



Parte A. DATOS PERSONALES		Fecha del CVA		27/03/2025
Nombre y apellidos	María Teresa Martín Valdivia			
DNI/NIE/pasaporte		Edad		
Núm. identificación del investigador	Researcher ID	ABF-1709-2021		
	Código Orcid	0000-0002-2874-0401		

A.1. Situación profesional actual

Organismo	Universidad de Jaén			
Dpto./Centro	Informática/Escuela Politécnica Superior de Jaén			
Dirección				
Teléfono		correo electrónico		
Categoría profesional	Catedrática de Universidad	Fecha inicio	30/07/2019	
Espec. cód. UNESCO	120304			
Palabras clave	Procesamiento de Lenguaje Natural, Modelos del Lenguaje, Aprendizaje Automático, Recuperación de información, Análisis de Sentimientos y Minería de Emociones, Generación de Recursos Lingüísticos en Español			

A.2. Formación académica (*título, institución, fecha*)

Licenciatura/Grado/Doctorado	Universidad	Año
Diplomada en Informática	Universidad de Granada	1990
Licenciada en Informática	Universidad de Granada	1992
Doctora en Informática	Universidad de Málaga	2004

A.3. Indicadores generales de calidad de la producción científica

Sexenios de investigación: 4 (1995-2003, 2004-2009, 2010-2015, 2016-2021).

Sexenios de transferencia: 1 (1995-2016).

Número de tesis dirigidas (últimos 10 años): 7 (5 con Mención Internacional; 1 premio nacional de investigación 2016 en Procesamiento del Lenguaje Natural; 2 premios extraordinarios de doctorado)

Parte B. RESUMEN LIBRE DEL CURRÍCULUM

M^a Teresa Martín Valdivia es Doctora en informática. Desde 1995 trabaja en la Universidad de Jaén como profesora del área de Lenguajes y Sistemas Informáticos del departamento de Informática. Desde 2019 es Catedrática de Universidad. Desde su creación en el año 2001 pertenece al grupo de investigación Sistemas Inteligentes de Acceso a la Información (SINAI) perteneciente al PAI (Plan Andaluz de Investigación). Su área de especialización es el Procesamiento del Lenguaje Natural y actualmente su investigación se centra en análisis de opiniones y emociones, minería de textos, aprendizaje automático, recuperación y extracción de información centradas en diferentes dominios de aplicación como biomedicina o detección de discurso de odio.

Tiene reconocidos 4 sexenios de investigación (1995-2003; 2004-2009; 2010-2015, 2016-2021) y 1 sexenio de transferencia (1995-2016). Ha publicado más de 200 trabajos nacionales e internacionales, de los que destacan 47 artículos en revistas JCR (25 Q1; 11 Q2; 7 Q3; 4 Q4). También son especialmente importantes los artículos publicados en la revista de la Sociedad Española para el Procesamiento de Lenguaje Natural (38 en total) que está acreditada por la FECYT con el sello de calidad desde el año 2011. Tiene una amplia experiencia en proyectos de investigación siendo investigadora principal de los 3 últimos proyectos del plan estatal de I+D+i desarrollados con su grupo de investigación. También ha participado activamente en los 5 proyectos anteriores consecutivos del plan estatal que su grupo de investigación ha conseguido ininterrumpidamente desde el año 2003. Además, ha sido la investigadora principal y coordinadora del equipo de la Universidad de Jaén en el proyecto europeo FIRST. Actualmente, es IP de otros dos proyectos nacionales, uno en la convocatoria TED y otro en una prueba de concepto. Ha pertenecido a importantes redes de investigación a nivel nacional e internacional, destacando la coordinación de la red temática de excelencia PLN.NET concedida por el Ministerio de Ciencia, Innovación y Universidades (2019-22). En el marco de esta red ha creado la comunidad "DiverTLES: Diversidad en Tecnologías del Lenguaje en España" formada por más de 150 mujeres que trabajan e investigan en PLN y que organiza actividades de promoción y difusión para hacer más visibles a las mujeres en este ámbito. De hecho, está involucrada en varias iniciativas para promover la tecnología en general y el PLN en particular. También tiene una dilatada experiencia en dirección y participación en contratos artículo 83 de la LOU (20 en total). Ha realizado 6 estancias de investigación, todas ellas financiadas con ayudas competitivas a nivel nacional y andaluz. Ha realizado actividades de transferencia mediante asesorías, informes técnicos de evaluación y

posicionamiento de tecnologías, así como desarrollo de recursos. Además, es socia fundadora de la Spin-off de la Universidad de Jaén YottaCode SL que se creó en el año 2016. En el año 2018 YottaCode S.L. obtuvo financiación a través de la convocatoria europea Instrumento PyME por valor de 50.000€.

