

Francielle Vargas
francielleavargas@usp.br
<https://francielleavargas.github.io/>

Education

PhD, Computer Science and Computational Mathematics
University of São Paulo
2019 - 2024

MSc, Computer Science and Computational Mathematics
University of São Paulo
Thesis: Semantic Clustering of Aspects for Opinion Mining [see](#)
2015 - 2017

BA, Linguistics
Federal University of Minas Gerais

BSc, Information Systems
Pontifical Catholic University of Minas Gerais

Keynote Speaker *Leibniz Institute for the Social Sciences (GESIS), Cologne, Germany*
[Conference on Harmful Online Communication \(CHOC2023\)](#), 2023.

Organizing and Program Committee

Organizing Committee

1. Data(set) Co-chair. *ICWSM 2023*. [see](#)
2. Accessibility Chair. *ICWSM 2022*. [see](#)
3. Diversity, Equity, and Inclusion Chair. *ICWSM 2021*. [see](#)
4. Steering Committee. 6th CASE@*RANLP 2023*. [see](#)
5. Steering Committee. 5th CASE@*EMNLP 2022*. [see](#)

Program Committee

- *ACL*
Computational Social Science and Cultural Analytics Track ([2023](#))
Discourse and Pragmatics Track ([2023](#))
- *EMNLP*
Computational Social Science and Cultural Analytics Track ([2022](#)) ([2023](#))
Discourse and Pragmatics Track ([2022](#)) ([2023](#))
- *LREC*
Offensive and Harmful Language Detection and Analysis Track ([2024](#))
Discourse and Pragmatics Track ([2024](#))
- *NAACL*
Computational Social Science and Cultural Analytics Track([2024](#))
Discourse and Pragmatics Track ([2024](#))
NAACL Student Research Workshop ([2022](#)) ([2024](#))
- *ICWSM*
Dataset and Demo Track ([2025](#))([2024](#))([2023](#))

- *CIKM*
Dataset and Demo Track ([2024](#))([2023](#))
- *HICSS*
Digital and Social Media Track ([2023](#))
- *WOAH*
Workshop on Online Abuse and Harms ([NAACL 2024](#))([ACL 2023](#)) ([NAACL-HLT 2022](#))
- *FEVER*
6th Fact Extraction and VERification ([EACL 2023](#))
- *CODI*
Workshop on Computational Approaches to Discourse ([EACL 2024](#))([ACL 2023](#)) ([COLING 2022](#))
- *ARGMINING-CMNA*
Workshop on Argument Mining and Workshop on Computational Models of Natural Argument ([ICLP 2023](#)) ([COLING 2022](#)) ([COMMA 2022](#))
- *SoCal NLP*
5th Southern California Natural Language Processing Symposium ([2022](#))

Awards & Honors	2024: <i>NAACL 2024 DI Award</i> , NAACL 2024 2024: <i>Latin America Research Awards (LARA)</i> , Google 2013: <i>Outstanding Academic Achievement</i> , UFMG. 2012: <i>Outstanding Academic Achievement</i> , UFMG.
--------------------------------	---

Publications	2024
---------------------	-------------

HausaHate: An Expert Annotated Corpus for Hausa Hate Speech Detection

Vargas, F., Guimarães, S., Alves, D., Muhammad, H. S., Abdulmumin, I., Ahmad, I. S., Mohamed, D., Pardo, T.A.S., Benevenuto, F.
8th Workshop on Online Abuse and Harms (NAACL), pp. 1-6. Mexico City, Mexico.
[accepted](#)

Context-Aware and Expert Data Resources for Brazilian Portuguese Hate Speech Detection

Vargas, F., Carvalho, I., Pardo, T.A.S., Benevenuto, F.
Natural Language Engineering Journal. pp.1-21. [accepted](#)

Discourse Annotation Guideline for Low-Resource Languages

Vargas, F., Schmeisser-Nieto, W., Rabinovich, Z., Benevenuto, F., Pardo, T.A.S.
Natural Language Engineering Journal. pp.1-33. [accepted](#)

Extended Multimodal Hate Speech Event Detection During Russia-Ukraine Crisis

Thapa, S, Rauniyar, K., Jafri, F. A., Veeramani, H., Jain, R., Jain, S., Vargas, F., Hürriyetoglu, A., Naseem, U.
7th International Workshop Challenges and Applications of Automated Extraction of Socio-Political Events from Text (EACL). pp.221–228. St. Julians, Malta. [see](#)

2023

Socially Responsible Hate Speech Detection: Can Classifiers Reflect Social Stereotypes?

