

Coleta, composição e etapas de pré-processamento de *corpus*: procedimentos para a anotação multimodal da FrameNet Brasil

Anna Beatriz C. Silva¹, Iasmin Rabelo¹, Igor M. Oliveira¹, Mariana Souza¹, Maucha Gamonal², Raquel Roza¹

¹Laboratório de Tradução (LETRA) — Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil

²Laboratório de Linguística Computacional (FrameNet Brasil) Universidade Federal de Juiz de Fora (UFJF) — Juiz de Fora, MG — Brasil

bbangz@ufmg.br, {igormoliveira7, iasminvaleria, marianassouza.mota, mauchaandrade, figueiredorozar}@gmail.com

Resumo: Este trabalho apresenta a preparação de um *corpus* voltado para a anotação multimodal na FrameNet Brasil. A anotação, desenvolvida a partir da teoria da Semântica de *Frames*, permite a integração de diferentes modos comunicativos, construindo uma base de tecnologia linguística aplicável a múltiplas áreas. As etapas de coleta, composição e pré-processamento do *corpus* são os primeiros passos para o desenvolvimento das pesquisas de anotação multimodal.

1. Introdução

A integração da multimodalidade dentro do panorama teórico-metodológico da rede semântico-computacional da FrameNet [RUPPENHOFER *ET AL* 2016] amplia as possibilidades de desenvolvimento de tecnologias linguísticas mais avançadas através da base teórica da Semântica de *Frames* [FILLMORE 1982]. Isso envolve não apenas a análise e o processamento de texto, mas também a capacidade de interpretar e utilizar informações provenientes de outras modalidades, áudios, vídeos e imagens que são transcritos, contribuindo para aplicações mais sofisticadas em áreas, como, por exemplo, a tradução automática, a análise de sentimentos e a indexação de mídia.

Dentro desse cenário, o projeto em andamento ReINVenTA (*Research and Innovation Network for Visual and Textual Analysis of Multimodal Objects*) tem avançado nos processos de anotação de imagens estáticas e vídeos por meio das ferramentas Webtool e Charon [BELCAVELLO *ET AL* 2022]. A iniciativa reforça a representação de contexto a partir de uma abordagem multidimensional e multimodal [TORRENT *ET AL* 2022]. Liderado pelo laboratório de Linguística Computacional FrameNet Brasil na Universidade Federal de Juiz de Fora, o projeto conta com a colaboração de outras instituições, como a Universidade Federal de Minas Gerais.

Partindo disso, este trabalho apresenta as etapas da anotação estrutural [ALUÍSIO e ALMEIDA 2021] de um *corpus* multimodal, que alinha as transcrições textuais aos trechos de áudio/vídeo através de marcadores temporais [XIAO 2010]. Por

meio do *corpus Audition*, esta pesquisa visa descrever os estágios envolvidos na coleta, composição e etapas de pré-processamento de um *corpus* multimodal para anotação segundo o suporte teórico-metodológico da Semântica de *Frames* e da FrameNet Brasil.¹ Esses procedimentos antecedem a anotação multimodal com imagem dinâmica. Para tanto, apresentaremos a plataforma Charon, utilizada na realização de tais tarefas de pesquisa.

A seguir, na seção 2, apresentaremos a Semântica de *Frames* e como ela é usada pela FrameNet Brasil para realizar as análises multimodais. Na seção 3, exploraremos a metodologia utilizada na pesquisa e concluiremos o trabalho na seção 4, ao fazer ponderações sobre o trabalho realizado e as futuras perspectivas para o projeto.

2. Semântica de *Frames* e FrameNet Brasil em uma abordagem multimodal

A Semântica de *Frames* é uma teoria linguística desenvolvida por Charles J. Fillmore que propõe o estudo do significado a partir de uma perspectiva empírica e cognitiva. A teoria tem como base o conceito de *frame*, definido pelo autor 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 1982, p. 111]

Dessa forma, a Semântica de *Frames* analisa como o conhecimento é “evocado” e expresso através da linguagem. Sob essa ótica, a compreensão de uma unidade lexical (UL) depende, necessariamente, da compreensão da cena, ou *frame*, em que ela está inserida. Visto que a teoria busca explorar a construção do significado considerando fatores sociais, culturais e cognitivos, surge a possibilidade de usá-la como auxílio no desenvolvimento de tecnologias linguísticas.

A FrameNet Brasil usa sua rede semântico-computacional para representar o conhecimento semântico e processar a linguagem. A multimodalidade, que abrange imagens, gestos e expressões faciais, enriquece o entendimento do significado, o que amplia o escopo do trabalho semântico-computacional. A abordagem multimodal da FrameNet Brasil, ao incorporar tipos de dados, como vídeos, programas de TV, imagens e textos [TORRENT ET AL 2022], oferece uma compreensão precisa do significado, considerando a interação complexa entre a linguagem e outras formas comunicativas.

