

## CONTACT INFORMATION

Barcelona, Spain

[daniel.ortiz.phd@gmail.com](mailto:daniel.ortiz.phd@gmail.com)

## EDUCATION

### MSc in Bioinformatics

[University of Valencia](#), Valencia, Spain

Graduation date: October 2016

- Thesis title: Systems Biology Strategies to Study Cancer Metabolism
- Advisors: Joaquín Dopazo Blázquez and Vicente Arnau Llobart

### PhD in Pattern Recognition and Artificial Intelligence

[Technical University of Valencia](#), Valencia, Spain

Graduation date: October 2011

- Thesis title: Advances in Fully-Automatic and Interactive Phrase-Based Statistical Machine Translation
- Advisors: Francisco Casacuberta Nolla and Ismael García Varea

### MSc in Pattern Recognition and Artificial Intelligence

[Technical University of Valencia](#), Valencia, Spain

Graduation date: November 2005

- Thesis topic: Search Algorithms for Phrase-based Statistical Machine Translation

### BSc in Computer Science Engineering

[University of Castilla La Mancha](#), Albacete, Spain

Graduation date: January 2003

- Specialization in program of Pattern Recognition and Artificial Intelligence
- Thesis topic: Stack Decoding Algorithms for Statistical Machine Translation

## PROFESSIONAL EXPERIENCE

### Research Assistant

April 2018 to today

[Institute for Research in Biomedicine \(IRB\)](#)

- Study cancer genomics under a data science and machine learning perspective

### Adjunct Professor

October 2019 to today

[University of Barcelona](#), Barcelona, Spain

- Computer Vision (undergrad level, Spanish): Course 2019/20
- Algorithms (undergrad level, Spanish): Course 2019/20

### Adjunct Professor

September 2018 to August 2019

[Technical University of Catalonia](#), Catalonia, Spain

- Programming (undergrad level, Spanish): Course 2018/19

### Natural Language Processing Engineer / Technical Leader

February 2016 to March 2018

[Webinterpret](#)

- Development of new natural language processing and statistical machine translation techniques

### Visiting Lecturer

March 2017

[University of Valencia](#), Valencia, Spain

- Introduction to big data in natural language processing at the Master's Degree in Data Science

### Research Consultant (Data Scientist) for [Webinterpret](#)

April 2015 to January 2016

- Introduction of natural language processing techniques into the Webinterpret's workflow

### Adjunct Professor

December 2010 to January 2016

[Technical University of Valencia](#), Valencia, Spain

- Statistics (undergrad level, Spanish and English): Courses 2011/12, 2012/13, 2013/14, 2014/15
- Operational Research (undergrad level, Spanish): Courses 2010/11, 2012/13

### Post-Doctoral Researcher

February 2012 to December 2014

[PRHLT Research Group](#), [Technical University of Valencia](#), Valencia, Spain

- [CASMACAT research project](#), funded by the [7th Framework Programme of the European Commission](#) (funding amount: €2 500 000)
  - Online and active learning techniques for statistical machine translation
  - Scalable and parallel estimation of HMM models using Map-Reduce

## European Commission

- Scientific reviewer of the [EuromatrixPlus research project](#) (budget: €5.94M), EuromatrixPlus belongs to the 7th Framework Programme of the European Union

## Research Assistant

July 2008 to January 2012

[Instituto Tecnol. de Informática](#), [Technical University of Valencia](#), Valencia, Spain

- [MIPRCV research project](#), part of the prestigious CONSOLIDER programme of the Spanish Government and the European Commission (funding amount: €4 500 000)
  - Application of the incremental EM algorithm to the estimation of HMM models
  - Design and implementation of stochastic error correction models for string pairs

## Computational Linguistic Researcher

March 2003 to June 2008

[Technical University of Valencia](#), Valencia, Spain

- Participation in several research projects funded by the Spanish Government

RESEARCH STAYS Three months research visit to the [Lershtul für Informatik VI](#) at [Aachen University of Technology](#) (Germany), from May 2006 to July 2006

SUPERVISED MSC Jesús González Rubio. Technical University of Valencia. [PhD Thesis](#). May 2014

AND PHD THESES Álvaro Peris Abril. Technical University of Valencia. MSc Thesis. September 2014

SELECTED **Daniel Ortiz-Martínez.** Online Learning for Statistical Machine Translation. *Computational Lin-*  
PUBLICATIONS *guistics*, 03/2016; DOI: 10.1162/COLI\_a\_00244

Antonio L. Lagarda, **Daniel Ortiz-Martínez**, Vicent Alabau, Francisco Casacuberta. Translating without In-domain Corpus: Machine Translation Post-Editing with Online Learning Techniques. *Computer Speech & Language Journal*, 11/2014; DOI: 10.1016/j.csl.2014.10.004

**Daniel Ortiz-Martínez**, Francisco Casacuberta. The New Thot Toolkit for Fully Automatic and Interactive Statistical Machine Translation. *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL) conference*, Gothenburg, Sweden, April 2014

