

CONTRIBUIÇÕES DA SEMÂNTICA DE FRAMES PARA A ANOTAÇÃO DE GÊNEROS MULTIMODAIS NA ÁREA DE TECNOLOGIA LINGUÍSTICA

CONTRIBUTIONS OF FRAME SEMANTICS FOR THE ANNOTATION OF MULTIMODAL GENRE IN THE FIELD OF LANGUAGE TECHNOLOGY

Mariane de Carvalho Pinto¹
Gabrielly Azalim Braz²
Natália Sathler Sigiliano³

DOI: 10.34019/2179-3700.2024.v24.46181

ENVIADO EM: 4/10/2024

APROVADO EM: 11/11/2024

RESUMO

O presente artigo apresenta os resultados do projeto Representação Semântica da Multimodalidade: definição de parâmetros analíticos para gêneros multimodais para a área de Tecnologia Linguística, realizado na FrameNet Brasil (FN-Br). Em tal estudo, os esforços de descrição lexicográfica computacional da FN-Br, ancorados na Semântica de Frames, foram expandidos para incluir outros modos de significação além do verbal. Foram feitas anotações de texto verbal e de imagens dinâmicas da primeira temporada do programa de viagens da GNT “Pedro pelo Mundo”, através de ferramentas próprias da FN-Br. Os resultados reforçam a necessidade de empreender anotações computacionais que considerem modos comunicativos além do texto verbal.

Palavras-chave: Tecnologia Linguística. Multimodalidade. Semântica de Frames.

ABSTRACT

The present article presents the results of the project Semantic Representation of Multimodality: defining analytical parameters for multimodal genres in the field of

¹Mestranda em Linguística na UFJF, bolsista PIBIC, e-mail: mariane.carvalho@estudante.ufjf.br.

²Graduanda em Letras na UFJF, bolsista VIC, e-mail: gabrielly.azalim@estudante.ufjf.br.

³Professora Associada da Faculdade de Letras da Universidade Federal de Juiz de Fora, doutora em Linguística pela Universidade Federal do Rio de Janeiro. E-mail: natalia.sigiliano@ufjf.br.

Linguistic Technology, conducted at FrameNetBrasil (FN-Br). In this study, the efforts of FN-Br's computational lexicographic description, anchored in Frame Semantics, were expanded to include other modes of signification beyond the verbal. Verbal text and dynamic images from the first season of GNT's travel program "Pedro pelo Mundo" were annotated using FN-Br tools. Results reinforce the need for annotations that take into consideration communicative modes other than verbal text.

1 INTRODUÇÃO

A Tecnologia Linguística é uma área multidisciplinar do conhecimento que une a linguística e a ciência da computação. Abrangendo diferentes campos de estudos, as pesquisas desenvolvidas pela área seguem múltiplas direções, envolvendo desde o desenvolvimento de ferramentas educacionais e de tradução automática à criação de modelos de língua. Uma das preocupações atuais dos estudos em Tecnologia Linguística é a multimodalidade (BATEMAN ET AL., 2017), mais especificamente a integração de texto verbal com outras formas de dados, como imagens, áudio e vídeo, para criar sistemas que possam entender e gerar informações em diferentes formatos (TORRENT ET AL., 2022).

Se, por um lado, as pesquisas linguísticas já se desenvolveram muito no que tange aos modelos de análise e anotação de textos verbais, o que auxiliou o avanço de tecnologia na área, por outro, nota-se uma lacuna na forma como dados de treinamento e, por consequência, os modelos deles derivados, representam os aspectos multimodais da significação. É nesse contexto que o laboratório de Linguística Computacional FrameNet Brasil (FN-Br) tem envidado esforços na composição de conjuntos de dados—*datasets*—multimodais do Português. Assim, a FrameNet Brasil (FN-Br), juntamente com outros laboratórios e grupos de pesquisa brasileiros, tem se dedicado, nos últimos anos, à construção e à avaliação de um modelo computacional para representação de objetos multimodais (BELCAVELLO, 2023).

