



Universidad Autónoma
de Madrid

MEMORIA DE CREACIÓN

LINTEC

**CENTRO DE INVESTIGACIÓN EN LENGUAJE Y
TECNOLOGÍAS**

(LANGUAGE AND TECHNOLOGY RESEARCH CENTER)

ÍNDICE

ÍNDICE	2
Resumen	3
Introducción: la lingüística en el mundo digital	4
Motivación: el papel de la UAM en el desarrollo de las tecnologías del lenguaje y de la IA	6
Líneas de Investigación	8
Objetivos y actividades de LINTEC	10
Estructura y personal de LINTEC	11
Financiación de las actividades de LINTEC	12
Código ético	13
ANEXO I: Propuesta de nombramientos	14

Resumen

La digitalización de los sistemas de comunicación, de generación de conocimiento y de los productos culturales ha provocado un cambio en cómo nos relacionamos con las lenguas en las sociedades contemporáneas. Algunas innovaciones evidentes son el uso de la escritura como modo de comunicación espontánea a través de aplicaciones de mensajería rápida y la aparición de interlocutores artificiales. Al mismo tiempo, la investigación lingüística se ha adaptado a este nuevo contexto: las innovaciones tecnológicas obligan a estudiar las formas de comunicación novedosas a la vez que se desarrollan aplicaciones que facilitan el análisis de cantidades ingentes de textos.

LINTEC nace con la vocación de aprovechar el momento especial de desarrollos tecnológicos relacionados con la lingüística para poner en valor los conocimientos y las capacidades de la universidad para realizar proyectos en lingüística de corpus y lingüística computacional, y transferir los resultados tanto al tejido productivo como a nivel formativo. De esta forma, LINTEC contribuirá a que la UAM sea un centro de referencia en un ámbito cuya relevancia es ampliamente reconocida tanto por la sociedad como la industria.

Abstract:

The digitalization of communication systems, of knowledge generation and cultural products has led to a profound transformation in the way we interact with languages in contemporary societies. Notable innovations include the use of writing as a means of spontaneous communication through instant messaging applications, as well as the emergence of artificial interlocutors. Concurrently, linguistic research has evolved to address this new landscape: technological advancements not only necessitate the study of emerging forms of communication but also enable the development of tools capable of analyzing vast volumes of textual data.

LINTEC is created with the purpose of capitalizing on the current wave of technological innovation in the field of linguistics, in order to showcase the university's expertise and capabilities in conducting projects in corpus linguistics and computational linguistics. It also seeks to facilitate the transfer of research outcomes to both the productive sector and the field of education. In doing so, LINTEC aims to position UAM as a leading center of excellence in a field whose significance is increasingly acknowledged by both society and industry.

Introducción: la lingüística en el mundo digital

Los desarrollos tecnológicos de las últimas décadas han provocado una digitalización progresiva de la producción lingüística. Esta cubre actualmente todos los ámbitos de nuestras vidas y sociedades, desde lo personal hasta lo laboral, desde la conversación espontánea hasta los comunicados oficiales. Esto está teniendo varias consecuencias importantes en la lingüística. Por un lado, ahora disponemos de cantidades ingentes de datos lingüísticos que facilitan su estudio. No obstante, ha sido necesario para ello el desarrollo de la lingüística de corpus, que utiliza herramientas informáticas para automatizar parte de los análisis. Por otro lado, algunas de las tecnologías más apreciadas en la actualidad se centran precisamente en la comunicación lingüística, desde los traductores automáticos hasta los agentes conversacionales basados en grandes modelos de lenguaje. El conocimiento lingüístico se incorpora a sus equipos de desarrollo a través de lingüistas computacionales.

Ambos aspectos abren grandes posibilidades para la lingüística y las humanidades. En el caso de la lingüística de corpus, la tecnología nos aporta nuevas formas de profundizar en el conocimiento de cómo funcionan las lenguas; en el de la lingüística computacional, es el conocimiento lingüístico el que nos ayuda a mejorar las tecnologías.

