

Plan TL

Estrategia Formativa en TL

Aitziber Atutxa (UPV/EHU), Nuria Bel (UPF), Arantza Díaz de Ilarraza (UPV/EHU),
Patricio Martínez (UA), Raquel Martínez (UNED), Koldo Gojenola (UPV/EHU),
Antoni Oliver (UOC), German Rigau (UPV/EHU), Alfonso Ureña (UJ)

Diciembre 2019



GOBIERNO
DE ESPAÑA

VICEPRESIDENCIA
TERCERA DEL GOBIERNO

MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN
DIGITAL

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

Plan TL

Plan de Impulso de las
Tecnologías del Lenguaje







Este estudio ha sido realizado dentro del ámbito del Plan de Impulso de las Tecnologías del Lenguaje con financiación de la Secretaría de Estado para el Avance Digital, que no comparte necesariamente los contenidos expresados en el mismo. Dichos contenidos son responsabilidad exclusiva de sus autores.

Reservados todos los derechos. Se permite su copia y distribución por cualquier medio siempre que se mantenga el reconocimiento de sus autores, no se haga uso comercial de las obras y no se realice ninguna modificación de estas.

ESTRATEGIA FORMATIVA EN TECNOLOGÍAS DEL LENGUAJE.....	5
ET1: OFERTA FORMATIVA EN TL A NIVEL NACIONAL, EUROPEO E INTERNACIONAL	14
ET2: DIRECTRICES GENERALES, PROPUESTA DE TEMÁTICAS Y ACCIONES FORMATIVAS PARA LA ESTRATEGIA DE FORMACIÓN EN TL.....	70
ET3: IDENTIFICACIÓN DE SECTORES PARA LA ESTRATEGIA DE FORMACIÓN TL. DIRECTRICES GENERALES PARA ABORDAR LA ESTRATEGIA DE FORMACIÓN.	88
ET4: DISEÑO DE LA ESTRATEGIA DE FORMACIÓN EN TL ESPECÍFICA PARA EL SISTEMA UNIVERSITARIO ESPAÑOL.....	113
ET5. IMPLANTACIÓN DE LA ESTRATEGIA DE FORMACIÓN EN TL ESPECÍFICA PARA EL SISTEMA UNIVERSITARIO ESPAÑOL.....	181

Estrategia formativa en Tecnologías del Lenguaje

INTRODUCCIÓN

El Gobierno de España puso en marcha en 2015 el Plan de Impulso de las Tecnologías del Lenguaje (Plan TL), un programa específico orientado al impulso de estas tecnologías en el marco de la Agenda Digital para España. El objetivo general del plan es fomentar el desarrollo del procesamiento del lenguaje natural, la traducción automática y los sistemas conversacionales en lengua española y lenguas cooficiales.

Para ello, establece medidas encaminadas a aumentar el número, calidad y disponibilidad de las infraestructuras lingüísticas en español y lenguas cooficiales; impulsar la Industria del lenguaje fomentando la transferencia de conocimiento entre el sector investigador y la industria; e incorporar a la Administración como impulsora del sector de procesamiento de lenguaje natural y la traducción automática.

El Plan TL se estructura en cuatro ejes. Uno de ellos, el eje II, de Impulso a la Industria de las tecnologías del lenguaje, tiene como objetivo apoyar la transferencia de conocimiento entre el sector investigador y la industria, así como favorecer la internacionalización de las empresas e instituciones que componen el sector. Como parte de dicho eje, desde Red.es y la SEAD (Secretaría de Estado para el Avance Digital) se enmarca la elaboración de estudios sobre la caracterización y análisis de la situación del Sector de las Tecnologías del Lenguaje.