Nesse cenário, o presente trabalho tem como objetivo apresentar contribuições da Semântica de Frames para a constituição de conjuntos de dados de treinamento anotados no âmbito da Tecnologia Linguística por meio dos resultados do projeto Representação Semântica da Multimodalidade: definição de parâmetros analíticos para gêneros multimodais para a área de Tecnologia Linguística, realizado na FN-Br. Em tal projeto, expandiram-se os

esforços de descrição lexicográfica computacional da FN-Br, ancorados na Semântica de Frames, para englobar outros modos de significação, que não apenas o verbal, em suas análises semânticas.

Para tanto, têm sido realizadas anotações dos textos multimodais em duas vias, de forma a se observar em que medida os distintos modos se entrecruzam para a produção de significados. Sendo assim, além de haver anotação do modo oral - por meio do texto verbal, com base na semântica de frames-, é realizada, também, anotação do texto visual.

As análises das anotações têm revelado a dependência dos múltiplos modos de significação para a produção de sentidos no texto multimodal e a ativação de distintas perspectivas sobre um mesmo frame, provocada pelos distintos modos de significação envolvidos na composição do texto.

De forma a apresentar como as anotações têm sido realizadas e a maneira como tais resultados puderam ser observados, a próxima seção deste trabalho se ocupará de apresentar as bases teóricas da proposta, abordando a Semântica de Frames (FILLMORE, 1982), e o modelo da FrameNet Brasil. Em seguida, trataremos da relevância dos estudos em multimodalidade (BATEMAN ET AL., 2017; GUALBERTO; SANTOS, 2019), destacando os esforços da FN-Br de abarcar, no processo de anotação, diferentes usos da língua e contemplar distintos modos de significação. Posteriormente, será apresentada a metodologia, em que discorreremos sobre como as anotações multimodais (que embasam os resultados a serem apresentados) ocorreram durante a execução do projeto de pesquisa intitulado “Representação Semântica da Multimodalidade: definição de parâmetros analíticos para gêneros multimodais para a área de Tecnologia Linguística”. Esse projeto teve como objetivo comparar em que medida os frames evocados pelo áudio de uma produção audiovisual (“Pedro pelo Mundo”, do canal televisivo GNT) interagem com as imagens do vídeo.

Por fim, serão apresentados alguns dos resultados da pesquisa. Os exemplos na seção mostram que restringir a análise dos objetos multimodais apenas ao texto verbal é prejudicial à análise semântica, já que esses objetos são capazes de complementar e enriquecer os padrões de evocação de frames registrados durante a anotação apenas do texto. Cabe destacar que o esforço de anotação relatado neste trabalho contribui para a constituição do primeiro corpus multimodal anotado para frames do Português Brasileiro.

2 A SEMÂNTICA DE FRAMES E A FRAMENET BRASIL

A Semântica de Frames é uma teoria linguística proposta por Charles J. Fillmore, que oferece, consoante o autor, uma perspectiva particular de se olhar para o significado das palavras (FILLMORE, 1982, p.111). A teoria está ancorada no conceito de frames, noção adotada por Fillmore que serve como uma espécie de construto fundador para outros conceitos usados no campo da Linguística, como esquema, script e cenário, por exemplo. Para o autor, o termo frame pode ser designado como

[...] qualquer sistema de conceitos relacionados de tal maneira que, para entender qualquer um deles, você precisa entender toda a estrutura na qual ele se encaixa; quando um dos conceitos dessa estrutura é introduzido em um texto ou em uma conversa, todos os outros são automaticamente disponibilizados. (FILLMORE, 1982, p.111)

Sob essa perspectiva, um frame é, portanto, um sistema de significados que engloba tanto os sentidos linguísticos das palavras quanto os sentidos que advêm das experiências dos falantes no mundo. Para o autor, o significado de uma palavra é relativizado à cena evocada por ela e, nesse sentido, só pode ser compreendido a partir do conhecimento dos falantes acerca dessa cena e dos participantes que a compõem (chamados, na teoria, de Elementos de Frame – doravante EFs). Para elucidar a questão, Fillmore (1982) apresenta o exemplo da transação comercial. De acordo com o autor, o significado da palavra “vender” só é compreendido porque se sabe que existe, na cena evocada por esse verbo (a de transação comercial), um “vendedor” e uma “mercadoria”, bem como outros elementos que podem ou não ser instanciados a depender da estrutura usada, como “comprador”, por exemplo. Assim, em análises semânticas baseadas em frames, é defendido que o significado de um termo é dado pela relação entre ele e o seu pano de fundo (FILLMORE, 1982).