3. Metodologia

Dois plataformas são utilizadas na anotação multimodal: a Webtool e a Charon. Na primeira, acontece a anotação linguística e, na segunda, a anotação de imagens estáticas e dinâmicas. O passo a passo para garantir que a anotação aconteça requer uma série de procedimentos, primeiramente, via Charon. Essa plataforma online é acessada mediante inscrição prévia. Tendo feito o login com usuário e senha, serão exibidas duas abas, “*Corpus*” e “*Annotation*”. Em “*Corpus*”, os *corpora* que compõem a pesquisa podem

¹ Agradecemos, pelo financiamento através de bolsas de iniciação científica, à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig) (PIBIC RED-00106/21) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (PIBIC 408269/2021-9 e 420945/2022-9)

ser acessados. Por meio desse ícone, estão os recursos que garantem as tarefas prévias à anotação em si. Nessa etapa, são inseridas informações básicas de identificação do material de análise, bem como o *upload* e pré-processamento do vídeo (Figura 1).

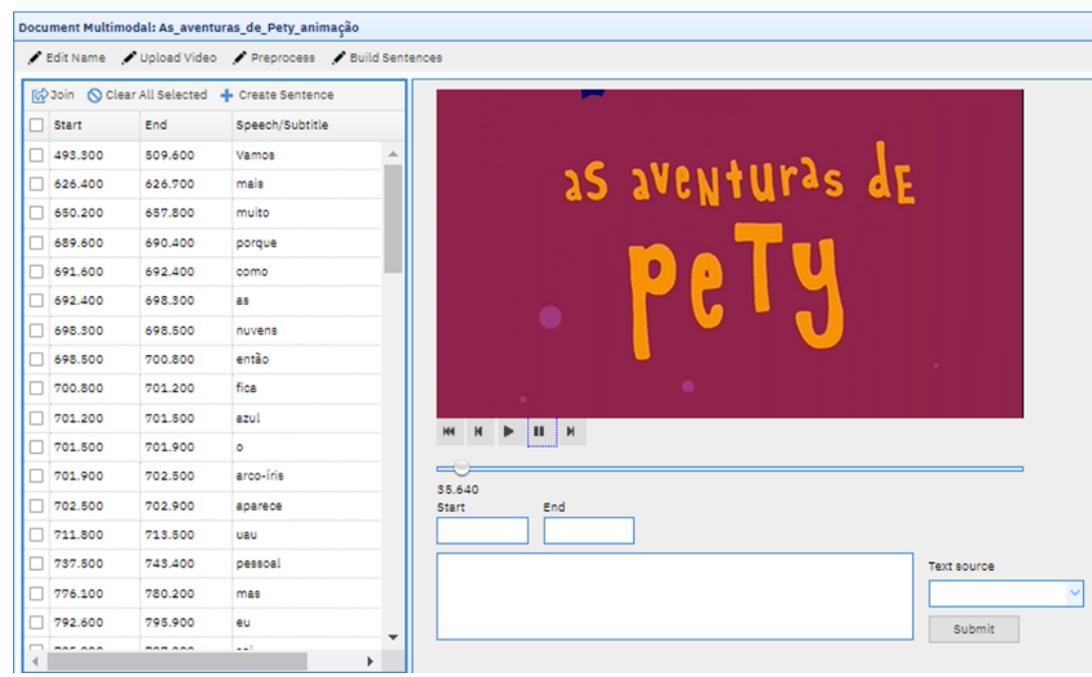


Figura 1. Página para a Edição das Sentenças do Curta-metragem

O *corpus Audition* inclui curtas animados e *live actions* com audiodescrição, como "As Aventuras de Pety", produzida por Aranhas Films, adicionado com a permissão da produtora e visível nas Figuras 1 e 2. A parte de texto corrido do *corpus* é composta pela extração dos áudio dos vídeos, o áudio é processado por um serviço em nuvem, que transcreve o que é dito e indica os intervalos de tempo. As legendas dos vídeos são capturadas e sincronizadas usando reconhecimento óptico de caracteres. O texto resultante do processo é exibido em um painel ao lado esquerdo da tela.

A Figura 1 mostra a etapa de revisão das sentenças reconhecidas automaticamente pela opção "**Build Sentences**". Nessa etapa, os anotadores acessam as informações e revisam o texto manualmente, fazendo correções e ajustes nos intervalos de tempo. Se necessário, eles podem criar novas sentenças ("**Create Sentence**"). As palavras são selecionadas no painel esquerdo e adicionadas ao painel direito através do comando "**Join**". Correções adicionais podem ser feitas na caixa de texto no centro da tela, como ilustra a Figura 2.

O tempo de início e fim das sentenças são ajustados usando as caixas "**Start**" e "**End**" acima da caixa de texto. O recurso "**Text source**" permite que o anotador classifique o texto que está sendo selecionado na caixa em "**Original audio**", para falas originais do vídeo, "**Subtitle**", para legendas, "**Text overlay**", para textos que aparecem na tela e não fazem parte da legenda, e "**Audio description**", para audiodescrição. Se um vídeo contém, por exemplo, áudio original, audiodescrição e legendas, três anotações serão feitas, cada uma com sua devida etiqueta. Para salvar as alterações, os anotadores selecionam a opção "**Submit**".