Entre sus líneas de actuación, se encuentra una línea de acción específica para la formación y divulgación de las tecnologías del lenguaje que se enmarca en el eje 2 y en particular en la acción II.1 “Mejora de la visibilidad y de la transferencia del sector” en el que se propone: A) Medida II.1.1- “Elaborar un plan de mejora de la visibilidad y la transferencia”; Medida II.1.2- “Elevar la formación: Inclusión de asignaturas específicas sobre tecnologías del lenguaje en el CV académico, impulso de la creación de contenidos formativos en línea (MooCs), *hackatones*, apoyo a programas de doctorado industrial, másteres y becas especializadas”. Enlazando con esa línea de actuación, el objetivo principal de este informe es:

- Identificación de los sectores educativos en los que definir una estrategia específica de formación en TL estableciendo recomendaciones para abordar, posteriormente, las respectivas estrategias en cada sector educativo.

- En referencia al sistema universitario, diseño de una estrategia formativa específica integrando los diferentes componentes de las TL.

Se trata de un estudio preliminar para la definición de un Libro Blanco para la formación en tecnologías del lenguaje en el estado español en el que se diseñe una estrategia formativa en TL específica para el sistema universitario.

Para ello, se han analizado detalladamente las iniciativas de formación existentes en España, Europa y en otros países de la OCDE en lo referente a la existencia de: i) asignaturas específicas sobre tecnologías del lenguaje en diferentes currículos académicos; ii) contenidos formativos en línea (MooCs); iii) programas de doctorado; iv) másteres; v) títulos propios universitarios (Especialista o Máster) y vi) becas especializadas. Además, se realizará una comparativa con la situación a nivel europeo y de los principales países del resto del mundo.

En concreto, de este informe se pretende obtener:

- Identificación de los sectores educativos en los que conviene definir una estrategia específica de formación en TL estableciendo recomendaciones para abordar, posteriormente, las respectivas estrategias en cada sector educativo. El objetivo trata de identificar los diferentes perfiles de usuario y colectivos que se corresponden con cada uno de los grupos educativos seleccionados.
- En lo que se refiere al sistema universitario, diseño de una estrategia formativa específica integrando los diferentes componentes de las TL: traducción automática, sistemas conversacionales, aprendizaje automático, gestión de contenidos, etc. aplicadas a grados, másteres y doctorados en las áreas de conocimiento relacionadas: informática, lingüística, telecomunicaciones, matemáticas, traducción, etc.

De esta forma se pretenden establecer las **primeras directrices generales de la estrategia global de formación en TL** en el estado español diseñando una estrategia formativa específica para el sistema universitario. Para ello hemos contemplado la realización de las siguientes tareas:

- Tarea 1: **Análisis previo:**
 - Estudio de la oferta formativa actual en TL a nivel nacional, europeo e internacional en países de referencia (EEUU, Reino Unido, Alemania, Países Bajos, Francia, Portugal, países latinoamericanos, etc.)



- Estudio de otras experiencias similares que hayan abordado la inclusión de nuevas temáticas en los diversos planes formativos en España (por ejemplo, las que hayan podido surgir sobre la temática de ciberseguridad, ...).
 - Estudio de la demanda de los empleadores mediante encuestas a los coordinadores de *clusters* del TL del entorno de cada uno de los participantes (País Vasco, Comunidad de Madrid, Andalucía, Comunidad Valenciana y Cataluña)
 - Estudio de informes internacionales de prestigio del ámbito que nos ocupa.
- Tarea 2: **Propuesta de temática**, acciones formativas y directrices generales para la estrategia de formación en el ámbito del plan TL.
 - Tarea 3: **Identificación de los sectores** a los que interesa la formación a nivel universitario de las TL (empresas, instituciones, etc.), así como directrices y recomendaciones para abordarlas en posibles estudios posteriores.
 - Tarea 4: **Diseño detallado de la estrategia de formación** en TL en el ámbito universitario.
 - Tarea 5: **Plan de acción a corto y medio plazo** para el impulso de la formación en TL en el sector universitario que sirva de guía detallada de prioridades, requisitos y plazos aconsejados para poner en marcha la iniciativa.

