PROPP - Pesquisa

Dados do Projeto e do(a) Coordenador do Projeto

Título do Projeto	REINVENTA - Research and Innovation network for Visual and Textual analysis of multimodal objects: Validação Psicolinguística do gold standart dataset – Fase II
Referência da Chamada:	( ) BIC/UFJF e PIBIC/CNPq ( ) PIBIC/CNPq AÇÕES AFIRMATIVAS (x) PROBIC/FAPEMIG ( ) PROBIC-JR/FAPEMIG ( ) Apoio ao Recém-Doutor ( ) Apoio a Grupos de Pesquisa ( ) Apoio à Instalação de Doutores ( ) Cadastro na Propesq
Coordenador do Projeto:	Aline Alves Fonseca
Equipe:	Tiago Timponi Torrent Marcelo Viridiano Ely Edison da Silva Matos Frederico Belcavello
Endereços para contato:	Eletrônico: aline.fonseca@ufjf.br Telefônico: (32) 998083525
Unidade/Departamento:	Faculdade de Letras / Departamento de Letras
Data:	16/10/2023

1. Justificativa/Caracterização do Problema

O projeto REINVENTA - Research and Innovation network for Visual and Textual analysis of multimodal objects, coordenado pelo professor Tiago Timponi Torrent e financiado pela FAPEMIG, é um grande projeto de rede de pesquisa multicentro que congrega pesquisadores de 4 renomadas instituições mineiras de ensino, a UFJF, a UFMG, a UFU e a PUC-Minas. O objetivo principal do projeto REINVENTA é implementar o modelo de cognição linguística da FrameNet no desenvolvimento de um dataset semanticamente anotado por humanos para modos comunicativos distintos, bem como modelos de Inteligência Artificial baseados nesse dataset. A partir deste objetivo geral, múltiplas frentes de trabalho estão operando em diferentes objetivos específicos que se coadunam. Dentre as atividades desenvolvidas pela equipe da UFJF está a tarefa de *[...] 2. Validar psicolinguisticamente a anotação semântica multimodal realizada para o gold standard dataset através de experimentos de eye-tracking, de modo a checar se as áreas anotadas na modalidade visual dos corpora concentra os pontos de fixação dos espectadores dos vídeos*. Esta tarefa é um dos objetivos específicos do projeto da REINVENTA, como consta em seu plano de trabalho (TERMO DE OUTORGA, p. 20-21) e é o objeto de investigação do presente projeto de Iniciação Científica.

Na Fase II deste projeto, continuaremos explorando o corpus Framed Multi30k dataset (TORRENT, et al. 2002) que consiste em um banco de imagens de mais de 30 mil itens com legendas e anotações em Inglês - corpus Flickr30k - (YOUNG et al., 2014) e legendas e anotações em Português criadas tanto a partir da tradução das legendas originais em Inglês, como a partir da anotação e legendagem originalmente feita em Português, realizada pelos anotadores da FrameNet-Br. A FrameNet-Br tem como base teórica a semântica de frames (FILLMORE, 1982) e sua implementação computacional. Segundo Viridiano (et al, 2002, p.108), *a ideia principal por trás da semântica de frames de Fillmore é que os seres humanos entendem o significado de uma expressão linguística a partir de um pano de fundo cognitivo de uma cena esquematizada, ou seja, um frame*. Para a semântica de frames, a noção de 'perspectiva' mostra-se de grande relevância. Viridiano (et al, 2002) cita um exemplo clássico de Fillmore que se refere à distinção entre os lexemas *coast.n* e *shore.n* em Inglês; as duas palavras se referem ao frame Relational_natural_features e descrevem o encontro entre a massa de terra dos continentes e o mar, mas *coast.n* pode ser entendido como a descrição do ponto de vista da terra, enquanto *shore.n* leva em consideração o ponto de vista do mar. No processo de anotação multimodal do corpus Framed Multi30k dataset, realizado pela FrameNet-Br, em que imagens e textos estão associados, os anotadores criam *bounding boxes* que relacionam partes das imagens com palavras das legendas descritivas a partir de entidades, considerando a relevância dos elementos para a descrição da cena e a perspectiva do próprio anotador. Uma das aplicações dos corpora anotados a partir da associação entre imagens e textos é o treinamento de Inteligências Artificiais (AI) para a automatização de atividades, como a visão assistida a partir de áudio-descrição de imagens, entre outras. Para a efetividade dos treinamentos de AI, a relação texto-imagem deve ser precisa e a relevância da descrição deve estar associada à perspectiva prevalecente do espectador/usuário.

