

PERSONAL INFORMATION

Laura Pérez-Mayos

📍 PhD Student in Natural Language Processing. UPF, Barcelona

✉ lpmayos@gmail.com

🌐 <https://lpmayos.github.io>

ACADEMIC EDUCATION

- | | |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2017 – 2022 | PhD on Natural Language Processing UPF
Universitat Pompeu Fabra, Barcelona, Spain
Thesis title (provisional): In-depth exploration of the syntactic capabilities of autoencoding language models for downstream applications.
Supervisors: Dr. Leo Wanner (ICREA, UPF) and Dr. Miguel Ballesteros (Amazon AI) |
| 2019 | 9th Lisbon Machine Learning School (LxMLS 2019) IST
Instituto Superior Técnico, Lisbon, Portugal
ML and their application to NLP. These areas include but are not restricted to: Classification, Structured Prediction (sequences, trees, graphs), Deep Learning and Reinforcement Learning. |
| 2015 – 2017 | Master on Computer Vision UAB, UPC, UPF, UOC (60 ECTS)
Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Universitat Pompeu Fabra, Universitat Oberta de Catalunya, Barcelona, Spain
Thesis title: Improving the quality of video-to-language models by an optimized annotation of the training material |
| 2014 – 2015 | Master on Intelligent Interactive Systems UPF (60 ECTS)
Universitat Pompeu Fabra, Barcelona, Spain
Thesis title: Co-speech gesture generation for the humanoid robot REEM-C |
| 2001 – 2005 | Technical Engineer in Computer Science UPC (183.6 ECTS)
Universitat Politècnica de Catalunya, Barcelona, Spain
Thesis title: Theclick production guide, visual directory of companies |

PROFESSIONAL EXPERIENCE

- | | |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| April 2021 – August 2021 | Applied Scientist Intern, Amazon Web Services (AWS)
https://aws.amazon.com/
Sentence ordering with Transformer-based autoencoding language models. |
| October 2017 – Present | NLP Pre-doctoral Researcher
http://taln.upf.edu
Syntactic capabilities of autoencoding language models. Web development for workshops and projects. Social media management. |
| July 2015 – September 2017 | NLP Research assistant at TALN, UPF
http://taln.upf.edu
Word Sense Disambiguation. UIMA pipelines. Web development for workshops and projects. |
| May 2014 – September 2014 | Web Architect at 4d-life
http://4d-life.net
Frontend and Backend Engineer. Analysis, architecture and development of web applications with Python and Javascript. |

March 2013 – February 2014	CTO at Iskra Desenvolupament S.C.C.L. http://www.iskra.cat Agile leader and project manager. Analysis, design and development of web applications with Python and Plone.
March 2010 – January 2013	Web applications developer at UPCnet Inc. http://www.upcnet.es Analysis, design and development of web applications with Python and Plone. Analysis, design and development of e-Government applications.
February 2008 – August 2009	Web applications developer at Guell Consulting Technologies Inc. http://www.guellconsulting.com Analysis, design and development of web applications with Python and Plone.
February 2006 – January 2008	Web developer at SAC Disseny i Comunicacions Inc. http://activaweb.com Analysis, design and development of web applications based on ASP and MySQL.

RESEARCH INTERESTS

Syntactic capabilities of autoencoding language models.
Deep Learning for natural language processing.
Natural language processing for social good.

PUBLICATIONS

- 2021 Pérez-Mayos, L.; Ballesteros, M.; Wanner, L. How much pretraining data do language models need to learn syntax?. In proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021). Punta Cana, Dominican Republic.
- 2021 Pérez-Mayos, L.; Táboas García, A.; Mille, S.; Wanner, L. Assessing the Syntactic Capabilities of Transformer-based Multilingual Language Models. In findings of the Association for Computational Linguistics (ACL 2021). Bangkok, Thailand.
- 2021 Pérez-Mayos, L.; Carlini, R.; Ballesteros, M.; Wanner, L. On the evolution of syntactic information encoded by BERT's contextualized representations. In proceedings of the European Chapter of the Association for Computational Linguistics (EACL 2021). Kyiv, Ukraine.
- 2021 Fortuna, P.; Cortez, V.; Sozinho Ramalho, M.; Pérez-Mayos, L. MIN_PT: An European Portuguese Lexicon for Minorities Related Terms. In proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021) on the ACL Anthology.
- 2021 Fortuna, P.; Pérez-Mayos, L.; AbuRa'ed, A.; Soler-Company, J.; Wanner, L. Cartography of Natural Language Processing for Social Good: Definitions, Statistics and White Spots. In proceedings of the 1st Workshop on NLP for Positive Impact (2021) on the ACL Anthology.
- 2019 Pérez-Mayos, L.; Farrús, M.; Adell, J. Part-of-Speech and Prosody-based Approaches for Robot Speech and Gesture Synchronization. Journal of Intelligent & Robotic Systems. <https://doi.org/10.1007/s10846-019-01100-3>
- 2018 Pérez-Mayos, L.; Sukno, F.; Wanner, L. Improving the Quality of Video-to-Language Models by Optimizing Annotation of the Training Material. International Conference on Multimedia Modeling. 1 ed. Bangkok: Cham Springer International; 2018. p. 279-290.

ACADEMIC SERVICES

Web Chair for COLING 2020.
Reviewer for COLING, EMNLP, EACL.
Secondary reviewer for NAACL-HLT 2019.

UNIVERSITY TEACHING

Teaching assistant in the Computational Logic subject of the Computer Science Degree. DTIC. Universitat Pompeu Fabra, Barcelona, Spain. Academic year: 2014-2015.

INVITED TALKS

MTUPC Research Seminar (Universitat Politècnica de Catalunya, Barcelona, Spain). Host: Javier Ferrando. 30th March 2022. Topic: In-depth exploration of the syntactic capabilities of autoencoding language models.

TALN Research Seminar (Universitat Pompeu Fabra, Barcelona, Spain). Host: Monica Domínguez. 9th Sep 2020. Topic: On the evolution of syntactic information encoded by BERT's contextualized representations.

TALN Research Seminar (Universitat Pompeu Fabra, Barcelona, Spain). Host: Monica Domínguez. 12th Dec 2018. Topic: Jupyter Notebook by example. Basic NLP with spaCy.

COMMUNITY

PyBCN President of PyBCN, the Barcelona Python Association, with 4k members on Meetup. Our mission is to promote, educate and advance the local Python community through monthly meetups, practice sessions and special events such as PyDay and PyData.

PyLadiesBCN Coordinator of PyLadiesBCN. Our mission is to promote, educate and advance a diverse Python community through outreach, education, conferences and events.

LANGUAGES

Mother tongue(s) Catalan and Spanish

Other language(s)	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken interaction	Spoken production	
English	C1	C1	C1	C1	C1
German	A2	A2	A2	A2	A2

Levels: A1/A2: Basic user - B1/B2: Independent user - C1/C2: Proficient user
Common European Framework of Reference (CEF) level

PERSONAL SKILLS

Computing Linux, OSX. Python. HTML, CSS, JS.

Scuba diving SSI Stress & Rescue diver. 2020. Barcelona.
SSI React Right. 2020. Barcelona.
PADI Advanced Open Water Certificate. 2018. Thailand.
PADI Open Water Certificate. 2016. Barcelona.

Interests Books. Oceans. Mountains. Photography. Cycling.