Los entregables técnicos intermedios que se han desarrollado para elaborar este **informe final** son los siguientes:

ET1: Resultado del estudio realizado en la tarea 1. Este resultado incluye:

Catálogo de las ofertas formativas de TL a nivel nacional, europeo e internacional. Para cada oferta identificada se indicará: i) localización; ii) estructura; iii) perfil alumnado; iv) tipo y nivel de formación; v) modo (presencial, *online*, *blended*). Resumen de recomendaciones derivadas de experiencias previas similares.

ET2: Propuesta de temáticas, acciones formativas y directrices generales para la estrategia de formación en TL.

ET3: Identificación de sectores de mayor interés en los que convendría definir una estrategia de formación específica en TL y metodología utilizada para la identificación de los sectores de

interés. Además incluirá directrices y recomendaciones para abordarlas en cada sector en estudios posteriores.

ET4: Diseño de la estrategia de formación en TL específica para el sistema universitario español en el que se incluirán las competencias y resultados de aprendizaje que deberían asegurarse al final de cada ciclo formativo propuesto; Además se incluirán: i) diseños preliminares de cada propuesta; ii) requisitos y trámites para ponerla en marcha; iii) modalidad de impartición; iv) infraestructura y herramientas necesarias; v) duración de cada propuesta formativa; v) créditos ECTS; vi) perfil del grupo destinatario; vii) nivel y tipo de formación; viii) tipo de certificado: propio, nacional, conjunto, etc.; ix) perfil del profesorado.

ET5: Plan de acción a corto y medio plazo para la formación en TL en el sector universitario. El objetivo de este plan de acción es servir de guía detallada de prioridades, requisitos y plazos aconsejados para poner en marcha cada iniciativa.

El presente informe final ha sido estructurado de la siguiente forma¹:

- A. El estudio preliminar (ET1, ET2 y ET3)
- B. Diseño de la estrategia (ET4)
- C. Plan de acción (ET5)

Los editores de este informe hemos tratado de trabajar con información lo más amplia y actual posible para poder incorporar el mayor número de perspectivas relevantes posible en el tema que nos ocupa; visiones y miradas referidas a áreas de conocimiento con las que las TLs están relacionadas: informática, lingüística, matemática, traducción, telecomunicaciones, etc. Para el ámbito universitario y académico, pero siempre pensando también en el mundo empresarial, la administración o la ciudadanía en general.

¹ Cada uno de los entregables incluye sus correspondientes anexos.



Estrategia Formativa en TL

ET1: Oferta formativa en TL a nivel nacional, europeo e internacional

Aitziber Atutxa (UPV/EHU), Nuria Bel (UPF), Arantza Díaz de Ilarraza (UPV/EHU),
Patricio Martínez (UA), Raquel Martínez (UNED), Koldo Gojenola (UPV/EHU),
Antoni Oliver (UOC), German Rigau (UPV/EHU), Alfonso Ureña (UJ)

Julio 2019

Este estudio ha sido realizado dentro del ámbito del Plan de Impulso de las Tecnologías del Lenguaje con financiación de la Secretaría de Estado para el Avance Digital, que no comparte necesariamente los contenidos expresados en el mismo. Dichos contenidos son responsabilidad exclusiva de sus autores.

Reservados todos los derechos. Se permite su copia y distribución por cualquier medio siempre que se mantenga el reconocimiento de sus autores, no se haga uso comercial de las obras y no se realice ninguna modificación de estas.