Neste ponto, destacamos a importância da análise Psicolinguística do processo de anotação e legendagem das imagens do corpus Framed Multi30k dataset e, posteriormente da validação com usuários/experienciadores. A Psicolinguística é o estudo das representações mentais e dos processos que envolvem o uso da linguagem, incluindo a produção, a percepção, a compreensão e o armazenamento da língua falada e escrita (WARREN, 2013. p. 4). Os campos de atuação da Psicolinguística vão desde a maneira como reconhecemos e armazenamos palavras em nossa mente, até como processamos e atribuímos significado a textos durante a leitura, passando também pela percepção e produção da fala, e pelos processos de associação entre sons, textos e imagens (FONSECA E MAIA, 2022. p. 60). A Psicolinguística se vale, em grande parte, de hipóteses sobre o

funcionamento da nossa mente em relação à linguagem e busca evidências empíricas, por meio de tarefas experimentais, para confirmar ou refutar tais hipóteses. A chegada das tecnologias que permitiram o rastreamento ocular foi um marco nas atividades experimentais da Psicolinguística, uma vez que possibilitou a medição mais precisa de respostas inconscientes dos participantes, o que espelha de forma mais fiel o funcionamento da nossa mente em relação ao processamento da linguagem, e a partir de tarefas consideradas mais naturais, sem grande manipulação por parte dos pesquisadores. Os rastreadores oculares atuais são equipamentos que monitoram os movimentos dos olhos através de um sistema de vídeo com emissão de luz infravermelha. Os olhos são localizados pelo equipamento através da detecção da pigmentação da pupila e/ ou da reflexão da córnea. Durante o monitoramento do olhar, o equipamento de rastreamento ocular mede as sacadas e as fixações. Durante a leitura, ou durante a visualização de imagens, por exemplo, nossos olhos realizam pequenos saltos de cerca de 20 a 35 milissegundos, chamados de **sacadas**. Entre uma sacada e outra, pousamos o olhar em um determinado ponto do texto ou da imagem por cerca de 200 a 250 milissegundos, o que chamamos de **fixação** (RAYNER; POLLATSEK, 1989). Diferentemente do que parece, não lemos "correndo os olhos" por todas as letras e todas as palavras de um texto. Também não reconhecemos e exploramos uma imagem com um "olhar único" que abrange toda a cena em questão. Na verdade, nossa visão se dá a partir das fixações que são separadas pelos saltos das sacadas. Durante as fixações, nosso sistema cognitivo processa a informação visual e planeja o momento e a distância da próxima sacada. Em contrapartida, durante os movimentos sacádicos, os olhos estão se movendo tão rapidamente que é quase impossível obter informação visual neste momento (RAYNER, 2009). Durante um experimento, seja com imagens ou texto, podemos medir a quantidade e a duração das fixações, além da direção das sacadas, que podem ser progressivas ou regressivas. Normalmente, usamos os termos em Inglês para essas medidas: *fixation count* (número de fixações), *first fixation duration* (duração da primeira fixação), *total fixation duration* (tempo total de fixação), *first path duration* (tempo da primeira passada), *regression path duration* (tempo da passada regressiva) e proportion of looks (proporção de olhares no tempo), além de outras. A técnica experimental, em Psicolinguística, que correlaciona o mundo visual à linguagem falada ou escrita é chamada de Paradigma do Mundo Visual (Visual Word Paradigm). Parte-se do princípio de que as medidas realizadas a partir do rastreamento ocular em imagens associadas a estímulos auditivos ou escritos podem revelar como as representações linguísticas tornam-se ativas na mente do participante em um determinado momento. Essa relação entre o processamento da linguagem falada/escrita e estímulos visuais é chamada por Tanenhaus e Trueswell (2006) de medida representacional.

