

Elena Álvarez Mellado

NLP & IR GROUP
SCHOOL OF COMPUTER SCIENCE
UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA (UNED)

✉ ealvarezmellado@gmail.com
💻 lirondos.github.io
🐦 [@lirondos](https://twitter.com/lirondos)
🐙 github.com/lirondos
in linkedin.com/in/elenaalvarezmellado/

Research interests

Computational Linguistics • Natural Language Processing • Corpus Linguistics • Pop Linguistics

Education

- 2021– *PhD program in Language Technologies*, UNED University (Madrid)
PhD student at the NLP&IR group at UNED School of Computer Science.
- 2018–2020 *Master of Science in Computational Linguistics*, Brandeis University (Grade: 9.56/10)
MS Thesis: *Lázaro: An Extractor of Anglicisms in Spanish Newswire*. Advisor: Constantine Lignos.
Coursework: *Fundamentals of Computational Linguistics; Statistical & Neural Approaches to NLP; NLP Systems: Named-Entity Recognition; NLP in Python; Information Extraction; Linguistic Annotation for Machine Learning; Data Structures & Algorithms; Advanced Programming in Java*.
- 2008–2010 *Licenciatura en Lingüística*, Universidad Complutense de Madrid (Grade: 7.59/10)
Erasmus scholarship at Université Paris Diderot (Paris 7) in 2009.
Final project: *AZRAEL. A-Z Reconocedor Automático de Español*. Advisor: Eduardo Basterrechea.
Coursework: *Phonetics; Morphology; Syntax; Semantics; Pragmatics; Sociolinguistics; Intro to NLP*.

OTHER STUDIES AND COURSES

- 2017–2018 *Máster Universitario en Tecnologías del Lenguaje*, Universidad Nacional de Educación a Distancia
3 courses taken (18 ECTS credits): *Procesamiento de Lenguaje Natural, Minería de información social, Descubrimiento de información en textos*.
- 2005–2008 *Primer Ciclo de la Licenciatura de Medicina*, Universidad Complutense de Madrid (Grade: 7.0/10)
Coursework: *Anatomy; Physiology; Biochemistry; Pathology; Genetics; Pharmacology; Biostatistics*.

Research & Professional Experience

- 07/2021 – RESEARCH STAFF (full time)
School of Computer Science, Universidad Nacional de Educación a Distancia (UNED) – Madrid, Spain
• Staff at the Natural Language Processing and Information Retrieval research group at UNED.
- 06/2020 RESEARCH PROGRAMMER (full time)
06/2021 *Information Sciences Institute, University of Southern California (USC ISI)* – MA, USA
• NLP programmer at USC ISI Center for Vision, Image, Speech and Text Analytics (VISTA).
• Maintained an NLP pipeline for extracting causal relationships in unstructured text.
• Built BERT-based classifiers to detect sentence intent.
• Defined and supervised annotation projects on modality and causality.
- 06/2019 RESEARCH DATA ANALYST (part time)
12/2019 *McLean Hospital, Harvard Medical School* – MA, USA
Supervised by Dr. Mei-Hua Hall, advised by prof. James Pustejovsky
• Applied Machine Learning and NLP techniques to health records to predict readmission risk.

06/2017	RESEARCH ASSISTANT (full time)
05/2018	<i>Digital Humanities Lab, Universidad Nacional de Educación a Distancia (UNED)</i> – Madrid, Spain Supervised by prof. Elena González-Blanco and prof. Salvador Ros <ul style="list-style-type: none"> • Researcher in POSTDATA, ERC project on Poetry Standardization & Linked Open Data • Developed annotation guidelines for Named Entity Recognition in medieval texts.
12/2014	LINGUISTIC DATA ANALYST (part time)
01/2016	<i>Fundación del Español Urgente (Fundéu)</i> – Madrid, Spain <ul style="list-style-type: none"> • Conducted <i>ARACNE</i>, a Corpus Linguistics study on the evolution of the Spanish language on the media during the 20th century.
10/2010	ANALYTICAL LINGUIST (full time)
1/2016	<i>Molino de Ideas</i> – Madrid, Spain <ul style="list-style-type: none"> • Developed, annotated and evaluated linguistic resources for Spanish language.
01/2010	INTERNSHIP STUDENT (part time)
06/2010	<i>Lingway-Eptica</i> – Paris, France <ul style="list-style-type: none"> • Annotated keywords for automatic CV matching.
06/2009	INTERNSHIP STUDENT (part time)
09/2009	<i>Molino de Ideas</i> – Madrid, Spain <ul style="list-style-type: none"> • Provided morphological annotation of a database of Spanish conjugated verbs.
10/2008	RESEARCH ASSISTANT (part time)
03/2009	<i>Department of Linguistics, Universidad Complutense de Madrid</i> – Madrid, Spain
09/2005	RESEARCH ASSISTANT (part time)
02/2006	<i>Department of Cell Biology, Universidad Complutense de Madrid</i> – Madrid, Spain

