições genéricas comparando eventos (da mesma forma). Isso também pode indicar características proeminentes do **Comprador** que afeta a ação (presunçosamente, friamente, deliberadamente, ansiosamente, cuidadosamente).
- Meio** O **Meio** pelo qual uma transação comercial ocorre.
- Taxa** Em alguns casos, o preço ou pagamento são descritos por unidades de **Mercadorias**.
- Tempo** Quando o evento ocorre.
- Unidade** Indica qualquer unidade na qual mercadorias ou serviços podem ser mensurados.
- Vendedor** O **vendedor** tem posse sobre as **Mercadorias** e as troca pelo **Dinheiro** de um **Comprador**.
- Explicação** A **Explicação** pela qual o evento ocorre.
- Finalidade_imposta** A finalidade do **Comprador** para as **Mercadorias**.
- Período de iteração** O período de tempo de quando o evento comercial começa a se repetir até quando é interrompido.

Fonte: captura de tela feita pelas autoras (FrameNet Brasil 4.0)⁹.

Durante o processo de anotação linguística, há sentenças em que Elementos de *Frame* nucleares não aparecem instanciados. Esses casos são chamados de Instanciações Nulas e estão categorizados em três tipos: (i) Instanciação Nula Definida (DNI, do inglês, *Definite Null Instantiation*), (ii) Instanciação Nula Indefinida (INI, do inglês, *Indefinite Null Instantiation*) e (iii) Instanciação Nula Construcional (CNI, do inglês, *Constructional Null Instantiation*).

No caso da DNI, a omissão do FE pode ser recuperada pelo contexto. Isso ocorre na sentença (3) em que não está explícito o FE ALVO, a expressão que diz onde o TEMA termina, porém pelo contexto compreende-se que se trata da casa de Léo.

(3) No portão da casa de Léo, os três **entram** ^{CHEGAR}.

A INI ocorre quando o FE não é mencionado na sentença, dessa forma, por mais que possa ser inferido, não há material linguístico com a informação. Por exemplo, na

⁹ Disponível em: <https://webtool.frame.net.br>. Acesso em: 05 de set. de 2024.

sentença (4) na anotação da LU *adolescente.n*, tem-se evocado o *frame* *Pessoa_por_Idade*, o FE nuclear *IDADE* não é expresso, por isso ele é classificado como *INI*.

(4) A garota olha o **adolescente**^{PESSOA_POR_IDADE}.

Já no caso da *CNI*, a construção gramatical da sentença causa a ausência do FE, como em sentenças do português brasileiro com sujeito desinencial. Um exemplo desse tipo de instanciação está apresentada no exemplo (5) em que a LU *acariciar.v* evoca o *frame* *Manipulação*, que tem três FE nucleares: *AGENTE*, *ENTIDADE* e *PARTE_DO_CORPO_DO_AGENTE*. Nesse exemplo, o FE *AGENTE* não está explícito, pois o sujeito é desinencial.

(5) **Acarícia**^{MANIPULAÇÃO}, pega e cheira o moletom.

Além dos casos de instanciações nulas, há a *Incorporação* (*INC*, do inglês, *Incorporation*), em que o elemento de *frame* incorpora a LU anotada conforme é visto na sentença (6) em que, na LU *máquina.n*, o FE *ELETROELETRÔNICO* se recupera na morfologia da própria LU.

(6) A **máquina**^{ELETROELETRÔNICO} braille.

Na metodologia adotada na anotação multimodal, assume-se como parâmetro a obra audiovisual na íntegra. Desse modo, quando há um caso de *INI* cuja informação não apresentada seja compreendida pelo modo visual a instanciação deixa de ser indefinida e pode passar a ser definida. No exemplo em (7), o FE *FONTE* não é identificado textualmente, porém se avaliado junto das imagens, percebe-se que eles estão na sala de aula. Desse modo, o FE *FONTE* pode ser assumido como *DNI* numa anotação de viés multimodal.

(7) Rapidamente, Gabriel pega sua mochila, o moletom e os três **saem** MOVIMENTO.