Además, ha participado en la organización de diferentes congresos y talleres. Asimismo, participa en el comité de programa en distintos congresos y revistas y es revisora de varias revistas de prestigio. Desde 2007 es miembro del consejo asesor de la SEPLN y desde 2007 hasta 2019 perteneció a la junta directiva (tesorera) de la SEPLN.

Desde el 27 de noviembre de 2017 hasta el 29 de abril de 2019 fue Directora del Secretariado de Investigación de la Universidad de Jaén. Desde el 30 de abril de 2019 hasta el 31 de octubre de 2022 ha sido Vicerrectora de Universidad Digital de la Universidad de Jaén. Actualmente, desde el 1 de noviembre de 2022 es gestora de investigación en el panel TIC para la Agencia Estatal de Investigación. Además, desde noviembre de 2021 pertenece al consejo asesor de la Agencia Digital Andaluza (ADA) habiendo participado en el diseño de la estrategia de Inteligencia Artificial para Andalucía. Desde mayo de 2023 forma parte de la comisión de Mujer en Informática de SCIE. Desde octubre de 2023 hasta julio de 2024 ha sido vocal de la comisión 12 de acreditación de profesores titulares de universidad y catedráticos de universidad.

Parte C. MÉRITOS MÁS RELEVANTES (ordenados por tipología)

C.1. Publicaciones (Últimos 10 años)

1. Collado-Montañez, J., López-Úbeda, P., Chizhikova, M., Díaz-Galiano, M. C., Ureña-López, L. A., Martín-Noguerol, T., ... & Martín-Valdivia, M. T. (2024). Automatic text classification of prostate cancer malignancy scores in radiology reports using NLP models. *Medical & Biological Engineering & Computing*, 1-11.
2. Chizhikova, M., López-Úbeda, P., Collado-Montañez, J., Martín-Noguerol, T., Díaz-Galiano, M. C., Luna, A., ... & Martín-Valdivia, M. T. (2023). CARES: A Corpus for classification of Spanish Radiological reports. *Computers in Biology and Medicine*, 154, 106581. (4 citas)
3. Plaza-del-Arco, F. M., Molina-González, M. D., Ureña-López, L. A., & Martín-Valdivia, M. T. (2022). Integrating implicit and explicit linguistic phenomena via multi-task learning for offensive language detection. *Knowledge-Based Systems*, 258, 109965. (5 citas)
4. López-Úbeda, P., Díaz-Galiano, M. C., Martín-Noguerol, T., Luna, A., Ureña-López, L. A., & Martín-Valdivia, M. T. (2021). Automatic medical protocol classification using machine learning approaches. *Computer Methods and Programs in Biomedicine*, 200, 105939. Q1 (23 citas)
5. Plaza-del-Arco, F. M., Molina-González, M. D., Ureña-López, L. A., & Martín-Valdivia, M. T. (2021). Comparing pre-trained language models for Spanish hate speech detection. *Expert Systems with Applications*, 166, 114120. Q1 (97 citas)
6. López-Úbeda, P., Díaz-Galiano, M. C., Ureña-López, L. A., & Martín-Valdivia, M. T. (2021). Combining word embeddings to extract chemical and drug entities in biomedical literature. *BMC bioinformatics*, 22, 1-18. Q2 (6 citas)
7. López-Úbeda, P., Díaz-Galiano, M. C., Martín-Noguerol, T., Luna, A., Ureña-López, L. A., & Martín-Valdivia, M. T. (2020). COVID-19 detection in radiological text reports integrating entity recognition. *Computers in Biology and Medicine*, 127, 104066. Q2 (36 citas)
8. Plaza-del-Arco, F. M., Martín-Valdivia, M. T., Ureña-López, L. A., & Mitkov, R. (2020). Improved emotion recognition in Spanish social media through incorporation of lexical knowledge. *Future Generation Computer Systems*, 110, 1000-1008. Q1 (44 citas)
9. López-Úbeda, P., Díaz-Galiano, M. C., Noguerol, T. M., Ureña-López, L. A., Martín-Valdivia, M. T., & Luna, A. (2020). Detection of Unexpected Findings in Radiology Reports: a Comparative Study of Machine Learning Approaches. *Expert Systems with Applications*, 113647. Q1 (13 citas)
10. Plaza-Del-Arco, F. M., Molina-González, M. D., Ureña-López, L. A., & Martín-Valdivia, M. T. (2020). Detecting misogyny and xenophobia in Spanish tweets using language technologies. *ACM Transactions on Internet Technology (TOIT)*, 20(2), 1-19. Q1 (70 citas)
11. Jiménez-Zafra, S. M., Morante, R., Teresa Martín-Valdivia, M., & Ureña-López, L. A. (2020). Corpora Annotated with Negation: An Overview. *Computational Linguistics*, 46(1), 1-52. Q2 (25 citas)
12. Jiménez-Zafra, S. M., Martín-Valdivia, M. T., Molina-González, M. D., & Ureña-López, L. A. How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain. *Artificial intelligence in medicine*. (2019) Q1 (75 citas)
13. Jimenez-Zafra, S. M., Martin-Valdivia, M. T., Molina-González, M.D., & Ureña-Lopez, L. A. Relevance of the SFU ReviewSP-NEG corpus annotated with the scope of negation for supervised polarity classification in Spanish, *Information Processing & Management*, Vol. 54, Issue 2 (2018). Q1 (9 citas)
14. Jimenez-Zafra, S. M., Martin-Valdivia, M. T., Martinez-Camara, E., & Ureña-Lopez, L. A. Studying the Scope of Negation for Spanish Sentiment Analysis on Twitter. *IEEE Transactions on Affective Computing*, (2017) DOI: 10.1109/TAFFC.2017.2693968. Q1 (62 citas)



