

Pablo Faria – *Curriculum vitae* (updated on September, 2017)

Faculty member, University of Campinas

Contact

Departamento de Linguística – IEL
Universidade Estadual de Campinas (Unicamp)
Rua Sérgio Buarque de Holanda, no 571, sala D.1.19
CEP 13.083-859 – Campinas – SP – Brazil
pablofaria@iel.unicamp.br

Education

- **Ph. D., Linguistics**
Universidade Estadual de Campinas, November 2013
 - Dissertation: *A computational model of first language acquisition*
Adviser: Ruth Lopes (Linguistics)
Committee: Marcelo Ferreira, Sergio Menuzzi, Edson Françaço, Leticia Sicuro Correa
- **M.S., Linguistics**
Universidade Estadual de Campinas, December 2009
 - Thesis: *Properties of natural languages and the acquisition process: reflections based on an implementation of the model in Berwick (1985)*
Adviser: Ruth Lopes (Linguistics)
Committee: Marcelo Ferreira, Gabriel Othero
- **B.S., Computer Science**
Sociedade de Ensino Superior Objetivo (UNIP/Objetivo), December 2002
- Additional: **Course on Information Systems** (360 class hours)
Universidade Federal de Goiás, 2003-2004
Database systems, programming languages, software engineering, graphical computation, computational complexity, and other topics.

Research and Teaching Experience

01/2014 – 07/2016

Post-doctoral fellow at Unicamp

Instituto de Estudos da Linguagem, Universidade Estadual de Campinas

Research on treebank inconsistency detection, entitled “Study and development of methods for automatic detection and correction of errors and inconsistencies in syntactically annotated corpora” (FAPESP no. 13/18090-6). Supervisor: Charlotte Galves. Co-supervisor: Jesus Garcia.

01/2015 – 09/2015

Visiting Scholar at University of Pennsylvania, Department of Linguistics

Research on inconsistency detection on parsed corpora using the Tree Adjoining Grammar based formalisms (FAPESP no. 14/17172-1). Supervisor: Anthony Kroch. Co-supervisor: Seth Kulick.

Since 08/2014

Lecturer

Instituto de Estudos da Linguagem, UNICAMP

Language Acquisition (HL311, HL422, HL550)

Computational Linguistics (introduction) (HL377, LL904)

Practices on Grammatical Analysis (HL220)

Professional Activities

- ***Software development, 1996 – 2007***
Various companies (chronologically): MT Móveis, ISA Informática, Politec Ltda. e Inloc Sistemas
Programming languages: Clipper, Java (desktop and web programming), Delphi
Operational Systems: Windows, Linux
- ***Software development for linguistic applications, 2007 – 2009***
Technical training TT4A (FAPESP no. 08/04312-9)
Instituto de Estudos da Linguagem/UNICAMP
- ***Other technical productions***
 - FARIA, P. Building of Annotated Corpora: E-Dictor, 2011. (Other, Short Term Course Taught) Knowledge areas: Corpus linguistics, Computational Linguistics. Additional references : Brasil/Portuguese. 20 hours.
 - FARIA, P. Building of Annotated Corpora: Tycho Brahe Parsed Corpus of Historical Portuguese and annotation and edition tools, 2011. (Other, Short Term Course Taught). Knowledge areas : Corpus linguistics, Computational Linguistics. Additional references : Brasil/Portuguese. 6 hours.
 - FARIA, P. Building Annotated Corpora using E-Dictor, 2011. (Other, Short Term Course Taught). Knowledge areas : Corpus linguistics, Computational Linguistics, Additional references : Brasil/Portuguese. 20 hours.
 - FARIA, P. II Workshop in Digital Edition and dissemination of ancient texts, 2011. (Other, Short Term Course Taught) Knowledge areas : Corpus linguistics, Computational Linguistics, Additional references : Brasil/Portuguese. 15 hours.

Other information

- ***Development of eDictor***
The *eDictor* tool (<http://manualedictor.wordpress.com/edictor/>) developed from 2007 to 2012 has been extensively used in research projects interested in text edition and annotation, such as the ones cited below. It provides an user-friendly interface for text tokenization, XML annotation of textual, lexical and part-of-speech information. It also includes a version of *vmmtagger*¹ for automatic part-of-speech tagging.

¹Kepler, Fábio N., and Marcelo Finger. "Comparing two markov methods for part-of-speech tagging of portuguese." Advances in Artificial Intelligence-IBERAMIA-SBIA 2006. Springer Berlin Heidelberg, 2006. 482-491.