Miguel Domingo

Curriculum Vitae

✉ *midobal@prhlt.upv.es*
mdomingo.me
midobal

Education

2015–2022 **Ph.D. in Computer Science**, *Universitat Politècnica de València, Departament de Sistemes Informàtics i Computació.*
I did a Ph.D. in computer science at the *PRHLT Research Center*. It was mainly focused on interactive machine translation and on applying machine translation to some tasks related with the processing of historical documents (language modernization and spelling normalization).

2014–2015 **M.Sc. in Artificial Intelligence, Pattern Recognition and Digital Imaging**, *Universitat Politècnica de València, Departament de Sistemes Informàtics i Computació.*
I did a M.Sc. in artificial intelligence, pattern recognition and digital imaging, focusing on interactive machine translation and automatic post-editing.

2008–2014 **B.Sc. in Computer Science (Computer Science Engineer)**, *Universitat Politècnica de València, Escola Tècnica Superior d' Enginyeria Informàtica.*
I did a B.Sc. in computer science (computer science engineer), focusing on artificial intelligence and formal languages. I spent half a year at *Edinburgh Napier University* taking courses on sound design and security and forensic computing. My Bachelor's thesis was focused on identifying classical melodies from a fragment of it, a hum, a whistle or an a cappella fragment (only for songs with lyrics).

Professional Experience

February 2022–Present **Ph.D. Researcher**, *Pattern Recognition and Human Language Technology (PRHLT) Research Center, Universitat Politècnica de València.*
I have been working on several machine learning research projects, mainly on the field of machine translation.

July 2015–January 2022 **Researcher**, *Pattern Recognition and Human Language Technology (PRHLT) Research Center, Universitat Politècnica de València.*
I worked on several machine learning research projects—mainly on the fields of machine translation and historical documents processing—while doing my Ph.D. studies.

Publications

Journal Articles

[1] Francisco Casacuberta, Alexandru Ceausu, Khalid Choukri, Miltos Deligiannis, **Miguel Domingo**, Mercedes García-Martínez, Manuel Herranz, Guillaume Jacquet, Vassilis Papavassiliou, Stelios Piperidis, Prokopis Prokopidis, Dimitris Roussis, and Marwa Hadj Salah. Findings of the Covid-19 MLIA machine translation task. *Under review at Language Resources and Evaluation*, 2022.

[2] **Miguel Domingo** and Francisco Casacuberta. Modernizing historical documents: A user study. *Pattern Recognition Letters*, 133:151–157, 2020.

[3] **Miguel Domingo** and Álvaro Peris and Francisco Casacuberta. Segment-based interactive-predictive machine translation. *Machine Translation Journal*, 31:163–185, 2017.

[4] Erik Tjong Kim Sang, Marcel Bollmann, Remko Boschker, Francisco Casacuberta, Feike Dietz, Stefanie Dipper, **Miguel Domingo**, Rob van der Goot, Marjo van Koppen, Nikola Ljubešić, Robert Östling, Florian Petran, Eva Pettersson, Yves Scherrer, Marijn Schraagen, Leen Sevens, Jörg Tiedemann, Tom Vanallemeersch, and Kalliopi Zervanou. The CLIN27 shared task: Translating historical text to contemporary language for improving automatic linguistic annotation. *Computational Linguistics in the Netherlands Journal*, 7:53–64, 2017.

[5] Álvaro Peris, **Miguel Domingo**, and Francisco Casacuberta. Interactive neural machine translation. *Computer Speech & Language*, 45:201–220, 2017.

Conference Proceedings

[1] Ángel Navarro, **Miguel Domingo**, and Francisco Casacuberta. PRHLT's submission to WLAC 2022. In *Proceedings of the Seventh Conference on Machine Translation*, pages 1182–1186, 2022.

[2] **Miguel Domingo** and Francisco Casacuberta. An interactive machine translation framework for modernizing the language of historical documents. In *Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis*, pages 41–53, 2022.

[3] **Miguel Domingo** and Francisco Casacuberta. A comparison of character-based neural machine translations techniques applied to spelling normalization. In *Proceedings of the International Conference on Pattern Recognition. International Workshop on Pattern Recognition for Cultural Heritage*, pages 326–338, 2021.

[4] **Miguel Domingo** and Francisco Casacuberta. Two demonstrations of the machine translation applications to historical documents. *arXiv preprint arXiv:2102.01417*, 2021. Presented at the Demos session of ICPR 2020: `https://www.micc.unifi.it/icpr2020/index.php/demos/`.

