

Permanência semântica entre áudio original e legenda: um estudo sobre anotação semântica multimodal em obra audiovisual

*Semantic Permanence between Original Audio and Subtitles:
A Study on Multimodal Semantic Annotation in
Audiovisual Work*

Mariana Mota Silva Souza

Universidade Federal de Minas Gerais
(UFMG) | Belo Horizonte | MG | BR
mariannassouza.mota@gmail.com
<https://orcid.org/0009-0005-0171-6312>

Maucha Andrade Gamonal

Universidade Federal de Juiz de Fora
(UFJF) | Juiz de Fora | MG | BR
CAPES
mauchaandrade@gmail.com
<https://orcid.org/0000-0002-5810-3913>

Adriana Silvina Pagano

Universidade Federal de Minas Gerais
(UFMG) | Belo Horizonte | MG | BR
CNPq
FAPEMIG
apagano@ufmg.br
<https://orcid.org/0000-0002-3150-3503>

Resumo: Com base na Semântica de *Frames* (Fillmore, 1982) e no Modelo da Primazia do Frame (Czulo, 2013), este estudo realiza uma análise multimodal das sentenças transcritas do áudio original em português, das legendas em inglês e das imagens correspondentes aos seis primeiros minutos do curta-metragem brasileiro *Eu não quero voltar sozinho*, de Daniel Ribeiro. A pesquisa objetiva investigar a similaridade semântica entre os *frames* evocados nas sentenças equivalentes, avaliar as escolhas tradutórias e analisar as relações semânticas estabelecidas entre texto e imagem. Adota a metodologia da FrameNet Brasil para anotações semânticas multimodais na plataforma Webtool, e a similaridade semântica entre *frames* é calculada por meio do algoritmo de *spread activation* e da métrica de similaridade de cossenos soft. As medidas organizam as sentenças em similaridade nula, parcial ou total, conforme a proposta de Samagaio (2023). As estratégias tradutórias de legendagem são analisadas conforme Díaz-Cintas e Remael (2021) e avaliadas quanto às relações semânticas entre texto e imagem. Os resultados evidenciam que as estratégias de tradução empregadas alteram a evocação de *frames* nas línguas analisadas, impactando a similaridade semântica entre sentenças. Este estudo contribui para o campo da tradução audiovisual e da



Semântica de *Frames*, ao explorar a construção de sentido em objetos multimodais e demonstrar como estratégias tradutórias influenciam na relação entre texto, imagem e *frames* evocados.

Palavras-chave: legendagem; anotação semântica multimodal; Semântica de *Frames*.

Abstract: Based on Frame Semantics (Fillmore, 1982) and the Frame Primacy Model (Czulo, 2013), this study performs a multimodal analysis of the sentences transcribed from the Portuguese original audio, the English subtitles and the images corresponding to the first six minutes of the Brazilian short film *Eu não quero voltar sozinho* (“I Don’t Want to Go Home by Myself”), by Daniel Ribeiro. This research aims to investigate the semantic similarity between the frames evoked in their equivalent sentences, evaluate the translation choices and analyze the semantic relations established between text and image. It adopts the FrameNet Brasil methodology for multimodal semantic annotations on the Webtool platform, and the semantic similarity between frames is calculated using the spread activation algorithm and the soft cosine similarity metric. The measures organize the sentences in null, partial or total similarity, as proposed by Samagaio (2023). The subtitling translation strategies are analyzed according to Díaz-Cintas and Remael (2021) and evaluated in terms of the semantic relations between text and image. Our results show that the translation strategies employed alter the evocation of frames in the languages analyzed, impacting the semantic similarity between sentences. This study contributes to the field of audiovisual translation and to Frame Semantics, by exploring the construction of meaning in multimodal objects and demonstrating how translation strategies influence the relationship between text, image and evoked frames.

Keywords: Subtitle; multimodal semantic annotation; Frame Semantics.

1 Introdução

A Semântica de *Frames*, desenvolvida por Fillmore (1982), é uma teoria linguística inserida nos estudos cognitivos da linguagem que postula que o significado das palavras é estruturado por meio de cenas, denominadas frames, evidenciando as relações entre experiência e linguagem. Essa abordagem tem sido utilizada nos Estudos da Tradução, em que Rojo e Ibarretxe-Antuñano (2013) afirmam que, assim como a linguagem, a tradução possui uma natureza cognitiva. Nesse contexto, Czulo (2017) propôs o Modelo da Primazia do Frame, segundo o qual o tradutor busca maximizar a similaridade semântica entre o texto fonte e o texto alvo.

Este artigo apresenta um estudo multimodal da tradução audiovisual, fundamentado na Semântica de Frames e no Modelo da Primazia do Frame, a partir da análise de um trecho do curta-metragem brasileiro *Eu não quero voltar sozinho*, produzido pela Lacuna Filmes em 2010 e dirigido por Daniel Ribeiro. A obra acompanha Leonardo, um adolescente com deficiência visual, em suas descobertas sobre sexualidade e amor ao conhecer Gabriel, um novo aluno na escola.

A pesquisa justifica-se pela escassez de estudos em português brasileiro que relacionam legendagem à Semântica de Frames. Investigações anteriores concentram-se na tradução de elementos culturais (Rojo, 2002; Bertoldi, 2016), audiodescrição filmica (Souza *et al.*, 2022; Dornelas *et al.*, 2024) e criação de recursos tradutórios (Boas, 2013). Em relação à legendagem, destacam-se os trabalhos de Freitas (2022), que anotou o áudio original e a legenda de um corpus sem incluir uma análise multimodal, e de Samagaio (2023), que analisou a similaridade semântica em textos transcritos de áudio e legenda de um episódio da série *Pedro pelo mundo*, mas também sem incorporar o aspecto multimodal.

A presente pesquisa expande esses estudos ao incluir a análise multimodal, comparando frames evocados no texto transscrito do áudio original, na legenda em inglês e no vídeo. Os objetivos gerais do trabalho incluem contribuir para os Estudos da Tradução, para as pesquisas em Semântica de Frames e para a análise semântica de objetos multimodais, especialmente no campo da tradução audiovisual.

Os objetivos específicos da pesquisa são: (i) verificar os frames evocados na anotação de texto corrido do áudio original transscrito e da legenda, bem como na anotação de imagens dinâmicas do excerto do curta-metragem; (ii) avaliar a permanência semântica na legenda em inglês ao comparar os frames do áudio original com os da legenda; (iii) determinar como estratégias de tradução de legendagem influenciam a métrica de similaridade semântica entre o texto original e a legenda; e (iv) analisar as relações semânticas entre texto e imagem.

A metodologia consiste na anotação semântica da FrameNet Brasil voltada à multimodalidade. Para isso, foi utilizado o software Webtool. Além disso, aplicaram-se o algoritmo de spread activation e o cálculo de similaridade de cossenos soft para aferir a similaridade semântica entre os frames evocados nas sentenças equivalentes do áudio original e da legenda.

Este trabalho contribui para a ampliação do diálogo entre Semântica de Frames e Estudos da Tradução, investigando a interação entre texto e imagem em legendagem com base em uma abordagem multimodal. Ao considerar a interseção entre essas áreas, espera-se fornecer subsídios teóricos e metodológicos para futuras pesquisas que explorem a tradução audiovisual sob a perspectiva da linguística cognitiva.

2 Abordagem multimodal nos Estudos da Tradução e na Semântica de *Frames* e FrameNet Brasil

2.1 Estudos da Tradução e Legendagem

Os Estudos da Tradução, conforme Holmes (1972), podem ser categorizados em estudos puros e aplicados. No âmbito dos estudos puros, destacam-se os teóricos e os descritivos. Os estudos descritivos podem ser orientados ao produto, ao processo ou à função. O presente trabalho se insere no campo dos Estudos da Tradução como um estudo descritivo orientado ao produto, pois o objetivo é analisar o produto final de uma tradução audiovisual: a legenda em inglês do curta-metragem. Além disso, o estudo contribui para o campo dos estudos aplicados ao auxiliar na construção de um *corpus* anotado que pode ser empregado em tarefas de processamento de linguagem natural, como a tradução automática.

