

# Jorge Diz Pico | NLP / ML Engineer

Barcelona – Spain

🏠 diz.es • ✉ jorge@diz.es • 📞 (+34) 667 657 447

## Work experience

---

### NLP Engineer / Computational linguist

Feb. 2015 - Dec. 2017

*Universitat Pompeu Fabra, Barcelona, Spain*

Hired as research staff for the following projects:

- **LPS-Bigger**: Researched, designed, tested and deployed an algorithm for **classification of social media texts** into the industry categories Marketing Mix and Purchase Funnel. The job required combined use and tuning of several techniques such as NLP, **word embeddings** (word2vec) and **machine learning classifiers** (SVM), and involved coordination with partners such as Indra Software Labs, HAVAS MEDIA Group and Universidad de Alcalá, among others.
- **Neómetro**: Collaborated in the gathering, selection and **description of Spanish neological units** with the goal of measuring neologicity and dictionariability. The job required informed **consultation of diverse lexicographical and corpora sources**. I also took charge of the assorted technical support needed for the project.
- **INSIGHT**: Selected, prioritized and designed **visualizations** of the relevant data from **agent-based experiments on contact-learning grammar**. Developed both outreach and academic d3.js animations of the evolving neurodynamics.

### Erlang software engineer

Nov. 2012 - Jan. 2015

*Nomasystems S.L., A Coruña and Madrid, Spain*

Erlang consultancy for international corporations:

- **Inditex**: Designed, built and deployed the initial prototype for a **parallel architecture** for text notification storage, queuing (RabbitMQ) and **mobile push delivery** (iOS/Android) still serving millions of customers.
- **Tuenti (Telefónica)**: Development and scalability of new architectural features for the central chat dispatching system (such as **VoIP calls**) as well as **performance optimization** reworks in both software and protocol (such as reducing login time by a factor of 7). The production cluster comfortably serves more than a **million concurrent users**.

### Intern machine learning researcher

Nov. 2011 - May 2012

*Jobssy S.L.*

Consulted with management, researched and developed a bidirectional **recommendation engine** to match job offers and job candidates. The software was based on Apache Mahout for document **clustering** and Apache Solr for fulltext search.

## Education

---

**Computer Science and Engineering degree** 300 credits

*Universidade da Coruña, A Coruña, Spain*

- Degree thesis: *Analysis, design and implementation of a push notification framework*

**MSc in Computer Science: Algorithms, Languages and Logic** 120 credits

*Chalmers University of Technology, Göteborg, Sweden*

- Master thesis: *A natural language interface for a music database*

**MA in Theoretical and Applied Linguistics** 60 credits

*Universitat Pompeu Fabra, Barcelona, Spain*

- Master thesis: *Lexical variation and innovation: a study of ideological factors*

**PhD in Translation and Language Sciences** (2016-2017)

*Universitat Pompeu Fabra, Barcelona, Spain*

- Doctoral thesis proposal: *Neology as seen by pragmatics: functions and transgressions*

## Publications and conferences

---

Bel, Núria / **Diz Pico, Jorge** / Marimon, Montserrat / Pocostales, Joel. (2017) "Classifying short texts for a Social Media monitoring system". *Procesamiento del Lenguaje Natural 59*, pp. 57-64. ISSN: 1135-5948

**Diz Pico, Jorge**. (2017) *The process of compiling a Twitter corpus for Galician*. Sociolinguistics Summer School 8 – COST Action IS1306 New Speakers in a Multilingual Europe Training School (4-7 July 2017, Barcelona, Spain).

**Diz Pico, Jorge**. (2017) *Neoloxía e ideoloxía: unha proposta universal a partir do galego*. 1st International Conference on Revitalization of Indigenous and Minoritized Languages (19-21 April 2017, Barcelona and Vic, Spain).

## Other merits

---

**2016:** 8.000€ grant from A Coruña's Provincial Government for the development of a prototypal NLP library for text processing and automatic neologism extraction.

**2012:** 1<sup>st</sup> place in the level generation track for the Mario AI Championship with an adaptive procedural algorithm based on a custom grammatical representation of gameplay elements.

**Certifications:** deeplearning.ai (Coursera) Specialization on Neural Networks and Deep Learning, University of Adelaide (edX) course on endangered language revivalistics.

**Developer experience:** Java, Python, Haskell, Erlang, HTML, CSS, Javascript, Weka, Freeling, TensorFlow/Keras, JIRA, Jenkins, GitHub...

**Languages spoken:** Galician and Spanish (native), English (very high, C2), Catalan and French (high, B2), Portuguese (medium), Swedish, Asturian and German (basic)