Um outro exemplo apresentado por Fillmore em *Frame Semantics* revela que o significado está ancorado em percepções culturalmente

compartilhadas pelos falantes. Trata-se do exemplo do “fim de semana”, no qual o linguista destaca que o entendimento desse conceito tem relação, obrigatoriamente, com a divisão feita em cada cultura entre dias de trabalho e final de semana (FILLMORE, 1982, p.116). Nesse sentido, as ideias de Fillmore (1982) se afastam das defendidas pela semântica vericondicional, que partem do pressuposto de que o significado de uma sentença é determinado exclusivamente pelas condições sob as quais ela seria verdadeira ou falsa (SAEED, 2003), sem considerar o contexto experiencial e cultural dos falantes.

Fundamentada no conceito de frame, assim, o objetivo da Semântica de Frames é investigar como os falantes da língua categorizam as coisas no mundo e, por consequência, atribuem sentido às expressões linguísticas nas línguas naturais. Para tanto, as investigações desenvolvidas na área são baseadas em dados reais de uso da língua.

Nesse contexto, em 1997, a FrameNet teve início com Fillmore no Instituto Internacional de Ciência da Computação, em Berkeley, na Califórnia. A iniciativa, desde o seu início, tem como referencial teórico a Semântica de Frames e investiga, primordialmente, a atribuição de sentido nas línguas naturais. Segundo Ruppenhoffer et al. (2016), o objetivo do projeto é

[...] documentar a variedade de possibilidades combinatórias semânticas e sintáticas — valências — de cada palavra em cada um de seus sentidos, por meio de anotação de sentenças assistida por computador e da tabulação automática e exibição dos resultados da anotação. (RUPPENHOFFER ET AL., 2016, p.7)

Antes com dados e anotações restritos à língua inglesa, a FrameNet tem sido expandida para diferentes idiomas ao longo dos anos. Exemplo disso é a inclusão de anotações no português brasileiro por meio da FrameNet Brasil (FN-Br), laboratório de linguística computacional localizado na Universidade Federal de Juiz de Fora (UFJF), em Minas Gerais (SALOMÃO, 2009).

3 A FRAMENET BRASIL E A MULTIMODALIDADE

De acordo com Gualberto e Santos (2019), a multimodalidade é uma característica de todos os textos, uma vez que todos eles, em maior ou menor grau, combinam diferentes modos de comunicação para transmitir significado (GUALBERTO; SANTOS, 2019, p.6). Segundo as autoras, o aumento da produção e disseminação de significados em diversos modos e meios fez com que as teorias que antes se voltavam apenas para a construção de sentidos a partir do modo verbal tivessem que ser repensadas para englobar outros modos semióticos. Acerca disso, citando Hodge e Kress (1988), Gualberto e Santos (2019) apontam, no texto, que contemplar a multimodalidade é uma tarefa desafiadora, haja vista que, mais que apenas acrescentar outros modos à análise, envolve o desenvolvimento de novas ferramentas e estratégias de estudo desses modos. Nessa mesma linha, Bateman, Wildfeuer e Hiipala (2017) defendem que a análise de textos multimodais deve ser feita de modo a explorar os multiplicadores que emergem da relação de combinação entre os modos semióticos, afirmando, ainda, que “o principal desafio da pesquisa multimodal encontrar formas de caracterizar a natureza dessas interdependências e desenvolver metodologias para investigarem-nas empiricamente”. (Bateman, Wildfeuer, Hiipala, 2017, p. 17).