No que diz respeito aos tipos de tradução, Pagano *et al.* (2020) apontam que eles podem ser divididos em tradução no sentido restrito e interpretação no sentido amplo. A tradução no sentido restrito abrange tanto a tradução escrita quanto a tradução audiovisual. Esta última subdivide-se em linguagem visual-gestual e linguagem vocal-auditiva. A linguagem vocal-auditiva, por sua vez, inclui modalidades como a revocalização e a legendagem, sendo a última o foco de estudo desta pesquisa.

A legendagem envolve diversos aspectos que influenciam o processo e o produto da tradução, como o tempo de exibição do texto na tela, a sincronia com a fala e as restrições de caracteres, além das adaptações culturais necessárias. Díaz-Cintas e Remael (2007) definem a legendagem como:

uma prática tradutória que consiste na apresentação de um texto escrito, geralmente na parte inferior da tela, que busca narrar o diálogo original dos locutores, bem como os elementos discursivos que aparecem na imagem (cartas, encartes, grafite, inscrições, cartazes e similares) e as informações contidas na trilha sonora (músicas, *voices off*) (Díaz-Cintas; Remael, 2007, p. 8).

A legendagem, portanto, não se limita a uma simples transposição de uma língua para outra. Como destacam Díaz-Cintas e Remael (2021), trata-se de um processo de tradução da linguagem oral para a linguagem escrita. Isso demanda reformulações e interpretações do discurso para adaptá-lo ao meio escrito, o que frequentemente resulta na perda de certas características da oralidade, por exemplo. Os limites temporais e espaciais impostos pelas legendas originam uma linguagem marcada pela simplificação. Díaz-Cintas e Remael (2021) observam que:

A gramática, a sintaxe e os itens lexicais tendem a ser simplificados e apagados, enquanto os aspectos interacionais e a entonação são mantidas apenas até certo ponto (por exemplo, através da ordem das palavras, perguntas retóricas, interjeições ocasionais e frases incompletas) (Díaz-Cintas; Remael, 2021, p. 89).

Essas restrições estruturais refletem-se nas convenções específicas seguidas pelos tradutores, e o resultado tem impacto na experiência do espectador. De acordo com as conven-

ções, para evitar que as legendas desviem o foco da narrativa audiovisual, elas devem ocupar até duas linhas de texto e ocupar no máximo um sexto da tela. Além disso, as legendas são localizadas no centro da tela, na parte inferior, podendo ser deslocadas para não se sobrepor a outros textos na tela. Devem ser sincronizadas com o áudio e ter duração de cinco a seis segundos. Para facilitar a leitura, as fontes não devem possuir serifas, como Arial ou Helvetica, e devem estar em cores como branco ou amarelo.

As reduções mencionadas inserem a necessidade de estratégias de redução textual. Segundo Díaz-Cintas e Remael (2021), a reprodução fiel das falas na legenda não é necessária, pois ela é interpretada em conjunto com as imagens e os sons do produto audiovisual. Além disso, como eles destacam, as legendas devem respeitar o tempo que o espectador leva para absorver o texto e permitir que experiencie também as imagens e a trilha sonora.

Para os autores, a redução pode ser parcial, condensando o texto, ou total, eliminando elementos textuais. No nível da palavra, Díaz-Cintas e Remael (2021) destacam a simplificação de perifrases verbais, generalização de enumerações, uso de sinônimos ou expressões equivalentes. Já no nível da sentença, observam-se várias mudanças, como no tipo de oração (se negativa ou afirmativa, interrogativa ou imperativa etc.), alteração de voz ativa para passiva, fusão de orações, uso de dêiticos no lugar de grupos nominais e junção de duas ou mais sentenças.

Omissões também desempenham um papel relevante na legendagem. No nível da palavra, itens como modificadores (adjetivos e advérbios), interjeições e vocativos podem ser omitidos sem comprometer a mensagem principal. No nível da sentença, exclusão de orações inteiras podem ocorrer em situações como sobreposição de diálogos ou quando as falas apresentam baixa contribuição para a narrativa. Segundo Díaz-Cintas e Remael (2021), o tradutor deve priorizar a integridade da mensagem e a experiência do espectador ao optar por tais omissões.

Com base nas estratégias listadas, este estudo reforça a importância de compreender a legendagem como um processo criativo e ao mesmo tempo técnico, que demanda decisões criteriosas capazes de conciliar a fidelidade ao conteúdo a partir das limitações do formato.

2.2 Semântica de *Frames*

A Semântica de *Frames*, teoria linguística proposta por Charles Fillmore em 1982 é descrita pelo próprio autor como “um programa de pesquisa sobre semântica empírica e um modelo descritivo para apresentar os resultados de tal pesquisa” (Fillmore, 2009, p. 25). Em termos práticos, a Semântica de *Frames* constitui tanto uma base estruturada para investigação sobre significação linguística, baseada em corpora, quanto um modelo para a apresentação e descrição sistemática dos resultados obtidos.

O objetivo central de Fillmore com a Semântica de *Frames* era compreender as relações entre a linguagem e a experiência, investigando como as sociedades categorizam o mundo e como essas categorias são codificadas linguisticamente. Para isso, Fillmore (2009) introduziu o conceito de *frame*, definido como:

qualquer sistema de conceitos relacionados de tal modo que, para entender qualquer um deles, é preciso entender toda a estrutura na qual se enquadram;

quando um dos elementos dessa estrutura é introduzido em um texto, ou em uma conversa, todos os outros elementos serão disponibilizados automaticamente (Fillmore, 2009, p. 25).

Os *frames*, portanto, são representações esquemáticas de situações que englobam elementos interconectados. Sua compreensão está intrinsecamente ligada ao entendimento de instituições sociais, experiências humanas e aspectos culturais. Assim, segundo a teoria, os significados linguísticos são inseparáveis de seu contexto, representado por *frames* semânticos.

Esses *frames* podem ser evocados pelo texto por meio de unidades lexicais (LU, do inglês *Lexical Unit*), que combinam uma forma linguística (lema) e um significado, ou podem ser invocados, por inferência, pelo intérprete a partir de pistas contextuais, mesmo na ausência de material linguístico.

A teoria de Fillmore está inserida no campo de estudos da linguística cognitiva, uma abordagem que integra a língua às demais faculdades cognitivas humanas, como a memória e o raciocínio (Rojo; Ibarretxe-Antuñano, 2013). Essa teoria refuta o gerativismo chomskyano, predominante na década de 1970, que se concentrava nos padrões linguísticos como internos à estrutura da língua e ignorava o papel da semântica e do contexto na construção de significados.

A Semântica de *Frames* apresenta uma aplicação particularmente relevante no campo dos Estudos da Tradução. Como explicam Rojo e Ibarretxe-Antuñano (2013):

A Linguística Cognitiva defende a natureza cognitiva da tradução como um processo de mediação entre dois mundos conceituais diferentes. Além disso, a sua visão integrada da língua e da cognição, juntamente com o papel crucial da cultura, ajuda a reforçar a ligação entre o comportamento do tradutor e as estratégias cognitivas que conduzem a tal comportamento, fortalecendo assim a ligação entre o produto e o processo de tradução (Rojo; Ibarretxe-Antuñano, 2013, p. 19).

De acordo com as autoras, o tradutor utiliza não apenas seu conhecimento linguístico das línguas de origem e de chegada, mas também suas experiências, seu conhecimento de mundo e sua compreensão das culturas que permeiam ambas as línguas. Esse processo envolve adaptar o texto de forma que ele seja acessível e compreensível à audiência da língua-alvo. Isso ressalta a importância de teorias como a Semântica de *Frames* para compreender a tradução como um processo cognitivo e culturalmente situado, indo além da transposição lexical.

Um modelo particularmente relevante que articula a teoria com os estudos da tradução é o Modelo de Primazia do *Frame* desenvolvido por Czulo (2017). Após extensa revisão da literatura, Czulo avalia que, no processo tradutório, o tradutor busca preservar o conteúdo semântico da língua de origem na língua de chegada. Sob essa ótica, o modelo analisa os processos cognitivos envolvidos na tradução e se propõe a mapear os *frames* evocados em cada idioma, visando a uma comparabilidade máxima.

O modelo sugerido pelo autor reforça o potencial da Semântica de *Frames* e da sua contraparte prática a FrameNet para contribuir tanto para a descrição dos processos tradutórios como também para o entendimento do processo de tradução como atividade cognitiva situada culturalmente.