Neste trabalho, as anotações de imagem dinâmica tiveram como objetivo encontrar, nas cenas do curta-metragem, aquilo que se referia aos *frames* evocados nas sentenças transcritas. Na Figura 4, aparece a tela de anotação, composta, no lado esquerdo, pelo vídeo do curta-metragem no qual é possível fazer o rastreamento do que se refere aos *frames* evocados nas sentenças, encontrando, desse modo, os *frames* nas imagens do vídeo.

Figura 4 – Anotação semântica do curta “Eu não quero voltar sozinho”.

The screenshot displays the FNBR Webtool 4.0 interface. On the left, a video player shows a scene with a young man. The video is paused at 643 [25.697s]. The current object is #none. The control panel on the right includes a search bar for frames (FE) and lemmas (LU), a 'Save' button, and a list of objects and sentences. The objects list includes 'Pessoas.pessoa.n', 'Partes_do_corpo.cabelo.n', and 'Pessoas.pessoa.n'. The sentences list includes '#13 589 / #316x', '#14 650 / #105x', '#15 650 / #105x', '#16 650 / #105x', '#17 650 / #105x', and '#18 650 / #105x'. The video player shows the video is paused at 643 [25.697s] and the current object is #none.

Fonte: captura de tela feita pelas autoras (FrameNet Brasil 4.0)¹⁰.

Em seguida, insere-se o *frame* e o Elemento de *Frame*. A LU também é inserida na anotação de imagem, ela corresponde ao objeto detectado por um algoritmo de visão computacional. Se, por exemplo, a entidade é uma casa serão sugeridos os *frames* com o lema *casa.n*, e o anotador sinaliza o *frame* adequado.

¹⁰ Disponível em: <https://webtool.frame.net.br>. Acesso em: 05 de set. de 2024.

7 Diretrizes para elaboração de audiodescrição

As questões técnicas do Guia para Produções Audiovisuais Acessíveis (Naves *et al.*, 2016) abordam pontos estratégicos de como a ADF deve ser realizada. Explica-se, no documento, que é necessário que seja fluida para colaborar com o imaginário do espectador e que seja inserida entre os diálogos evitando sobreposições. Se houver duas ou mais falas e sons ao mesmo tempo, a compreensão dos fatos será prejudicada. Entretanto, se a informação visual for mais importante do que uma fala ou som, neste caso, poderá haver sobreposição.

Já as questões linguísticas orientam a forma em que a linguagem deve ser usada para aproximar o espectador da história, dos personagens e do ambiente. Essas orientações apontam o tempo e o modo verbal, a importância de utilizar adjetivos, advérbios e verbos para descrever as ações, caracterizar os personagens e os ambientes, além de instruir o uso de uma linguagem, considerada, “simples” e “imaginativa”.

As questões tradutórias, por sua vez, abrangem orientações sobre a descrição dos atributos físicos dos personagens, das reações emocionais, como narrar cenas de romance, a importância de mencionar os nomes dos personagens, citar o ambiente da cena e localizá-lo no decorrer da trama. Além disso, é orientado situar o espectador sobre quem está na cena, e, se houver algum som fora do diálogo como barulho de trânsito ou latido de cachorro, é importante esclarecer para o ouvinte o contexto, caso seja necessário.

Além das diretrizes sobre personagens, ações, ambientes e emoções, são abordados os critérios de plano de enquadramento para orientar o audiodescritor a definir os pontos mais importantes a serem narrados. Para analisar esses critérios, foi utilizada a investigação dos planos, e não das cenas. Os planos consistem nos momentos de cada cena, ou seja, a segunda cena, em que está sendo mostrado apenas o adolescente, é o primeiro plano, quando passa a mostrar os outros alunos é o

segundo plano, e, ao mostrar a professora, é o terceiro plano. Assim, a segunda cena tem 29 planos. No total, nesse curta-metragem, há 81 planos e 22 cenas.