Por essa dificuldade, até pouco tempo, as análises desenvolvidas pela FN-Br se concentravam apenas na modalidade textual da língua. No entanto, ao se considerar o aspecto inerentemente multimodal da comunicação humana, viu-se a necessidade de que outras modalidades fossem também contempladas pelos estudos desenvolvidos no projeto, a fim de que fossem realizadas análises semânticas mais completas da linguagem em uso. Para que isso fosse feito, a FN-Br passou a buscar meios de anotação multimodais, o que foi impulsionado com a criação da ReINVenTA—Research and Innovation Network for Vision and Text Analysis of Multimodal Objects—, que assume como uma de suas preocupações o desenvolvimento de um modelo semântico-computacional que lide com informações multimodais, como imagens estáticas e dinâmicas. Atualmente, a FN-Br possui uma proposta de anotação

multimodal utilizando ferramentas próprias, o que permite investigar a interação entre diferentes modalidades da linguagem humana e seu impacto na construção de sentidos (BELCAVELLO ET AL., 2022; LUZ ET AL., 2023).

4 METODOLOGIA DO PROJETO: ANOTAÇÃO DOS DADOS

Esta seção apresenta, de forma breve, o dataset Frame²(BELCAVELLO ET AL., 2024) e o passo a passo das anotações multimodais desenvolvidas pela FrameNet Brasil a partir dele.

Como já foi mencionado, a Semântica de *Frames* defende a ideia de que os falantes interpretam o mundo através de *frames*, sendo que esses *frames* são evocados por expressões linguísticas no uso da língua. Ao considerar a multimodalidade, a FN-Br passou a investigar se existiria a possibilidade de que, assim como as palavras, os elementos visuais em um vídeo fossem capazes de evocar *frames* ou complementar o que foi evocado pela narração (BELCAVELLO ET AL., 2020). A partir daí, a FN-Br buscou fornecer (e, atualmente, tem refinado) uma proposta de anotação multimodal por meio de ferramentas próprias.

Inicialmente, foi necessária a formação de um dataset multimodal, o qual foi produzido no âmbito da ReINVenTA. Esse dataset, o Frame², é composto por 10 episódios de 23 minutos cada do programa de TV *Pedro pelo Mundo*, exibido a partir de 2016 pelo canal GNT. A série de viagens, apresentada por Pedro Andrade, aborda aspectos sociais, culturais e econômicos dos diferentes países onde os episódios são ambientados. Para a formação do dataset, cada um dos episódios da primeira temporada teve as suas falas transcritas automaticamente e revisadas por um grupo de 12 anotadores. Em seguida, esse grupo anotou, para *frames*, as sentenças manualmente, usando a FrameNet Brasil WebTool (TORRENT ET AL., 2024)—vide 4.1—de acordo com as diretrizes da FN-Br. Depois, em outra ferramenta, a Charon (BELCAVELLO ET AL., 2022)—vide 4.2—, o mesmo anotador foi designado para anotar os elementos visuais presentes no episódio.

Como já mencionado, ancorada na Semântica de Frames (FILLMORE, 1982), a FrameNet é um modelo semântico-computacional que tem seus itens lexicais organizados em frames (FILLMORE, BAKER, 2009). O frame que será mencionado adiante, no passo a passo que exemplifica a anotação, é o de Ingestão (Figura 1), que tem como Elementos de Frame (EF) nucleares os micropapéis de INGESTOR e INGERÍVEL. O frame ainda apresenta os micropapéis de DURAÇÃO, FINALIDADE, FONTE, GRAU, INSTRUMENTO, LUGAR, MANEIRA e TEMPO como EFs não nucleares.

Figura 1 — O frame de Ingestão

Ingestão [@Action][@Food][@Lexical][#238]

Definição	
Um Ingestor consome comida ou bebida (Ingeríveis), o que implica colocar os Ingeríveis na boca para a entrega ao sistema digestivo. Isso pode incluir o uso de um Instrumento . Sentenças que descrevem a provisão de comidas para outros NÃO estão incluídas nesse frame.	
Elementos de Frame Nucleares	
FE Core:	
Ingeríveis	Os Ingeríveis são entidades que podem ser consumidas pelo Ingestor .
Ingestor semantic_type: @sentient	O Ingestor é a pessoa que come ou bebe.