2.3 FrameNet Brasil numa abordagem multimodal

A FrameNet é um projeto de Lexicografia Computacional iniciado por Fillmore em 1997 no *International Computer Science Institute*, em Berkeley, com o objetivo de descrever a língua inglesa tanto semântica quanto sintaticamente. Essa descrição é baseada na análise linguística de sentenças extraídas de corpora, orientada pelos princípios da Semântica de *Frames*. O projeto tem ampla aplicabilidade em áreas como tradução automática e Processamento de Língua Natural (PLN) (Torrent *et al.*, 2024).

Com os anos, o recurso foi adaptado para diversas línguas, incluindo o português brasileiro. A versão brasileira, a FrameNet Brasil (FN.Br), é desenvolvida no laboratório de Linguística Computacional da Universidade Federal de Juiz de Fora. A FN.Br se destaca pelas iniciativas multimodais e pela colaboração com instituições parceiras no desenvolvimento de aplicações baseadas na teoria.

O projeto ReINVenta (*Research and Innovation Network for Visual and Textual Analysis of Multimodal Objects*) se destaca nesse sentido. Conduzido em colaboração com universidades parceiras, os objetivos incluem a entrega de *dataset* de objetos multimodais anotados semanticamente e a criação de algoritmos de inteligência artificial para rotulação automática e descoberta de conhecimento em objetos multimodais (Torrent *et al.*, 2022).

No glossário de uma framenet, destacam-se alguns conceitos, como *frames*, elementos de *frame*, relações entre *frames* e unidades lexicais. Os *frames* são representações linguístico-computacionais do conhecimento humano motivados lexicograficamente. Para que sejam modelados, são definidos por meio de seus elementos, denominados Elementos de *Frame* (FE, do inglês *frame element*). Estes podem ser nucleares e não nucleares. Os FE nucleares são essenciais para a compreensão do *frame*, diferenciando-o de outros, enquanto os FE não nucleares acrescentam informações complementares. A Figura 1 ilustra o *frame* Chegar¹ modelado na FN.Br.

Figura 1: Definição e FE nucleares do *frame* Chegar

The screenshot shows the FrameNet webtool interface for the 'Chegar' frame. At the top, there is a red header with the frame name 'Chegar'. To the right of the header are buttons for '@Generic', '#48', 'Arriving [en]', and a PDF download link. Below the header, there is a 'Definition' section containing the following text: 'Um **Tema** se move na direção de um **Alvo**. O **Alvo** pode ser expresso ou pode ser entendido a partir do contexto, mas é sempre implícito no próprio verbo.' Under the 'Definition' section is a 'Frame Elements' section with a 'Core' sub-section. The 'Core' table has two rows. The first row, with 'Alvo' in the first column, contains the text: 'O **Alvo** é qualquer expressão que diz onde o tema acaba, ou iria acabar, como resultado do movimento.' and the code '@location' in the third column. The second row, with 'Tema' in the first column, contains the text: 'O **Tema** é o objeto que se move. Pode ser uma entidade que se move sob seu próprio poder, mas não precisa ser.' and the code '@physical_object' in the third column.

Fonte: Captura de tela da webtool 4.02

¹ Por convenção, o nome do *frame* é registrado em fonte Courier New.

O evento de chegar é definido por um Tema que se move em direção de um Alvo. Alvo e Tema são FE nucleares, como ilustrados na sentença (1) em que a LU *chegar.v* em destaque é antecedida pelo FE Tema “eu”. Depictivo e Frequência são FE não nucleares instanciados nesta sentença.

(1) Às vezes, raro, esse ano eu acho que [eu _{TEMA}] **CHEGUEI** ^{CHEGAR} [atrasada _{DEPICTIVO}] [duas vezes _{FREQUÊNCIA}], porque com essas obras do Metrô, com congestionamen ... congestionamento [#7873]

Quando um FE nuclear não está expresso em uma sentença, ele pode ser classificado em tipos de instanciação nula, que é uma maneira de registrar a motivação de sua ausência. Na sentença (1), o FE Alvo não é instanciado linguisticamente e não é recuperável na sentença. Desse modo, prevê-se essa anotação semântica como uma Instanciação Nula Indefinida (INI, do inglês *Indefinite Null Instantiation*). Os outros dois casos de instanciação nula previstos numa FrameNet é a Nula Definida (DNI) e a Construcional (CNI). A DNI ocorre quando o elemento de *frame* pode ser recuperado pelo contexto linguístico, como é o caso de anáforas e elipses. E a CNI ocorre em casos em que a construção linguística licencia a ausência do FE, como no modo imperativo.

Um aspecto fundamental da FrameNet é a capacidade de estabelecer relações semânticas entre si, o que define o projeto em uma rede semântica. As relações tradicionais previstas entre os *frames* são: Herança, Uso, Subframe, Perspectiva, Precedência e Causativo_de / Incoativo_de.

Na relação Herança, um *frame* gera outro mais específico. *Percepção_ativa*, por exemplo, é herdeiro dos *frames* *Agir_intencionalmente* e *Percepção* e é herdado por *Fazer_turismo*. Uma consequência da relação de herança é que os frames mãe e filhos possuem FE correspondentes iguais ou semelhantes entre si. *Percepção_ativa* tem por FE nucleares Fenômeno e Perceptor, enquanto os FE nucleares de *Percepção* são Fenômeno e Perceptor.

Na relação Uso, um *frame* faz referência a outro, mais abstrato. *Percepção_ativa* estabelece relação de uso com *Atenção*, pois considera-se que não seja possível compreender o primeiro sem o conhecimento sobre o segundo. Nesse tipo de relação, não há, necessariamente, correspondência entre os FE dos dois *frames*.

Já na relação de *Subframe*, um *frame* é subevento de um que representa um evento mais complexo. Frequentemente, há uma relação de sequência temporal ou estados de coisas entre mãe e filho. A relação de Precedência é um tipo de relação Subframe que ocorre quando os subeventos são ordenados temporalmente. *Atividade_em_andamento*, por exemplo, é precedido por *Atividade_iniciar* e precede *Atividade_terminar*. Todos os três são Subframes de *Atividade*.

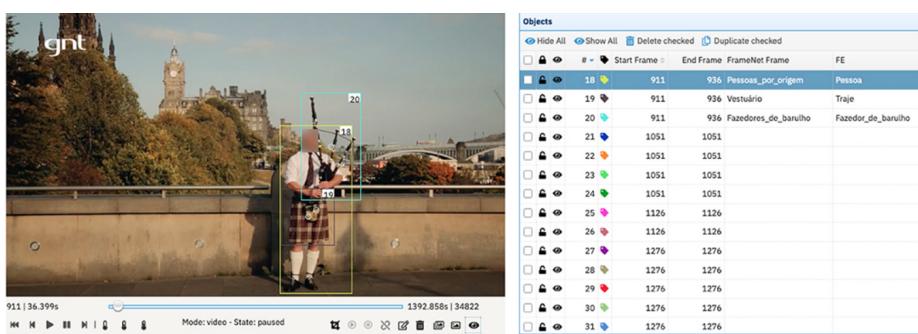
Na relação de perspectiva, os *frames* compartilham um mesmo cenário, mas focalizam aspectos diferentes. Por exemplo, *Comércio_transferência_de_mercadorias* pode servir de perspectiva do comprador em *Comércio_comprar* ou do vendedor em *Comércio_vender*.

As relações Causativo_de/ Incoativo_de ocorrem entre os *frames* chamados de causativos, incoativos e estativos. Nos *frames* causativos, o sujeito é o agente causador da ação; nos incoativos, o sujeito não é o causador da ação, e os estativos representam os casos em que o

sujeito é o experienciador de um estado. A relação *Causativo_de* ocorre entre *frames* que causam outros *frames*, como é o caso de *Causar_emoção*, um *frame* causativo, que é causativo de *Emoção_com_foco_no_experienciador*, um *frame* incoativo. Já a relação *Incoativo_de* ocorre em *Experienciar_ferimento_corporal*, que é incoativo do estativo *Estar_ferido*.

Ao adotar uma abordagem multimodal, a FB.Br assume que a linguagem verbal assim como imagens, gestos e outros modos semióticos atuam de forma interdependente na construção de sentido. Bateman *et al.* (2017) propõem uma taxonomia de classificação dos gêneros multimodais e discutem a centralidade de ferramentas computacionais que auxiliam na análise de conjuntos de dados multimodais. Nesse sentido, Torrent *et al.* (2022) apresentam um panorama do aperfeiçoamento da FN.Br de modo a incluir uma abordagem multimodal e multidimensional para representar sentido e garantir inferências contextuais na integração de objetos modais diversos.

