PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA

# *PROPP - Pesquisa*

## Dados do Projeto e do(a) Coordenador do Projeto

| | |
|---|---|
| **Título do Projeto** | Gêneros Textuais Multimodais em um Contexto Tecnológico: definições de novas camadas de anotação para a representação semântica computacional da deixis narrativa em objetos multimodais |
| **Referência da Chamada:** | (X) BIC/UFJF e PIBIC/CNPq<br>( ) PIBIC/CNPq AÇÕES AFIRMATIVAS<br>( ) PROBIC/FAPEMIG<br>( ) PROBIC-JR/FAPEMIG<br>( ) Apoio ao Recém-Doutor<br>( ) Apoio a Grupos de Pesquisa<br>( ) Apoio à Instalação de Doutores<br>( ) Cadastro na Propesq |
| **Coordenador do Projeto:** | Natália Sathler Sigiliano |
| **Equipe:** | Ely Edison da Silva Matos, Frederico Belcavello, Tiago Timponi Torrent |
| **Endereços para contato:** | Eletrônico: natalia.sigiliano@ufjf.br<br>Telefônico: (32) 98821-5723 |
| **Unidade/Departamento:** | Faculdade de Letras / Departamento de Letras |
| **Data:** | 12/06/2023 |

## 1 . Justificativa/Caracterização do Problema

Este projeto insere-se no projeto estruturante Representação Semântica Computacional da Multimodalidade: aliando visão computacional e compreensão de língua natural através da Semântica de Frames. Em específico, busca delinear a conceituação de gênero textual multimodal no que concerne ao seu uso pela área de Tecnologia Linguística. Assim, busca uma conceituação – e os critérios analíticos dela decorrentes – que olhe para a multimodalidade fora dos dois contextos em que ela costuma ser estudada. De um lado, afasta-se de seu tratamento no âmbito escolar (ROJO & MOURA, 2012; 2019), em que a preocupação se assenta em especial na questão da exploração didática da multimodalidade; enquanto, de outro, vai além, também, dos estudos em comunicação multimodal (BATEMAN, WILDFEUER & HIIPPALA, 2017), os quais se voltam em especial para análises qualitiativas de contextos multimodais amplos.

No contexto tecnológico em questão, gêneros multimodais são objetos que compõem um dataset a ser utilizado no treinamento de algoritmos de Inteligência Artificial (IA). A área de interface entre Linguística e Inteligência Artificial é uma das mais profícuas no que concerne à produção de inovação e novas tecnologias de grande impacto na sociedade. A disponibilização de grandes corpora em formato digital, aliada aos desenvolvimentos em aprendizagem de máquina, tem levado a um sensível avanço nos modelos linguísticos para Processamento de Língua Natural. Entretanto, referidos modelos não raro encontram como barreira a indeterminação da significação linguística. De forma análoga, as pesquisas em Visão Computacional avançam no uso de técnicas de aprendizagem de máquina sobre dados rotulados. Contudo, as intrincadas relações que se estabelecem entre as diferentes semioses ainda passam ao largo dos modelos computacionais.

Um exemplo tradicional de uso de Visão Computacional para geração de língua são os modelos que trabalham com geração de legendas para fotos. Tais modelos dão ênfase ao aprimoramento da capacidade de detecção de elementos, seguida da descrição das imagens (FANG et al., 2015; DEVLIN et al. 2015; NIKOLAUS et al. 2019). Numa perspectiva de melhor integração entre os campos de Visão Computacional e Processamento de Língua Natural, o trabalho de geração automática de legendas para fotojornalismo é um exemplo. Nesse caso, a expectativa sobre as legendas é de texto mais interpretativo do que descritivo. Para isso, os sistemas de geração costumam ser alimentados por matérias jornalísticas bem como suas respectivas imagens ilustrativas (BATRA; HE; VOGIATZIS, 2016). Há ainda os trabalhos que se dedicam à geração de legendas para vídeos, seja por meio da descrição dos elementos e cenas visuais (AKSOY et al., 2017) ou pela interpretação de áudio presente no vídeo em contraste com as imagens síncronas (SUN et al., 2019).

No que concerne aos modelos para tratamento linguístico de objetos multimodais, Cohn (2016) apresenta teoria de compreensão de narrativas visuais sequenciais adaptada às especificidades das narrativas fílmicas. O modelo propõe diretrizes para uma análise das sequências narrativas em textos multimodais, articulando linguagem verbal e visual. Entretanto, não prevê de que forma as relações semânticas locais estabelecidas entre esses elementos possam ser tratadas ou modeladas.