Fonte: WebTool

2.1 Anotação de texto corrido

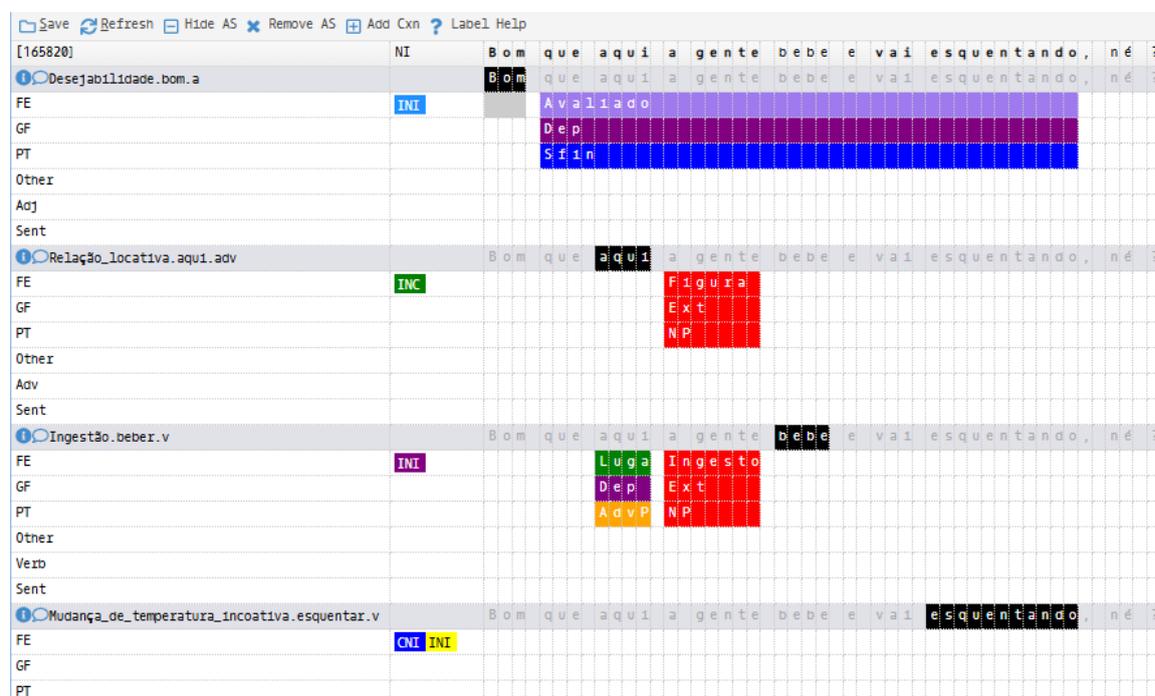
A tarefa de anotação de texto corrido é feita por meio da WebTool, da FrameNet Brasil. Nessa tarefa, são atribuídos diferentes lotes de sentenças para os anotadores do laboratório. Em contato com as sentenças que deve anotar, o anotador precisa atribuir um *frame* a cada unidade lexical (UL) presente nelas. Para isso, na interface gráfica, ele é orientado a clicar na palavra que deseja anotar, o que fará com que um quadro com os *frames* associados àquela UL e disponíveis na FN-Br seja carregado em sua tela.

Quando o anotador seleciona o *frame* mais adequado para a UL naquele contexto, é gerada uma camada de anotação de Elemento de Frame (EF), na qual ele deve categorizar os demais itens da sentença, segundo o *frame* escolhido anteriormente. Na FN-Br, a anotação pode contemplar mais camadas,

no entanto, pela natureza do presente texto, elas não serão abordadas neste trabalho.

A Figura 2 apresenta uma sentença anotada. Como pode ser percebido nela, a UL *beber.v* foi anotada no Frame Ingestão. A partir dessa anotação, o anotador pôde atribuir os valores de LUGAR e INGESTOR, respectivamente, aos elementos *aqui* e *a gente*. Note que um dos EFs nucleares, o de INGERÍVEL, não está instanciado na sentença, o que será diferente na anotação de vídeo.