Figura 2: Anotação de vídeo na FN.Br



Fonte: Torrent *et al.* (2022, p. 12)

A Figura 2 é um exemplo em que os autores mostram a anotação semântica da imagem do vídeo de um programa televisivo brasileiro, associada à sentença (2) proferida pelo apresentador do programa.

(2) Quando a gente pensa na Escócia, a primeira coisa que vem à mente é homem de saia, uísque escocês e gaita de fole.

Os autores discutem que a sentença é proferida alguns segundos antes de a imagem ser exibida na tela, o que torna necessário um ajuste de *frames*, homem [de saia], que, primeiramente, é anotado no *frame Pessoa*, a partir da exibição da imagem, sinaliza a necessidade de reajuste de anotação para outro mais específico, o *Pessoa_por_etnia*. Esse exemplo reforça a relevância de representar informações contextuais capturadas por informações não advindas de textos verbais, como é o caso do enquadramento situacional proporcionado pela imagem mostrada.

3 Material de análise e anotação semântica multimodal

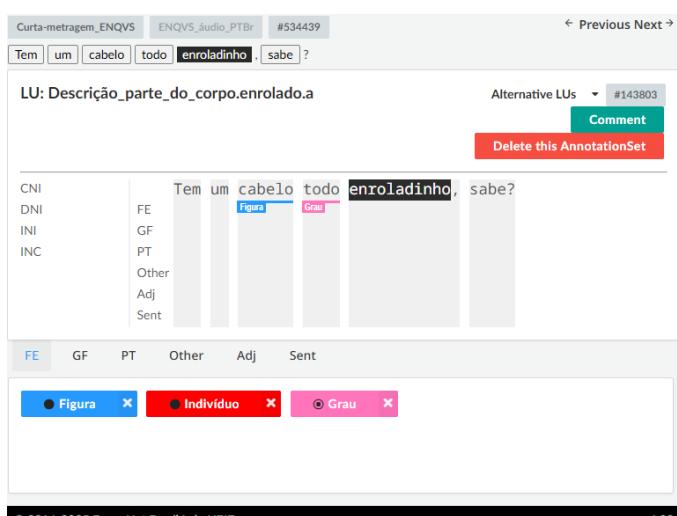
O material de análise utilizado neste estudo consiste no texto transscrito do áudio original em português, nas legendas em inglês e nas imagens dinâmicas do curta-metragem brasileiro *Eu não quero voltar sozinho*, de 2010, dirigido e roteirizado por Daniel Ribeiro e produzido pela Lacuna Filmes. A trama acompanha Leonardo, um adolescente com deficiência visual que, ao longo da narrativa, explora questões relacionadas ao amor, à amizade e à descoberta de sua sexualidade. A duração total do curta é 17 minutos, e o excerto analisado neste trabalho se concentra nos seis minutos e quatro segundos iniciais.

A coleta e o processamento do material incluem a transcrição do áudio original e a organização das legendas em inglês em um editor de texto, segmentando-as em sentenças alinhadas. Posteriormente, as sentenças foram inseridas na ferramenta para anotação multimodal da FN.Br, atualmente integrada na Webtool 4.02. Por meio dela, foram realizadas a anotação estrutural das sentenças do áudio em português. Nesta etapa, cada sentença recebe uma etiqueta especificando a origem do texto (no caso, áudio original e legenda) e faz-se a delimitação do tempo inicial e final.

Feita a anotação estrutural, o próximo passo é a anotação semântica de texto. Para essa etapa, foi feita a anotação de texto corrido (no inglês, *full text annotation*), que consiste em anotar todas as LU de cada sentença. Uma vez concluída a anotação de texto verbal, inicia-se a anotação de imagens dinâmicas. Nesta etapa, selecionam-se objetos visuais nas imagens do vídeo com base na anotação de texto corrido. Para cada objeto visual, identificam-se *frame*, FE e LU, esta última associada a um modelo de visão computacional utilizado para tarefas de reconhecimento automático de objeto visual nas imagens que vem sendo refinado com informações dos *frames* semânticos.

A Figura 3 exibe a anotação de texto da LU *enroladinho.a* na sentença “*Tem o cabelo todo enroladinho*”. Para essa LU, o *frame* *Descrição_parte_do_corpo* foi evocado.

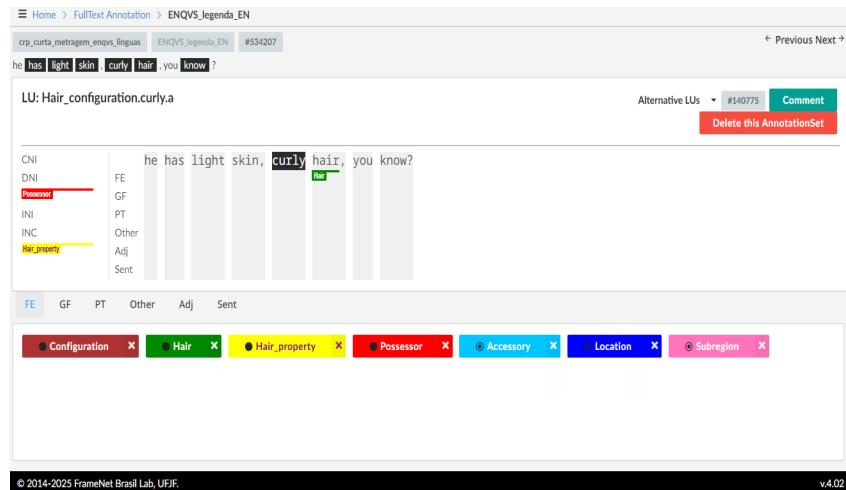
Figura 3: Anotação semântica da LU *enroladinho.a*



Fonte: Captura de tela da FN.Br 4.02

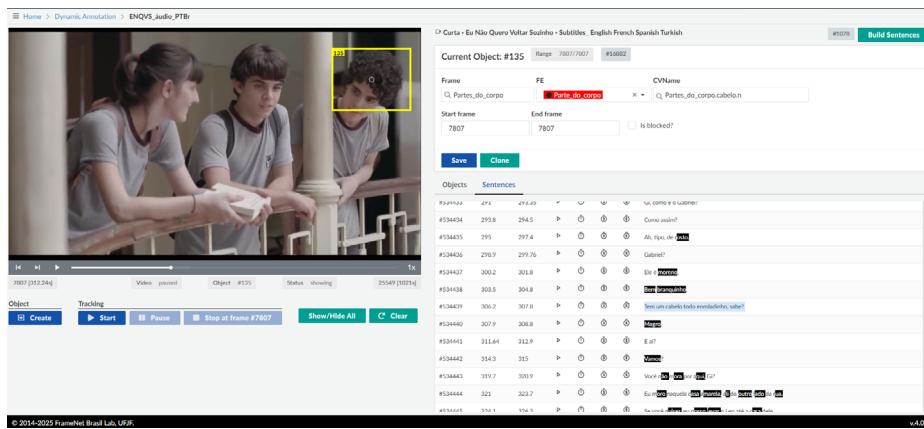
A legenda dessa mesma cena é *He has light skin, curly hair, you know?* A Fig. 4 mostra a LU *curly.a* anotada em *frame* mais específico que o escolhido para a sentença em pt.br, o *Hair_configuration*.

Figura 4: Anotação semântica da LU *curly.a*



Fonte: Captura de tela da FN.Br 4.02

Figura 5: Anotação de imagem dinâmica do objeto visual *cabelo.n*



Fonte: Captura de tela da FN.Br 4.02

A Figura 5 apresenta uma tela do *software* de anotação, em que o cabelo do personagem Gabriel foi marcado como objeto visual, tendo como referência a sentença da Fig. 3. Para isso, foi selecionado o *frame* *Partes_do_corpo* e *FE* *Parte_do_corpo*. A LU é *cabelo.n* e está vinculada ao *frame* *Partes_do_corpo*.