Vargas, F., Carvalho, I., Hürriyetoglu, A., Pardo, T.A.S., Benevenuto, F.
Recent Advances in Natural Language Processing (RANLP). pp.1-10. Varna, Bulgaria. [see](#)

Predicting Sentence-Level Factuality of News and Bias of Media Outlets

Vargas, F., Jaidka, K., Pardo, T.A.S., Benevenuto, F.
Recent Advances in Natural Language Processing (RANLP). pp.1-10. Varna, Bulgaria. [see](#)

NoHateBrazil: A Brazilian Portuguese Text Offensiveness Analysis System

Vargas, F., Carvalho, I., Schmeisser-Nieto, W., Benevenuto, F., Pardo, T.A.S.
Recent Advances in Natural Language Processing (RANLP). pp.1-7. Varna, Bulgaria. [see](#)

Multimodal Hate Speech Detection

Thapa, S., Jafr, F. A., Hürriyetoglu, A., Vargas, F., Lee, R. K., Naseem, U.
6th International Workshop Challenges and Applications of Automated Extraction of Socio-Political Events from Text (RANLP). pp.1-9. Varna, Bulgaria. [see](#)

2022

Rhetorical Structure Approach for Online Deception Detection: A Survey

Vargas, F., D'Alessandro, J., Rabinovich, Z., Benevenuto, F., Pardo, T.A.S.
13th Conference on Language Resources and Evaluation (LREC). pp.5906-5915. Marseille, France. [see](#)

HateBR: A Large Expert Annotated Corpus of Brazilian Instagram Comments for Abusive Language Detection

Vargas, F., Carvalho, I., Góes, F.R., Pardo, T.A.S., Benevenuto, F.
13th Conference on Language Resources and Evaluation (LREC). pp.7174-7183. Marseille, France. [see](#)

Studying Dishonest Intentions in Brazilian Portuguese Texts

Vargas, F., Pardo, T.A.S.
Deceptive AI. Springer, vol 1296. pp.166-178. [see](#)

Extended Multilingual Protest News Detection

Hurriyetoglu, A., Mutlu, O., San, F. D., Uca, O., Gurel, A. S., Radford, B., Dai, Y., Hettiarachchi, H., Stoeck, N., Nomoto, T., Slavcheva, M., Vargas, F., Javid, A., Beyhan, F., Yoruk, E.
5th International Workshop Challenges and Applications of Automated Extraction of Socio-Political Events from Text (EMNLP). pp.223-228. Abu Dhabi, Arab Emirates. [see](#)

2021

Contextual-Lexicon Approach for Abusive Language Detection

Vargas, F., Góes, F.R., Carvalho, I., Pardo, T.A.S., Benevenuto, F.
Recent Advances in Natural Language Processing (RANLP). pp.1442-1451. Held Online.[see](#)

Towards Discourse-Aware Models for Multilingual Fake News Detection
Vargas, F., Benevenuto, F., Pardo, T.A.S.
Recent Advances in Natural Language Processing (RANLP). pp.210-218. Held Online. [see](#)

Implicit Opinion Aspect Clues in Portuguese Texts: Analysis and Categorization
Machado, M. T., Pardo, T.A.S., Ruiz, E. E. S., Di Felippo, A., Vargas, F.
15th International Conference on the Computational Processing of Portuguese (PROPOR). pp.68-78. Fortaleza, Brazil. [see](#)

2020

Linguistic Rules for Fine-Grained Opinion Extraction
Vargas, F., Pardo, T.A.S.
5th International Workshop on Social Sensing: Special Edition on Narrative Analysis on Social Media (ICWSM). pp.1-6. Held Online. [see](#)

Identifying Fine-Grained Opinion and Classifying Polarity on Coronavirus Pandemic
Vargas, F., Santos, R.S.S., F., Rocha, P.R.
9th Brazilian Conference on Intelligent Systems (BRACIS). pp.511-520. Rio Grande, Brazil. [see](#)

2019 and before

Aspect Clustering Methods for Sentiment Analysis
Vargas, F., Pardo, T.A.S.
13th International Conference on the Computational Processing of Portuguese (PROPOR). pp.365-374. Canela, Brazil. [see](#)