Dessa forma, as diretrizes direcionam pontos importantes de serem incluídos no texto da ADF como adjetivos, local, características dos personagens etc. Assim, cabe ao audiodescritor escolher quais adjetivos irá colocar no texto da ADF, assim como, quais características dos personagens, como descrever o local, entre outros. Neste sentido, o produto de uma AD assume a perspectiva do audiodescritor sobre a obra audiovisual por mais que os critérios a serem seguidos orientem o especialista a não inserir uma valoração subjetiva acerca do objeto narrado.

8 Levantamento das anotações de texto e imagem dinâmica

O levantamento feito com os dados anotados na plataforma Webtool trouxe várias informações relevantes para a anotação semântica multimodal. A anotação de texto da transcrição da ADF somou 489 *frames* identificados e analisados semanticamente; desses, 320 estão classificados nas categorias raízes, conforme demonstrado na Tabela 1.

Tabela 1 – Categorias raízes e levantamento dos *frames* do texto da ADF.

Categoria Raiz	Quantidade de <i>Frames</i> Anotados
Atributo	23
Entidade	98
Estado	71
Evento	116
Relação	12
Total	320

Fonte: elaborado pelas autoras.

Já a anotação de imagem dinâmica somou 299 *frames*, sendo que 189 estão classificados nas categorias de *frames* mencionadas, como especificado na Tabela 2.

Tabela 2 – Categorias raízes e levantamento dos *frames* das imagens dinâmicas.

Categoria Raiz	Quantidade De <i>Frames</i> Anotados
Atributo	7
Entidade	85
Estado	4
Evento	92
Relação	1
Total	189

Fonte: elaborado pelas autoras.

A Tabela 3 apresenta exemplos de *frames* recorrentes na anotação do texto transcrito da ADF e na anotação de vídeo a partir das categorias raízes ou os *frames* de topo como também são chamados.

Tabela 3 – Categorias raízes e *frames* recorrentes na anotação multimodal.

Categoria Raiz	Exemplos de <i>Frames</i>
Atributo	Atributos_mensuráveis, Cor
Entidade	Pessoas_por_Idade, Locais_por_uso, Texto, Móveis
Estado	Contato_espacial, Relação_locativa
Evento	Causar_movimento, Percepção_ativa, Movimento, Manipulação
Relação	Acompanhamento, Partitivo

Fonte: elaborado pelas autoras.

Já na Tabela 4, há os *frames* mais evocados na anotação de texto e, na Tabela 5, as cinco Unidades Lexicais mais evocadas no texto da ADF, como *adolescente.n* e *olhar.v*.

Tabela 4 – Os dez *frames* mais evocados na anotação de texto.

<i>Frames</i> mais evocados na anotação de texto	
Partes_Do_Corpo	30
Movimento	20
Percepção_Ativa	18
Peça_Arquitetônica	16
Movimento_Corporal	14

Finalidade	13
Manipulação	12
Expressão_Facial	11
Parte_Orientacional	11
Pessoas	11
Pessoas_Por_Idade	11

Fonte: elaborado pelas autoras.

Tabela 5 – Unidades Lexicais mais frequentes na anotação de texto.

ULs mais evocadas		
5 ULs mais evocadas	<i>Frame</i>	Ocorrências
<i>adolescente.n</i>	Pessoa_por_idade	11
<i>olhar.v</i>	Percepção_Ativa	10
<i>sentar.v</i>	Movimento_corporal	10
<i>em,</i> <i>no(em+o).prep</i>	Contato_espacial	19
<i>para.prep</i>	Alvo	13

Fonte: elaborado pelas autoras.

A respeito dos Elementos de *Frame* na anotação de vídeo, eles fornecem informações importantes na efetivação da anotação multimodal. A sua marcação sinaliza qual FE se enquadra entre o *frame* correspondente à LU do texto e a entidade identificada na imagem dinâmica.

Tabela 6 – Cinco FEs mais frequentes na anotação de vídeo.

Elementos de <i>Frame</i> recorrentes na anotação de vídeo	
AGENTE	61
POSSUIDOR	39
PESSOA	26
TEMA	25
PERCEPTOR_AGENTIVO	18

Fonte: elaborado pelas autoras.