4 A medida de similaridade semântica entre áudio original e legenda

Após a anotação do material linguístico, a medida de similaridade semântica foi realizada utilizando o algoritmo de *spread activation* em conjunto com a similaridade de cossenos. O algoritmo calculou o decaimento de energia entre nós em um gráfico semântico com valores no intervalo de $[0,1]$. A metodologia seguiu dois estágios principais: o pré-ajuste, no qual foi calculado o valor de energia inicial para cada nó, e a propagação de energia, em que o decaimento foi determinado ao longo da rede semântica. O estágio de pós-ajuste não foi realizado, pois o gráfico da FrameNet é acíclico, como explicado por Viridiano *et al.* (2022). Nesse modelo, cada nó representa um *frame* evocado diretamente por uma sentença ou indiretamente por meio de relações entre *frames*, e os valores atribuídos refletem a proximidade entre eles na rede semântica.

A similaridade de cossenos foi então utilizada para medir o grau de similaridade semântica entre sentenças equivalentes no áudio original em português e nas legendas em inglês. Os resultados foram categorizados em três níveis. A similaridade total de valor 1 ocorreu quando todos os *frames* evocados pelas duas sentenças foram equivalentes. A similaridade parcial foi registrada quando alguns *frames* eram equivalentes e outros não, com valores entre 0 e 1. Por fim, a similaridade nula foi subdividida em dois tipos: o valor 0 foi atribuído a sentenças sem qualquer coincidência ou comparabilidade entre os *frames* evocados, enquanto o valor -1 foi aplicado a casos específicos, como sentenças sem anotação de *frames* ou que não foram traduzidas para o inglês. Essa distinção refina a categorização proposta por Samagaio (2023), que havia atribuído valor 0 a todos esses casos.

A análise dos resultados levou em conta estratégias de tradução descritas por Díaz-Cintas e Remael (2021), garantindo um olhar detalhado sobre as relações semânticas nas legendas. A classificação geral como similaridade semântica nula foi mantida para as sentenças sem tradução ou sem ativação de *frames*, segundo Samagaio (2023). No entanto, ao adotar valores distintos de 0 e -1, a pesquisa oferece maior precisão ao modelo proposto, sem comprometer a comparabilidade com estudos anteriores.

4 Resultados e discussão

4.1 Anotação semântica multimodal em números: áudio original, legenda e vídeo

O total de *frames* anotados para o texto da legenda em inglês foi inferior ao número registrado na anotação do áudio original. Esse resultado era esperado considerando que o número de sentenças na legenda do curta é menor que o total de falas transcritas do áudio original, como mostra a Tabela 1.

Tabela 1: Levantamento das anotações de texto corrido

Anotação do áudio original		Anotação da legenda	
Frames distintos	Frames totais	Frames distintos	Frames totais
105	235	104	203
45%	100	51%	100%

Fonte: Elaborada pelos autores

As tabelas 2 e 3 mostram os *frames* mais frequentes nas anotações de texto. Foram considerados mais frequentes os *frames* com 5 ocorrências ou mais.

Tabela 2: *Frames* mais frequentes no áudio original

Frame	Ocorrências	Porcentagem
Negação	16	7%
Desejar	9	4%
Frequência	8	3%
Evento_desejável	5	2%
Dizer	5	2%
Conhecimento	5	2%
Outros	188	80%
Total	235	100%

Fonte: Elaborada pelos autores

Tabela 3: *Frames* mais frequentes na legenda

Frame	Ocorrências	Porcentagem
Motion	8	4%
Direction	6	3%
Frequency	6	3%
Body_parts	6	3%
Experiencer_focused_emotion	5	2%
Certainty	5	2%
Animals	5	2%
Outros	162	81%
Total	203	100%

Fonte: Elaborada pelos autores

A partir das tabelas 2 e 3, é possível observar que estratégias de tradução audiovisual como a omissão de itens lexicais e o uso de contrações interferiram na permanência semântica entre áudio original e legenda do curta. O *frame* Negação, por exemplo, foi o mais frequente na anotação do português, mas sua contraparte para o inglês no *frame* Negation não está na lista dos mais frequentes.

Tabela 4: Levantamento da anotação de vídeo

Frames	Ocorrências	Porcentagem
Partes_do_corpo	13	8%
Pessoas	8	5%
Conhecimento	7	5%
Conhecer_pessoas	7	5%
Familiaridade	6	4%
Atividade	6	4%
Atividade_pausar	5	3%
Desejar	5	3%
Transportar	5	3%
Outros	93	60%
Total	155	100%

Fonte: Elaborada pelos autores

Quanto à anotação semântica de imagens do vídeo, foram registrados 155 frames semânticos no recorte escolhido. A Tabela 4 especifica os *frames* mais frequentes atribuídos aos objetos visuais. A respeito dos *frames* registrados a partir das LUs identificadas nas imagens, o *frame* Pessoas assume destaque com 120 ocorrências em comparação com 8 ocorrências na anotação de objetos visuais. Isso ocorreu, pois a maioria dos objetos visuais anotados foram os personagens do curta, descritos como pessoas pela LU.

4.2 Similaridade semântica em números: áudio original e legenda

Segundo classificação de Samagaio (2023), similaridades com valores entre 0,01 e 0,40 foram consideradas baixas; similaridades entre 0,41 e 0,69 foram consideradas médias, e similaridades entre 0,70 e 0,99 foram consideradas altas. Samagaio (2023) constatou que a similaridade nula ocorreu em três casos: quando não houve comparabilidade ou equivalência entre os *frames* evocados nas sentenças, quando uma das sentenças não evocou *frame*, e quando a sentença não foi traduzida na legenda. No trabalho de Samagaio (2023), todos esses casos foram representados pelo valor 0, no entanto, nesta pesquisa, os dois últimos casos receberam o valor -1. A similaridade total foi representada pelo valor 1.

A Tabela 5 mostra a frequência dos níveis de similaridade semântica obtidos no cálculo da similaridade de cossenos. A similaridade parcial representou 54% dos casos. A similaridade nula foi a segunda mais frequente, com 33% do total, somando as similaridades de

valor 0 e -1. No entanto, considerando apenas a similaridade nula de valor 0, em que ambas as sentenças evocaram *frames*, esta passa a ser a menos frequente, com 4% dos casos.

Tabela 5: Níveis de similaridade semântica

Similaridade	Ocorrências	Porcentagem
1	13	13%
de 0,01 a 0,40	20	19%
de 0,41 a 0,69	22	21%
de 0,70 a 0,99	15	14%
0	4	4%
-1	30	29%
Total	104	100%

Fonte: Elaborada pelos autores

Nota-se que, embora considerado por Czulo (2017) no modelo da Primazia do *Frame* que o tradutor busque a permanência semântica máxima entre texto original e texto traduzido, o nível de similaridade menos frequente foi o de similaridade total de valor 1, com 13 ocorrências, equivalentes a 13% do total. Somada à porcentagem das sentenças com similaridade alta, 27% do total obteve similaridade igual ou maior que média, indicando comparabilidade parcial entre os *frames* das sentenças.

4.2.1 Similaridade semântica total

A similaridade semântica total foi indicada com o valor 1. Nesses casos, os *frames* evocados foram idênticos nas duas línguas, como se pode observar no exemplo (2):

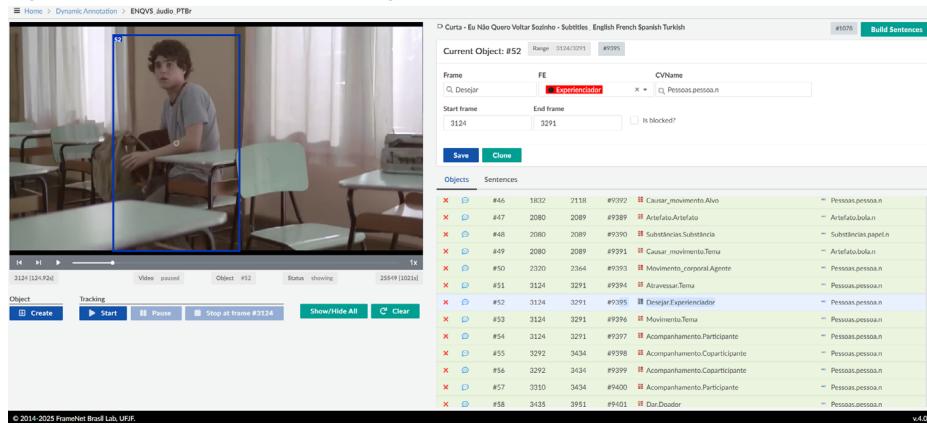
(2) **QUER**^{DESEJAR} **VIR**^{MOVIMENTO} com a gente? (628267)²

Do you WANT^{DESIRING} to COME^{MOTION} with us? (628060)

Nesse caso, a interação entre a sentença e a imagem foi a mesma para áudio original e legenda, em termos de *frames* evocados. A Figura 6 mostra a anotação para a sentença 628267.