1. INTRODUCCIÓN	14
2. OBJETIVOS Y ORGANIZACIÓN DEL DOCUMENTO	15
3. METODOLOGÍA DE TRABAJO	15
4. CATÁLOGO DE ESTUDIOS UNIVERSITARIOS.....	17
5. OTRAS EXPERIENCIAS SIMILARES QUE HAYAN ABORDADO LA INCLUSIÓN DE NUEVAS TEMÁTICAS	24
6. RECOMENDACIONES DE ENTIDADES NACIONALES Y EUROPEAS SOBRE PLN E IA	25
7. ESTUDIO DE LA DEMANDA DE LOS EMPLEADORES	26
7.1 ANÁLISIS DE OFERTAS DE EMPLEO EN TL	26
7.2 ANÁLISIS DE ENCUESTAS SOBRE NECESIDADES DE EMPRESAS EN TL	30
8. CONCLUSIONES.....	36
9. REFERENCIAS	38
10. GLOSARIO DE SIGLAS Y ACRÓNIMOS.....	39
11. ANEXOS	40

ÍNDICE DE FIGURAS

FIGURA 1. DISTRIBUCIÓN DE PERFILES DE OFERTAS DE EMPLEO.....	27
FIGURA 2. PERFILES DE TITULACIONES DE GRADO PARA EXPERTOS EN TL.	28
FIGURA 3. CATEGORÍAS DE ESTUDIOS REQUERIDAS.	29
FIGURA 4. SECTORES QUE DEMANDAN FORMACIÓN EN TL.	30
FIGURA 5. PROPORCIÓN DE RESPUESTAS DE ENTIDADES DE TL.	30
FIGURA 6. DISTRIBUCIÓN DE ENTIDADES TL.....	31
FIGURA 7. TAMAÑO DE ENTIDADES DE TL.	31
FIGURA 8. DISTRIBUCIÓN DE EXPERTOS EN TL.....	32
FIGURA 9. DIFICULTAD PARA CONTRATAR EXPERTOS EN TL (1: NINGUNA DIFICULTAD, 5: DIFICULTAD ALTA).....	33
FIGURA 10. AYUDAS A CONTRATOS DE INVESTIGADORES EN EMPRESAS.....	34
FIGURA 11. INTERÉS EN EVENTOS DE TL.	35

ÍNDICE DE TABLAS

TABLA 1. INFORMACIÓN APORTADA POR CADA TITULACIÓN.....	16
TABLA 2. GRADOS Y ASIGNATURAS SOBRE PROCESAMIENTO DEL LENGUAJE NATURAL EN ESPAÑA.....	17
TABLA 3. GRADOS EN PROCESAMIENTO DE LENGUAJE NATURAL EN ESTADOS UNIDOS.	18
TABLA 4. MÁSTERES EN PROCESAMIENTO DEL LENGUAJE NATURAL EN ESPAÑA.	19
TABLA 5. MÁSTERES EN PROCESAMIENTO DEL LENGUAJE NATURAL EN EUROPA Y EL RESTO DEL MUNDO.....	20
TABLA 6. MÁSTERES EN INTELIGENCIA ARTIFICIAL EN ESPAÑA Y OTROS PAÍSES DE HABLA HISPANA.	21
TABLA 7. MÁSTERES EN INTELIGENCIA ARTIFICIAL EN EUROPA Y RESTO DEL MUNDO.....	22
TABLA 8. MÁSTERES EN BIG DATA O CIENCIA DE DATOS.....	22
TABLA 9. PROGRAMAS DE DOCTORADO EN PROCESAMIENTO DEL LENGUAJE NATURAL EN ESPAÑA.	23
TABLA 10. CURSOS MOOC EN PROCESAMIENTO DEL LENGUAJE NATURAL.	24
TABLA 11. TIPOLOGÍA DE PERFILES REQUERIDOS DE TL.	27
TABLA 12. <i>COMPETENCIAS REQUERIDAS EN EMPLEOS DE TL</i>	34

ET1: Oferta formativa en TL a nivel nacional, europeo e internacional

1. INTRODUCCIÓN

El Gobierno de España puso en marcha en 2015 el Plan de Impulso de las Tecnologías del Lenguaje (Plan TL), un programa específico orientado al impulso de estas tecnologías en el marco de la Agenda Digital para España. El objetivo general del plan es fomentar el desarrollo del procesamiento del lenguaje natural, la traducción automática y los sistemas conversacionales en lengua española y lenguas cooficiales.