El estudio de corpus se aplica a todos los niveles del lenguaje, desde la fonética hasta la pragmática, y también puede hacer aportaciones relevantes al resto de ámbitos de las Humanidades y las Ciencias sociales en los que tengan importancia los textos, como son el estudio de la literatura o la historia. Lo denominamos *humanidades digitales* cuando aplicamos el procesamiento de datos digitales para estudiar fenómenos humanísticos.

En cuanto a la lingüística computacional, hay una clara y creciente demanda en el panorama actual de incorporar especialistas en lingüística en empresas tecnológicas con el objetivo de responder a los retos que plantean la digitalización y la inteligencia artificial. La mejor prueba de ello es el éxito del Máster de formación permanente en Lenguaje e Inteligencia Artificial (LeIA) de la UAM, promovido y desarrollado por miembros de LINTEC. Las empresas que colaboran con el título han puesto de manifiesto desde el principio el problema de no encontrar suficientes perfiles de este tipo para poder completar sus equipos de trabajo. Este título ha recibido el premio a mejor iniciativa formativa del año por la asociación Humanidades Digitales Hispánicas.

Otro ejemplo es la elección por parte de la Comunidad de Madrid de este ámbito como uno de los cuatro de desarrollo más relevante para la región, creando el Clúster de Innovación Tecnológica y Talento de Humanidades Digitales y Tecnologías del Español. La UAM ha estado representada en este clúster desde su creación por medio de miembros de LINTEC por delegación de la Vicerrectora de Investigación y Biblioteca. En el ámbito internacional, este mismo interés ha sido mostrado por la Unión Europea creando dos redes ERIC para esta temática: DARIAH (Digital Research Infrastructure for the Arts and Humanities) y CLARIN

(Common Language Resources and Technology Infrastructure). La UAM también está representada en ambas redes con miembros de LINTEC.

La mencionada irrupción de la digitalización en la comunicación lingüística y en el estudio del lenguaje justifican, tal y como expondremos a continuación, la creación de un Centro de Investigación en Lenguaje y Tecnologías (LINTEC) que aborde tanto el estudio del lenguaje con herramientas de lingüística de corpus como las aportaciones lingüísticas a los nuevos desarrollos tecnológicos, desde una perspectiva transdisciplinar en ambos casos.

Motivación: el papel de la UAM en el desarrollo de las tecnologías del lenguaje y de la IA

La UAM posee una larga experiencia de varias líneas de investigación relacionadas con las tecnologías del lenguaje, que se desarrollan en el seno de diferentes grupos de investigación de la Facultad de Filosofía y Letras y de la Escuela Politécnica Superior. Para ilustrarlo, incluimos a continuación una muestra de las líneas de trabajo desarrolladas por los grupos de investigación de los que forman o han formado parte los miembros de LINTEC:

- Desarrollo de corpus lingüísticos escritos y orales del español peninsular y sus dialectos: grupo “Historia de la Lengua y Dialectología del Español” (HLDE).
- Anotación semántica de corpus: grupo “Discurso y Lengua Española” (DiLE).
- Creación de diccionarios multilingües digitales: grupo “Unidad de estudio de la palabra. Estructura interna y relaciones sintácticas” (UPSTAIRS).
- Digitalización inclusiva: grupo Wor(l)ds Lab.
- Uso de recursos digitales para la enseñanza del español como segunda lengua: grupo “Metodología de la Enseñanza del Español como Lengua Extranjera / L2” (UAM-MELE).
- Sistemas educativos adaptativos: “Grupo de Herramientas Interactivas Avanzadas” (GHIA).
- Respuestas regulatorias a la desinformación: grupo “Política y Gobierno en la Unión Europea” (PolGobUE)

Este desarrollo también se evidencia en las iniciativas desarrolladas hasta el momento por algunos miembros del futuro centro:

- Revista internacional *CHIMERA: Journal of Romance Corpora and Linguistic Studies*, fundada en 2014.
- Máster de Formación Permanente en Lenguaje e Inteligencia Artificial (LeIA), impartido desde el curso 2024-2025.
- Actividades formativas transversales de la Escuela de Doctorado (EDUAM) centradas en el uso de los sistemas de consulta de corpus y de otras herramientas digitales para la elaboración de los trabajos científicos, impartidas anualmente desde el curso 2018-2019.

