

Pablo Malvar Fernández - Curriculum Vitae

Personal Data:

Address: 7945 Cinnabar Drive
91941 La Mesa, California
United States of America

Phone number (Spain): (619) 303 7288

e-mail: paulomal@gmail.com

Webpage: <http://paulomalvar.net>

Citizenship: Spanish

US Work Status: Green Card Holder

Education:

- 2008 M.A. Computational Linguistics at
San Diego State University (CA), USA
Master Thesis: *Improving Word-to-Word Alignments Using
Morphological Information.*
- 2006 M.A. Applied Linguistics
Universitat Pompeu Fabra (Barcelona)
- 2005 Ph.D. Doctoral Candidate in Linguistics
Universidade de Santiago de Compostela, Spain
Second year thesis: *Corpus paralelo bilingüe (portugués-español)
de textos literarios das Literaturas Africanas en Lingua
Portuguesa* [Bilingual Parallel Corpus (Portuguese-Spanish) of
literary texts of Portuguese Language African Literature]
- 2003 B.A. Portuguese Philology
Universidade de Santiago de Compostela

Professional Development:

- 2011 Certificate Machine Learning (Basic Track, 77 points out of 80)
Stanford University

Professional experience:

- 2011-2012 **Freelance Computational Linguist, imaxin | software**, Santiago de
Compostela (Spain)
Description of Duties: As a Computational Linguist I redesigned an Opinion
Mining prototype (called COATI) to classify Spanish texts according to their

sentiment orientation. As part of my duties, I compiled a sentiment-tagged on-line corpus. I trained several Machine Learning Algorithms (Neural Networks and Support Vector Machines) to detect the sentiment orientation.

2011-2012 **Freelance Software Localizer, imaxin | software**, Santiago de Compostela (Spain)

Description of duties: I am currently working on the coordination and localization of a major office suite.

2011 **European Portuguese Transcriber**, Appen Butler Hill

Description of Duties: As a European Portuguese transcriber I had to transcribe voice recordings into proper European Portuguese spelling for the development of Speech Recognition and Voice-over systems.

2011 **European Spanish Transcriber**, Appen Butler Hill

Description of Duties: As a European Spanish transcriber I had to transcribe voice recordings into proper European Spanish spelling for the development of Speech Recognition and Voice-over systems.

2008-2011 **Computational Linguist, imaxin | software**, Santiago de Compostela (Spain)

Description of Duties: As a Computational Linguist I worked on the development of a BioNLP application for normalizing natural texts from the domain of cardiology using UMLS terminology codes.

As a Computational Linguist I developed a Spanish-Chinese SMT. As part of my duties, I compiled the Spanish-Chinese corpus necessary for training the SMT system as well as other resources to minimize the impact of proper names sparseness.

As a Computational Linguist I designed and developed an Opinion Mining prototype (called COATI) to classify according to their sentiment orientation Spanish and Galician journalistic texts. As part of my duties, I compiled a sentiment-tagged corpus and trained two SVM models (one for each language) for tagging the sentiment orientation of above-cited texts.

As a Computational Linguist I worked on the improvement two RBMT systems (English-Portuguese and Spanish-Portuguese). As part of my duties, I am extended the English-Portuguese and Spanish-Portuguese dictionaries (monolingual and bilingual dictionaries) necessary for RBMT systems to work.

As a Computational Linguist I coordinated a team of linguists and computer scientists for the development of one SMT and one RBMT systems (both English-Galician). As part of my duties, I tagged with morpho-syntactic information the English and Galician corpora necessary for the PoS tagger of the RBMT system and I compiled the English-Galician parallel corpus necessary for the SMT system.

As a Computational Linguist I worked for Imaxin | software optimizing the search engine of the Galician General Network of Libraries. For this purpose my job was the development of two rich thesauri, one in Galician and another in Spanish.

