

STIL 2023

14th Symposium in Information and Human
Language Technology

Proceedings of the Conference, Vol. 1

December 29, 2023

About the workshop

The Proceedings of the XIV Brazilian Symposium on Information Technology and Human Language (STIL 2023) present the selected papers presented at the event held from September 25 to 29, 2023, in the city of Belo Horizonte, MG, in conjunction with the XII Brazilian Conference on Intelligent Systems (BRACIS 2023), XX National Meeting on Artificial and Computational Intelligence (ENIAC 2023), and VI Brazilian Competition on Knowledge Discovery In Databases (KDD-BR 2023). In this edition, the proceedings compile the selected papers for the Main Conference and the works accepted in its satellite events:

- XIV Brazilian Symposium on Information Technology and Human Language (STIL 2023) received a total of 73 paper submissions, of which 25 were accepted, resulting in an acceptance rate of 34.2%. These works were selected through a double-blind peer review process;
- VIII Jornada de Descrição do Português (JDP 2023) received a total of 21 paper submissions, of which 12 were accepted, resulting in an acceptance rate of 57.1%. These works were selected through a double-blind peer review process;
- VIII Workshop on Scientific Initiation in Information Technology and Human Language (TILic 2023) received a total of 15 paper submissions, of which 13 were accepted, resulting in an acceptance rate of 86.6%. These works were selected through a double-blind peer review process.

Acknowledgments

The Program Committee chairs of STIL and collocated events acknowledge the support to the conference provided by the Brazilian Computer Society (SBC), and the North American Chapter of Association for Computational Linguistics (NAACL). We thank the Program Committees of the XIV Brazilian Symposium in Information and Human Language Technology and Collocated Events for the reviews that they produced. Last but not least, we are grateful to the local organization led by Gisele Pappa (UFMG) and Wagner Meira Jr (UFMG), general chairs of BRACIS 2023.

December 2023

Helena de Medeiros Caseli (UFSCar, Brazil)
Maria José Bocorny Finatto (UFRGS, Brazil)

Program chairs

- **STIL**

Helena Caseli (UFSCar, Brazil)

Maria José Bocorny Finatto (UFRGS, Brazil)

- **JDP**

Cláudia Freitas (ICMC/USP)

Gabriela Wick Pedro (IBICT/NILC)

- **TILIC**

Roney Lira de Sales Santos (UFBA)

Raimundo Santos Moura (UFPI)

Rafael T. Anchiêta (IFPI)

Program Comittee

Adriana Pagano (UFMG)

Adriana Silvina Pagano (UFMG)

Alexandre Rademaker (IBM)

Aline Evers (UFRGS)

Amanda Rassi (Red. Nota 1000)

Ariani Di Felippo (UFSCar)

Arnaldo Candido Junior (UTFPR)

Celso Ferrarezi (UNIFAL-MG)

Clarissa Xavier (SiDi)

Cláudia Freitas (PUC-Rio)

Daniela Claro (UFBA)

Diana Santos (Univ. of Oslo)

Eloize Rossi Marques Seno (IFSP)

Eric Laporte (Uni. Gustave Eiffel)

Evandro Fonseca (Take Blip)

Guilherme Fromm (UFU)

Heliana Mello (UFMG)

Jackson Souza (UFBA)

Joakim Nivre (Uppsala University)

Jorge Baptista (Univ. do Algarve)

Juliano Desiderato Antonio (UEM)

Larissa Freitas (UFPEL)

Leonardo Zilio (UFRGS)

Lucene Lopes (USP)

Magali Duran (USP/ICMC)

Magnun Madruga (UFMG)

Marcela Couto (UFMG)

Marcio Inácio (Univ. de Coimbra)

Marcos Lopes (USP)

Maria Cantoni (UFMG)

Maria G. Volpe Nunes (USP)

Maria José Finatto (UFRGS)

Marlo Souza (UFBA)