La importancia de estas líneas de trabajo se demuestra por su inclusión dentro de los ejes estratégicos del Proyecto Estratégico para la Recuperación y Transformaciones Económica (PERTE) de la Lengua, aprobado por el Consejo de Ministros el 1 de marzo de 2022. Por su papel relevante en este ámbito, la UAM está en el proceso de ser incluida en el Registro de Entidades Interesadas en dicho PERTE.

A pesar de su papel relevante en el contexto regional, nacional e internacional en el ámbito de las Humanidades Digitales, la UAM no dispone de un centro de investigación que aglutine las iniciativas de los diferentes grupos de investigación, les dé visibilidad institucional y les proporcione el impulso necesario para optar a fuentes de financiación. En este sentido, cabe señalar que otras universidades públicas madrileñas disponen de entidades que cumplen estas funciones. Por citar solo algunas, se pueden mencionar la Oficina CLARIAH-CM de la Universidad Complutense de Madrid (que lidera el nodo CLARIAH-CM), la Plataforma de Humanidades Digitales y Tecnologías del Lenguaje de la Universidad de Alcalá, o el Instituto de Cultura y Tecnología y el Seminario Permanente en Humanidades Digitales de la Universidad Carlos III de Madrid. La existencia de una estructura similar en la UAM va a conceder ventajas importantes en cuanto a prestigio y competitividad a los grupos e investigadores que trabajan en este ámbito en crecimiento.

En cuanto a la formación, un centro de referencia como el que se propone permitirá ofrecer formación a egresados de estudios de lingüística y filología en conocimientos y competencias del ámbito de las tecnologías del lenguaje, poco presentes en los títulos oficiales, así como transferir estos conocimientos a las empresas a través de formaciones específicas. Sirviendo de puente entre lo lingüístico y lo tecnológico, les abrirá oportunidades al estudiantado en varias direcciones. Por un lado, podrán incorporarse a los equipos de desarrollo tecnológico gracias a los conocimientos y competencias ofrecidas (corpus, PLN, lenguajes de etiquetado, etc.) y a las prácticas en empresas del sector. Por otro, podrán incorporar las herramientas digitales a su trabajo lingüístico, aprovechando las ventajas del tratamiento automatizado de datos para poder estudiar corpus de grandes dimensiones. El *Máster de formación permanente en Lenguaje e Inteligencia Artificial (LeIA)*, en su segunda edición en el curso 2025-2026, es un primer desarrollo en este sentido que ha sido bien acogido tanto por los estudiantes (ha vuelto a tener más candidatos que plazas ofertadas) como por las empresas.

Líneas de Investigación

El escenario descrito anteriormente permite hacerse una idea de las líneas de investigación que tiene previsto desarrollar LINTEC. Las incluimos a continuación y en la sección 4 (“Objetivos y actividades de LINTEC”) nos referiremos a las actividades concretas que se llevarán a cabo dentro de cada una de estas líneas.

- **Lingüística de corpus y uso de tecnologías para la investigación lingüística:** Esta línea abarca el análisis de los textos con las herramientas digitales, la elaboración de modelos teóricos del lenguaje testados en amplios conjuntos de datos empíricos, compilación y anotación de corpus, y el uso de la tecnología en la enseñanza de primeras y segundas lenguas.
- **Integración del conocimiento lingüístico en las tecnologías del lenguaje y los programas de lingüística computacional:** Esta línea se centra en la investigación y el desarrollo de tecnologías del lenguaje desde el punto de vista de la lingüística. Incluye todo el conocimiento lingüístico aplicable a las tareas de procesamiento del lenguaje y a los Grandes Modelos del Lenguaje (LLM) que subyacen en los programas de Inteligencia Artificial. Dicho conocimiento abarca diversas formas de interacción, no limitadas al texto: por ejemplo, la voz, las imágenes y los gestos. Es decir, se adopta una perspectiva multimodal de la interacción comunicativa.
- **Análisis crítico e impacto social de la interacción entre el lenguaje y la tecnología:** Esta línea de trabajo profundiza en los fenómenos provocados por los desarrollos de las tecnologías lingüísticas que afectan al modo en que vivimos y nos comunicamos en las sociedades actuales. Se fundamenta en una mirada ética tanto sobre las innovaciones digitales como sobre los desarrollos descritos en las otras dos líneas de investigación del centro.