Figura 2 — Anotação semântica para a Unidade Lexical *beber* no frame deIngestão



Fonte: WebTool

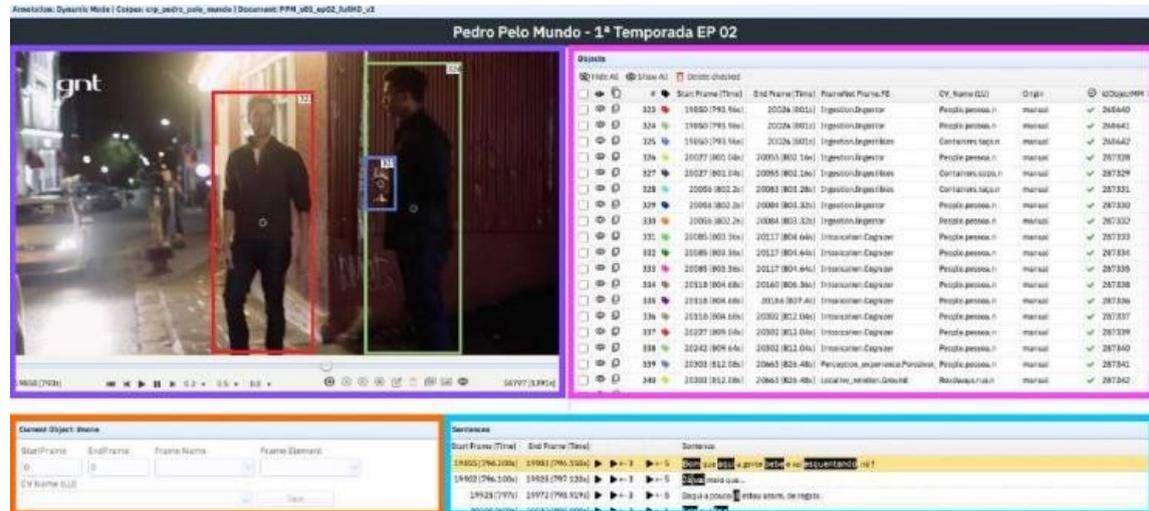
2.2 Anotação de imagens dinâmicas (vídeo)

A anotação de imagens dinâmicas, por sua vez, acontece em uma outra ferramenta, a Charon, que conta com três painéis e um arquivo de vídeo, o qual congrega as entidades que serão anotadas. Na ferramenta, o anotador, guiando-se pelos *frames* já anotados no texto verbal (mas não se limitando a eles), deve manualmente delimitar com *bounding boxes* os objetos multimodais que serão anotados. Depois, o anotador deve atribuir, assim como fez para o texto, um

frame para o objeto marcado a partir da lista de *frames* disponíveis na base da FN-Br. Escolhido o *frame*, o anotador define qual EF aparece na imagem e atribui um *Computer Vision Name* (CV Name) a ele. Esta última etiqueta tem por objetivo indicar a categoria do elemento mostrado na imagem, associando-o a um *frame* da base de dados da FrameNet Brasil.

A título de exemplificação, a Figura 3 apresenta uma anotação de imagem dinâmica. Como pode ser observado nela, do lado esquerdo da interface está localizado o vídeo para a anotação, no qual estão marcadas três *bounding boxes*: uma vermelha, uma verde e uma azul. A *bounding box* vermelha delimita o apresentador do programa, Pedro, enquanto a verde delimita a pessoa com quem ele conversa no vídeo. Guiando-se pela anotação de texto corrido, o anotador atribuiu aos dois o EF nuclear de INGESTOR na cena, ao passo que, no CV Name, associou as duas *bounding boxes* na tela à UL *pessoa.n*, já que era o que ele estava vendo objetivamente na cena (dois homens). A UL *pessoa.nevoca*, na FrameNet Brasil, o *frame* *Pessoas*. Em seguida, o anotador atribuiu à taça o EF nuclear de INGERÍVEL, enquanto, no CV Name, associou a sua *bounding box* à UL *taça.n*, que evoca o *frame* *Recipiente*.

Figura 3 — Anotação de imagem dinâmica



Fonte: Charon

5A SEMÂNTICA DE FRAMES, A FRAMENET BRASIL E A MULTIMODALIDADE

Todos esses esforços de anotação, norteados pela Semântica de *Frames*, fizeram parte de diferentes pesquisas na FrameNet Brasil. O objetivo desta seção é apresentar dois dos resultados do estudo *Representação Semântica da Multimodalidade: definição de parâmetros analíticos para gêneros multimodais para a área de Tecnologia Linguística*, que teve o objetivo de verificar, a partir da hipótese de que imagens também podem evocar *frames* (BELCAVELLO, 2023), se os *frames* evocados pelos elementos visuais do programa *Pedro pelo Mundo* de fato complementavam ou enriqueciam os padrões de anotação de texto verbal percebidos ao longo do *corpus*. Isso foi feito por meio da comparação entre a anotação de texto corrido e a de imagens dinâmicas.