A Tabela 6 faz o registro dos FEs mais anotados no vídeo. A identificação de um ser senciente (como AGENTE e POSSUIDOR) e de outras possibilidades de entidade (como TEMA) indica a prioridade desses elementos na estruturação de uma ADF.

9 Análise das diretrizes da audiodescrição

Para esta pesquisa, foi feita a análise das questões linguísticas e tradutórias, pois são as orientações que colaboram com o estudo de *frames* na construção do sentido em ambientes multimodais. Para tanto, as cenas 1 e 2 estão apresentadas sequencialmente expondo o acontecimento narrativo e os *frames* evocados pelo texto da audiodescrição e pelas imagens dinâmicas, e quais diretrizes são utilizadas na cena.

Cena 1: Créditos iniciais

Frames evocados pelo texto da ADF, conforme o Quadro 1:

Quadro 1 - *Frames* evocados pelo texto da Cena 1.

Frames evocados pelo texto da Cena 1
Acompanhamento
Contato espacial
Cor
Criar intencionalmente
Desejar
Intérpretes e papéis
Ler
Movimento
Negação
Parte orientacional

Fonte: elaborado pelas autoras.

Frames evocados pela imagem dinâmica conforme demonstrado no Quadro 2:

Quadro 2 - *Frames* evocados pela imagem dinâmica da Cena 1.

Frames evocados pela imagem dinâmica da Cena 1
Cor
Texto

Fonte: elaborado pelas autoras.

Nesta primeira cena, temos as seguintes narrações de AD:

- a) No fundo acinzentado, lê-se: produção Lacuna Filmes, co-produção Cine Pro.
- b) Apresentando Guilherme Lobo, Tess Amorim, Fábio Audi.
- c) Eu não quero voltar sozinho.

Diante dessas sentenças, constata-se a seguinte diretriz quanto à questão tradutória:

Recomenda-se que elementos visuais verbais, tais como créditos, textos, títulos, legendas e intertítulos, sejam lidos. Assim, sua leitura deve ser feita em momento que não se sobreponha à audiodescrição de cenas, o que pode acontecer de forma corrida logo no início, ou então deixá-los todos para o final. (Naves *et al.*, 2016, p. 22)

Cena 2: Os alunos estão sentados na sala de aula estudando enquanto a professora está sentada na frente da sala. Ela pede ao novo aluno que se apresente para a turma.

No Quadro 3, há o levantamento dos *frames* evocados pelo texto da ADF quanto a essa cena:

Quadro 3 - *Frames* evocados pelo texto transcrito da Cena 2.

Frames semânticos	
Acompanhamento	Movimento
Agregado	Movimento corporal
Alvo	Mudar direção
Arrumar	Números cardinais
Artefato	Partes do corpo
Atenção	Parte orientacional
Atividade pausar	Parte todo
Atravessar	Partitivo
Atributos mensuráveis	Peça arquitetônica
Causar movimento	Pegar
Cercanias	Percepção ativa
Certeza	Pessoas
Descrição parte do corpo	Pessoas por etnia
Distinção	Pessoas por idade
Eletroeletrônicos	Pessoas por vocação
Emoção direcionada	Proximidade não graduável
Entidade	Relação locativa
Expectativa	Ser localizado
Expressão facial	Sons
Incremento	Subpartes de prédios
Levar tempo	Texto
Local	Texto criação
Manipulação	Usar
Móveis	Vestuário

Fonte: elaborado pelas autoras.

No Quadro 4, há o levantamento dos *frames* evocados pela imagem dinâmica:

Quadro 4 - *Frames* evocados pela imagem dinâmica da Cena 2.

Frames semânticos	
Agregado	Manipulação
Arrumar	Móveis
Artefato	Movimento
Atividade em andamento	Movimento corporal
Atividade iniciar	Mudar direção
Atividade pausar	Partes do corpo
Atributos mensuráveis	Peça arquitetônica
Causar movimento	Percepção ativa
Dar	Pessoas
Eletroeletrônicos	Pessoas por etnia
Entidade	Pessoas por idade
Expectativa	Pessoas por vocação
Expressão facial	Vestuário
Fechamento	

Fonte: elaborado pelas autoras.