Estas líneas de trabajo son las que confieren entidad propia y bien definida al LINTEC: aunque existen en el contexto madrileño, español e internacional otras entidades centradas en las Humanidades Digitales, estos ejes específicos y su combinación se han definido partiendo de las lagunas que han detectado los promotores de LINTEC en las iniciativas existentes y teniendo en cuenta la especialización y las actividades previas de sus miembros.

Los estudios centrados en estas líneas de investigación tendrán difusión en distintas publicaciones científicas con indicadores de impacto que refrendan el interés de esta materia. Puede tratarse de publicaciones tanto generalistas del ámbito de la lingüística –que se basan cada vez más en corpus de distinta naturaleza y utilizan herramientas automáticas para identificar y analizar las estructuras lingüísticas estudiadas–, como revistas específicas que, debido a la importancia de esta área de conocimiento, han visto la luz y, entre las cuales, cabe destacar las siguientes:

- *Computational Linguistics* (MIT Press)
- *Corpus Linguistics and Linguistic Theory* (De Gruyter/ Brill)
- *Corpora* (Edinburgh University Press)
- *International Journal of Computational Linguistics* (CSC Journals)
- *Journal for Language Technology and Computational Linguistics* (Gesellschaft für Sprachtechnologie und Computerlinguistik, GSC)
- *Natural Language & Linguistic Theory* (Springer)
- *Linguistic Issues in Language Technology* (University of Colorado Boulder)
- *Frontiers in AI: Language and Computation* (Frontiers)
- *Revista de la Sociedad Española del Procesamiento del Lenguaje Natural* (SEPLN)
- *Revista Española de Lingüística Aplicada* (Asociación Española de Lingüística Aplicada)
- *CHIMERA: International Journal of Romance Corpora and Linguistic Studies* (UAM y Università di Firenze, codirigida por Manuel Alcántara Pla, miembro de LINTEC)

La trayectoria de los miembros que inician la tramitación y forman parte del centro de investigación propuesto (ver Anexo I) queda reflejada en los currículos de todos ellos que se encuentran publicados en las páginas web de los respectivos departamentos y áreas de conocimiento de la Facultad de Filosofía y Letras, la Facultad de Derecho y la Escuela Politécnica Superior.

Objetivos y actividades de LINTEC

La finalidad de LINTEC es servir como plataforma para fomentar la investigación, la formación, la transferencia, el debate y las publicaciones en los ámbitos en los que coexisten las lenguas y las tecnologías. Con ello se quiere contribuir a que la UAM se identifique como un centro de referencia en este ámbito tanto a nivel nacional como internacional. Para alcanzar estos fines, las principales actividades que LINTEC desarrollará serán las siguientes:

- La promoción y realización de proyectos y contratos de investigación en Lingüística de corpus, Lingüística computacional, y análisis crítico de las tecnologías del lenguaje y la comunicación.
- La difusión de investigaciones, trabajos y materias de interés mediante publicaciones, ciclos de conferencias, seminarios, congresos o cualquier otro medio que se considere oportuno.
- El diseño y desarrollo de actividades orientadas a la transferencia de conocimiento en la forma de enseñanzas propias de especialización, cursos monográficos, talleres y el ya existente Máster de Formación Permanente en Lenguaje e Inteligencia Artificial.
- El asesoramiento científico y técnico a grupos de investigación de cualquier ámbito que quieran incorporar análisis (semi)automático de textos a sus investigaciones.
- El asesoramiento científico y técnico a otras instituciones públicas o privadas en materia de tecnologías del lenguaje.
- La representación de la UAM en el nodo madrileño de la red de infraestructuras CLARIAH-ES, que engloba las redes europeas (ERIC) *Digital Research Infrastructure for the Arts and Humanities* (DARIAH) y *Common Language Resources and Technology Infrastructure* (CLARIN). La representación ha sido llevada por miembros de este centro desde su creación.
- La representación de la UAM en el Clúster de Innovación Tecnológica y Talento (CITT) de Humanidades Digitales y Tecnologías del Español de la Comunidad de Madrid.
- Difusión de las actividades de LINTEC a través de una página web del propio centro de investigación y de las redes sociales para lograr el mayor alcance e impacto posibles de sus actividades.

Estructura y personal de LINTEC

La estructura de LINTEC consta de dos órganos unipersonales (director o directora y secretario o secretaria) y de dos órganos colegiados (el Consejo del Centro y el Comité de Asesoramiento Científico).

Los órganos unipersonales de gestión de LINTEC (director o directora y secretario o secretaria) tienen como misión la administración del centro y velar por el cumplimiento de sus objetivos, observando los principios de transparencia y eficacia.

El Comité de Asesoramiento Científico estará integrado por investigadores y profesionales de reconocido prestigio y se configura como un órgano externo que asesora y evalúa las actividades de LINTEC.

El Reglamento de Régimen interno del centro incluye en sus disposiciones generales los criterios para formar parte de LINTEC y una descripción del procedimiento de inclusión.

Financiación de las actividades de LINTEC

Las actividades de LINTEC serán autofinanciadas. Sus fondos procederán de convenios o contratos de investigación, tanto competitivos como artículo 60 de la Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario, así como de tareas de transferencia de conocimiento con instituciones públicas y privadas en particular, y con la ciudadanía en general.

LINTEC fomentará la investigación de grupos y equipos existentes y de nueva creación para que puedan acceder a convocatorias competitivas de I+D+i nacionales, europeas e internacionales de Administraciones Públicas, así como de Empresas y Fundaciones públicas o privadas en diversos ámbitos de la Economía Digital.

Asimismo, LINTEC aspira a celebrar convenios de colaboración con entidades públicas y privadas vinculadas al área del procesamiento del lenguaje natural y la IA en base a los cuales estas entidades puedan actuar como patrocinadoras de actividades promovidas por el centro.

En todo caso, el régimen de financiación del centro se ajustará a las obligaciones previstas en el Reglamento de Institutos Universitarios y Centros de Investigación propios de la UAM aprobado en el Consejo de Gobierno de la UAM de 27 de marzo de 2020 y modificado mediante el Acuerdo de Consejo de Gobierno de la UAM de 12 de marzo de 2021 y Acuerdo de Consejo de Gobierno de la UAM de 12 de noviembre de 2021.

Código ético

LINTEC asume el compromiso del estricto cumplimiento de todas las normas de comportamiento ético elaboradas por la Universidad Autónoma de Madrid relativas a la gestión y el funcionamiento de centros de investigación (véase el Acuerdo 20/CG de Consejo de Gobierno de 16-07-2020, por el que se aprueba el Código Ético de la UAM).

ANEXO I: Propuesta de nombramientos

1. Dirección del Centro

Olga Batiukova Belotserkovskaya. Profesora Titular de Universidad. Facultad de Filosofía y Letras. Departamento de Filología Española.

2. Secretaría del Centro

Se elegirá un/a secretario/a a propuesta del Consejo del Centro, siempre que su estatus laboral no incumpla la normativa legal vigente.

3. Consejo del Centro

Formado por los miembros permanentes y no permanentes de LINTEC.

