

ARTHUR SANTANA, Ph.D.

Computational Linguist | Quality Controller | NLP | NLU | ASR | TTS

arthurpereirasantana@gmail.com

<http://www.arthursantana.net>

+55 (11) 94621-6111

<https://www.linkedin.com/in/arthursantana>

Bio

Computational linguist with 7+ years of experience developing NLU grammars, linguistic data pipelines, and QA frameworks for AI and speech systems (ASR/TTS).

SKILLS

- Linguistics

Phonetics, Phonology, Syntax, Morphology, Portuguese

- Computational Linguistics

Language Data Analysis, Experiment Design, Language Modeling, Ontology

- NLP

NLU, NLG, ASR, TTS

- Technical

Python, Version Control (Git), R, Praat, SQL

EDUCATION

- Ph.D in Linguistics

University of São Paulo (2015-2019)

- Visiting Ph.D. Scholar

University of Southern California (2017-2018)

- M.A in Linguistics

University of São Paulo (2013-2015)

- B.A in Letters

Federal University of Maranhão (2008-2013)

WORK EXPERIENCE

☐ Quality Controller at Telus Digital

(Jun/2023 - current)

- Lead quality assurance for language data projects, ensuring accuracy, consistency, and alignment with client specifications;
- Collaborate with project managers and subject matter experts to refine task guidelines and resolve quality issues;
- Conduct audits and performance reviews of annotators and reviewers;
- Analyze error patterns and generate insights to improve data quality and team efficiency;
- Support the onboarding and training of new team members, ensuring alignment with quality standards;
- Create, update, and edit annotation and QA guidelines to reflect evolving project needs and improve clarity for annotators
- Use quality tracking tools and dashboards to report on KPIs, monitor SLA compliance, and drive continuous improvement

☐ Senior Linguist at Google Brazil (via Telus Digital)

(oct/2023 - Jun/2025)

- Support the Project Manager on management tasks;
- Perform review of work authored by Linguists;
- Elaborate and conduct training activities of Linguists and new-hires
- Provide extensive linguistic expertise for developing an Artificial Intelligence product using NLP techniques;
- Develop complex grammars for Natural Language Understanding;
- Create complex templates for Natural Language Generation;
- Evaluate current system outputs;
- Annotate and review linguistic data.
- Follow and fulfill project KPIs

☐ Researcher at Center for Curriculum Redesign

(oct/2024 - apr/2025)

- Developed and tested automated workflows to compare and contrast curricular frameworks;
- Designed and implemented scoring systems to measure key competencies, such as creativity, curiosity, and metacognition;
- Collaborated on the creation of competency assessment and measurement instruments;
- Contributed to the development of workflows and quality assurance processes for integrating AI into organizational tasks (automation, agent design, prompt engineering);
- Authored detailed documentation of experiments, methodologies, and results throughout the research and development cycle.

☐ NLU Language Developer at Cerence (via Maize)

(jun/2022 - aug/2024)

- Developed NLP components: write rules, build grammars/FSTs
- Contributed to the design of new features by extending existing annotation schemas to cover
- new areas
- Used modeling tools to bootstrap and test new functionalities
- Created annotation systems and mappings for Portuguese

LANGUAGES

- Brazilian Portuguese

Native

- English

Fluent/Bilingual

- Spanish

Intermediate Level

GRANTS

- Ph.D. Research

Conselho Nacional de Desenvolvimento Científico e Tecnológico.
Process n. 140139/2015-0
(2015-2019)

- International Visiting Scholar Grant

Conselho Nacional de Desenvolvimento Científico e Tecnológico.
(2017-2018)

- M.A. Research

Conselho Nacional de Desenvolvimento Científico e Tecnológico.
Process n. 130194/2013-1
(2013-2015)

- Analyzed system performance and identify areas for improvement
- Collaborated with other Engineers and annotators in data collection and development efforts
- Developed specific guidelines for Brazilian Portuguese phonetic/phonemic transcription
- Supported script design used to train speech synthesis models
- Worked on localization (en-ptBR-en)

☐ Associate Linguist at Google Brazil (via Telus Digital) (sep/2018 - sep/2023)