Nesse sentido, com o estudo, foi verificado que os elementos visuais da produção são capazes de instanciar elementos que faziam parte do *frame* escolhido, e não foram mencionados ao longo da narração. Como pode ser notado ao se compararem as Figuras 2 e 3, o anotador escolheu, em uma das sentenças do sexto episódio, por exemplo, o *frame* de *Ingestão*, que seria evocado por *beber.v*. Esse *frame*, como já mencionado, tem dois elementos considerados principais no português: quem ingere (o *INGESTOR*) e o que é ingerido (o *INGERÍVEL*). Note que este último elemento não aparece na sentença da Figura 2 e, portanto, é marcado como uma instanciação nula na anotação do texto. No entanto, durante a anotação de imagens dinâmicas, esse elemento aparece como um elemento visual: a bebida na taça. Nesse caso, o anotador atribuiu o EF *INGERÍVEL* à imagem, o que mostra a função dos elementos visuais na complementação do que está sendo narrado. Além disso, a taça, presente no vídeo, também foi associada a um novo *frame*: o de *Recipiente*. Isso porque, durante a anotação de imagens dinâmicas, o espaço CV Name permite, conforme vimos, que o anotador indique, também, aquilo que ele está vendo objetivamente na cena. Dessa forma, várias informações podem ser associadas a uma mesma parte do vídeo, o que resulta em muito mais pontos de dados semânticos—os quais seriam perdidos, caso a análise não considerasse a multimodalidade constituinte do objeto de análise.

Os dados da pesquisa também permitiram perceber que os *frames* evocados pelas imagens são capazes de apontar para novas perspectivas em relação às cenas apresentadas nos episódios. No segundo episódio (Figura 4), por exemplo, o anotador atribuiu o *frame* `Turismo_de_atração` à expressão *atrações turísticas.n*, presente na sentença. No entanto, na cena que faz parte desse momento da narração, o anotador escolheu o *frame* `Atrair_turistas` ao marcar a catedral presente no vídeo. A diferença de escolha coloca em perspectiva elementos diferentes da cena: no primeiro caso, o foco é dado ao turista, que vai ao encontro da atração; no segundo, a atração é colocada em foco.

Figura 4 — Mudança de perspectiva na anotação

O centro de Reykjavik é famoso pela arte de rua, pelas casas coloridas e por algumas **atrações turísticas** `Attraction_tourism.42`



Fonte: Charon

Sendo assim, os resultados das investigações do estudo mostram que restringir a análise semântica de objetos multimodais, como o programa de

televisão analisado, apenas ao texto verbal pode ser prejudicial a uma análise dos efeitos de sentido, já que as imagens anotadas enriquecem de forma muito significativa a anotação. Por meio da análise de *frames* na anotação multimodal, foi possível realizar comparações entre texto e vídeo e analisar como os elementos de ambos os modos se correlacionam para a construção de sentidos. A Semântica de *Frames*, nesse sentido, fornece uma estrutura organizada e consistente para essa análise, facilitando a integração das informações que esses diferentes modos apresentam.

6 CONCLUSÃO

O presente trabalho teve como objetivo apresentar algumas das contribuições da Semântica de Frames para análises no âmbito da Tecnologia Linguística por meio dos resultados do estudo Representação Semântica da Multimodalidade: definição de parâmetros analíticos para gêneros multimodais para a área de Tecnologia Linguística. Nesse sentido, abordaram-se os esforços da FrameNet Brasil, ancorados na Semântica de Frames, para englobar outros modos semióticos, que não apenas o verbal, em suas análises semânticas.

As anotações dos episódios de “Pedro pelo mundo” evidenciaram que a produção de significados de um texto não se restringe ao aspecto verbal. Dessa forma, a análise das anotações apontou para a necessidade de que os textos sejam abordados, nas pesquisas linguísticas, em sua integralidade multissemiótica, a fim de dar conta dos fenômenos de linguagem que hoje compreendem os textos multimodais.