Na fase II deste projeto de iniciação científica, continuaremos adotando a metodologia experimental de Paradigma do mundo visual, com rastreamento ocular, em partes previamente selecionadas a partir de categorias/entidades comuns, do corpus Framed Multi30k dataset, criando estímulos que vão explorar a relação entre imagem, anotação e legendas, do ponto de vista do processamento da linguagem visual e textual. Temos o intuito de explorar elementos secundários das imagens, a partir de um incremento nas legendagens que, por hipótese, irão direcionar o olhar dos participantes para pontos específicos da imagem. Com este novo método experimental, pretendemos verificar a influência direta da linguagem verbal no reconhecimento visual das imagens associadas. Os experimentos serão aplicados inicialmente apenas com os anotadores da FrameNet-Br que são alunos de graduação e pós-graduação, bolsistas e voluntários, que trabalham no laboratório da FrameNet-Br, na Faculdade de Letras da UFJF.

O uso de rastreadores oculares tem trazido grandes avanços não só para as pesquisas em Linguística Teórica, mas também para áreas relacionadas como a Linguística Computacional. Três aspectos positivos sobre o uso de rastreadores oculares como instrumento de investigação científica merecem destaque: (a) a versatilidade da ferramenta que pode ser usada com diferentes metodologias experimentais e para diferentes finalidades; (b) a naturalidade dos experimentos que captam os movimentos naturais e espontâneos que qualquer ser humano faz para ler ou observar cenas; e (c) a

grande precisão espaço-temporal de suas medidas. As atividades experimentais aplicadas com a ajuda de rastreadores oculares possibilitam que os estímulos linguísticos sejam utilizados de forma integral, sem cortes ou manipulações que os tornem antinaturais, e ainda, sem que seja necessário fazer uso de julgamentos metalinguísticos por parte dos participantes. Isso faz com que os resultados encontrados possam ser mais fielmente associados a fatores automáticos do nosso processamento mental da linguagem.

2. Objetivos

O objetivo geral deste projeto é atender a etapa de validação psicolinguística do gold standart dataset do projeto de rede REINVENTA - Research and Innovation network for Visual and Textual analysis of multimodal objects, financiado pela FAPEMIG. Como objetivos específicos, podemos citar, principalmente:

I. Investigar a relação representacional entre linguagem escrita e visual nas legendas e imagens de parte previamente selecionada do corpus Framed Multi30k dataset, a partir da associação entre as teorias de Semântica de Frames e Psicolinguística;
II. Contribuir para o treinamento dos anotadores e, consequentemente, para a maior acurácia na relação texto imagem dos itens que compõem o dataset do projeto REINVENTA;
III. Ajudar na criação de objetos anotados com maior relevância do ponto de vista do espectador, o que gera elementos mais precisos para os treinamentos de Inteligência Artificial;
IV. Contribuir para a produção de materiais de boas práticas de autodescrição semanticamente orientada.

3. Metodologia e Estratégias de Ação

Realizaremos testes de leitura e testes de paradigma do mundo visual (Tanenhaus & Trueswell, 2006) com rastreamento ocular de objetos selecionados do corpus Framed Multi30k dataset. Nos testes de paradigmas do mundo visual, estímulos linguísticos são apresentados aos participantes junto com imagens projetadas em uma tela de computador que estão associadas a questões de interpretação. No âmbito deste projeto de iniciação científica, participarão das atividades experimentais apenas os anotadores que serão voluntariamente recrutados. As tarefas experimentais não diferem das tarefas de anotação e observação dos dados já comumente realizadas pelos voluntários. A diferença se dá apenas pela presença do equipamento Tobii TX300, pertencente ao laboratório do Núcleo de estudos em aquisição da Linguagem e Psicolinguística (NEALP) da UFJF, que será utilizado nos processos que captação dos movimentos oculares durante as atividades dos anotadores. Mediremos a movimento ocular dos participantes sobre as imagens e sobre as legendas. Tais medidas são feitas em termos de sacadas e fixações (Rayner e Pollatsek, 1989). As medidas coletadas são analisadas estatisticamente com testes de regressão linear e logística de efeitos mistos. Utilizamos o software R Studio (R CORE TEAM, 2020) para realizar as análises estatísticas.