- 2008-2011 **Translator (English-Galician), imaxin | software**, Santiago de Compostela (Spain)
Description of duties: I was integrated into the localization team responsible for translation of Office 2010 and Windows 7 into Galician. I coordinated the localization (translation) team responsible for the localization into Spanish and Galician of several pieces of software originally written in English.
- 2009 **Linguistic Consultant**, Butler Hill Group
Description of Duties: As a Linguistic Consultant I worked for Butler Hill Group expanding and polishing a Speech-recognized Portuguese lexicon.
- 2009 **Linguistic Consultant**, Vantage Linguistics
Description of Duties: As a Linguistic Consultant I worked for Vantage Linguistics morpho-syntactically tagging a Spanish lexicon.
- 2007-2008 **Research Assistant** of the Computational Linguistics Laboratory, SDSU (San Diego, USA)
Description of Duties: As the RA of the Computational Linguistics Laboratory my main duty is the maintenance of the lab's hardware and software, which runs under Fedora Linux.
- 2005 **Linguistic Consultant**, Asociación Galega de Antropoloxía Social e Cultural, Santiago de Compostela (Spain)
Description of Duties: My duties as a Linguistic Assistant were to check the terminology, the grammar and the style of the book, Mariño Costales, M. and Casado Neira, D. (2005). *Cadernos dos Obradoiros Abertos de Antropoloxía tres*. Santiago de Compostela (Spain): Asociación galega de Antropoloxía Social e Cultural.
- 2004 **Translator (English-Galician), imaxin | software**, Santiago de Compostela (Spain)
Description of duties: I collaborated in the project of localization from English to Galician of Windows XP and Office XP. My duties were to localize the software using both the the original source language (English) and also the Spanish and Portuguese translations (both languages closely related to Galician)
- Translator (English-Galician), imaxin | software**, Santiago de Compostela (Spain)
Description of duties: I coordinated the localization from English to Galician of Windows XPSP2. My duties were to localize the software (in the same way explained above) and, as the main person responsible, I coordinated the localization team.

Selected Fellowships and Honors:

- 2006-2008 Grant from the Ministerio de Educación y Ciencia and the Fulbright Foundation to earn a M.A. in Computational Linguistics at San Diego State University (CA, USA)
- 2005-2006 Grant from the Fundación Caixa Galicia-Claudio San Martín “Para Estudios de Postgrado en Universidades y Centros de Investigación de Excelencia Académica” to earn a M.A. in Applied Linguistics at Universitat Pompeu Fabra (Barcelona)
- 2005 Grant from the Galician Government to attend a course of English at King's School in Oxford (UK)
- 2003-2004 Award from the Galician Government as the highest ranked student of my class (Portuguese Philology)
- 2003 Grant from the Universidade de Santiago de Compostela to do an internship at Imaxin | software in the field of Linguistic Engineering

Languages:

Spanish: Native Speaker

Galician: Native Speaker

Portuguese: Proficiency level in listening, speaking, reading and writing

Catalan: High level in listening speaking, reading and intermediate level in writing

English: Superior level in listening, speaking, reading and writing

Programming Languages:

Python: Medium-high level

Perl: Basic level

Prolog: Basic level

Objective-C: Basic level (for Mac OS and iPhone OS)

Selected Courses/Workshops Attended:

- 2008 Johns Hopkins University Summer School on Human Language Technology

Selected Papers and Publications:

Papers:

- 2011 “Generación semiautomática de recursos de Opinion Mining para el gallego a partir del portugués y el español.” In ICL 2011: Workshop on Iberian Cross-Language NLP tasks.
- 2010 “Obtaining computational resources for languages with scarce resources from closely related computationally-developed languages. The Galician and Portuguese case.”. In II Congreso Internacional de Lingüística de Corpus (CILC10) at Universidade da Coruña
- 2008 “Improving Word-to-Word Alignment Using Morphological Enrichment”. In Linguistics Spring Colloquium at SDSU
- 2005 “Descrição e análise do panorama actual das traduções para castelhano de romances e livros de contos produzidos nos sistemas literários das Literaturas Africanas em Língua Portuguesa”. In *VIII Congresso da Associação Internacional de Lusitanistas*
- 2004 “Estudo quantitativo da distribuição diastrática da vibrante múltipla na região de Coimbra”. In *6º Congresso de Lingüística General*

Projects at Conferences:

- 2009 Pichel Campos, J. R., Malvar Fernández, P., Senra Gómez, Ó., Gamallo Otero, P. & García, A. “Carvalho SMT English-Galician from EuroParl English-Portuguese parallel corpus”. In The 25th edition of the Annual Conference of the Sociedad Española para el Procesamiento del Lenguaje Natural 2009 (SEPLN'09).

Articles:

- 2012 “Métodos semiautomáticos de generación de recursos de Opinion Mining para el gallego a partir del portugués y el español.” In *Novatica* (accepted and pending publication.)
- 2010 Malvar, P. et al. “Vencendo a escassez de recursos computacionais. Carvalho: Tradutor Automático Estatístico Inglês-Galego a partir do corpus paralelo Europarl Inglês-Português”. In *Linguamática*, vol 2, n. 2. 31-38. <<http://linguamatica.com/index.php/linguamatica/article/view/57>>.
- 2008 “Autonomismo vs Reintegracionismo. Um conflito normativo visto desde a Análise Crítica do Discurso Especializado”. In *Revista AGÁLIA*, Nº. 91-92, Compostela: AGAL.