The Coreference Annotation of the CSTNews Corpus
Pardo, T.A.S., Baptista, J., Duran, M.S., Nunes, M.G.V., Nóbrega, F.A.A., Aluísio, S.M., Di Felippo, A., Seno, E.R.M., Silva, R.R., Anchieta, R.T., Brum, H.B., Dias, M.S., Martins, R.S.O., Maziero, E.G., Souza, J.W.C., Vargas, F.
2nd Workshop on Evaluation of Human Language Technologies for Iberian Language (SEPLN). pp.102-112. Murcia, Spain. [see](#)

Brazilian Students Connected with Social Media and High Schools
Sales, S., Ferreira, A., Vargas, F.
Acta Científica XXIX Congreso de la Asociación Latinoamericana de Sociología (ALAS). pp.1-15. Santiago de Chile, Chile. [see](#)

Book Published / **Por que amo Llansol?** (*Why are I love Llansol?*). [pdf](#)
Organized Bittencourt, A., Costa, F. A., Vargas, F. A.

Viva Voz, FAE/UFGM: Belo Horizonte: 34p. ISBN 978-85-7758-227-3.

Conference Participation NAACL 2024. to appear
RANLP 2023. [see](#)
ACL 2023. [see](#)
LREC 2022. [see](#)

RANLP 2021. [see](#)
ICWSM 2021. [see](#)
ACL 2020. [see](#)
EMNLP 2020. [see](#)
LREC 2020. [see](#)
BRACIS 2020. [see](#)

Experience

Research experience

2024/2024: Researcher Visiting, University of Southern California (USC)
2015/2023: Graduate Researcher, University of São Paulo (USP)
2011/2014: Undergraduate Researcher, Federal University of Minas Gerais (UFMG)

Teaching experience

SCC5809: Neural Networks and Deep Learning University of São Paulo, 2021. [see](#)
SCC0605: Theory of Computation and Compilers, University of São Paulo, 2020. [see](#)
SCC0227: Computation Topics Seminars, University of São Paulo, 2016. [see](#)

Industry experience

2021/2023: Doctoral Researcher, Sinch
2021/2022: Data Scientist, Cisco Webex
2014/2015: System Analyst, Unisys

Adviser

CS Undergraduate: Lucas Sobral Fontes Cardoso, University of São Paulo, 2020.
Thesis: *Web resources for an aspect identification and clustering tool in fine-grained opinion mining.* [see](#)

Summer School

2024-2024: 2nd Mexican NLP Summer School, NAACL 2024, Mexican. [to appear](#)
2023-2023: 2nd Summer School on Deep Learning in NLP, RANLP 2024, Bulgaria. [see](#)
2022-2022: 2nd Advanced NLP School, Université Grenoble Alpes, France. [see](#)
2021-2021: 4th Advanced School in Big Data Analysis, University of São Paulo, Brazil. [see](#)
2020-2020: 10th Lisbon Machine Learning School, INESC, Portugal. [see](#)

Courses

2019-2019: Introduction on the Stars - Astrophysics, University of São Paulo, Brazil. [see](#)
2014-2020: Software Engineering, Pontifical Catholic University of Minas Gerais, Brazil. [see](#)

Computer Programs

[OPCluster](#): Automatic extraction and clustering of fine-grained opinions.
[NoHateBrazil](#): A Brazilian Portuguese text offensiveness analysis system.
[FACTual](#): Automated fact-checking and news credibility.

Data Resources

[FactNews](#): Sentence-level annotated dataset to predict factuality and media bias.
[HateBR](#): Large-scale abusive language dataset of Brazilian Instagram comments.
[HausaHate](#): A hate speech dataset of Facebook comments for the Hausa language.
[OPCovidBR](#): Aspect-based sentiment analysis dataset of Covid-19 Tweets.
[AspectPRO](#): Aspect-based annotated dataset of web consumer reviews.
[Deceiver](#): Multilingual discourse-annotated dataset of fake and true news.
[PRO](#): Taxonomies for aspect-based sentiment analysis in Portuguese.
[MOL](#): Multilingual offensive lexicon.

Recommendation Letters	Prof. Dr. Fabricio Benevenuto. <i>DCC-UFMG.</i> letter Dra. Norah Abokhodair. <i>Microsoft Research.</i> letter Prof. Dra. Roseli Romero. <i>ICMC-USP.</i> letter Prof. Dr. Carlos Ramisch. <i>LIS-CNRS-Aix Marseille Université.</i> letter
-------------------------------	---

Coding	Python, SQL, RDF(s), OWL, XML, JSON, Unix/Bash.
---------------	---