Skills

Coding: Python (spaCy, NLTK, sklearn, Keras, huggingface), Java. Familiar with: Prolog, MySQL.
Languages: Spanish (native), English (fluent; CEFR: C2; TOEFL: 112; CPE: A), French (basic).

Publications

CONFERENCE PAPERS

2021	Elena Álvarez Mellado. Extracting English Lexical Borrowings from Spanish Newswire. <i>Proceedings of the Society for Computation in Linguistics 4 (SCiL)</i> (peer-reviewed). [DOI]
2020	Elena Álvarez Mellado. Two models for Named Entity Recognition in Spanish. <i>Proceedings of the Iberian Languages Evaluation Forum (IberLEF)</i> (peer-reviewed). [DOI]
2020	Elena Álvarez Mellado. A corpus of Spanish political speeches from 1937 to 2019. <i>Proceedings of The 12th Language Resources and Evaluation Conference (LREC)</i> (peer-reviewed). Marseille, France. ISBN: 979-10-95546-34-4. [DOI] QUALITY INDEX: GGS Rating: B.
2020	Elena Álvarez Mellado. An Annotated Corpus of Emerging Anglicisms in Spanish Newspaper Headlines. <i>Proceedings of the 4th Workshop on Computational Approaches to Code Switching</i> (peer-reviewed), at LREC 2020 (GGS Rating: B). Marseille, France. ISBN: 979-10-95546-66-5. [DOI]
2019	Elena Álvarez Mellado, Eben Holderness, Nicholas Miller, Fyonn Dhang, Philip Cawkwell, Kirsten Bolton, James Pustejovsky and Mei-Hua Hall. Assessing the Efficacy of Clinical Sentiment Analysis and Topic Extraction in Psychiatric Readmission Risk Prediction. <i>Proceedings of the 10th International Workshop on Health Text Mining and Information Analysis</i> (peer-reviewed), colocated with EMNLP 2019 (GGS Rating: A+). Hong Kong, China. ISBN 978-1-950737-77-2. [DOI]

- 2017 **Elena Álvarez Mellado**, Leticia Martín-Fuertes Moreno. Aracne: estudio de la variación lingüística en la prensa española entre 1914 y 2014. In *III Congreso de la Sociedad Internacional de Humanidades Digitales Hispánicas* (peer-reviewed). Málaga, Spain. ISBN: 978-84-697-5692-8. [\[DOI\]](#)