² As sentenças nos exemplos contêm o número identificador no *corpus* da FN.Br.

Figura 6: Anotação semântica de imagem dinâmica da sentença id 628267



Fonte: Captura de tela de anotação semântica multimodal na FN.Br 4.02

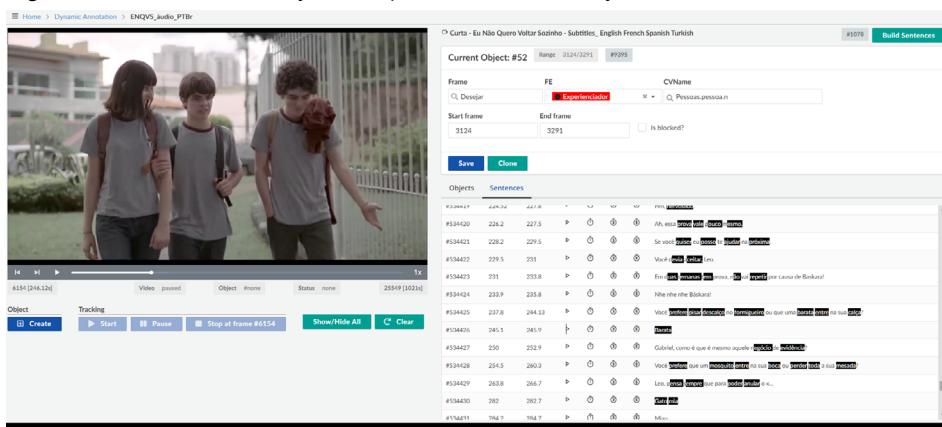
Na Figura 6, a personagem Gabriel foi delimitada como um objeto no *frame* *Desejjar*, instanciando o elemento de *frame* *Experienciador*, pois, na sentença que embasa essa anotação, esse FE é instanciado no sujeito “[você] quer vir com a gente?”, que se refere à personagem. Como todos os *frames* estão presentes nas duas línguas, as duas sentenças interagem com o vídeo da mesma maneira, do ponto de vista dos *frames* evocados.

(3) BARATA ANIMAIS . (628313)

The COCKROACH ANIMALS! (628106)

Da mesma forma, houve sentença com similaridade total que não gerou anotação no vídeo, pois as entidades instanciadas pelos FE não possuíam objetos visuais que as representassem na tela, como é o caso no exemplo (3).

Figura 7: Cena sem demarcação de objeto visual na sentença de id 628313



Fonte: Captura de tela de anotação semântica multimodal na FN.Br 4.02

Nessa cena, mostrada na Figura 7, as personagens discutem se preferem pisar descalços no formigueiro ou que uma barata entre em suas calças. Como é uma discussão de uma situação hipotética, a barata não aparece na tela.

4.2.2 Similaridade semântica nula

A similaridade semântica nula ocorre em três casos: (i) não existe convergência de *frames* entre as sentenças; (ii) uma das sentenças não evoca *frame*, e (iii) não existe uma sentença equivalente em uma das línguas.

O primeiro caso resulta em similaridade semântica com valor o e pode ser observada no exemplo (4):

(4) **SUBIR** ATRAVESSAR. (628266)

DOWN DIRECTION. (628059)

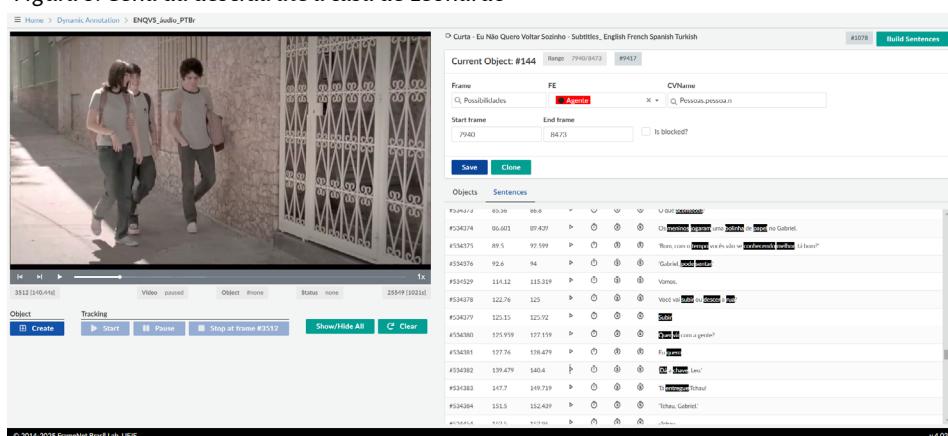
O exemplo (4) ilustra a natureza multimodal da tradução audiovisual, que deve levar em consideração as imagens na tela em busca de sentidos na língua alvo. No áudio original, com intenção de convidar o estudante novato a se juntar aos dois, a personagem Giovana faz uma pergunta a ele, exibida no exemplo 5 com a respectiva legenda produzida para o inglês:

(5) Você vai **SUBIR** ATRAVESSAR ou **DESCER** ATRAVESSAR a **RUA** VIAS?

Is your **HOUSE** BUILDINGS **UP** DIRECTION or **DONW** DIRECTION the **STREET** ROADWAYS?

As respostas em português e em inglês, mostradas no exemplo (5), parecem contraditórias se se leva em conta a cena seguinte, que mostra os personagens descendo a rua. Em português, o áudio, certamente, justifica-se por eles terem realizado um caminho mais longo, inicialmente subindo a rua, mas, em inglês, a ênfase, certamente, foi dada à cena que aparece em seguida, como mostra a Figura 8.

Figura 8: Cena da descida até a casa de Leonardo

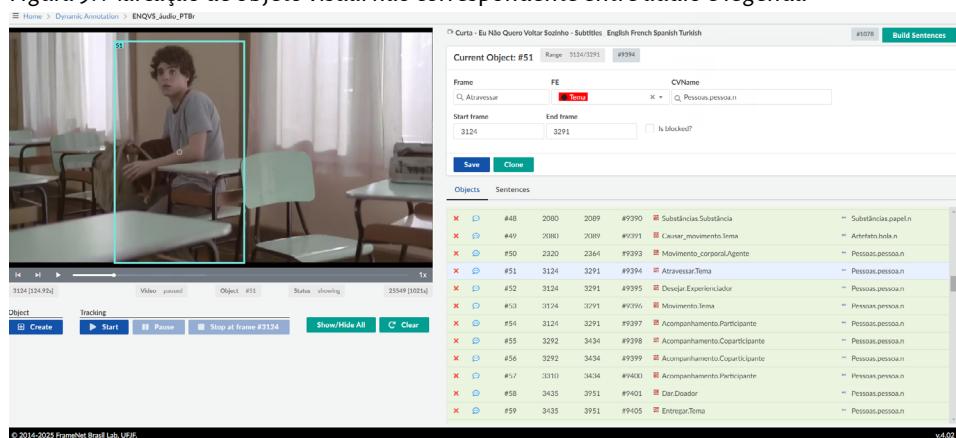


Fonte: Captura de tela de anotação semântica multimodal na FN.Br 4.02

O exemplo (4) apresentado anteriormente ilustra caso em que a anotação dos *frames* das imagens do curta não se alinha com aqueles evocados na legenda, como mostra a Figura 9. Nesse caso, a personagem Gabriel instancia o FE Tema do *frame* Atravessar evocado pela LU *subir.v*. De verbo no áudio, a opção tradutória na legenda foi um advérbio, que evocou

Direction, por esse motivo, Gabriel não é mais marcado como entidade que instancia um FE, pois o *frame* em questão apresenta como FE nucleares Ponto_de_referência e Trajetória.