Na segunda cena, há as seguintes narrações de AD:

- a) O olhar fixo marca a parte superior do rosto de um adolescente branco, cabelos lisos e curtos.
- b) Outros adolescentes uniformizados em carteiras à sua volta.
- c) Com uma colega à esquerda escrevendo num caderno, ao fundo, o adolescente usa máquina braille.
- d) Um dos adolescentes olha para outro que balança positivamente a cabeça.
- e) A professora sentada diante de sua mesa olha a turma.
- f) A máquina braille.
- g) O adolescente sorri.
- h) A garota olha o adolescente.
- i) A voz de Gabriel chama atenção do adolescente.
- j) Desconcertado, se levanta e vai para a frente da sala.
- k) Gabriel passa por colegas que estão saindo e dirige-se para seu lugar que fica atrás do adolescente.
- l) A turma sai da sala.
- m) O adolescente, a garota e Gabriel ficam arrumando suas coisas.
- n) O adolescente sentado com sua mochila no ombro, espera.
- o) Segura acima do cotovelo da garota e levanta.
- p) Se dirigem à porta, param e a garota vira para Gabriel.
- q) Rapidamente Gabriel pega sua mochila, o moletom e os três saem.

Diante dessas sentenças, as seguintes diretrizes são constatadas para as questões linguísticas e tradutórias:

Questão linguística:

Os adjetivos descritivos são muito importantes na audiodescrição, pois tornam cenas, ações, características dos personagens e ambientes mais claros para o espectador. Os adjetivos devem expressar estados de humor e de emoções condizentes com os construtos universais sem valoração subjetiva por parte do audiodescritor. Também se recomenda que as cores sejam referidas. (Naves *et al.*, 2016, p.15)

Usar verbos específicos que indiquem a maneira de realização das ações. (Naves *et al.*, 2016, p.16)

Questão tradutória:

Não é necessário descrever o figurino de todos os personagens em todas as cenas, pois o excesso de informação torna a audiodescrição cansativa e tira o foco do ponto principal. (Naves *et al.*, 2016, p. 19)

Descrever os elementos que levam o espectador a inferir o estado emocional dos personagens pode funcionar em alguns casos, como “Ela leva as mãos ao rosto e chora” ao invés de “Ela está triste” (Naves *et al.*, 2016, p. 19)

É necessário localizar sempre os ambientes, dizer que o personagem volta a um determinado ambiente em que já esteve; deixar claro caso um mesmo ambiente tenha sofrido mudanças e descrever quais. Além do ambiente, outra informação importante para o entendimento da cena é dizer quantos estão em cena e quem são. (Naves *et al.*, 2016, p. 21)

Como exposto na seção *Diretrizes para elaboração de audiodescrição*, além das diretrizes, são abordados os planos de enquadramento em que há orientações para elaboração do roteiro de ADF, conforme mostrado no Quadro 5.

Quadro 5 – Diretrizes dos planos de enquadramento e pontos de vista.

Diretrizes dos planos de enquadramento e pontos de vista
“Grande plano geral (GPG) enquadra uma grande área de ação, na qual o ambiente é mostrado de maneira ampla e é captado a longa distância, o que apresenta o local onde a história ocorrerá naquele momento e situa os personagens da trama. Geralmente é utilizado no início da história ou quando há mudança de local. Por meio desse plano, o audiodescritor descreverá o ambiente, a fim de situar o espectador com relação ao espaço que é apresentado no filme (ex.: cidade em que a história se passa)”. (p.17)