JOURNAL ARTICLES

- 2021 **Elena Álvarez Mellado**, M. Luisa Díez Platas, Pablo Ruiz Fabo, Helena Bermudez, Salvador Ros Muñoz, Elena González-Blanco. TEI-friendly annotation scheme for medieval named entities. *Language Resources and Evaluation*. Springer. ISSN: 1574-0218.
QUALITY INDEX: JCR Impact factor: 1.014. JCR: Q4 (Computer Science). SJR: Q1 (Language and Linguistics). [\[DOI\]](#)
- 2021 María del Carmen Méndez Santos, **Elena Álvarez Mellado**. Origen y comportamiento del formante *-exit* en español. *Tonos Digital. Revista Electrónica de Estudios Filológicos*. Universidad de Murcia. ISSN 1577-6921. SJR: 0.12. SJR: Q3 (Language and Linguistics). [\[DOI\]](#)
- 2020 M. Luisa Díez Platas, Salvador Ros Muñoz, Elena González-Blanco, Pablo Ruiz Fabo, **Elena Álvarez Mellado**. Medieval Spanish (12th–15th centuries) Named Entity Recognition and attribute annotation system based on contextual information. *Journal of the Association for Information Science and Technology*. ISSN: 2330-1643. [\[DOI\]](#)
QUALITY INDEX: JCR Impact Factor: 2.41. JCR: Q2 (Information Science & Library Science). SJR: Q1 (Information Science & Library Science).
- 2018 **Elena Álvarez Mellado** and Leticia Martín-Fuertes Moreno. Aracne: estudio de la variación lingüística en la prensa española entre 1914 y 2014. *Artnodes*, (22), Fundacio per la Universitat Oberta de Catalunya. ISSN: 16955951. [\[DOI\]](#)
QUALITY INDEX: JCR Impact factor: 0.66. JCR: Q2 (Humanities). SJR: 0.13. SJR: Q2 (Literature & Literary Theory).

BOOKS

- 2016 **Elena Álvarez Mellado**, Eduardo Basterrechea, *Anatomía de la Lengua*, Larousse-Vox, ISBN 978-84-9974-216-8, 256 pages.
- 2012 **Elena Álvarez Mellado**, Daniel Ayuso de Santos, Eduardo Basterrechea, Elisa Borsari, José Calvo Tello, César José Gutiérrez, Miguel Ortega, Alejandro López de Pablos, Rubén Pérez Ramón, Luz Rello Sánchez, Aitor Ruiz. *Aprende a conjugar el verbo en español*. Molino de Ideas. ISBN: 978-84-937706-1-7, 140 pages.

BOOK CHAPTERS

- 2022 (in press) **Elena Álvarez Mellado**, Aplicaciones tecnológicas de la Lingüística, in Marqueta, B., Lopez N., Arino A. (eds.), *Avances de la Lingüística y su aplicación didáctica*, Ed. Akal.

Scholarships & Awards

- 2021 [Premio HDH 2021](#) (Hispanic Digital Humanities Association Award)
Awarded by: Asociación de Humanidades Digitales Hispánicas
- Awarded yearly to Digital Humanities projects in the Hispanic world.
 - Awarded to [Observatorio Lázaro](#) as best DH resource created in 2020.

- 2021 *Ayuda para contratos predoctorales FPI*
 Awarded by: UNED University
- Government-funded scholarship for PhD studies at the Natural Language Processing & Information Retrieval group in the School of Computer Science at UNED University.
- 2020 *International Outstanding Corpus Thesis Award (MS level)*
 Awarded by: Institute for Corpus Research, Incheon National University (South Korea)
- Annual award given to recognise candidates that have written an outstanding BA/MSc/PhD thesis in the field of interest of corpus research.
 - 3 international master students selected yearly.
- 2020 *Karen Spärck Jones Award for Outstanding Achievement in NLP*
 Awarded by: Computer Science Department, Brandeis University
- Given yearly to one student among the graduating students from the Master of Science in Computational Linguistics program for excellence in the practice of NLP.
- 2018 *Premio Nacional de Periodismo Miguel Delibes* (Miguel Delibes National Journalism Award)
 Awarded by: Asociación de Prensa de Valladolid, Fundación Miguel Delibes
- Awarded yearly to one individual for journalistic works on language and linguistics.
- 2017 *Beca LaCaixa para estudios de máster en EE.UU.* (LaCaixa fellowship for graduate studies in the US)
 Awarded by: Fundación LaCaixa
- Fully-funded scholarship to pursue master studies in the US.
 - 40 students selected yearly (4% success rate).
- 2009 *VIII Certamen Universitario Arquímedes de Introducción a la Investigación Científica*
 Awarded by: Ministerio de Educación (Spanish Ministry of Education)
- First award at the Arquímedes contest for young researchers.
 - Humanities and Social Sciences category.
- 2005 *Beca de Excelencia*
 Awarded by: Comunidad de Madrid (Madrid regional government)
- Awarded to the best students in Madrid region.