15. Jiménez-Zafra, S.M., Taulé, M., Martín-Valdivia, M.T. et al. SFU ReviewSP-NEG: a Spanish corpus annotated with negation for sentiment analysis. A typology of negation patterns. *Language Resources & Evaluation* (2017). Q3 (62 citas)
16. Jiménez-Zafra, S.; Martín Valdivia, M.T.; Martínez-Cámara, E.; Ureña-López, L.A. Combining resources to improve unsupervised sentiment analysis at aspect level. *Journal of Information Science*, Volumen: 42, 213-229 (2016). Q2 (61 citas)
17. Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2015). A Spanish semantic orientation approach to domain adaptation for polarity classification. *Information Processing & Management*, 51(4), 520-531. Q2 (41 citas)
18. Martínez, E., Martín, M. T., Ureña, L. A., & Mitkov, R. (2015). Polarity classification for Spanish tweets using the COST corpus. *Journal of Information Science*. Q2 (48 citas)
19. Barbu, E., Martín-Valdivia, M. T., Martínez-Cámara, E., & Ureña-López, L. A. (2015). Language technologies applied to document simplification for helping autistic people. *Expert Systems with Applications*, 42(12), 5076-5086. Q1 (33 citas)

C.2. Proyectos

1. **PID2021-122263OB-C21**. CONSENSO: CONtEnt for Semantic Nice and SOcial Communication through NLP. Plan nacional I+D+I. 01/09/2022-31/08/2025. IP: L. Alfonso Ureña y **M^a Teresa Martín Valdivia**. 223.003,00€.
2. **TED2021-130145B-I00**. MODERATES: MODERATION of ConTEnts in Social networks using Language Technologies. Plan Estatal de Investigación Científica y Técnica y de Innovación 2021-2023. Ministerio de Ciencia e Innovación. IP: L. Alfonso Ureña y **M^a Teresa Martín Valdivia**. 295.205,00€
3. **PDC2022-133146-C21**. SocialTox: Assessing toxicity in digital media. Plan Estatal de Investigación Científica y Técnica y de Innovación 2021-2023. Ministerio de Ciencia e Innovación. Ministerio de Ciencia e Innovación. IP: L. Alfonso Ureña y **M^a Teresa Martín Valdivia**. 54.855,00€
4. **P20_00956**. **BigHug**: Artificial intelligence for the protection of digital societies. Proyectos de excelencia Junta Andalucía PAIDI 2020. IP. Arturo Montejo Ráez (85.400€) 11/10/21-31/12/22
5. **REF 1380939**. WeLee: Web legible para todos mediante simplificación y enriquecimiento de textos, Proyectos FEDER Andalucía 2014-2020 IP: **M^a Teresa Martín Valdivia** & Arturo Montejo Ráez (42.195,78€) 1/11/2021-31/12/2022. Prorrogado hasta 30/6/2023
6. **RED2018-102418-T**. PLN.NET II: Red de dinamización de actividades en tecnologías de procesamiento del lenguaje natural. Plan nacional I+D+I. IP: **M^a Teresa Martín Valdivia**. 10.000€ 1/01/2020-31/12/2021.