Outra proposta de modelo para o tratamento de corpora multimodais é feita por Loschky et al. (2020). Os autores propõem uma teoria de percepção de cena e compreensão de evento que

mapeia processos cognitivos complexos colocados em ação em atividades de compreensão de narrativas visuais. O modelo indica a necessidade de um apararto semântico para a interpretação refinada dos eventos, mas não o detalha.

Nesse contexto, o projeto estruturante supracitado alia as abordagens do Processamento de Língua Natural e da Visão Computacional, à luz de um modelo semântico curado por humanos, para propor um modelo de representação computacional da semântica de objetos multimodais tais como programas de TV e reportagens em que se apresentam conjuntos de fotos e legenda, tipicamente usados em textos da esfera jornalística, dentre outros. Para tanto, parte do aporte teórico da Semântica de Frames (FILLMORE, 1982), consubstanciado em sua aplicação computacional: a FrameNet (BAKER, 2017). Uma FrameNet é um modelo computacional da cognição linguística no qual itens lexicais, bem como outros tipos de estruturas linguísticas, são modelados em termos dos frames - ou sistemas de conceitos - que evocam.

A hipótese perseguida é a de que, assim como os itens linguísticos podem evocar frames, elementos visuais presentes em imagens podem igualmente fazê-lo ou, ainda, atuar de maneira complementar aos frames evocados pela língua. A metodologia de trabalho envolve a criação de um corpus multimodal anotado (“gold standard dataset”) que possa ser posteriormente empregado para tarefas de aprendizagem de máquina focadas na rotulação semântica automática de objetos multimodais, tanto na semiose linguística quanto em outras. Como resultados esperados, pretende-se entregar o referido "gold standard dataset" multimodal, bem como uma ferramenta computacional de identificação de frames em objetos multimodais.

Entretanto, para que tais objetivos sejam atingidos, é necessário, antes, definir o que caracteriza os gêneros multimodais do ponto de vista de suas propriedades semântico-pragmáticas e estruturais, com especial atenção para os aspectos que sejam relevantes para a anotação computacional e para a posterior aplicação desta anotação ao treinamento de modelos de IA. A primeira etapa deste projeto, desenvolvida no período de setembro de 2022 a agosto de 2023, debruçou-se sobre a definição das características gerais constitutivas dos gêneros multimodais com relevância para aqueles aspectos que precisariam ser tratados para a representação dos níveis mais imediatos da semântica de conteúdo dos objetos multimodais. Para o próximo ano, o escopo do projeto se volta para a representação semântico-computacional da deixis narrativa (DUCHAN et al., 1995), em específico, no sentido de especificar os requisitos para que mudanças no centro dêitico das narrativas fílmicas possam ser indicadas na anotação multimodal.

## 2 . Objetivos

Este projeto tem como objetivo principal especificar os requisitos para o tratamento linguístico-computacional da deixis narrativa em objetos multimodais, em especial em narrativas fílmicas. Apresenta, ainda, os seguintes objetivos correlatos:

- Delineamento de um corpus piloto para identificação de mudanças de centro dêitivo em narrativas fílmicas.
- Levantamento de características definidoras das alternâncias de centros déiticos em gêneros multimodais, em especial nas narrativas fílmicas.

- Proposição de categorias de análise e especificação de requisitos para sua implementação na ferramenta de anotação computacional de objetos multimodais da FrameNet Brasil (BELCAVELLO et al., 2020).

## 3 . Metodologia e Estratégias de Ação

Para alcançar os objetivos listados na seção 3, este projeto adotará as seguintes metodologia e estratégias de ação:

- Levantamento bibliográfico sobre deixis narrativa em gêneros multimodais: os bolsistas de IC que atuarão no projeto realizarão pesquisa bibliográfica acerca da temática da deixis multimodal, compilando propostas de conceituação.
- Cotejamento das características e *frameworks* levantados com o arcabouço teórico-metodológico da Semântica de Frames e da FrameNet: uma vez feito o levantamento bibliográfico, serão comparados os arcabouços teórico-metodológicos que se debruçam sobre a multimodalidade com aquele da Semântica de Frames e de sua implementação computacional, a FrameNet.
- Constituição de corpus piloto para análise: definido o conjunto preliminar de critérios analíticos, será realizada seleção de um pequeno corpus piloto.
- Levantamento preliminar de fenômenos de alteração do centro dêitico no corpus: uma vez definido o corpus piloto, uma análise preliminar, sem uso de ferramenta computacional será realizado para indicar os tipos de alternância de centro dêitico nas narrativas fílmicas e as pistas que as sinalizam.
- Especificação de requisitos: uma vez finalizada a anotação preliminar, os critérios propostos para a análise da dêixis narrativa no contexto da Tecnologia Linguística serão especificados para posterior implementação na ferramenta de anotação multimodal da FrameNet Brasil.