Para ello, establece medidas encaminadas a aumentar el número, calidad y disponibilidad de las infraestructuras lingüísticas en español y lenguas cooficiales; impulsar la Industria del lenguaje fomentando la transferencia de conocimiento entre el sector investigador y la industria; e incorporar a la Administración como impulsor del sector de procesamiento de lenguaje natural y la traducción automática.

El Plan TL se estructura en cuatro ejes. Uno de ellos, el eje II, de Impulso a la Industria de las tecnologías del lenguaje, tiene como objetivo apoyar la transferencia de conocimiento entre el sector investigador y la industria, así como favorecer la internacionalización de las empresas e instituciones que componen el sector. Como parte de dicho eje, desde Red.es y la SEAD (Secretaría de Estado para el Avance Digital) se enmarca la elaboración de estudios sobre la caracterización y análisis de la situación del Sector de las Tecnologías del Lenguaje.

Entre sus líneas de actuación, se encuentra una línea de acción específica para la formación y divulgación de las tecnologías del lenguaje que se enmarca en el eje 2 y en particular en la acción II.1 “Mejora de la visibilidad y de la transferencia del sector” en el que se propone: A) Medida II.1.1- “Elaborar un plan de mejora de la visibilidad y la transferencia”; Medida II.1.2- “Elevar la formación: Inclusión de asignaturas específicas sobre tecnologías del lenguaje en el CV académico, impulso de la creación de contenidos formativos en línea (MooCs), *hackatones*, apoyo a programas de doctorado industrial, másteres y becas especializadas”. Enlazando con esa línea de actuación, el objetivo principal de este informe es:

- Identificación de los sectores educativos en los que definir una estrategia específica de formación en TL estableciendo recomendaciones para abordar, posteriormente, las respectivas estrategias en cada sector educativo.

- En referencia al sistema universitario, diseño de una estrategia formativa específica integrando los diferentes componentes de las TL.

2. OBJETIVOS Y ORGANIZACIÓN DEL DOCUMENTO

El objetivo de este documento es establecer un análisis previo para las directrices generales de la formación en TL en España, diseñando una estrategia formativa específica para el sistema universitario, concretamente:

1. Presentar el resultado de la oferta formativa actual principalmente en el ámbito universitario en el área de las tecnologías lingüísticas a nivel nacional, europeo e internacional en países de referencia (EEUU, Reino Unido, Alemania, Países Bajos, Francia, Portugal, países latinoamericanos, etc.)
2. Estudio de otras experiencias similares que hayan abordado la inclusión de nuevas temáticas en los diversos planes formativos en España. Concretamente, se ha examinado la propuesta temática de ciberseguridad.
3. Estudio de las ofertas de empleo en el ámbito de las TL, así como la demanda de los empleadores mediante encuestas a los coordinadores de clusters del TL del entorno de cada uno de los participantes (País Vasco, Comunidad de Madrid, Andalucía, Comunidad Valenciana y Cataluña).

3. METODOLOGÍA DE TRABAJO

La metodología de trabajo que se ha seguido para la realización del presente estudio se ha basado en:

1. La recogida de información a través de lo que cada Universidad, Centro o Institución ha publicado en su web
2. Entrevistas y contactos personales con expertos en la materia a nivel nacional, europeo e internacional.

Como resultado de esta recopilación de información, presentaremos en el apartado 3 un censo de grados, másteres y cursos de formación que contienen temas relacionados con las TLs. La información recopilada se ha discutido en reuniones de trabajo y cada uno de los participantes se ha

encargado de profundizar en si se trata de: i) máster profesional (de carácter interuniversitario, Erasmus Mundus, etc.), ii) títulos propios; iii) formación continua; iv) programas de doctorado.

La información que se ha recogido para cada titulación se recoge en la Tabla 1.