3.1. Miembros Permanentes

Miembros que inician el procedimiento de aprobación:

- Manuel Alcántara Plá. Profesor Titular de Universidad. Facultad de Filosofía y Letras. Departamento de Lingüística General, Lógica y Filosofía de la Ciencia, Lenguas Modernas, Teoría de la Literatura y Literatura Comparada y Estudios de Asia Oriental.
- Olga Batiukova Belotserkovskaya. Profesora Titular de Universidad. Facultad de Filosofía y Letras. Departamento de Filología Española.

Otros miembros permanentes:

- Luis Bouza García. Profesor Titular de Universidad. Facultad de Derecho. Departamento de Ciencia Política y Relaciones Internacionales.
- Iván Cantador Gutiérrez. Profesor Titular de Universidad. Escuela Politécnica Superior. Departamento de Ingeniería Informática.
- Rosa Castro Prieto. Profesora Contratada Doctora. Facultad de Filosofía y Letras. Departamento de Filología Española.
- Inés Fernández-Ordóñez Hernández. Catedrática de Universidad. Facultad de Filosofía y Letras. Departamento de Filología Española.

- Jacinto González Cobas. Profesor Titular de Universidad. Facultad de Filosofía y Letras. Departamento de Filología Española.
- Francisco Jurado Monroy. Profesor Titular de Universidad. Escuela Politécnica Superior. Departamento de Ingeniería Informática.
- Sebastián López Maza. Profesor Titular de Universidad. Facultad de Derecho. Departamento de Derecho Privado, Social y Económico.
- Gemma Minero Alejandre. Profesora Titular de Universidad. Facultad de Derecho. Departamento de Derecho Privado, Social y Económico.
- Antonio Revuelta Puigdollers. Profesor Titular de Universidad. Facultad de Filosofía y Letras. Departamento de Filología Clásica.
- Ana Ruiz Sánchez. Profesora Titular de Universidad. Facultad de Filosofía y Letras. Departamento de Lingüística General, Lógica y Filosofía de la Ciencia, Lenguas Modernas, Teoría de la Literatura y Literatura Comparada y Estudios de Asia Oriental.
- Santiago Urbano Sánchez Jiménez. Profesor Titular de Universidad. Facultad de Filosofía y Letras. Departamento de Filología Española.
- Jesús Vega Encabo. Catedrático de Universidad. Facultad de Filosofía y Letras. Departamento de Lingüística General, Lógica y Filosofía de la Ciencia, Lenguas Modernas, Teoría de la Literatura y Literatura Comparada y Estudios de Asia Oriental.

3.2. Miembros no permanentes

1. Carlota de Benito Moreno. Investigadora del Programa de Atracción de Talento César Nombela de la Comunidad de Madrid, UAM.
2. Alejandro Flores Moleón. Personal Docente e Investigador en Formación. Facultad de Filosofía y Letras y Facultad de Derecho, UAM.
3. Filip Zeman. Investigador del Programa Ayudas para Contratos Predoctorales para la Formación de Doctores (FPI del Ministerio de Ciencia, Innovación y Universidades), UAM.

4. Comité de Asesoramiento Científico

- Elena Battaner Moro. Catedrática de Universidad. Departamento de Artes y Humanidades. Facultad de Artes y Humanidades. Universidad Rey Juan Carlos (Madrid, España).
- Alessandro Lenci. Catedrático de Universidad (Professore Ordinario). Dipartimento di Filologia, Letteratura e Linguistica. Univesità di Pisa (Pisa, Italia).
- Rogelio Nazar, Grupo “Technologies for Linguistic Analysis” (Tecling).
- Elena Montiel Ponsoda. Profesora Titular de Universidad. Departamento de Lingüística Aplicada a la Ciencia y la Tecnología. Universidad Politécnica de Madrid.
- James Pustejovsky. Catedrático de Universidad (Full Professor). Computer Science Departament. Brandeis University (Massachusetts, Estados Unidos de América).
- Irene Renau Araque. Lectora. Facultad de Filosofía y Letras. Departamento de Filología Española (Barcelona, España). Miembro correspondiente de la Academia Chilena de la Lengua.
- Cristina Vela Delfa. Catedrática de Universidad. Departamento de lengua española. Universidad de Valladolid (Valladolid, España).