7. **RTI2018-094653-B-C21**. LIVING-LANG: Tecnologías del lenguaje humano para entidades digitales vivas. Plan nacional I+D+I. IP: L. Alfonso Ureña y **M^a Teresa Martín Valdivia** 146.047€ 1/01/2019-31/12/2021. Prorrogado hasta 1/9/2022
8. **TIN2015-65136-C2-1-R**. REDES: Reconocimiento de Entidades Digitales: Enriquecimiento y Seguimiento Mediante Tecnologías del Lenguaje. Plan nacional I+D+I. IP: L. Alfonso Ureña. y **M^a Teresa Martín Valdivia**. 136.972 € 1/01/2016-31/12/2018 (prórroga 31/12/2019).
9. **TIN2012-38536-C03-1**. ATTOS: Análisis de Tendencias y Temáticas a través de Opiniones y Sentimientos. Ministerio de Economía y Competitividad. Plan nacional I+D+I. IP: L. Alfonso Ureña. 1/01/2013-31/12/2015. 71.680,00 €.
10. **TIC 07684**. **AORESCU**: Análisis de Opinión en REdes Sociales y Contenidos generados por Usuarios Proyectos de Excelencia 2010. Consejería de Economía, Innovación, Ciencia y Empleo. Junta de Andalucía. IP: Jose Antonio Troyano. (U.Sevilla). 2012-2016. 151.685,00€
11. **FP7-ICT-287607**. **FIRST: Flexible interactive reading Suport Tool**. European Commision. FP7 STREP. IP: **M^a Teresa Martín** (U. Jaén). 01/10/2011 - 30/09/2014. Total 2.507.218,00€ (UJA 470.000,00€).
12. **TIN2009-13391-C04-02**. **TEXT-COOL 2.0: TEXT PROCESSING FROM COLLABORATIVE AND ON-LINE SOURCES**. Ministerio de Ciencia e Innovación. Plan nacional I+D+I. IP: L. Alfonso Ureña (U. Jaén). 01/01/2010 - 31/12/2013. 85.305,00 €.

C.3. Contratos, méritos tecnológicos o de transferencia

- ONCOPLES: clasificación automática de estudios con sospecha de patología oncológica mediante PLN. TÉCNICAS MÉDICAS MILENIUM S.L. **PI: M.Teresa Martín-Valdivia**. 3/10/2022-3/6/2024. 65.340€
- Proceso de análisis de la contranarrativa, para preparar respuestas al discurso de odio online, dentro del Proyecto Europeo REAL UP. MINISTERIO DE INCLUSIÓN, SEGURIDAD SOCIAL Y MIGRACIONES (OBERAXE). **PI: M.Teresa Martín-Valdivia**. 26/4/2023-31/1/2024. 11.427,77€



- Sistema de información basado en IA para identificación de estructuras empresariales instrumentales (Grupos-Nextgen). Axesor-EXPERIAN. PI: L. Alfonso Ureña. Universidad de Jaén. 24/04/2023-24/10/2023. 143.947,26€
- Detección, extracción de entidades y relaciones en documentos mercantiles AXESOR Conocer para decidir S.A. PI: L. Alfonso Ureña. 1/07/2020-30/04/2021. 96.324,26€.
- Estudio y Desarrollo de un sistema software denominado eBookSync. Grammata (Papyre). IP: Dr. Arturo Montejo Ruez. 4/2010-3/2011. 21.111€.
- Asesoramiento técnico-científico y transferencia de conocimiento al proyecto DELEGA. GUADALTEL S.A. Universidad de Jaén. IP: L. Alfonso Ureña. 10/01/2011-31/12/2011. 50.000 €.