Identificación del título
Denominación
Especialidad
Tipo (oficial/título propio)
Universidad coordinadora
Carácter (internacional/europeo/nacional)
Otros participantes
Tipo de enseñanza (Presencial/Distancia/Semipresencial)
Nº créditos
Duración oficial
Idiomas en los que se imparte
Orientación (Profesional/Investigación)
Participación de empresas
Página web

Tabla 1. Información aportada por cada titulación.

4. CATÁLOGO DE ESTUDIOS UNIVERSITARIOS

Si nos referimos a los grados universitarios y a España, hay que recalcar que no existe ningún título de grado, ni en el ámbito de las humanidades ni de las ingenierías, que contemple esta disciplina como tal. Podemos encontrar materias aisladas que se imparten en algunos grados de ingeniería informática o de lingüística, en varios casos como asignaturas optativas, pero siempre se trata de aspectos parciales de las tecnologías lingüísticas y además, no es generalizado, es decir, un gran número de esas titulaciones no tienen ningún contenido relativo al PLN. En la Tabla 2 se muestra la información básica de estos grados y en el anexo 1 se presentan los detalles de los grados españoles analizados. Si examinamos el resto de titulaciones no correspondientes a lingüística o informática, tanto de humanidades, ámbito jurídico, como STEM, la ausencia de contenidos relativos a PLN es casi general, con alguna excepción como algunos grados en ciencia de datos (ver Tabla 2).

Denominación Grado	Universidad	Asignatura	Idioma
Grado de Lingüística	Universitat Barcelona (UB)	Lingüística Computacional	Catalán
Lingüística y Lenguas Aplicadas	Universidad de Cádiz (UCA)	Industrias de la Lengua. Ingeniería del Lenguaje. Lingüística Computacional.	Español
Lingüística y Lenguas Aplicadas	Universidad Complutense de Madrid (UCM)	Lingüística Computacional	Español
Traducción, interpretación y lenguas aplicadas	Universitat Oberta de Catalunya (UOC)	Tecnologías del Lenguaje. Traducción automática y posesición	Español y catalán
Lenguas Aplicadas	Universitat Pompeu Fabra (UPF)	Procesamiento automatico de textos	Español y catalán
Ingeniería Informática	Universidad del País Vasco (UPV/EHU)	Procesamiento del Lenguaje Natural (opt.)	Español
Ingeniería Informática	Universitat Pompeu Fabra (UPF)	Lenguaje Natural (opt.)	Español y catalán
Grado en Ingeniería Matemática en Ciencias de Datos	Universitat Pompeu Fabra (UPF)	Introducción a las técnicas de Procesamiento del Lenguaje Natural para aplicaciones cotidianas (opt.)	Español y catalán
Ingeniería Informática	Universidad Politécnica de Cataluña	No	
Ingeniería Informática	Universidad Carlos III	No	
Ingeniería Informática	Univ. de Alicante	No	
Ingeniería Informática	Uni. de Granada	No	

Tabla 2. Grados y asignaturas sobre Procesamiento del Lenguaje Natural en España.

Es diferente la situación en otros países donde sí encontramos grados especializados en tecnologías lingüísticas. La Tabla 3 enumera los grados de lingüística computacional dentro de las universidades estadounidenses de mayor calidad basado en el ranking US News and World Report². Todos los grados se imparten en inglés. En el anexo 2 se incluye información adicional sobre estos grados.

Denominación	Universidad	País	Idioma
Joint Major in CS + Linguistics	Universidad de Stanford	Estados Unidos	Inglés
NLP and CL	Carnegie Mellon	Estados Unidos	Inglés
Cognitive science, NLP & Speech	Johns Hopkins	Estados Unidos	Inglés
Computational Linguistics	Columbia University	Estados Unidos	Inglés
Computational Linguistics	University of Pennsylvania	Estados Unidos	Inglés
Computational Linguistics	Cornell University	Estados Unidos	Inglés
Computational Linguistics	University of Washington	Estados Unidos	Inglés
Computational Linguistics	University of Southern California	Estados Unidos	Inglés
Linguistics	University of Illinois Urbana Champaign	Estados Unidos	Inglés
Linguistics	University of California, Berkeley	Estados Unidos	Inglés
Linguistics	Massachusetts Institute of Technology	Estados Unidos	Inglés

Tabla 3. Grados en Procesamiento de Lenguaje Natural en Estados Unidos.