Invited talks & Seminars

- Apr 2021 *Why we lingcomm*, closing plenary at LingComm 2021 (online).
- Nov 2020 *Lingüística computacional en español*, Universität Zürich.
- Jun 2018 *Nacimiento, vida y muerte de una metáfora*, Natural Language Processing & IR Group, UNED.
- Jun 2018 *Aplicaciones del PLN para en humanidades*, Digital Humanities Summer School, UNED.
- Mar 2018 *Radiografía lingüística de la prensa*, Jornadas de Estudios de Lingüística, Universidad de Alicante.
- Mar 2017 *La lematización de textos*, Seminario Filologías Digitales Hoy, Universidad de Extremadura.
- Dic 2016 *Prolog, estructura silábica y reconocimiento automático del idioma*, Lingwars meetup, Madrid.
- Nov 2016 *Lingu ¿qué? Compu ¿cómo?*, Universidad de Alicante.
- Dec 2015 *La lengua a disección*, Fundación Sierra Pambley.
- Oct 2015 *Desayuno con Molino de Ideas*, 3CICTE (Madrid).
- Jun 2015 *A History of Speaking Devices*, APIdays San Francisco, California.
- Mar 2015 *Vivir (de) la lengua*, Universitat Rovira i Virgili.
- Mar 2014 *Cacharreando con la lengua. Recursos informáticos para filología*, Universidad de Zaragoza.
- Jun 2014 *From 1856 to 2014: resurrecting the dictionary with APIs*, APIdays Mediterranea, Barcelona.
- Dec 2013 *Todo lo que un filólogo no sabe que sabe hacer*, grupo Psylex, Universidad de Zaragoza.
- Jun 2013 *Cómo crear una app sin programar*, Euskal Encounter (Bilbao).
- Dec 2012 *Nuevas tecnologías aplicadas a la enseñanza de español*, Universidad de Alcalá de Henares.
- Nov 2011 *Lenguas y Tecnología*, XI Semana de la Ciencia de la Comunidad de Madrid.

Personal Projects & Open Source contributions

- 2020 – [Observatorio Lázaro](#)
- An observatory of anglicism usage in the Spanish press.
 - Awarded with the [Premio HDH](#) (Hispanic Digital Humanities Association Award) in 2021.
- 2018–2019 [Corpus of political speeches from 1937 to 2019](#)
- Analysis, visualizations and Python interface of a corpus of Spanish political speeches.
- 2009–2016 [AZRAEL: A-Z RECONOCEDOR AUTOMATICO DE ESPAÑOL](#)
- An automatic language detector in Prolog based in the syllable structure.
 - Current supported languages: Spanish, French, Italian, Portuguese, Catalan, Latin and Basque.
 - Awarded with the [first award at the Certamen Arquímedes de Introducción a la Investigación Universitaria](#) in 2009.

Dissemination & Outreach

- 2018 – [Revista Archiletras](#) – Columnist & Editorial board member
- [Archiletras](#) is a panhispanic magazine about language and Linguistics directed by Arsenio Escolar and published both in digital and paper format.
 - Contributions: 6 columns published and 3 interviews (including [one interview to Noam Chomsky](#)).
- 2017 – [elDiario.es](#) – Columnist
- Columnist for Spanish newspaper [elDiario.es](#) with a column about Linguistics.
 - 50+ published columns since January 2017.
 - Awarded with the Premio Nacional de Periodismo Miguel Delibes for the column [Metáforas peligrosas: el cáncer como lucha](#) published on October 23rd, 2017.
- 2012–2014 [La noche en vela, Radio Nacional de España](#) – Radio contributor
- Radio contributor at RNE (Spanish National Radio) on a weekly section about Linguistics.
 - Contributed to 50 programs (aprox)