<p>“O plano geral (PG) possui um ângulo de visão menor do que o GPG. Por meio dele, o local é apresentado de forma mais precisa e é mostrada a posição do personagem em cena. Com esse plano, o audiodescritor poderá descrever locais mais específicos em que os personagens se encontram, como suas casas, locais de trabalho, etc.” (p.17)</p>
<p>“O plano médio (PM) tem uma função descritiva e, para isso, os personagens são enquadrados da cintura para cima, dando destaque para a figura humana. Nesse momento, o audiodescritor pode fazer uma descrição mais precisa sobre as características físicas dos personagens e de suas vestimentas”. (p.17-18)</p>
<p>“O primeiro plano (PP) enquadra o personagem do busto para cima. Seu objetivo é mostrar os diálogos entre os personagens e suas expressões faciais, que podem ser mais bem detalhadas pelo audiodescritor. Seu objetivo é mostrar os diálogos entre os personagens e suas expressões faciais, que podem ser mais bem detalhadas pelo audiodescritor”. (p.18)</p>
<p>“O primeiríssimo plano (PPP) enquadra somente a cabeça dos personagens. É utilizado para ressaltar as expressões dos personagens, a fim de revelar suas emoções. É utilizado para ressaltar as expressões dos personagens, a fim de revelar suas emoções”. (p.18)</p>
<p>“O <i>close-up</i> ou plano detalhe enquadra apenas o que é essencial para a compreensão do que está sendo apresentado, destacando-o do resto da cena”. (p.18)</p>
<p>“O <i>plongée</i> e o <i>contraplongée</i> também são muito significativos, aumentando ou diminuindo o tamanho dos personagens ou objetos, não só física mas também simbolicamente, o que deverá ser enfatizado na audiodescrição, como por exemplo: “Vista de baixo para cima de uma igreja”. (p.18)</p>
<p>“Já os planos-ponto-de-vista mostram diferentes pontos-de-vista, que podem ser o do autor, o do narrador, ou o de um personagem e podem também ser explicitados na audiodescrição, como no exemplo: “Lentamente a imagem se aproxima do braço do homem e foca no relógio de pulso. Detalhe dos ponteiros do relógio em movimento”. (p.18)</p>

Fonte: adaptado, pelas autoras, a partir de Naves *et al.* (2016, p. 17-18).

O Plano Geral (PG) e o Plano Médio (PM) foram os planos de enquadramento mais utilizados neste curta-metragem. Para o PG, as diretrizes orientam a descrição dos ambientes e locais, e, para o PM, as descrições de características físicas e expressões faciais. No curta, conforme mostrado no Quadro 6, foram evocados os seguintes *frames* semânticos:

Quadro 6 - *Frames* que expressam o Plano Geral e o Plano Médio no curta.

Frames semânticos
Locais por uso
Local
Pessoas
Pessoas por etnia
Pessoas por idade
Pessoas por vocação
Vestuário
Ser localizado
Relação locativa
Descrição parte do corpo
Expressão facial
Partes do corpo
Peça arquitetônica
Subpartes de prédios

Fonte: elaborada pelas autoras.

O Primeiro Plano (PP), o Primeiríssimo Plano (PPP) e o Plano Detalhe indicam menos planos do que o PG e PM. O PP aborda as expressões faciais e o PPP as emoções, os *frames* com esse enfoque são especificados no Quadro 7. Apesar de terem um plano para PP e um para PPP não significa que apenas esses apresentaram audiodescrições de expressão facial e emoção. Estas descrições foram encontradas em outros planos que seguiam as outras diretrizes, PG, PM e Plano Detalhe.

Quadro 7 - *Frames* que denotam emoções e expressões faciais.

Frames semânticos
Desejar
Emoção com foco no experienciador
Emoção direcionada
Expressão facial

Fonte: elaborada pelas autoras.

No critério do Plano Detalhe, não é feita uma orientação direta ao audiodescritor como ocorre no PG, PM, PP e PPP. A sentença “A máquina braille”, apresentada anteriormente, no exemplo em (8), expressa o Plano Detalhe. Ao destacar apenas o essencial, faz-se uma narração sobre ele.

10 Considerações finais

Este estudo identificou os *frames* semânticos evocados na transcrição da audiodescrição e nas imagens do curta-metragem "Eu não quero voltar sozinho" em alinhamento às diretrizes do Guia para Produções Audiovisuais Acessíveis (2016). As análises confirmaram que a ADF em questão segue essas diretrizes, contribuindo para experiência fílmica de espectadores com deficiência visual.