Nathan Hartmann (Itaú Unibanco)

Norton Trevisan Roman (USP)

Oto Vale (UFSCar)

Paula C. Figueira Cardoso (UFLA)

Raquel Freitag (UFS)

Roana Rodrigues (UFS)

Rogério Sousa (IFPI)

Stella Tagnin (USP)

Thiago Pardo (USP/ICMC)

Tiago Timponi Torrent (UFJF)

Valeria de Paiva (Topos Institute)

Proceedings organization

Jackson Wilke da Cruz Souza (UFBA, Brazil)

Helena de Medeiros Caseli (UFSCar, Brazil)

Maria José Bocorny Finatto (UFRGS, Brazil)

Contents

STIL

Less is More? Investigating Meta-Learning's Suitability in Sentence Compression for Low-Resource Data

Luis Gustavo C. Rêgo., José Antônio F. de Macêdo, Ticiane L. Coelho da Silva..... 10-19

Predição de transtorno depressivo em redes sociais: BERT supervisionado ou ChatGPT zero-shot?

Wesley Ramos dos Santos, Ivandré Paraboni..... 20-30

Semantic Textual Similarity: In Defense of Wordnet-Based Methods

Eduardo Corrêa Gonçalves..... 31-40

Contextual stance classification using prompt engineering

Felipe Penhorate Carvalho de Fonseca, Ivandré Paraboni, Luciano Antonio Digiampietri..... 41-51

Classificação de gêneros a partir de letras de músicas em português

Matheus Bastos de Oliveira, João Baptista de Oliveira e Souza Filho..... 52-61

Sexismo no Brasil: análise de um Word Embedding por meio de testes baseados em associação implícita

Fernanda Tiemi de S. Taso, Valéria Quadros dos Reis, Fábio Viduani Martinez..... 62-71

Etiquetagem morfossintática multigênero para o português do Brasil segundo o modelo "Universal Dependencies"

Emanuel Huber Silva, Thiago Alexandre Salgueiro Pardo, Norton Trevisan Roman..... 72-82

Automated question answering via natural language sentence similarity: Achievements for Brazilian e-commerce platforms

Víctor Jesús Sotelo Chico, Luiz Zucchi, Daniel Ferragut, Rodrigo Caus, Victor Hochgreb de Freitas, Julio Cesar dos Reis..... 83-92

Classificação de Polaridade Orientada aos Alvos de Opinião em Comentários sobre Debate Político em Português

Eloize R. Marques Seno, Fábio S. Igarashi Anno, Lucas Lazarini, Helena M. Caseli..... 93-102

How Good Is ChatGPT For Detecting Hate Speech In Portuguese?

Amanda S. Oliveira, Thiago C. Cecote, Pedro H. L. Silva, Jadson C. Gertrudes, Vander L. S. Freitas, Eduardo J. S. Luz..... 103-112

When Tweets Get Viral - A Deep Learning Approach for Stance Analysis of Covid-19 Vaccines Tweets by Brazilian Political Elites

Lorena Guadalupe Barberia, Pedro Henrique de Santana Schmalz, Norton Trevisan Roman..... 113-123