Event organization & Community service

- 2021– [Automatic Detection Of Borrowings shared task \(ADoBo 2021\)](#) – Organization committee
- [ADoBo](#) is a shared task on automatic detection of borrowings.
 - Celebrated as part of IberLEF 2021.
- 2021 [Computational Approaches to Linguistic Code-Switching \(CALCS 2021\)](#) – Program Committee
- CALCS is a yearly workshop on NLP for mixed language data.
 - Celebrated as part of NAACL 2021.
- 2018 – [MariCorners](#) – Scientific committee member
- [MariCorners](#) is a conference on language, communication and LGTB issues.
 - Celebrated yearly at Universidad Politécnica de Madrid.
- 2016–2018 [Lingwars](#) – Organization team
- Madrid-based meetup and virtual community on NLP and Computational Linguistics
- 2013–2015 [Lenguando](#) – Organization team
- Events for language professionals in Madrid, Valencia, Vigo, Logroño, London.
- 2013–2015 [APIdays Mediterranea](#) – Organization team
- Events on API technology in Madrid and Barcelona.

Attended conferences & courses

July 2021	<i>11th Lisbon Machine Learning Summer School</i> (online).
July 2021	<i>Deep Learning for Natural Language Processing</i> , HiTZ research center (online).
Apr 2021	<i>LingComm 2021: Communicating Linguistics to Broader Audiences</i> (online).
Feb 2021	<i>SCiL 2021: Society for Computation in Linguistics conference</i> (online).
Sep 2020	<i>SEPLN 2020: Congreso Anual de la Sociedad Española de Procesamiento de Lenguaje Natural</i> (online).
Jul 2020	<i>ACL 2020: The 58th Annual Meeting of the Association for Computational Linguistics</i> (online).
Jul 2018	<i>Digital Humanities Summer School</i> , ETSII Informatica, UNED.
May 2018	<i>XIV Encuentro de Morfólogos</i> , Facultad de Filología, UNED.
Oct 2017	<i>III Congreso de la Sociedad Internacional Humanidades Digitales Hispánicas</i> , (Málaga).
Oct 2017	<i>Trabalengua 2017: trabajamos con la lengua</i> (Logroño).
Sep 2017	<i>SEPLN 2017: Congreso Anual de la Sociedad Española de Procesamiento de Lenguaje Natural</i> (Murcia).
Jul 2017	<i>2nd Summer Datathon on Linguistic Linked Open Data (SD-LLOD-17)</i> (Cercedilla, Madrid).
Jul 2017	<i>Escuela de Verano Humanidades Digitales</i> (Madrid).
Oct 2015	<i>3CICTE Congreso Internacional de Correctores de Textos</i> (Madrid).
Oct 2012	<i>VII Jornadas MAVIR</i> , Universidad Carlos III (Madrid).

Dissemination articles

2021	Donde nacen las preposiciones , <i>elDiario.es</i>
2021	600 formas de mirar una pandemia , <i>elDiario.es</i>
2021	¿Para qué sirven las lenguas? , <i>elDiario.es</i>
2021	Réquiem por un ‘cuyo’ , <i>elDiario.es</i>
2021	Radiografía del anglicismo en la prensa española , <i>Archiletras</i>
2020	Instrucciones para hacer una pregunta trampa , <i>elDiario.es</i>
2020	Hamburguesa de espinacas, leche de soja , <i>elDiario.es</i>
2020	La belleza de decir “los Rolling” , <i>elDiario.es</i>
2020	Sobre la ortografía , <i>elDiario.es</i>
2020	Un cerdo a la izquierda , <i>elDiario.es</i>
2020	¿El covid o la covid? , <i>elDiario.es</i>
2020	Palabras monógamas , <i>elDiario.es</i>
2020	Entrevista a Noam Chomsky , <i>Archiletras</i>
2020	Entrevista a Gretchen McCulloch , <i>Archiletras</i>
2020	La irresistible agramaticalidad del teclado predictivo , <i>Archiletras</i>
2019	La falacia etimológica , <i>Archiletras</i>
2019	Siete mitos sobre lengua que la Lingüística desmiente , <i>Archiletras</i>
2019	‘Marrona’ y los adjetivos ‘wannabe’ , <i>elDiario.es</i>
2019	De ser un grammar nazi también se sale , <i>Archiletras</i>
2019	Baia baia. La irreverencia ortográfica del meme , <i>elDiario.es</i>
2018	‘Preveyó’ y los cantos de sirena , <i>elDiario.es</i>
2018	Más allá de la economía del lenguaje , <i>elDiario.es</i>
2018	Las trabajadoras y la cooperación , <i>elDiario.es</i>
2018	Palabras de ida y vuelta: ‘sororidad’ , <i>elDiario.es</i>
2018	La RAE y el lujo de definir , <i>elDiario.es</i>
2018	Sobre las “portavozas” , <i>elDiario.es</i>
2018	James Rhodes, ‘ancabuela’ y la mirada extraterrestre , <i>elDiario.es</i>
2018	‘Aprovechategui’ y el arte del insulto , <i>elDiario.es</i>
2018	‘Viejoven’, ‘juernes’, ‘trabacaciones’. La plaga de las palabras siamesas , <i>elDiario.es</i>
2018	Ultras, sí. Pero ¿ultra qué? , <i>elDiario.es</i>