[5] **Miguel Domingo**, Mercedes García-Martínez, Álvaro Peris, Alexandre Helle, Amando Estela, Laurent Bié, Francisco Casacuberta, and Manuel Herranz. A user study of the incremental learning in NMT. In *Proceedings of the Annual Conference of the European Association for Machine Translation*, pages 319–328, 2020.

[6] **Miguel Domingo** and Francisco Casacuberta. Enriching character-based neural machine translation with modern documents for achieving an orthography consistency in historical documents. In *Proceedings of the International Conference on Image Analysis and Processing. International Workshop on Pattern Recognition for Cultural Heritage*, pages 59–69, 2019.

[7] **Miguel Domingo**, Mercedes García-Martínez, Álvaro Peris, Alexandre Helle, Amando Estela, Laurent Bié, Francisco Casacuberta, and Manuel Herranz. Incremental adaptation of NMT for professional post-editors: A user study. In *Proceedings of the Machine Translation Summit*, pages 219–227, 2019.

[8] **Miguel Domingo**, Mercedes García-Martínez, Amando Estela, Laurent Bié, Alexandre Helle, Álvaro Peris, Francisco Casacuberta, and Manuel Herranz. Demonstration of a neural machine translation system with online learning for translators. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 70–74, 2019.

[9] **Miguel Domingo**, Mercedes García-Martínez, Alexandre Helle, Francisco Casacuberta, and Manuel Herranz. How much does tokenization affect neural machine translation? *arXiv preprint arXiv:1812.08621*, 2018. Accepted for publication at CICLing 2019.

[10] **Miguel Domingo** and Francisco Casacuberta. A machine translation approach for modernizing historical documents using back translation. In *Proceedings of the International Workshop on Spoken Language Translation*, pages 39–47, 2018.

[11] **Miguel Domingo** and Francisco Casacuberta. Spelling normalization of historical documents by using a machine translation approach. In *Proceedings of the Annual Conference of the European Association for Machine Translation*, pages 129–137, 2018.

[12] **Miguel Domingo**, Mara Chinea-Rios, and Francisco Casacuberta. Historical documents modernization. In *Proceedings of the Annual Conference of the European Association for Machine Translation*, pages 295–306, 2017.

[13] **Miguel Domingo**, Álvaro Peris, and Francisco Casacuberta. Interactive-predictive translation based on multiple word-segments. In *Proceedings of the Annual Conference of the European Association for Machine Translation*, pages 282–291, 2016.

Thesis

[1] **Miguel Domingo**. *Some Contributions to Interactive Machine Translation and to the Applications of Machine Translation for Historical*

Documents. PhD thesis, Universitat Politècnica de València, 2022. Advisor: Francisco Casacuberta.

[2] **Miguel Domingo**. Interactive post-editing in machine translation. Master's thesis, Master's Degree in Artificial Intelligence, Pattern Recognition and Digital Imaging. Universitat Politècnica de València, 2015. Advisor: Francisco Casacuberta.

[3] **Miguel Domingo**. Reconocedor automático de melodías de música clásica. Bachelor's thesis, Universitat Politècnica de València, 2014. Advisor: Carlos D. Martínez.

Awards and Honors

2016 19th annual conference of the European Association for Machine Translation (EAMT) best paper award.

Conference Activity

Papers Presented

2022 **An Interactive Machine Translation Framework for Modernizing the Language of Historical Documents**, *10th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, May 4–6.

2021 **Two Demonstrations of the Machine Translation Applications to Historical Documents**, *25th International Conference on Pattern Recognition (ICPR)*, January 10–15.

2021 **A Comparison of Character-based Neural Machine Translation Techniques Applied to Spelling Normalization**, *2nd International Workshop on Pattern Recognition for Cultural Heritage (PatReCH)*, January 11.

2020 **A User Study of the Incremental Learning in NMT**, *22nd annual conference of the European Association for Machine Translation (EAMT)*, November 3–5.

2019 **Enriching Character-Based Neural Machine Translation with Modern Documents for Achieving an Orthography Consistency in Historical Documents**, *1st International Workshop on Pattern Recognition for Cultural Heritage (PatReCH)*, September 9.

2019 **Demonstration of a Neural Machine Translation System with Online Learning for Translators**, *57th Annual Meeting of the Association for Computational Linguistics (ACL)*, July 28–August 2.

2019 **How Much Does Tokenization Affect Neural Machine Translation?**, *20th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*, April 7–13.

2018 **A Machine Translation Approach for Modernizing Historical Documents Using Back Translation**, *15th International Workshop on Spoken Language Translation (IWSLT)*, October 29–30.

2018 **Spelling Normalization of Historical Documents by Using a Machine Translation Approach**, *21st annual conference of the European Association for Machine Translation (EAMT)*, May 28–30.