The Dawn of the Porttinari Multigenre Treebank: Introducing its Journalistic Portion <i>Magali S. Duran, Lucelene Lopes, Maria das Graças V. Nunes, Thiago A. S. Pardo</i>	124-133
Explorando variações no tagset e na anotação Universal Dependencies (UD) para Português: Possibilidades e resultados com base no treebank PetroGold <i>Elvis de Souza, Cláudia Freitas</i>	134-143
Yauti: A Tool for Morphosyntactic Analysis of Nheengatu within the Universal Dependencies Framework <i>Leonel Figueiredo de Alencar</i>	144-154
Albertina in Action: An Investigation of its Abilities in Aspect Extraction, Hate Speech Detection, Irony Detection, and Question-Answering <i>Júlia da Rocha Junqueira, Claudio Luis Junior, Félix Leonel V. Silva, Ulisses Brisolara Côrrea, Larissa A. de Freitas</i>	155-164
Studying the Dependence of Embedding Representations on the Target of NLP Tasks <i>Bárbara Stéphanie Neves Oliveira, Ticiane L. Coelho da Silva, José A. F. de Macêdo</i>	165-175
Viés de gênero na tradução automática do GPT-3.5 turbo: avaliando o par linguístico inglês-português <i>Tayane Arantes Soares, Yohan Bonescki Gumiel, Rafael Junqueira, Tácio Gomes, Adriana Pagano</i>	176-185
CDJUR-BR - Uma Coleção Dourada do Judiciário Brasileiro com Entidades Nomeadas Refinadas <i>Maurício Brito, Vlândia Pinheiro, Vasco Furtado, João Araújo Monteiro Neto, Francisco das Chagas Jucá Bomfim, André Câmara Ferreira da Costa, Raquel Silveira</i>	186-195
A call for a research agenda on fair NLP for Portuguese <i>Luiz Fernando F. P. de Lima, Renata Mendes de Araujo</i>	196-201
Avaliação do senso comum em modelos de linguagem através de benchmarks: Desafio de Winograd aplicado ao ChatGPT em português brasileiro <i>Thiago Gomes do Nascimento, Diogo Cortiz</i>	202-207
A Sentiment Analysis Benchmark for Automated Machine Learning Applications and a Proof of Concept in Hate Speech Detection <i>Marília Costa Rosendo Silva, Vitor Augusto de Oliveira, Thiago Alexandre Salgueiro Pardo</i>	208-215
Proposta e Avaliação Linguística de Técnicas de Aumento de Dados <i>Arthur Scalercio, Cláudia Freitas</i>	216-232
Semantic Textual Similarity for Abridging Clinical Notes in Brazilian Electronic Health Records <i>Lucas T. Bandeira, Bernardo S. Consoli, Renata Vieira, Rafael H. Bordin</i>	233-237

Towards analysis on textual inference at ASSIN-2 dataset

Felipe O. da Silva, Giovana M. Craveiro, Vinícius F. da Silva, Vinícius João B. Vanzin.....238-243

Previsão de Utilidade de Avaliações de Produtos Online na Língua Portuguesa Brasileira

Larissa F. S. Britto, Luciano D. S. Pacífico, Teresa B. Ludermit.....244-248

JORNADA DE DESCRIÇÃO DO PORTUGUÊS

Tipologia de fenômenos ortográficos e lexicais em CGU: o caso dos tweets do mercado financeiro

Clarissa Lenina Scandarolli, Ariani Di Felippo, Norton Trevisan Roman, Thiago A. S. Pardo.....249-257

Sinalizadores retórico-discursivos: revisitando a anotação RST no cópua CSTNews

Roana Rodrigues, Jackson Wilke da Cruz Souza, Paula Christina Figueira Cardoso.....258-266

Complexidade textual em narrativas orais produzidas por informantes de diferentes níveis de escolaridade

Juliano Desiderato Antonio.....267-276

Aposições anafóricas e catafóricas no português e sua anotação no esquema Universal Dependencies

Magali Sanches Duran, Maria das Graças Volpe Nunes.....277-286

Um pronome com muitas funções: Descrição e resultados da anotação do pronome -se em um treebank segundo o esquema Universal Dependencies (UD) para Português

Elvis de Souza, Cláudia Freitas.....287-296

A funcionalidade dos adjetivos em dois gêneros discursivos: uma investigação com base nas dependências universais

Andre V. Lopes Coneglian, Adriana Pagano, Carlos Perini.....297-306

Indução Gramatical para o Português: a Contribuição da Informação Mútua para Descoberta de Relações de Dependência