Start	End	Sentence
		esquecer.
313.40	318.0	E, lembre-se: para não se perder, volte antes do anoitecer.
318.759	321.123	Então o que você está esperando, Madureira?
321.680	324.835	Como assim? Eu também vou junto?
324.839	330.959	Lógico. Vamos apressar o passo, gente, antes que alguém chegue primeiro.
330.959	331.639	Eles atravessam a ponte.
331.639	333.560	Abaixo, um rio.
340.870	341.805	Entram no bosque.
343.900	346.79	E caminham entre as árvores.
352.300	355.600	Pety come uma banana e Zezinho uma maçã.
357.400	361.439	Pety joga a casca no chão, Zezinho o resto de maçã e outros lixos.
357.400	361.439	Pety joga a casca no chão, Zezinho o resto de maçã e outros lixos.
361.879	364.199	A casca de banana vai até o cabelo Pety.

Figura 2. Espaço para Edição das Sentenças do Curta-metragem

Após o processo de transcrição do curta-metragem, passamos para a anotação de texto corrido, que acontece pela plataforma Webtool, onde fazemos as análises das sentenças e atribuímos os *frames* e elementos de *frame* ao texto. Feita a anotação linguística, fazemos a anotação do vídeo, anotação de imagem dinâmica, associando os *frames* evocados no texto, aos *frames* dispostos sobre o vídeo. As imagens são delimitadas por quadrados que as acompanham *frame* por *frame*, e são anotadas conforme o *frame* evocado, cv name (UL que identifica o objeto no vídeo) e unidade lexical.

4. Conclusão

A incorporação da multimodalidade no contexto da FrameNet Brasil proporciona um avanço significativo nas tecnologias linguísticas. Ao integrar informações de diferentes modalidades, – áudio, imagem e texto – é possível obter uma compreensão mais abrangente e precisa do significado na linguagem natural. Isso abre novas possibilidades de aplicações avançadas em áreas como tradução automática, análise de sentimentos e treinamento de IAs (Inteligências Artificiais). A anotação multimodal de *corpora*, por meio das plataformas Webtool e Charon, desempenha um papel fundamental nesse processo, permitindo a anotação de informações linguísticas, imagens estáticas e dinâmicas.

Com base nos avanços obtidos e no trabalho realizado com o *corpus Audition*, são planejadas ações para explorar a variedade de materiais analisados e avançar no desenvolvimento de tecnologias linguísticas. A área da audiodescrição, por exemplo, é grande beneficiária das pesquisas de anotação multimodal, visto que a associação entre o conteúdo linguístico verbal e não verbal atuam diretamente na experiência do usuário.

5. Referências

- Aluísio, S. M. e Almeida, G. M. de B. (2021). “O que é e como se constrói um *corpus*? Lições aprendidas na compilação de vários *corpora* para pesquisa linguística”, *Calidoscópico*, 4(3), p. 156–178. Disponível em: <https://revistas.unisinos.br/index.php/calidoscopio/article/view/6002>. Acesso em: 1 de jul. 2023.
- Belcavello, F., Viridiano, M., Matos, E. E. d. S., e Torrent, T. T. (2022). Charon: a FrameNet Annotation Tool for Multimodal *Corpora*. In *Proceedings of the 16th Linguistic Annotation Workshop*, páginas 91–96, Marseille, France, June. European Language Resources Association (ELRA). Disponível em: <http://www.lrec-conf.org/proceedings/lrec2022/workshops/LAWXVI/pdf/2022.lawvi-1.11.pdf>. Acesso em: 28 de jun. 2023.
- Charon [FNBr]. Disponível em: <https://charon.frame.net.br/>. Acesso em: 21 de jun. 2023.
- Fillmore, C. J. Semântica de Frames. In *Cadernos de Tradução*. Porto Alegre, nº 25, jul-dez, 2009.
- Pety; As Aventuras de. Direção: Anahí Borges. Produção: Anahí Borges. YouTube. 16 de jun. 2021. 14 min. Disponível em: <https://www.youtube.com/watch?v=h0TbaPIDkFI>. Acesso em: 21 de jun. 2023.
- Torrent T. T., Matos E.E. dos S., Belcavello F., Viridiano M., Gamonal M.A., Costa A.D. da, e Marim M.C. (2022). Representing Context in FrameNet: A Multidimensional, Multimodal Approach. *Front. Psychol.*
- Webtool [FNBr]. Disponível em: <https://webtool.framenetbr.ufjf.br/>. Acesso em: 21 de jun. 2023.
- Xiao, Richard. Empirical and Statistical Approaches. In: *Handbook of Natural Language Processing*. Nova York, 2ª Edição, 2010, p. 161.