Figura 9: Marcação de objeto visual não correspondente entre áudio e legenda



Fonte: Captura de tela de anotação semântica multimodal na FN.Br 4.02

O segundo caso de similaridade nula possui valor -1 e ocorre quando não há equivalente de uma sentença na outra língua, como nos exemplos (6) e (7):

(6) **QUARTO** NÚMEROS_ORDINAIS **ANDAR** SUBPARTES_DE_PRÉDIOS! (628246)

(7) Eu **VOU** MOVIMENTO no **DÉCIMO** NÚMEROS_ORDINAIS! (628247)

Essas sentenças não foram traduzidas para o inglês, configurando uma estratégia de tradução de legendagem prevista por Díaz-Cintas e Remael (2021) chamada omissão. As falas acontecem em uma cena na qual os alunos brincam sobre o barulho da máquina de escrever de Leonardo parecer o som de um elevador. Além de a sentença ser proferida rapidamente, no meio de várias outras falas, e não ter destaque na composição da cena, ela é dita por personagens figurantes que não aparecem na tela. Por esse motivo, as sentenças não estão atreladas a objetos visuais evocadores de *frames*, portanto a ausência na legenda se justifica por fatores inerentes à multimodalidade da obra audiovisual.

(8) Oi Gabriel! (628258)

Hi Gabriel! (628052)

(9) **DESENCANA** PREOCUPAR_SE. (628290)

Nevermind. (628082)

(10) Como assim? (628321)

O terceiro caso também é indicado com o valor -1 e ocorre quando pelo menos uma das sentenças não evoca *frame*. De acordo com a metodologia de anotação da FN.Br, expre-

sões com enfoque pragmático³ e nomes próprios são tipos de casos não anotados. As sentenças em (8), (9) e (10) mostram alguns exemplos.

4.2.3 Similaridade semântica parcial

Similaridades de cosseno entre o e 1 indicam uma convergência parcial dos *frames*. Vários casos de similaridade parcial ocorreram por consequência das estratégias de tradução da legendagem. O exemplo (11) mostra um caso de simplificação de perífrase verbal:

(11) E aí, **JÁ** VETOR_TEMPO tá **CONSEGUINDO** CAPACIDADE_DE_AÇÃO se **ACHAR** LOCALIZAR na **ESCOLA**
LOCAIS_POR_USO? (628293)

So, are you **FINDING** LOCATING your **WAY** SELF_MOTION around **SCHOOL** LOCALE_BY_USE? (628086)

Consequentemente, o frame Capacidade_ação não é evocado pela legenda, devida à omissão da LU *conseguir.v*. Da mesma forma, a LU *já.adv* é omitida na legenda, eliminando, assim, *Vetor_tempo*. Sua ausência não altera a relação estabelecida entre a sentença e as imagens, pois não houve FE correspondente a um objeto visual nesse caso. A similaridade entre as sentenças foi média, equivalente a 0,68.

(12) **SE** OCORRÊNCIA_CONDICIONAL você **QUISER** DESEJAR eu **POSSO** POSSIBILIDADES te **AJUDAR** ASSISTÊNCIA
na **PRÓXIMA** SEQUÊNCIA. (628308)

I CAN POSSIBILITY HELP ASSISTANCE you for the **NEXT** RELATIVE_TIME **TEST** EXAMINATION. (628100)

O exemplo (12) ilustra a estratégia de simplificação de oração composta e obteve similaridade baixa, equivalente a 0,31. Essa estratégia eliminou a marcação de Leonardo como objeto visual instanciador do FE Experienciador de *Desejar*. Outro exemplo de similaridade baixa (0,32) exibe a junção de sentenças, o que aproveita o limite de caracteres por segundo em uma sentença no lugar de duas. Além disso, as LU *próximo.a* e *next.a*, apesar de estabelecerem relação de equivalência nas sentenças do áudio original e da legenda, não evocaram os mesmos *frames*. A primeira evocou Sequência, e a última evocou Relative_time. No entanto, os dois são comparáveis pela rede semântica: tanto Sequence, que é o equivalente de Sequência no inglês, quanto Relative_Time herdam de Relation. Dessa forma, as duas sentenças evocam o *frame Relation* (ou Relação) indiretamente.

4.3 Discussão

A análise dos resultados revelou que há diferentes níveis de permanência semântica na legendagem e diferentes maneiras em que as sentenças podem interagir com as imagens na tela. Em casos de permanência semântica máxima, a interação das sentenças com os objetos

³ Para informações sobre modelagem de *frames* ditos pragmáticos, ver o trabalho de Andrade e Matos (2025) publicado neste dossiê.

na tela é a mesma, em termos de *frames* evocados. A similaridade nula em caso de omissão de sentença faz com que não haja *frames* evocados na imagem correspondentes na legenda.

Nos casos de total divergência dos *frames*, ou parcial como na similaridade média, os *frames* na imagem pela legenda serão diferentes dos *frames* evocados pelas falas do áudio original. Assim como constatado por Samagaio (2023) na análise do corpus Pedro Pelo Mundo, as estratégias de tradução audiovisual descritas por Díaz-Cintas e Remael (2021) influenciaram a permanência semântica entre texto original e traduzido.

Os pares de sentenças com similaridade semântica diferente de total mostraram que estratégias como a omissão de conteúdo menos importante, a simplificação de sentenças compostas, a junção de sentenças, a simplificação de perífrase verbal e a mudança do tipo de oração provocaram divergências na evocação de *frames*, resultando na similaridade parcial ou nula.

Além disso, a análise evidenciou que o processo de tradução audiovisual envolve a análise do conteúdo não verbal em conjunto com o linguístico. Em alguns casos, as imagens justificam a mudança do significado de uma sentença e a omissão de sentenças, como em instâncias em que não há objetos visuais correspondentes aos *frames* evocados nas falas.

Com esses resultados, avalia-se que as estratégias para a tradução de uma obra audiovisual influenciam a permanência semântica máxima entre texto original e traduzido almejada pelo tradutor, segundo o modelo da Primazia do Frame (Czulo, 2017).

Nota-se que as relações estabelecidas entre os *frames* de sentenças equivalentes através da rede semântica da FrameNet possibilitaram uma similaridade média ou alta em 35% dos casos, enquanto 52% do total, mais da metade, obtiveram similaridade baixa ou nula, sendo significativamente afetados pelas estratégias utilizadas.

A constatação dos casos em que houve omissão total ou parcial da sentença foi compatível com os resultados de Freitas (2022), que observou uma diminuição do número de *frames* do texto da legenda em comparação com o texto em português referente ao áudio original. Em adição a essa discussão, constatou-se através dos exemplos apresentados que a relação estabelecida entre o texto da legenda em inglês e as imagens do vídeo foi diferente da estabelecida entre as imagens e o texto original em português, em decorrência dessas omissões, bem como da evocação de *frames* diferentes em ambos os textos.

5 Conclusão

Em concordância com os achados de Samagaio (2023), conclui-se por meio da discussão dos resultados que as estratégias de tradução audiovisual utilizadas no processo de legendagem do curta-metragem, como listadas por Díaz-Cintas e Remael (2021), interferem diretamente no nível de permanência semântica da legenda em relação ao áudio original.

Tendo em vista o modelo da Primazia do *Frame* proposto por Czulo (2017), notou-se que, apesar de a similaridade total ter sido a menos frequente, com 4% dos casos, houve comparabilidade parcial em 54% dos casos, aproximando semanticamente as sentenças do áudio original e da legenda mesmo quando não houve permanência semântica máxima. Além disso, o uso do *frame* como uma representação cognitiva no processo tradutório evidencia a natureza cognitiva da tradução, como afirma Rojo e Ibarretxe-Antuñano (2013).

Ademais, as relações semânticas estabelecidas entre texto e imagem foram diferentes para o áudio original e para a legenda. No caso analisado de similaridade total, os *frames* evocados no vídeo foram os mesmos evocados no áudio original e na legenda. Porém, nos casos de similaridade parcial, os objetos visuais anotados no vídeo seriam diferentes caso a

anotação tivesse sido feita com base na legenda, em vez do áudio original. Em alguns casos, as mudanças feitas no texto na legenda, como a omissão de falas, foram justificadas pela análise da interação imagem-texto entre o áudio original e o vídeo.

Por esse motivo, concluiu-se também que o processo da legendagem deve levar em consideração a multimodalidade da obra. Além de estar limitada por fatores espaciais e temporais, como o espaço máximo que a legenda deve ocupar na tela e o limite de tempo de cada legenda para que o espectador possa acompanhar a obra cinematográfica, a legenda deve estabelecer uma relação de coerência com as imagens. Além disso, visto que a redução do texto original é inevitável, as relações semânticas entre a imagem e o texto podem servir para guiar o tradutor no momento de decidir o que é relevante ou não para ser traduzido na legenda.

