

<b>Fecha del CVA</b>	29/04/2024
----------------------	------------

### Parte A. DATOS PERSONALES

Nombre y apellidos	Paloma Martínez Fernández		
DNI/NIE/pasaporte			
Núm. identificación del investigador	Researcher ID		
	Código Orcid	<a href="http://orcid.org/0000-0003-3013-3771">http://orcid.org/0000-0003-3013-3771</a>	

#### A.1. Situación profesional actual

Organismo	Universidad Carlos III de Madrid		
Dpto./Centro	Departamento de Informática		
Dirección	Avda. Universidad 30, 28911 Leganés, Madrid		
Teléfono		correo electrónico	
Categoría profesional	Catedrática de Universidad	Fecha inicio	28/12/2017
Espec. cód. UNESCO	120317 Informática, 120318 Sistemas de Información, 120311 Software de Ordenadores		
Palabras clave	Procesamiento del Lenguaje Natural, Extracción de Información, Accesibilidad Web		

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

Licenciatura/Grado/Doctorado	Universidad	Año
Licenciada en Informática	Universidad Politécnica de Madrid	1992
Doctora en Informática	Universidad Politécnica de Madrid	1998

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

**Nº de sexenios:** 5 (4 de investigación y uno de transferencia). Fecha último sexenio de investigación: 2020)

**Tesis doctorales dirigidas:** 11 tesis (5 premios extraordinarios, 5 con mención europea y 1 premio nacional de investigación 2011 en Procesamiento del Lenguaje Natural)

#### Citas e índice H (consulta 29/04/2024)

Número de citas Google Scholar: 5547

Índice H (Google Scholar): 32

Índice I10 (Google Scholar): 118

### Parte B. RESUMEN LIBRE DEL CURRÍCULUM

Soy Licenciada en Informática por la Universidad Politécnica de Madrid desde el año 1992. Cursé el doctorado en informática en la misma universidad y en 1998 defendí la tesis doctoral. Desde el año 1996 pertenezco al Grupo Human Language and Accessibility Technologies del Departamento de Informática de la Universidad Carlos III de Madrid (ver [hulat.inf.uc3m.es](http://hulat.inf.uc3m.es)).

Durante mi etapa predoctoral trabajé en dos líneas de investigación confluyentes que no he abandonado hasta la fecha, por un lado el procesamiento del lenguaje natural (PLN) y por otro la disciplina de metodologías de desarrollo en Ingeniería del software que culminaron en el año 1998 en mi trabajo de tesis doctoral consistente en la aplicación de técnicas de procesamiento del lenguaje natural aplicadas a sistemas de información, en particular en el modelado conceptual a partir de especificaciones de requisitos en lenguaje natural.

Los proyectos más directamente relacionados con Ingeniería del Software como PANDORA (TIC1999-0215) y GPS (TIN2004-07083) han tenido también una componente relacionada con las tecnologías del lenguaje y en el marco de estos proyectos se han llevado a cabo las tesis doctorales de Dolores Cuadra (2003), Ana M. Iglesias (2004), Harith Al-Jumaily (2006) y José L. Martínez (2010, con mención europea) de las cuales he sido directora.

Desde el año 2005 mi investigación ha continuado en (1) el área del procesamiento del lenguaje natural y (2) enfoques metodológicos para aplicaciones en web accesibles así como accesibilidad a productos y servicios. En la primera de ellas trabajo en la aplicación de tecnologías del lenguaje humano en sistemas de búsqueda de respuestas sobre todo en contenidos en castellano (proyecto BRAVO TIN2007-67407-C03-01, coordinado con el grupo GSI de la UPM y el Laboratorio de Lingüística Computacional de la UAM) y en sistemas de extracción de información especialmente en textos científicos que tratan interacciones entre fármacos así como en notas clínicas (proyectos ISSE FIT-350300-2007-75 y MULTIMEDICA TIN2010-20644-C03-01 proyecto también coordinado con UPM y UAM) y también en el dominio de contenidos audiovisuales (como OPI de la empresa DAEDALUS en el proyecto CENIT BUSCAMEDIA). En la línea de PLN también he aplicado las tecnologías del lenguaje humano en medios sociales en el dominio de salud y farmacovigilancia en el proyecto TrendMiner (FP7-ICT 287863). En esta línea he dirigido las tesis doctorales de Isabel Segura (premio extraordinario de doctorado 2010 con mención europea), César de Pablo (en 2010 con mención europea y premio extraordinario de doctorado), María Herrero (en 2015 con mención europea y premio extraordinario de doctorado) y Julián Moreno Schneider (en 2015 con mención europea y premio extraordinario de doctorado).