Jesús González-Rubio, **Daniel Ortiz-Martínez**, Francisco Casacuberta. Active learning for interactive machine translation. *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL) conference*, Avignon, Paris, 2012

**Daniel Ortiz-Martínez**, Ismael García-Varea, Francisco Casacuberta. Online Learning for Interactive Statistical Machine Translation. *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT) conference*, Los Angeles, US, 2010

**Daniel Ortiz-Martínez**, Ismael García-Varea, Francisco Casacuberta. Phrase-level alignment generation using a smoothed loglinear phrase-based statistical alignment model. *Proceedings of the XII European Association for Machine Translation (EAMT) conference*, Hamburg, Germany, October 2008 (**Best paper award**)

[Link to full publication list](#) (currently comprises more than 50 research papers)

SELECTED INVITED “[Big Data and Natural Language Processing](#)”, Introductory talk in the MSc Degree in Data Science  
TALKS from University of Valencia, Valencia, Spain, March 2017

“Incremental Learning for Statistical Machine Translation”, Workshop on Future Directions on Translation Research, Nara, Japan, December 2012

“Thot: New Features to Deal with Larger Corpora and Long Sentences”, In TC-STAR OpenLab on Speech Translation Workshop, Trento, Italy, March 2006

SCIENTIFIC AND Scientific reviewer of international conferences and journals (IUI, ACL, COLING, Computer Speech  
PROGRAM and Language Journal, Knowledge Engineering Review Journal)

COMMITTEES Member of the program committee of conferences and workshops (CAEPIA, ACL, AMTA)

## OPEN SOURCE SOFTWARE

### Thot Toolkit for Statistical Machine Translation

- Hosted on github <https://daormar.github.io/thot/>
- Provides search algorithms and scalable and parallel estimation of statistical models
- Training of statistical models from huge corpora using MapReduce
- Funded by the European Commission and by the Spanish Government
- +50 000 lines of code

### Flux Capacitor Toolkit for Systems Biology

- Hosted on github <https://daormar.github.io/flux-capacitor/>
- Provides systems biology algorithms to study metabolism
- Integration of transcriptomic and metabolic information by means of flux-balance analysis
- Metabolic network visualization and reduction techniques

### snptools Package

- Hosted on github <https://github.com/daormar/snptools>
- Provides tools to work with single nucleotide polymorphisms (SNPs)
- Uses the SNPedia database to generate reports about SNPs of interest
- Applies natural language processing techniques to work with unstructured information

### PanPipe Package

- Hosted on github <https://daormar.github.io/panpipe>
- Implements a fully-fledged workflow manager

### Bio-PanPipe Package

- Hosted on github <https://daormar.github.io/bio-panpipe>
- Implements pipeline modules that can be executed with PanPipe
- Incorporates the `bam_analysis` module to analyze genomic variants

## AWARDS AND SCHOLARSHIPS

Qualification as Associate professor issued by the The Spanish Ministry of Science and Innovation's National Agency for Quality Assessment and Accreditation (ANECA), Spain 2016

Extraordinary Master Award for outstanding MSc thesis in the Master's degree of Bioinformatics issued by the University of Valencia, Spain 2016

Award for the best academic record in the Master's degree of Bioinformatics issued by the Engineering School of the University of Valencia, Spain 2016

Co-supervisor of the PhD Thesis "On the Effective Deployment of Current Machine Translation Technology" written by Jesús González Rubio, which received the [LRC best PhD Thesis award of 2014](#)

Honorable mention award in the Information and Communications Technology category at Valencia Idea competition, Spain, 2010

Best paper award for "Phrase-level alignment generation using a smoothed loglinear phrase-based statistical alignment model" issued by the European Association for Machine Translation. Germany, 2008

Predoctoral Fellowship (FPI), public competitive call, Spanish Government, (2003-2007)

5th course of Computer Science studied in the Technical University of Valencia under the Seneca scholarship programme of the Spanish Government (2001-2002)

**TECHNICAL SKILLS** *Programming Languages:* C, C++, Python, R, AWK, UNIX shell scripting

*Scientific Libraries:* StatsModels, NumPy, SciPy, Scikit-learn, Scikit-image

*Deep Learning Frameworks:* PyTorch

*Optimization Software:* CPLEX, COIN-OR, OR-Tools

*Bioinformatics Tools:* fastqc, bowtie, tophat, qualimap, samtools, BLAST, COBRA toolbox, etc.

*R packages for Bioinformatics:* affy, limma, genefilter, samr, GO.db, GSA, piano, etc.

*Somatic Variant Detection Tools:* Manta, Strelka, Platypus, CNVkit, Lumpy, Delly, MSIsensor, etc.

*Machine Learning Toolkits for Natural Language Processing:* NLTK, SRILM, OpenNMT, Moses

*Databases and Database Libraries:* MongoDB, LevelDB, Berkeley DB

*Workload Managers:* Slurm (experience as administrator), SGE

*Code Repositories:* GitHub, GitLab (experience as administrator), SourceForge