# 2<sup>nd</sup> Edition of the
# Universal Dependencies Brazilian Festival

# Proceedings of the Conference, Vol. 1

September 25, 2023

## About the workshop

The 2nd Edition of the Universal Dependencies Brazilian Festival (UDFest-BR 2023) takes place with the 14th Symposium in Information and Human Language Technology (STIL 2023). It is a forum in which researchers involved with the study and application of the Universal Dependencies (UD) model, along with its adaptation to Portuguese, can meet and discuss best practices and strategies, as well as problems and solutions related to this topic.

UD is a cross-language international proposal for grammatical annotation (morphosyntactic tags, lexical/morphosyntactic features and syntactic dependencies) that resulted from the efforts of an open community of hundreds of researchers. Building upon a common framework, the model captures idiosyncrasies and similarities between different languages, fostering contrastive linguistic studies, along with the development of multilingual technologies for Natural Language Processing (NLP). In recent years, this model has drawn considerable attention from the NLP community. Currently, there are nearly 200 UD annotated corpora available in more than 100 languages.

Although there are some annotated corpora and lexical resources for Portuguese, computational linguistic research involving UD and this language is still limited. As such, this workshop helps to fill in this gap, giving more visibility to current efforts in this area and fostering new initiatives and collaborations.

This edition of the workshop includes ten papers. There are theoretical and practical discussions, reports on corpus annotation and the related challenges, and software for UD-related processing. The papers and their authors are:

- *Lexical noun phrase chunking with Universal Dependencies for Portuguese* (Aleksander Tomaz de Souza, Evandro Eduardo Seron Ruiz)

- *Construções sintáticas do português que desafiam a tarefa de parsing: uma análise qualitativa* (Magali Sanches Duran, Maria das Graças Volpe Nunes, Thiago A. S. Pardo)

- *Annotation of fixed Multiword Expressions (MWEs) in a Portuguese Universal Dependencies (UD) treebank: Gathering candidates from three different sources* (Elvis de Souza, Cláudia Freitas)

- *Verifica-UD: a Verifier for Universal Dependencies Annotation for Portuguese* (Lucelene Lopes, Magali Sanches Duran, Thiago Alexandre Salgueiro Pardo)

- *Enhanced dependencies para o português brasileiro* (Adriana S. Pagano, Magali Sanches Duran, Thiago Alexandre Salgueiro Pardo)

- *A dependency-based study of medicine package inserts in Brazilian Portuguese* (Adriana S. Pagano, André V. Lopes Coneglian, Lucas Emanuel Silva e Oliveira)

- *Um estudo das construções dar + para + V [infinitivo] nas Universal Dependencies* (Marcella M. Lemos Couto, Oto Araújo Vale)

- *Insights into the UD Tagset: Unveiling its Intricacies* (Magali Sanches Duran)

- *Em Direção à Anotação Sintática - UD de Tweets do Mercado Financeiro* (Bryan K. S. Barbosa, Ariani Di Felippo)

- *Universal Dependencies and Language Contact Annotation: Experience from Warao refugees signs in Brazil* (Dalmo Buzato)

## Organizing committee

This workshop is promoted by the POeTiSA project, which is a long term initiative that aims at growing syntax-based resources and developing related tools and applications for Brazilian Portuguese language, looking to achieve world state-of-the-art results in this area. The project is part of the Natural Language Processing initiative (NLP2) of the Center for Artificial Intelligence (C4AI) of the University of São Paulo, sponsored by IBM and FAPESP. The workshop is also supported by the Ministry of Science, Technology and Innovation, according Law N. 8,248, of October 23, 1991.

The following researchers organized this edition of the workshop:

Thiago Alexandre Salgueiro Pardo, Magali Sanches Duran, Lucelene Lopes
*Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo*
São Carlos/SP, Brazil

## Program committee

Adriana Silvina Pagano (UFMG)
Alexandre Rademaker (IBM)
Ariani Di Felippo (UFSCar)
Cláudia Freitas (PUC-Rio)
Joakim Nivre (Uppsala University)
Jorge Baptista (*Universidade do Algarve*)
Leonardo Zilio (UFRGS)
Maria das Graças Volpe Nunes (USP)
Norton Trevisan Roman (USP)
Roana Rodrigues (UFS)
Valeria de Paiva (Topos Institute)