C.4. Patentes

- Martín Valdivia, M.T. y otros “Herramienta basada en procesamiento del lenguaje natural para la asignación automática de protocolos de adquisición en radiología”, N° registro/patente: 2302023325848. Fecha: 13/02/2023. País: España. En explotación.
- Martín Valdivia, M.T. y otros “Herramienta basada en PLN para la detección de hallazgos sugestivos de COVID-19 en informes radiológicos”, N° registro/patente: 2204291032407. Fecha: 26/04/2022. País: España. En explotación.
- Martín Valdivia, M.T. y otros “Herramienta basada en PLN para la detección automática de informes con hallazgos Inesperados”, N° registro/patente: 2102267030764. Fecha: 26/02/2021. País: España. En explotación.
- García Cumbereras M.A., Ureña López L.A., Martín Valdivia, M.T., García M., Martínez F., Díaz M.C., Montejo A. 04/2012/9572. SINTRAM. País: España. Fecha: 24/02/2010. Entidad Titular: Universidad de Jaén. Tipo: Programa de ordenador para traducción.

C.5. Otros méritos

- Tesorera de la Sociedad Española para el Procesamiento del Lenguaje Natural (sep. 2007 - sep 2019).
- Desde 2007 es miembro del consejo asesor de la SEPLN
- Directora Secretariado de Investigación de la Universidad de Jaén (nov 2017-abril 2019)
- Vicerrectora de Universidad Digital de la Universidad de Jaén (abril 2019-nov 2022)
- Gestora de investigación en el panel TIC para la Agencia Estatal de Investigación (nov 2022 y continúa)
- Miembro del consejo asesor de la Agencia Digital de Andalucía (nov 2021 y continúa).
- Miembro en la Comisión de Mujer en Informática de SCIE (mayo 2023 y continúa)
- Miembro en la Comisión 12 de acreditación de profesorado universitario (TU y CU) (oct 2023 -jul 2024)
- Coordinadora y promotora de la comunidad “DiverTLES: Diversidad en Tecnologías del Lenguaje en España”

C.6 Estancias de investigación

- Research Institute for Information and Language Processing. University of Wolverhampton. UK. 1/06/2012-1/09/2012 (Financiada por la Junta de Andalucía) y 1/06/2013-1/09/2013 (Financiada por el Ministerio de Educación, Cultura y Deporte)
- Computational Lexicology & Terminology Lab (CLTL). VU Amsterdam, The Netherlands. 1/05/2016-1/08/2016. Financiada por el Ministerio de Educación, Cultura y Deporte
- Discourse Processing Lab at the Simon Fraser University, Vancouver, Canadá. 1/06/2018-31/08/2018. Financiada por el Ministerio de Educación, Cultura y Deporte
- Grupo de PLN de la Universidad de Córdoba, Argentina. 15/02/23-15/03/23. Financiada por AUIP
- Grupo de PLN de la Universidad de la República, Uruguay. 15/10/24-15/11/25. Financiada por AUIP

C7. Premios

- Primer premio en el hackathon «Neologismos y tecnicismos en el ámbito energético y medioambiental» con el proyecto POEMA (2021)
- 2º premio en el II Hackathon de Tecnologías del Lenguaje en la categoría de ‘Corpus General’ por el prototipo ‘Monitor de dispersión geográfica de enfermedades’ (2018)
- ISCA Award for the best paper published in Computer Speech & Language 2012-2016 por el artículo “Ranked WordNet graph for sentiment polarity classification in Twitter”
- Historias de Luz 2014. Finalistas en el premio a la innovación por la aplicación “OpenBook: Lectura accesible a personas con autismo”.
- Best Poster Award MICAI 2007 -Mexican International Conference on Artificial Intelligence and the Mexican Society of Artificial Intelligence.