En cuanto a la línea de accesibilidad a productos, servicios y a la web, he sido la investigadora principal del proyecto eGovernAbility (TIN2014-52665-C2-2-R). Desde el año 2005 he pertenecido al CESyA (Centro Español de Subtitulado y Audiodescripción, [www.cesya.es](http://www.cesya.es)) liderado por la Universidad Carlos III de Madrid que me ha permitido investigar en accesibilidad a los medios audiovisuales y a la web; proyectos del Plan Avanza como DISUIPA y SAGAS nos han permitido colaborar estrechamente con empresas con fin de trasladar buenas prácticas en desarrollos de aplicaciones accesibles. En esta línea he dirigido la tesis de Lourdes Moreno (premio extraordinario de doctorado 2010).

Fui IP del proyecto “DeepEMR: Extracción de información clínica usando deep learning y técnicas de Big Data” (TIN2017-87548-C2-1-R) cuyo objetivo es la aplicación de técnicas de PLN al procesamiento de la información no estructurada textual de la historia clínica electrónica.

Por otro lado, desde el año 2006 hasta el 2013 el grupo de investigación formó parte del consorcio MAVIR (Mejorando el acceso y la visibilidad de la información multilingüe en red para la Comunidad de Madrid, [www.mavir.net](http://www.mavir.net)), una red de investigación co-financiada por la Comunidad de Madrid y el Fondo Social Europeo. Hemos participado en las ediciones del Cross Language Evaluation Forum (CLEF) desde el año 2004, en el Text Analysis Conference (TAC) desde 2009 y en las tareas de Semantic Evaluation (SEMEVAL) del 2007 y 2010, principalmente trabajando con el idioma español así como en las ediciones de Iberlef 2018 y 2019.

Fuí Directora del Departamento de Informática de la Universidad Carlos III de Madrid desde diciembre de 2017 hasta marzo 2024 y secretaria de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN, [www.sepln.org](http://www.sepln.org)). Formo parte de la [Red de excelencia de Dinamización de Actividades en Tecnologías de Procesamiento del Lenguaje Natural](#)

+info: <http://hulat.inf.uc3m.es>

## **Parte C. MÉRITOS MÁS RELEVANTES** *(ordenados por tipología)*

### **C.1. Publicaciones en revistas JCR**

Se listan a continuación las publicaciones indexadas en Journal Citation Report más recientes:

1. García Subies, Guillem, Álvaro Barbero Jiménez, and Paloma Martínez Fernández. "A comparative analysis of Spanish Clinical encoder-based models on NER and classification tasks." *Journal of the American Medical Informatics Association* (2024), ocae054.
2. Alarcon, Rodrigo, et al. "EASIER system. Evaluating a Spanish lexical simplification proposal with people with cognitive impairments." *International Journal of Human-Computer Interaction* 40.5 (2024): 1195-1209.
3. Rodrigo Alarcon, Lourdes Moreno, Paloma Martínez. EASIER corpus: A lexical simplification resource for people with cognitive impairments. *Plos one*, 2023.
4. Jose Manuel Masiello, Belén Ruiz-Mezcua, Paloma Martínez, Israel Gonzalez-Carrasco. Synchro-Sub, an adaptive multi-algorithm framework for real-time subtitling synchronisation of multi-type TV programmes. *Computing*, 2023
5. Areej Jaber, Paloma Martínez. Disambiguating clinical abbreviations using a one-fits-all classifier based on deep learning techniques. *Methods of Information in Medicine*, 2022
6. Renzo M. Rivera Zavala, Paloma Martínez (2021): Analyzing transfer learning impact in biomedical cross-lingual named entity recognition and normalization. *BMC Bioinform.* 22(1): 601
7. Rodrigo Alarcón, Lourdes Moreno, Paloma Martínez ( 2021): Lexical Simplification System to Improve Web Accessibility. *IEEE Access* 9: 58755-58767
8. Liliya Akhtyamova, Paloma Martínez Fernández, Karin Verspoor, John Cardiff. (2020). Testing Contextualized Word Embeddings to Improve NER in Spanish Clinical Case Narratives. *IEEE ACCESS.* 8, 164717-164726.
9. Jose Luis Lopez-Cuadrado, Israel González Carrasco, Leonardo Lopez Hernandez, Paloma Martínez Fernández, José Luis Martínez Fernández. (2020). Automatic learning framework for pharmaceutical record matching. *IEEE Access.* 8, 171754-171770. *IEEE*, 2169-3536. 2020, Septiembre. 10.1109/ACCESS.2017.D
10. Lourdes Moreno, Paloma Martínez. The Harmonization of Accessibility Standards for Public Policies. *IEEE Computer* 52(7): 57-66 (2019).
11. Victor Suarez Paniagua, Renzo Mauricio Rivera Zavala, Isabel Segura Bedmar, Paloma Martínez Fernández. (2019). A two-stage deep learning approach for extracting entities and relationships from medical texts. *Journal of Biomedical Informatics.* 99. 1532-0480. 2019, Noviembre. <https://doi.org/10.1016/j.jbi.2019.103285>. [Computer Science, Interdisciplinary Applications, 2.95, Q2]
12. Lourdes Moreno, Paloma Martínez, Javier Muguerza, Julio Abascal (2018). Support resource based on standards for accessible e-Government transactional services. *Computer Standards & Interfaces* 58: 146-157 (**JCR Impact Factor (2016):1.633, Q2 (Software Engineering)**).
13. Isabel Segura-Bedmar, Paloma Martínez (2017). Simplifying drug package leaflets written in Spanish by using word Embedings. *Journal of Biomedical Semantics*, in press. (**JCR Science, Impact Factor (2016):1.845, Area: MATHEMATICAL & COMPUTATIONAL BIOLOGY, Rank: 18/57, Q2.**)
14. Víctor Suarez Paniagua, Isabel Segura-Bedmar, Paloma Martínez (2017). Exploring Convolutional Neural Networks for Drug-Drug Interaction Extraction. *DATABASE. Volumen 2017, 1, pp:1-15 (JCR Science, Impact Factor (2016):2.627, Area: MATHEMATICAL & COMPUTATIONAL BIOLOGY, Rank: 8/56, Q1).*
15. Julián Moreno Schneider, Paloma Martínez, José Luis Martínez-Fernández (2017). Combining heterogeneous sources in an interactive multimedia content retrieval model (2017). *Expert Syst. Appl.* 69: 201-213. (**JCR Science, Impact Factor (2016):2.981, Area: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE, Rank: 19/130, Q1**)
16. María Herrero-Zazo, Isabel Segura-Bedmar, Paloma Martínez (2016). Conceptual models of drug-drug interactions: A summary of recent efforts. *Knowl.-Based Syst.* 114: 99-107.