As anotações de texto e de imagem dinâmica evidenciaram uma preferência por determinadas categorias de *frames*, refletindo as diferenças na natureza das anotações (no caso, texto verbal e vídeo). Em particular, na anotação de imagem dinâmica, a relevância dos Elementos de *Frame* foi destacada pela correlação direta dos FEs marcados no vídeo e os FEs e LUs na anotação textual.

Este trabalho contribui para os Estudos da Tradução Audiovisual Acessível ao explorar a articulação e complementação das modalidades audiovisuais com base na Semântica de *Frames* e na FrameNet Brasil. Além disso, os resultados expandem o *dataset* multimodal da Reinventa, que integra linguística computacional e Semântica de *Frames* no processamento de objetos multimodais. A geração de anotações semânticas refinadas de texto e imagem fortalece o desenvolvimento de algoritmos para a rotulação semântica e a análise de sentido em materiais multimodais.

Essas contribuições não são apenas teóricas, mas tem aplicações práticas significativas. Ferramentas de inteligência artificial baseadas nesses avanços podem transformar a Tradução Audiovisual Acessível, promovendo uma inclusão social efetiva, que é um direito de todos os cidadãos. Assim, este trabalho oportuniza novas pesquisas que integrem os estudos cognitivos da linguagem numa abordagem social e multimodal.

Agradecimentos

As autoras agradecem as contribuições dos avaliadores anônimos que realizaram uma revisão detalhada deste artigo. Agradecemos também à Coordenação de Pessoal de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio de

pesquisa (88887.683333/2022-00, 88887.936139/2024-00). Adriana Pagano é bolsista de produtividade do CNPq, processo nº 313103/2021-6.

Referências

CABITZA, F.; CAMPAGNER, A.; BASILE, V. Toward a Perspectivist Turn in Ground Truthing for Predictive Computing. *In: Proceedings of the AAAI Conference on Artificial Intelligence*, 37(6), 6860-6868, 2023. DOI <https://doi.org/10.1609/aaai.v37i6.25840>

BATEMAN, J.; WILDFEUER, J.; HIIPPALA, T. **Multimodality**: Foundations, research and analysis—A problem-oriented introduction. Walter de Gruyter GmbH & Co KG. 2017. DOI <https://doi.org/10.1515/9783110479898>

BELCAVELLO, F.; VIRIDIANO, M.; MATOS, E.; TORRENT, T. T. Charon: A FrameNet Annotation Tool for Multimodal Corpora. *In: Proceedings of the 16th Linguistic Annotation Workshop (LAW-XVI) within LREC2022*. Marseille: ELRA, 2022. p. 91-96.

BELCAVELLO, F. **FrameNet annotation for multimodal corpora**: devising a methodology for the semantic representation of text-image interactions in audiovisual productions. 135 f. Tese (Doutorado). Universidade Federal de Juiz de Fora, Faculdade de Letras. Programa de Pós-graduação em Linguística, Juiz de Fora. 2023.

CZULO, O. Frame Semantics. *In: ENTI (Enciclopédia de Tradução e Interpretação)*. Granada. Associação Ibérica dos Estudos de Tradução e Interpretação. 2024. Disponível em https://www.aieti.eu/enti/frame_semantics_ENG/.

DE SOUZA, D. S.; PAGANO, A. S.; GAMONAL, M. A. A audiodescrição sob a perspectiva da Semântica de Frames: um estudo exploratório. *Revista Gatilho*, 23, 2022. DOI <https://doi.org/10.34019/1808-9461.2022.v23.38564>

EU NÃO QUERO VOLTAR SOZINHO. Direção: Daniel Ribeiro. Produção: Lacuna Filmes. Brasil: [s. n.], 2010. Disponível em: <https://www.youtube.com/watch?v=FkNoXubidmk>.

FILLMORE, C. Frame semantics. *In: Linguistics in the Morning Calm*. Seul: Hanshin Publishing Co., 1982. p.111-137.