“Estudo etimológico, diacrónico e diatópico do topónimo *Cunha* (e variantes)”. In *Georgetown University Luso-Hispanic Linguistic Papers*, 2nd Issue (accepted for publication)

2007 “Aproximação à linguística de corpus como metodologia de base empírica. Compilação e anotação do Corpus Paralelo PALOP (português-espanhol) de Narrativa Pós-colonial”. In *Revista AGÁLIA*, N.º. 89-90, Compostela: AGAL. 9-80

2006 Mari-Carmen Marcos, et al. “Evaluación del posicionamiento web en sistemas de información terminológicos online” [on line]. In *Hipertext.net*, n.º 4. 2006 <<http://www.hipertext.net/web/pag267.htm>>.

Mari-Carmen Marcos, et al. “Evaluación de la usabilidad en sistemas de información terminológicos on-line” [published on-line]. In *Hipertext.net*, n.º 4. 2006 <<http://www.hipertext.net/web/pag268.htm>>.

“Análise comparativa dos sistemas fonético-fonológicos do galego e do portugués. Sugestões didáticas dirigidas a portugués-falantes para a aprendizagem do padrão oral galego”. In *Revista AGÁLIA*, n.º 85/86, Compostela: AGAL. 89-114.

2005 “Distribuição distrática e diafásica do R na região de Coimbra”. In *Revista AGÁLIA*, n.º 81/82, Compostela: AGAL. 143-170.

Conference Proceedings:

2011 “Generación semiautomática de recursos de Opinion Mining para el gallego a partir del portugués y el español.” In Proceedings of ICL 2011: Workshop on Iberian Cross-Language NLP tasks <<http://ceur-ws.org/Vol-824/>>.

2010 Malvar, P. et al. “Obtaining computational resources for languages with scarce resources from closely related computationally-developed languages. The Galician and Portuguese case.”. In *Language Windowing through Corpora*. A Coruña: Universidade da Coruña. 529-536.

2009 Pichel Campos, J. R., Malvar Fernández, P., Senra Gómez, Ó., Gamallo Otero, P. & García, A. “Carvalho SMT English-Galician from EuroParl English-Portuguese parallel corpus”. In *Procesamiento del Lenguaje Natural*, 43. Jaen: Sociedad Española para el Procesamiento del Lenguaje Natural. 379-381.

1998 “Galiza, esse lugar da Europa”. In Castro Seixas, P. e Mota Santos, p. (eds.) (1998). *Actas del 6º Congreso Internacional de Estudiantes de Antrpologia*. Porto: Departamento de Ciências Políticas e do Comportamento da Universidade Fernando Pessoa. 33-37.

Workshops Attended:

2010 BioSEPLN10: First Workshop on Language Technology applied to Biomedical and Health Documents at Universidad Politécnica de Valencia.

Reviews:

2006 “Educação em Língua Materna – A Sociolingüística na Sala de Aula”. In *Grial (Revista Galega de Cultura)*, nº 170. Vigo: Editorial Galaxia. 122-123.

2005 “Sociolingüística Rural. A freguesia de Almalaguês”. In *Verba – Anuario Galego de Filoloxía*, vol 32. Santiago de Compostela: USC. 419-423.

““O português são dois...” - Novas fronteiras, velhos problemas”. In *Cadernos de Lingua*, 27. Vigo: Editorial Galaxia. 199-202.

Professional Organizations and Honor Societies:

Association for Computational Linguistics (ACL)

Phi Kappa Phi

ProLNat (NLP research group at University of Santiago de Compostela, Spain)

Dissertation Advisors and References:

Co-advisor (Ph.D.) Professor Ana Isabel Codesido García, Universidade de Santiago de Compostela
E-mail: fekuna@usc.es
Phone number: (0034) 981563100-11776

Co-advisor (Ph.D.) Researcher Pablo Gamallo Otero, Universidade de Santiago de Compostela
E-mail: pablogam@usc.es
Phone number: (0034) 212948300-10743

Mailing Address: Departamento de Literatura Española, Teoría da Fac. de Filoloxía - Campus Universitario Norte.
15782 Santiago de Compostela
(Spain)

Advisor (MA) Professor Jean Mark Gawron, San Diego State University
E-mail: gawron@rohan.sdsu.edu
Mailing Address: Department of Linguistics and Oriental Languages BAM 32
San Diego State University
5500 Campanile Dr.

San Diego, CA 92182-7727
(USA)

Working reference José Ramom Pichel Campos
Imaxin | software
E-mail: jramompichel@imaxin.com
Mailing Address: Salgueiriños de abaixo nº11 L6, 15891
15891 Santiago de Compostela (Spain)
Phone nuber: (0034) 981 554 068