Essas novas necessidades indicaram diversificados caminhos de pesquisa a serem seguidos, em que se destaca uma segunda rodada de anotação em curso, a qual tem por objetivo anotar os eventos que caracterizam as cenas observadas no decorrer dos episódios do programa “Pedro pelo Mundo”. Além disso, a partir dos dados anotados, um modelo de Inteligência Artificial será gerado, o qual pode servir, por exemplo, para realizar busca semântica de vídeos e de outros conteúdos multimodais, por meio da detecção semiautomática de objetos visuais.

REFERÊNCIAS

BATEMAN, J., WILDFEUER, J., & HIIPPALA, T. **Multimodality: Foundations, research and analysis—A problem-oriented introduction**. Berlin: Walter de Gruyter GmbH & Co KG, 2017.

BELCAVELLO, F. **FrameNet Annotation for Multimodal Corpora: devising a methodology for the semantic representation of text-image interactions in audiovisual productions**. 135f. Tese (Doutorado em Linguística) — Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2023.

BELCAVELLO, F.; TORRENT, T. T.; MATOS, E. E.; PAGANO, A. S.; GAMONAL, M. A.; SIGILIANO, N. S.; DUTRA, L. V.; ABREU, H. A.; SAMAGAI, 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. **Frame2: A FrameNet-based Multimodal Dataset for Tackling Text-image Interactions in Video**. In: **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**. Torino: ELRA/ICCL, 2024, p.7429 – 7437.

BELCAVELLO, F.; VIRIDIANO, M.; DINIZ DA COSTA, A.; MATOS, E. E.; TORRENT, T. T. **Frame-Based Annotation of Multimodal Corpora: Tracking (A)Synchronies in Meaning Construction**. In: **Proceedings of the LREC International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet**. Marseille, France: ELRA, 2020, p. 23-30.

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, France: ELRA, 2022, p. 91-96.

GUALBERTO, C. L.; SANTOS, Z. B. dos. Multimodalidade no contexto brasileiro: um estado de arte. **DELTA: Documentação E Estudos Em Linguística Teórica e Aplicada**, 35(2), 2019.

FILLMORE, C. J. Frame semantics. In.: **Linguistics in the morning calm**. Korea: Hanshin Publishing Company, 1982, p.111-137.

FILLMORE, C. J.; BAKER, C. A Frames Approach To Semantic Analysis. In: HEINE, B.; NARROG, H. (Orgs.). **The Oxford Handbook Of Linguistic Analysis**.(p.313–340). Oxford: Oxford University Press, 2009.

LUZ, A.; BRAZ, G.; RUIZ, L.;PINTO, M. C.; BELCAVELLO, F.; SIGILIANO, N. S.; TORRENT, T. Anotação do Dataset Multimodal da ReINVenTA. In: **Anais do Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL)**, 14.Porto Alegre: Sociedade Brasileira de Computação, 2023, p. 352-356.

RUPPENHOFER, J.; ELLSWORTH, M.; PETRUCK, M. R. L.; JOHNSON, C. R.; SCHEFFCZYK, J. **FrameNet II: Extendedtheoryandpractice**. Berkeley, International Computer Science Institute, 2016. Disponível em: <http://framenet.icsi.berkeley.edu/>. Acesso em: 14 jul. 2024.

SAEED, J.Sentences relations and truth. In: _____. **Semantics**. Malden (MA): Blackwell, 2003.

SALOMÃO, M. FrameNet Brasil: um trabalho em progresso. **Calidoscópico**,7(3), 2009, p. 171–182.

TORRENT, T. T., MATOS, E. E. D. S., BELCAVELLO, F., VIRIDIANO, M., GAMONAL, M. A., COSTA, A. D. D., & MARIM, M. C. Representingcontext in FrameNet: A multidimensional, multimodal approach. **Frontiers in Psychology**, 13, 838441, 2022.

TORRENT, T. T.; MATOS, E. E. S.; COSTA, A. D.; GAMONAL, M. A.; PERON-CORREA, S. R.; PAIVA, V. R. L. A Flexible Tool for a Qualia-Enriched FrameNet:

the FrameNet Brasil WebTool. **Language Resources and Evaluation**, 58(1), 2024, p. 1-29.