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

FILLMORE, C. J.; BAKER, C. A frames approach to semantic analysis. *In: HEINE, B.; NARROG, H. (ed.). The Oxford Handbook of Linguistic Analysis*. 1. ed. Oxford: Oxford University Press, 2010. p. 313-340. DOI <https://doi.org/10.1093/oxfordhb/9780199544004.013.0013>

FRAMENET BRASIL: **Webtool** 4.0, 2024. Plataforma de anotação. Disponível em: <https://webtool.frame.net.br>. Acesso em: 02, set. de 2024.

GAMONAL, M. **Copa 2014 FrameNet Brasil**: diretrizes para a constituição de um dicionário eletrônico trilingue a partir da análise de frames da experiência turística. 2013. 145 f. Dissertação (Mestrado em Linguística). Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

HIIPPALA, T. Genre analysis. **Multimodal genre analysis**, University of Helsinki, p. 2-14, 2014. DOI <https://doi.org/10.1515/9781614511175.111>

HOLMES, J. S. The Name and Nature of Translation Studies. *In: VENUTI, L. The Translation Studies Reader*. Londres: Routledge, 2000. p. 172-185.

MAYER, F. **Imagem como símbolo acústico**: a semiótica aplicada à prática da audiodescrição. 2012. 147 f. Dissertação (Mestrado em Comunicação) - Pontifícia Universidade Católica de Minas Gerais, [S. l.], 2012.

MUNDAY, J. **Main issues of translation studies**. *In: MUNDAY, J. (ed.) Introducing Translation Studies: Theories and Applications*. 4. ed. Londres: Routledge, 2016. p. 7-27. DOI <https://doi.org/10.4324/9781315691862>

NAVES, S. B.; M., C.; ALVES, S. F.; ARAÚJO, V. L. S. (org.). **Guia para Produções Audiovisuais Acessíveis**. Brasília: Secretaria do Audiovisual do Ministério da Cultura, 2016.

PAGANO, A. S.; TEIXEIRA, A. L. R.; MAYER, F. A. Accessible Audiovisual Translation. *In: The Oxford Handbook of Translation and Social Practices*. Oxford University Press, 2020, p. 66-82. DOI <https://doi.org/10.1093/oxfordhb/9780190067205.013.4>

RUPPENHOFER, J.; ELLSWORTH, M.; PETRUCK, M. R. L.; JOHNSON, C. R.; SCHEFFCZYK, J. **FrameNet II**: Extended theory and practice. Berkeley: International Computer Science Institute, 2016.

SANTOS, Z. B.; PIMENTA, S. M. O. Da semiótica social à multimodalidade: a orquestração de significados. **CASA: Cadernos de Semiótica Aplicada**, [S. l.], p. 295 - 324, 2014. DOI <https://doi.org/10.21709/casa.v12i2.7243>

TORRENT, T. T.; LORENZI, A.; MATOS, E. E. S.; BELCAVELLO, F.; VIRIDIANO, M.; GAMONAL, M. A. Lutma: a Frame-Making Tool for Collaborative FrameNet Development. *In: Proceedings of the 1st Workshop on Perspectivist Approaches to NLP*, Marselha, p. 100-107, 2022.

TORRENT, T. T.; ELLSWORTH, M. Behind the Labels: Criteria for Defining Analytical Categories in FrameNet Brasil. **Veredas: Frame Semantics and Its Technological Applications**, Juiz de Fora, v. 14, n. 1, p. 44-65, 2013.

VIRIDIANO, M.; ALMEIDA, A. L.; TORRENT, T. T.; MATOS, E. E.; PAGANO, A. S.; SIGILIANO, N. S.; GAMONAL, M. A.; ABREU, H. A.; DUTRA, L. V.; SAMAGAIO, M. M.; PINTO, M. C.; CAMPOS, F.; BRAZ, G. A.; MAZZEI, B.; OLIVEIRA, M. F.; LUZ, A. C. L.; RUIZ, L. P.; XAVIER, J. B.; PESTANA, A.; COSTA, J.; RABELO, I.; SILVA, A. B.; ROZA, R.; MOTA, M. S.; OLIVEIRA, I.; FREITAS, M. H. P. 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: ELRA/ICCL, p. 7438 – 7449, 2024.