Nesse sentido, esta pesquisa contribui para o melhor entendimento da legendagem e do processo de tradução audiovisual, por meio da Semântica de *Frames*, e este trabalho criou um corpus paralelo multimodal anotado semanticamente, que pode ser utilizado em tarefas de processamento de língua natural e em tarefas multilíngues.

Ao desenvolver um *corpus* paralelo multimodal de legendagem anotado semanticamente com potencial para uso em tarefas de processamento de língua natural, a pesquisa amplia a rede semântica da FrameNet Brasil e aumenta o seu conjunto de *corpora* para tarefas multilíngues, além de promover melhor entendimento sobre a legendagem e sobre o processo de tradução audiovisual. Dessa forma, o estudo contribui para os Estudos da Tradução no âmbito da tradução audiovisual e para o campo da Semântica de *Frames* e convida novas investigações que aprofundem as evocações semânticas de vídeo partindo da anotação de texto da legenda em vez do áudio original.

Referências

- BATEMAN, J.; WILDFEUSER, 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>. Disponível em: <https://www.degruyterbrill.com/document/doi/10.1515/9783110479898/html>. Acesso em: 10 jan. 2025.
- BERTOLDI, A. Semântica de *Frames* e tradução: um estudo da equivalência de tradução de termos culturalmente marcados. *Letras & Letras*. 2016, v.32. 149. 10.14393/LL63-v32n1a2016-8. Disponível em: <https://seer.ufu.br/index.php/letraseletras/article/view/33045>. Acesso em: 10 jan. 2025.
- BOAS, H. C. Frame semantics and translation. In: ROJO, A; IBARRETXEANTUNANO, I. *Cognitive Linguistics and Translation: Advances in Some Theoretical Models and Applications*. Berlin/Boston: De Gruyter Mouton, 2013. p. 125-158. Disponível em: <https://www.degruyterbrill.com/document/doi/10.1515/9783110302943/html>. Acesso em: 10 jan. 2025.
- CINTAS, J. D.; REMAEL, A. *Translation Practices Explained*. Oxfordshire: Routledge, 2007.
- CINTAS, J. D.; REMAEL, A. *Subtitling: Concepts and Practices*. Oxfordshire: Routledge, 2021. (Translation Practices Explained).

CZULO, O. Aspects of a Primacy of Frame Model of Translation. In: HANSEN-SCHIRRA, S.; CZULO, O.; HOFMANN, S. (ed.). *Empirical Modelling of Translation and Interpreting*. Berlin, Germany: Language Science Press, 2017. p. 465.

DORNELAS, L. D., GAMONAL, M. A., PAGANO, A. S. Análise semântica de audiodescrição em curta metragem: uma abordagem multimodal a partir da Semântica de Frames. *Domínios de Linguagem*, Uberlândia, v. 1866, p. 2-30, 2024. DOI: <https://doi.org/10.34019/1808-9461.2022.v23.38564>. Disponível em: <https://seer.ufu.br/index.php/dominiosdelinguagem/article/view/75272>. Acesso em: 10 jan. 2025.

FILLMORE, C. Frame Semantics. In: THE LINGUISTIC SOCIETY OF KOREA (ed.). *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co, 1982. p. 111-137.

FILLMORE, C. J. Semântica de Frames. *Cadernos de Tradução*. Porto Alegre, n. 25, jul-dez, 2009.

FREITAS, K. C. G. *A tradução audiovisual sob a perspectiva da semântica de frames: Uma análise dos significados evocados pelo texto do áudio original e sua legenda em L2*. 2022. Trabalho de Conclusão de Curso. Faculdade de Letras, Universidade Federal de Minas Gerais (UFMG). Belo Horizonte, 2022.

HOLMES, J. S. The Name and Nature of Translation Studies. *Translation Section of the Third International Congress of Applied Linguistics*, Copenhagen, p. 66-80, ago. 1972. Disponível em: <https://archive.org/details/Holmes1972TheNameAndNatureOfTranslationStudies>. mode/2up. Acesso em: 10 jan. 2025.

PAGANO, A. S.; TEIXEIRA, A. L. R.; MAYER, F. A. Accessible Audiovisual Translation. In: JI, M.; LAVIOSA, S. (ed.). *The Oxford Handbook of Translation and Social Practices*. Oxford: Oxford University Press, 202. p. 66-82. DOI: <https://doi.org/10.1093/oxfordhb/9780190067205.013.4>.

ROJO, A. Applying Frame Semantics to Translation: a practical example. *Meta: Translators' Journal*, 47(3), p. 312-350. 2002. DOI: <https://doi.org/10.7202/008018ar>.

ROJO, A.; IBARRETXE-ANTUÑANO, I. Cognitive linguistics and translation studies: Past, present and future. *Cognitive Linguistics and Translation*, Berlin, p. 3-30, 2013. Disponível em: <https://www.degruyterbrill.com/document/doi/10.1515/9783110302943/html>. Acesso em: 10 jan. 2025.

SAMAGAO, M. M. *A permanência semântica na tradução audiovisual: uma análise da legendagem pela ótica da Semântica de Frames*. 2023. 184p. Dissertação (Mestrado em Linguística) – Universidade Federal de Juiz de Fora, Juiz de Fora, Minas Gerais.

SOUZA, D., GAMONAL, M. A., PAGANO, A. S. A audiodescrição sob a perspectiva da Semântica de Frames: um estudo exploratório. *Revista Gatilho*, Juiz de Fora, v. 23, p. 101-125, 2022. DOI: <https://doi.org/10.34019/1808-9461.2022.v23.38564>. Disponível em: <https://periodicos.ufjf.br/index.php/gatilho/article/view/38564>. Acesso em: 10 jan. 2025.

TORRENT, T. T., MATOS, E., BELCAVELLO, F., VIRIDIANO, M., GAMONAL, M. A., COSTA, A. D. & MARIM, M. C. Representing Context in FrameNet: A Multidimensional, Multimodal Approach. *Frontiers in Psychology*, v. 13, 1-20, 2022. DOI: <https://doi.org/10.3389/fpsyg.2022.838441>. Disponível em: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.838441/full>. Acesso em: 10 jan 2025.

TORRENT, T. T.; HOFFMANN, T.; ALMEIDA, A. L.; TURNER, M. *Copilots for Linguists: AI, Constructions, and Frames*. Cambridge, Reino Unido: Cambridge University Press, 2024.

WEBTOOL 4.02. Definição e FE nucleares do *frame* Chegar. In: WEBTOOL 4.02. *Banco de dados*. [S. I.]: [s. n.]. Disponível em: <https://webtool.frame.net.br/report/frame/1455>. Acesso em: 10 jan. 2025.

WEBTOOL 4.02. Anotação Semântica da LU enroladinho.a. *In: WEBTOOL 4.02. Banco de dados. [S. I.]: [s. n.]*. Disponível em: <https://webtool.frame.net.br/annotation/fe/sentence/534439>. Acesso em: 10 jan. 2025.

WEBTOOL 4.02. Anotação semântica da LU curly.a. *In: WEBTOOL 4.02. Banco de dados. [S. I.]: [s. n.]*. Disponível em <https://webtool.frame.net.br/annotation/dynamicMode/1078>. Acesso em: 10 jan. 2025.

WEBTOOL 4.02. Anotação de imagem dinâmica do objeto visual cabelo.n. *In: WEBTOOL 4.02. Banco de dados. [S. I.]: [s. n.]*. Disponível em: <https://webtool.frame.net.br/annotation/dynamicMode/10785>. Acesso em: 10 jan. 2025.

WEBTOOL 4.02. Anotação semântica de imagem dinâmica da sentença id 628267. *In: WEBTOOL 4.02. Banco de dados. [S. I.]: [s. n.]*. Disponível em: <https://webtool.frame.net.br/annotation/dynamicMode/573>. Acesso em: 10 jan. 2025.

WEBTOOL 4.02. Cena sem demarcação de objeto visual na sentença de id 628313. *In: WEBTOOL 4.02. Banco de dados. [S. I.]: [s. n.]*. Disponível em: <https://webtool.frame.net.br/annotation/dynamicMode/573>. Acesso em: 10 jan. 2025.

WEBTOOL 4.02. Cena da descida até a casa de Leonardo. *In: WEBTOOL 4.02. Banco de dados. [S. I.]: [s. n.]*. Disponível em: <https://webtool.frame.net.br/annotation/dynamicMode/573>. Acesso em: 10 jan. 2025.