2017 **Historical Documents Modernization**, *20th annual conference of the European Association for Machine Translation (EAMT)*, May 29–31.

2016 **Interactive-Predictive Translation based on Multiple Word-Segments**, *19th annual conference of the European Association for Machine Translation (EAMT)*, May 30–June 1.

Other Presentations

2022 **Overview of the "Machine Translation" Task**, *Covid-19 MLIA: Round 2 Virtual Meeting*, February 17.

2021 **Covid-19 MLIA task 3**, *CHIST-ERA Challenge Call*, July 21.

2021 **Overview of the "Machine Translation" Task**, *Covid-19 MLIA: Round 1 Virtual Meeting*, January 12–14.

Research Experience

March–August 2022 **Machine Learning Applications**.
Research line from the PRHLT Research Center.
This research line included several research projects ongoing at the PRHLT Research Center on the fields of big data and deep learning, speech processing and dialog systems, handwritten text recognition, computer vision, language translation and natural language processing.
I worked as a researcher on the field of machine translation, mainly in the are of interactive machine translation.

October 2020–February 2022 **Covid-19 MultiLingual Information Access (Covid-19 MLIA)**.
Community effort to boost the development of language resources and multilingual information access systems specifically tailored on Covid-19. It is promoted by several communities which are closely working together. Supported by the European Commission, the European Language Resources Coordination, the European Language Resources Association, the European Research Infrastructure for Language Resources and Technology and the CLEF Initiative.
The machine translation task aims to organize a community evaluation effort with the intents of accelerating the creation of resources and tools for improving the generation of MT systems focused on Covid-19 related documents.
Co-organizer of the machine translation shared task.

December 2019–Present **Deep learning for adaptative and multimodal Interaction in pattern recognition (DeepPattern)**.

Research project supported by the Generalitat Valenciana (GVA) under the reference PROMETEO/2019/121.

This project follows past PROMETEO projects. Its goal is to incorporate deep neural networks and recurrent neural networks to interactive and multimodal pattern recognition from a deep learning approach. And to apply these techniques to several tasks such as interactive machine translation, handwritten text recognition, information retrieval and indexing from images, recognition of mathematical expressions, author profiling and multimodal speech recognition.

I work as a researcher developing new interactive protocols for the processing of historical data.

June–November 2019 **Machine Learning Applications**.

Research line from the PRHLT Research Center.

This research line included several research projects ongoing at the PRHLT Research Center on the fields of big data and deep learning, speech processing and dialog systems, handwritten text recognition, computer vision, language translation and natural language processing.

I worked as a researcher on the field of machine translation and its applications to the processing of historical data.

February 2018–May 2019 **Desarrollo de una Plataforma de Traducción Automática Utilizando Técnicas de Hibridación Sobre un Núcleo de Traducción Basado en Redes Neuronales**.

Research project registered under the project Hybrid Neural Machine Translation Platform, developed by the company Pangeanic with support by the Centro Para el Desarrollo Tecnológico Industrial (CDTI) and the European Union under Programa Operativo de Crecimiento Inteligente (reference IDI-20170964).

The goal of this project was to create a hybrid neural machine translation platform that profits from the company's current machine translation framework to create a new translation core based on neural machine translation that created added value for the professional users and the clients.

I worked as a researcher helping in the development of this platform and incorporating an incremental learning framework.

January 2016–December 2017 **Adaptive Learning and MultimodAlity in MAchine Translation and tExt tRanscription (ALMAMATER)**.

Research project supported by the Generalitat Valenciana (GVA) under the reference PROMETEOII/2014/030.

This project followed past PROMETEO projects based on profiting from the user's feedback on interactive pattern recognition approaches. Its goal was to retrain the system using the data produced during the feedback process to adapt the system to the user's behavior and the task's particularities; explore the multimodality inherent in the interaction in order to improve the overall behavior and usability of the system; and apply the developed techniques to the fields of handwritten text recognition, machine translation and image retrieval.

I worked as a researcher developing new interactive machine translation protocols.

July–December 2015 **Desarrollo de una plataforma tecnológica orientada a la eficiencia de los recursos en el campo de las nuevas tecnologías Internet of Things (SmartWays)**.
Research project supported by the Ministerio de Economía y Competitividad (MINECO) under the reference RTC-2014-1466-4.
This project was devoted to develop a data platform which would allow for implementing a Smart City environment, including features for Smart Light management and Smart Parking management.
I worked as a researcher helping improve the in-car speech recognition and dialog systems for Smart Parking.