## 4 . Resultados e os impactos esperados

Ao final do projeto, terão sido entregues os seguintes produtos:

- Corpus piloto para análise do fenômeno da alteração do centro dêitico nas narrativas fílmicas;
- Anotação multimodal preliminar do corpus piloto;
- Especificação de requisitos para implementação de novas categorias de análise da ferramenta de anotação multimodal da FrameNet Brasil;
- Participações em eventos;
- Publicação de artigo com os resultados finais do projeto.

## 5 . Cronograma

As atividades previstas para o projeto serão executadas conforme o cronograma a seguir:

| ATIVIDADES | MESES | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 |
| Levantamento bibliográfico sobre dêixis narrativa multimodal | | | | | | | | | | | | |
| Cotejamento com o arcabouço teórico-metodológico da Semântica de Frames e da FrameNet | | | | | | | | | | | | |
| Constituição de corpus piloto multimodal para análise | | | | | | | | | | | | |
| Análise preliminar do corpus | | | | | | | | | | | | |
| Especificação de requisitos para implementação de novas categorias de análise na ferramenta da FrameNet Brasil | | | | | | | | | | | | |
| Participação em eventos | | | | | | | | | | | | |
| Publicação de artigo com os resultados finais do projeto | | | | | | | | | | | | |

## 6. Orçamento

O projeto insere-se no projeto estruturante Representação Semântica Computacional da Multimodalidade: aliando visão computacional e compreensão de língua natural através da Semântica de Frame, o qual foi financiado no âmbito da Chamada Universal CNPq nº 18/2021 – Faixa B – Grupos Consolidados. O financiamento é de R$ 68.000,00. O mesmo projeto estrutura rede de colaboração entre ICTs mineiras, a ReINVenTA – Research and Innovation Network for Text and Visual Analysis of Multimodal Objects –, a qual recebeu recentemente financiamento da FAPEMIG, no âmbito da Chamada nº 07/2021, da ordem de R$ 1.000.000,00.

## 7. Referências Bibliográficas


AKSOY, E. E. et al. Unsupervised linking of visual features to textual descriptions in long manipulation activities. *IEEE Robotics and Automation Letters*, v. 2, n. 3, p. 1397-1404, 2017.

BAKER, C. F. FrameNet: Frame Semantic Annotation in Practice. In: IDE, N. & PUSTEJOVSKY, J. *Handbook of Linguistic Annotation*. Dordrecht: Springer, 2017. p. 771-811.

BATEMAN, J.; WILDFEUER, J.; HIIPPALA, T. *Multimodality: foundations, research and analysis.* Berlin: De Gruyter, 2017.

BATRA, V.; HE, Y.; VOGIATZIS, G. Neural caption generation for news images. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC 2018). Miyazaki, 2018.

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

COHN, N. From Visual Narrative Grammar to Filmic Narrative Grammar: The narrative structure of static and moving images. In: WILDFEUER, J. & BATEMAN, J. *Film text analysis: New perspectives on the analysis of filmic meaning*, 2016, p. 94-117.

DEVLIN, J. et al. Language Models for Image Captioning: The Quirks and What Works. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Beijing, 2015. P. 100-105.

DUCHAN, J., BRUDER, G. A., HEWITT, L. E. (eds.). *Deixis in narrative: a cognitive science perspective.* London: Psychology Press, 1995.

FANG, H. et al. From captions to visual concepts and back. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Boston, 2015. p. 1473-1482.

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

LOSCHKY, L. C. et al. The scene perception & event comprehension theory (SPECT) applied to visual narratives. *Topics in cognitive science*, v. 12, n. 1, 2020. p. 311-351.

NIKOLAUS, M. et al. Compositional Generalization in Image Captioning. In: *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*. 2019. p. 87-98.

ROJO, R.; MOURA, E. *Multiletramentos na escola*. São Paulo: Parábola Editorial, 2012

_____. *Letramentos, mídias, linguagens*. São Paulo: Parábola, 2019.

SUN, C. et al. Videobert: A joint model for video and language representation learning. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019. p. 7464-7473.