**(JCR Science, Impact Factor (2016): 3,325, Area: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE, Rank: 17/130, Q1).**

17. Paloma Martínez, José L. Martínez, Isabel Segura-Bedmar, Julián Moreno Schneider, Adrián Luna, Ricardo Revert (2016). Turning user generated health-related content into actionable knowledge through text analytics services. Computers in Industry 78: 43-56 **(JCR Science, Impact Factor (2016): 1,685, Area: COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS, Rank: 44/104, Q2).**
18. María Herrero-Zazo, Isabel Segura-Bedmar, Janna Hastings, Paloma Martínez, (2015). DINTO: Using OWL ontologies and SWRL rules to infer drug-drug interactions and their mechanisms. Journal of Chemical Information and Modeling, 55(8): 1698-1707. **(JCR Science, Impact Factor (2015): 3,657, Area: COMPUTER SCIENCE, INFORMATION SYSTEMS, Rank: 6/144, Q1).**
19. Isabel Segura-Bedmar, Paloma Martínez, Ricardo Revert, Julián Moreno-Schneider (2015). Exploring Spanish Health Social Media for detecting drug effects. BMC Medical Informatics and Decision Making, 15(Suppl 2):S6, **(JCR Science, Impact Factor (2015): 2.042, Area: MEDICAL INFORMATICS, Rank: 9/20, Q2).**
20. Martin Krallinger et al (2015). The CHEMDNER corpus of chemicals and drugs and its annotation principles. Journal of Cheminformatics 7(S-1): S2, **(JCR Science, Impact Factor (2015): 3.949, Area:**

## **C.2. Investigadora Principal en Proyectos Competitivos**

Referencia: TIN2017-87548-C2-1-R

Título: DeepEMR: Extracción de información clínica usando deep learning y técnicas de Big Data”

Entidad Financiadora: Ministerio de Economía y competitividad

Duración : 1/01/2018-31/12/2020

Financiación: 116.000 €,

Referencia: FP7-ICT 287863

Título: TRENDMINER: Large-scale, Cross-lingual Trend Mining and Summarisation of Real-time Media Streams

Entidad Financiadora: Unión Europea

Duración: 1/11/2013-31/12/2014

Financiación: 94.000 €,

Referencia: Plan Nacional de I+D+I 2008-2011, TIN2010-20644-C03-01

Título: MULTIMEDICA: Extracción de Información Multilingüe en Sanidad y su aplicación a documentación divulgativa y científica

Entidad Financiadora: Ministerio de Economía y competitividad

Duración: 1/01/2011-30/06/2014

Financiación: 86.152 €,

Referencia: S2009/TIC-1542

Título: MEJORANDO EL ACCESO, EL ANÁLISIS Y LA VISIBILIDAD DE LA INFORMACION Y LOS CONTENIDOS MULTILINGÜES Y MULTIMEDIA EN RED PARA LA COMUNIDAD DE MADRID (MAVIR)

Entidad Financiadora: Comunidad de Madrid

Duración: 1/01/2010-31/05/2014

Financiación: 234.245,95 €,

Referencia: TIN2007-67407

Título: BRAVO: Búsqueda de respuestas avanzada multimodal y multilingüe

Entidad Financiadora: Ministerio de Educación y Ciencia

Duración: 01/10/2007-31/12/2010

Financiación: 113.740 €,