2018 El extraño caso de las décadas sin nombre, [elDiario.es](#)
 2017 De qué hablamos cuando hablamos de ‘consentir’, [elDiario.es](#)
 2017 Las lenguas como castigo, [elDiario.es](#)
 2017 ‘Piolines’: el nacimiento de una palabra, [elDiario.es](#)
 2017 ‘Apología’ y las malas compañías, [elDiario.es](#)
 2017 Madrid, campos de fútbol y otras formas de medir el mundo, [elDiario.es](#)
 2017 “Iros”: el bueno, el feo y el malo, [elDiario.es](#)
 2017 ‘Invent’ y el arte de contar mentiras, [elDiario.es](#)
 2017 Nadie hablará de nosotras cuando nos eliminen del diccionario, [elDiario.es](#)
 2017 Todas, tod@s, todxs, todes: historia de la disidencia gramatical, [elDiario.es](#)
 2017 Elogio de ‘la calor’, [elDiario.es](#)
 2017 Teoría marica o el insulto como bandera, [elDiario.es](#)
 2017 El mito de las palabras que no están en la RAE, [elDiario.es](#)
 2017 “Andó” y el doloroso camino a la irregularidad, [elDiario.es](#)
 2017 Metáforas peligrosas: el cáncer como lucha, [elDiario.es](#), (Premio Nacional de Periodismo Miguel Delibes)
 2017 El poder delator de las comillas españolas, [elDiario.es](#)
 2017 Dejád de pedirle a la RAE que elimine palabras, [elDiario.es](#)
 2017 Comas: el infierno de la puntuación, [elDiario.es](#)
 2017 Cuando “literalmente” ya no es literal, [elDiario.es](#)
 2017 No hablarás con acento andaluz en el telediario de las 9, [elDiario.es](#)
 2017 Señoría, no le entiendo, [elDiario.es](#)
 2017 Hechos alternativos: el peligroso arte de construir realidades con palabras, [elDiario.es](#)
 2017 “Solo” y la tilde de la nostalgia, [elDiario.es](#)

Selected media appearances

2020 *20 anglicismos cada día*. Article by Álex Grijelmo at *El País* about Observatorio Lázaro.
 2020 Radio interview for *La Tarde de la COPE* about Observatorio Lázaro.
 2020 Radio interview for *Julia en la Onda at Onda Cero* about Observatorio Lázaro.
 2020 Interview for *Con la lengua fuera* podcast about Computational Linguistics.
 2020 *En la punta de la lengua*. Radio section by Raúl Urbina for *La SER* about Observatorio Lázaro.
 2019 *Así han evolucionado los discursos de Navidad*. Corpus project featured at [elDiario.es](#)
 2018 Interview at [elDiario.es](#) headquarters (ft. Ignacio Escolar), Asociación de Prensa de Valladolid.
 2018 Speech at the Miguel Delibes National Journalism Award, Asociación de Prensa de Valladolid.
 2018 *Los ordenadores son tontos. El lingüista computacional les permite bregar con el lenguaje humano*. Interview for Universidad Nacional de Educación a Distancia.
 2016 *¿Es la lengua un asunto de los hablantes o de los académicos?*. Interview by Mar Abad for *Yorokobu*.
 2016 Interview for Fundación del Español Urgente (Fundéu) about *Proyecto Aracne*.