Teaching Experience

September 2022–February 2023 **Machine Translation**, *Master's Degree in Artificial Intelligence, Pattern Recognition and Digital Imaging*, Departament de Sistemes Informàtics i Computació - Universitat Politècnica de València.
The goal of this course is to train students in techniques based on automatic learning that allow the construction of translation systems using datasets formed by sentences and their corresponding translations. First, students will study the fundamentals of translators based on statistical alignment models and translators based on dynamic neural networks. Then, they will implement and evaluate automatic translators using public tools.
Teaching assistant.

September 2021–February 2022 **Machine Translation**, *Master's Degree in Artificial Intelligence, Pattern Recognition and Digital Imaging*, Departament de Sistemes Informàtics i Computació - Universitat Politècnica de València.
The goal of this course is to train students in techniques based on automatic learning that allow the construction of translation systems using datasets formed by sentences and their corresponding translations. First, students will study the fundamentals of translators based on statistical alignment models and translators based on dynamic neural networks. Then, they will implement and evaluate automatic translators using public tools.
Teaching assistant.

February 2021–July 2021 **Programming**, *Bachelor's Degree in Data Science*, Escola Tècnica Superior d' Enginyeria Informàtica - Universitat Politècnica de València.
Programming course taught in Python. Its goal is to deep into the study of dictionary-type data structures; dataframes; object-oriented programming; recursivity programming; algorithm analysis; and linear data structures (heaps, queues and lists).
Teaching assistant.

September 2020–February 2021 **Machine Translation**, *Master's Degree in Artificial Intelligence, Pattern Recognition and Digital Imaging*, Departament de Sistemes Informàtics i Computació - Universitat Politècnica de València.
The goal of this course is to train students in techniques based on automatic learning that allow the construction of translation systems using datasets formed by sentences and their corresponding translations. First, students will study the fundamentals of translators based on statistical alignment models and translators based on dynamic neural networks. Then, they will implement and evaluate automatic translators using public tools.
Teaching assistant.

October–November 2020 **Computer Science**, *Bachelor's Degree in Geomatic and Surveying Engineering*, Escola Tècnica Superior d' Enginyeria Geodèsica, Cartogràfica i Topogràfica - Universitat Politècnica de València.
Programming course taught in Python. Its goal is to instill to students the ability to program, since programming is more a skill than a knowledge. The aim is for students to be able to carry out small programs in a high level language based on problem-solving in an informal language. Given the basic fundamentals of programming through this course, students should be able to make programs in other languages/environments after consulting the relevant manuals in a self-taught manner. The teaching of the basic concepts of programming allows its applicability in the context of the career in which it is located.
Teaching assistant.

January–June 2020 **Programming**, *Bachelor's Degree in Data Science*, Escola Tècnica Superior d' Enginyeria Informàtica - Universitat Politècnica de València.
Programming course taught in Python. Its goal is to deep into the study of dictionary-type data structures; dataframes; object-oriented programming; recursivity programming; algorithm analysis; and linear data structures (heaps, queues and lists).
Teaching assistant.

January–June 2020 **Computer Science**, *Bachelor's Degree in Electrical Engineering*, Escola Tècnica Superior d' Enginyeria del Disseny - Universitat Politècnica de València.
Programming course taught in C. Its goal is to introduce students to the fundamentals of computer science. The course will focus mainly on computer-related aspects of computer programming, since the large number of applications that programming has in engineering makes it an essential requirement in the training of every engineer. In addition, students will be introduced to the use of computer applications related to engineering.
Teaching assistant.

Service to the Community

Conference Reviewer and Program Committee Member

2022 The Seventh Conference on Machine Translation (WMT22).

2022 The 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (AACL-IJCNLP 2022).

2022 The 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022).

2022 ACL Rolling Review: January, March and September cycles.

2022 Language Resources and Evaluation Conference (LREC 2022).

2021 ACL Rolling Review: November cycle.

2021 Sixth Conference on Machine Translation (WMT21).

2021 The 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021).

2021 The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021).

2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).

2020 Fifth Conference on Machine Translation (WMT20).

2020 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing (AACL-IJCNLP 2020).

2019 Fourth Conference on Machine Translation (WMT19).

2019 The 17th Machine Translation Summit (MT Summit).

2018 Third Conference on Machine Translation (WMT18).

Journal Reviewer

2022 Artificial Intelligence.

Award Reviewer

2020 Anthony C Clarke Award: EAMT 2019 best thesis award.

Languages

Spanish Native speaker.

Catalan Native speaker.

English Proficiency[1] (listening, reading), advanced[1] (speaking, writing).

First Certificate in English.

French Beginner[1].

[1] *Common European Framework of Reference Level (CEFRL).*

Current Memberships

▷ Asociación Española de Reconocimiento de Formas y Análisis de Imágenes. *Member since 2015.*