4. Resultados e os impactos esperados

A proposta ora submetida prevê resultados em termos de desenvolvimento científico-tecnológico e formação de recursos humanos. No que se refere desenvolvimento científico-tecnológico, esperamos contribuir com a criação de um gold standard dataset de objetos multimodais em português brasileiro

validado psicolinguisticamente. A precisão deste gold standart dataset é essencial para a produção de modelo(s) de IA treinado(s) e de software(s) off-the-shelf que podem ser utilizados por outros grupos e projetos debruçados sobre a compreensão de objetos multimodais por máquinas. No que se refere à formação de recursos humanos, a iniciação científica é um caminho de formação em nível superior que transfere conhecimento e prática para além da teoria científica pura. O projeto de iniciação científica que ora se apresenta é parte integrante de um grande projeto de rede de pesquisa, o que proporcionará ao bolsista selecionado o contato com pesquisadores, professores, graduandos e pós-graduandos da UFJF e de outras instituições de ensino superior parceiras, estreitando laços de cooperação, expandindo o conhecimento e enriquecendo as experiências adquiridas para além da UFJF.

5. Cronograma

Atividades	Jan 2024	Fev-mar 2024	Abr- jun 2024	Jul-set 2024	Out-dez 2024
Seleção do bolsista	x				
Revisão bibliográfica	x	x	x		
Seleção dos itens do corpus			x		
Montagem do experimento no rastreador ocular			x	x	
Aplicação do experimento de validação psicolinguística com os anotadores da FrameNet-Br				x	
Compilação dos resultados e análise estatística				x	
Discussão e divulgação dos resultados				x	x
Escrita do relatório					x

6. Orçamento

A Faculdade de Letras da UFJF, unidade acadêmica que abriga os laboratórios da FrameNet-Br e do NEALP (Núcleo de Estudos em Aquisição da Linguagem e Psicolinguística), possui toda a infraestrutura necessária para a execução deste projeto de iniciação científica, como computadores do tipo PC com acessórios (monitor, teclado e mouse), fones de ouvido do tipo headfones, gravadores de áudio, o aparelho de Eye-tracker modeloTobii TX300, softwares de anotação, de tratamento de imagens, e de montagem de experimentos de rastreamento ocular, entre outros. O orçamento deste projeto se resume, então, a materiais de consumo como impressão e fotocópias de materiais que serão financiados com recursos da Unidade.

7. Referências Bibliográficas

FILLMORE, C. J. (1982). Frame Semantics. In Linguistics Society of Korea, editor, Linguistics in the morning calm. Hanshin Publishing Co., Seoul, South Korea. pages: 111 – 138.

FONSECA, A. A.; MAIA, M. (2022) Na trilha do processamento da linguagem: o uso de rastreadores oculares na análise de dados linguísticos. In: OLIVEIRA, C. S. F.; SÁ, T. M. M. (Orgs.) Métodos experimentais em psicolinguística [recurso eletrônico]. - 1. ed. - São Paulo: Pá de Palavra. p. 55-74.

R CORE TEAM. (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RAYNER, K. (1998) Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124(3). p. 372-422. doi:10.1037/0033-2909.124.3.372

RAYNER, K. (2009) Eye movements and attention in reading, scene perception, and visual search. Quarterly Journal of Experimental Psychology, 62(8). p. 1457-1506.

RAYNER, K.; POLLATSEK, A. (1989) The psychology of Reading. Englewood Cliffs, NJ: PrenticeHall.

TANENHAUS, M. K.; TRUESWELL, J. C. (2006) Eye movements and spoken language comprehension. In: TRAXLER, M. J.; GERNSBACHER, M. A (ed.), Handbook of Psycholinguistics (2nd edition). Amsterdam: Elsevier. p. 863-900.

TORRENT, T. T., MATOS, E. E. D. S., BELCAVELLO, F., VIRIDIANO, M., GAMONAL, M. A., COSTA, A. D. D., AND MARIM, M. C. (2022). Representing context in framenet: A multidimensional, multimodal approach. Frontiers in Psychology, 13.

VIRIDIANO, M., TORRENT, T. T., CZULO, O., ALMEIDA, A. L., MATOS, E. E. S., BELCAVELLO, F. (2022) The Case for Perspective in Multimodal Datasets. *Proceedings of the 1st Workshop on Perspectivist Approaches to NLP @LREC2022*. p. 108–116.

YOUNG, P., LAI, A., HODOSH, M., AND HOCKENMAIER, J. (2014). From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. Transactions of the Association for Computational Linguistics, 2:67–78.

WARREN, P. (2013) Introducing Psycholinguistics. Edinburgh: Cambridge University press.