Diego Pedro Gonçalves da Silva, Thiago Alexandre Salgueiro Pardo.....307-316

Características gerais das leis federais brasileiras: um estudo exploratório sob o enfoque da Análise Multidimensional

Carolina Godoi de Faria Marques, Lúcia de Almeida Ferrari.....317-326

Processamento da Linguagem Natural e Complexidade Sintática

Bruna R. da Silva.....327-336

Formal features for a syntactic-semantic classification of predicative adjectives in Brazilian Portuguese

Ryan Marçal Saldanha Magaña Martinez, Oto Araújo Vale.....337-346

Building a Frame-Semantic Model of the Healthcare Domain: Towards the identification of gender-based violence in public health data

Lívia Dutra, Arthur Lorenzi, Lorena Larré, Frederico Belcavello, Ely Matos, Amanda Pestana, Kenneth Brown, Mariana Gonçalves, Victor Herbst, Sofia Reinach, Renato Teixeira, Pedro de Paula, Alessandra Pellini, Cibele Sequeira, Ester Sabino, Fábio Leal, Mônica Conde, Regina Grespan, Tiago Torrent.....347-355

WORKSHOP DE IC EM TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA

Gramáticas Locais para Reconhecimento de Construções com Verbo Suporte em Português

Luis Enrique Santos Prado Vereau, Juliana Pinheiro Campos Pirovani.....356-360

Anotação do Dataset Multimodal da ReINVenTA

Ana Carolina Loçasso Luz, Gabrielly Braz, Lívia Pádua Ruiz, Mariane de Carvalho Pinto, Frederico Belcavello, Natália Sathler Sigiliano, Tiago Torrent.....361-365

Pipeline para identificação de erros lexicais e geração de sugestões de correção

Luana Q. Garcia, Miguel H. Chinellato, Helena de M. Caseli, Leandro H. M. Oliveira.....366-370

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

Anna B. C. Silva, Iasmin Rabelo, Igor M. Oliveira, Mariana Souza, Maucha Gamonal, Raquel Roza.....371-375

Abordagens Baseadas em Léxicos para a Classificação de Sentimentos Orientada aos Alvos de Opinião em Comentários do Domínio Político

Lucas Lazarini, Fábio S. Igarashi Anno, Eloize R. Marques Seno, Helena M. Caseli.....376-381

Aryon: um aplicativo Shiny para documentação e análise de línguas indígenas brasileiras

Mateus Zaparoli, Katiuska Rowe, Magnun Rochel Madruga.....382-385

Desambiguação dos termos do Atlas Linguístico do Brasil através da OpenWordnet-PT-ALiB

Augusto Sampaio Barreto, Daniela Barreiro Claro.....386-390

TransAlign: tradução e alinhamento de corpora para a língua portuguesa

Alan Rios Melo, Daniela Barreiro Claro.....391-396

Desafios da tarefa de Extração de Informação Aberta: uma abordagem metodológica de um corpus automatizado até o corpus manual

Beatriz Paixão Queiroz, Rerisson Cavalcante, Daniela Barreiro Claro.....397-401

Técnicas de sumarização de textos jurídicos para suporte à classificação de documentos de decisões judiciais

Hellen Harada, Fabíola Pereira, Alex Almeida, Daniela Freire, Márcio Dias, Nádia Silva, Pedro Andrade, André Carvalho.....402-406

Uso de modelagem de tópicos para agrupamento de notícias: uma abordagem usando BERTopic

Pedro Henrique Pereira, Ticiano Linhares Coelho da Silva.....407-411

Proposta de avaliação da percepção dos impactos da inteligência artificial generativa na educação superior

Ana Luíza Ferreira Vieira, Maria Cecília Zanon De Amorim, Evandro Cunha.....412-416

Explorando Redes Neurais Profundas para Tarefa de Aceitabilidade Linguística

Henrique Santos, Késia P. Alencar, Rogério F. de Sousa, Rafael T. Anchiêta.....417-422