- NLU Grammar development
- NLG development
- Elaborate Guidelines for NLP projects
- Annotate and review linguistic data
- Provide linguistic analysis and support for NLU and NLG projects
- Response modeling
- Labeling text for disambiguation, expansion, and text normalization
- Annotating lexicon entries according to guidelines
- Deriving NLP data

☐ Computational Linguist at Portuguese International Institute (jan/2022 - jul/2022)

- Support the Scientific and Technical Portuguese Terminologies platform development
- Data cleaning and wrangling
- Flexional paradigm script design

☐ Visiting PhD Scholar (USC, Los Angeles, United States) (2017-2018)

☐ Teacher's Assistant (USP, São Paulo, Brazil) (2015-2019)

☐ English teacher (English for Business, São Paulo, Brazil) (2013-2017)

☐ English teacher (Cultura Inglesa, São Luís, Brazil) (2011-2013)

ACADEMIC EXPERIENCE

PUBLICATIONS

Journal Articles

2018. *The status of unstressed lax-mid vowels in northern dialect of Brazilian Portuguese*. Revista Linguística.

2018. *The role of Phonological processes to discussing vowel inventories in Brazilian Portuguese*. Acta Linguistica Academica.

2016. *Uma abordagem fonológica para as postônicas não-finais*. Domínio de Linguagem.

2012. *A síncope revisitada: uma análise com base no corpus do ALiMA*. Littera.

2011. *O livro didático do português: uma análise da oralidade*. Littera.

Book Chapters

2014. *A redução de proparoxítonas no falar maranhense*. (with José Bezerra) In: In: RAZKY, Abdelhak et. al (Orgs.). Estudos sociodialetais do português brasileiro. Editora Pontes.

2013. *A permanência da tônica em variantes proparoxítonas no falar maranhense*. In: RAMOS, Conceição et. al. (Orgs.) O português falado no Maranhão: múltiplos olhares. EDUFMA. Conference Proceedings.

2012. *O apagamento das postônicas não-finais no falar maranhense*. IV Seminário Internacional de fonologia.

2012. *A neutralização das postônicas não finais no falar maranhense: uma análise preliminar.* II Congresso Internacional de Dialetologia e Sociolinguística.

2011. *Variação de proparoxítonas: traços da identidade popular no falar maranhense.* III Seminário Linguagem e identidades: múltiplos olhares.

2010. *'Risca o fósqui que a lâmpida apagô': um estudo geosociolinguístico das proparoxítonas no falar maranhense.* I Congresso Internacional de Dialetologia e Sociolinguística.

2010. *Apócope em proparoxítonos: um estudo preliminar.* I Congresso Internacional de Dialetologia e Sociolinguística.

CONFERENCE PRESENTATIONS

2018 *Is there reduction via laxing in northern dialects of Brazilian Portuguese?* Generative Linguistics in the Old World - GLOW 41. (Budapest, Hungary. April, 2018)

2018 *The status of unstressed lax mid vowels in the northern dialect of Brazilian Portuguese.* Experimental Portuguese Linguistics Workshop (Toronto, Canada. April, 2018)

2018 *Syncope and the canonical foot in Brazilian Portuguese.* 7th Hispanic and Luso-Brazilian Linguistics Conference (Tempe, Arizona. March, 2018)

2016. *The role of phonological processes to discussing vowel inventories in Brazilian Portuguese.* Old World Conference in Phonology - OCP 13 (Budapest, Hungary. January, 2016)

2015. *Uma abordagem fonológica para as postônicas não-finais.* II Encontro intermediário do GT de Fonética e Fonologia da ANPOLL. (Uberlândia, Brazil. September, 2015)

2015. *Neutralization of non-final postonic mid vowels in Brazilian Portuguese: assimilation or feature deletion?* Phonetics and Phonology in Europe - PAPE 2015 (Cambridge, UK. July, 2015)

2015. *Vogais medias em dialetos nordestinos: as postônicas não-finais em São Luís.* 63 Seminário do GEL. (Campinas, Brazil. June, 2015).

Last updated:
Jun, 2025