En el mapa de títulos de postgrado en tecnologías del lenguaje en España la situación es diferente ya que encontramos: i) títulos de posgrado oficiales en los que se imparte esta disciplina como tal y ii) títulos de posgrado especializados en otra materia relacionada pero que imparten alguna asignatura o bloque de asignaturas que podríamos decir son propias de las tecnologías lingüísticas; nos referimos a títulos de posgrado en Inteligencia Artificial, “Big Data” y “Data Science” que, dada su estrecha relación con las tecnologías lingüísticas, imparten algunas materias relacionadas; sin embargo, no se trata de enseñanzas que tengan en el centro de su interés las tecnologías lingüísticas.

En el primer grupo de másteres universitarios en esta disciplina encontramos amplia información en la página de la asociación española para el procesamiento del lenguaje natural. En la Tabla 4 que se

² <https://www.usnews.com/best-graduate-schools/top-science-schools/computer-science-rankings>

adjunta se incluye esta información sobre los másteres en España (la información se ampliará en el anexo 3).

Denominación	Universidad	Idioma
Lingüística Aplicada a las Tecnologías del Lenguaje y Gestión de Textos	Universidad de La Laguna (ULL)	Español e inglés
Master on Cognitive Science and Language	Universidad de Barcelona	Inglés
Tecnologías de la información y la comunicación en la enseñanza y el tratamiento de lenguas	Universidad Nacional de Educación a Distancia (UNED)	Español, inglés e italiano
Análisis y Procesamiento del lenguaje	Universidad del País Vasco	Euskara e inglés
Tecnologías de Lenguaje	Universidad Nacional de Educación a Distancia (UNED)	Español
Traducción y tecnologías	Universitat Oberta de Catalunya (UOC)	Español y catalán
Lingüística teórica y aplicada	Universitat Pompeu Fabra (UPF)	Español, catalán e inglés
Lenguas y tecnologías	Universitat Politècnica de València (UPV)	Español, valenciano e inglés
Intelligent Interactive Systems	Universitat Pompeu Fabra	Inglés
Máster Universitario en letras Digitales	Universidad Complutense	Español

Tabla 4. Másteres en Procesamiento del Lenguaje Natural en España.

En la Tabla 5 y anexo 4) presentamos distintos másteres que tienen como materia central las tecnologías lingüísticas que se ofrecen en universidades de Europa y el resto del mundo.

Denominación	Universidad	País	Idioma
Language and Communication Technologies	Saarbruecken University	Alemania	Inglés
Language, Science and Technology	Saarland University	Alemania	Inglés
Digital Humanities	KU Leuven	Bélgica	Inglés
Lingüística Aplicada	Pontificia Universidad Católica de Valparaiso	Chile	Español
IT and Cognition	Copenhagen University	Dinamarca	Inglés
Computational Linguistics	Brandeis University	Estados Unidos	Inglés
Natural Language Processing	Columbia University	Estados Unidos	Inglés
Language Technologies	Carnegie Mellon University	Estados Unidos	Inglés
Natural Language Processing and Human Language Technology	Franche-Comté	Francia	Inglés, francés, portugués, español
Computer Science with Speech and Language Processing	University of Sheffield	Reino Unido	Inglés
Speech and Language Processing	University of Edinburgh	Reino Unido	Inglés
Language Technology	University of Göteborg	Suecia	Inglés
Language Technology	Uppsala University	Suecia	Inglés

Tabla 5. Másteres en Procesamiento del Lenguaje Natural en Europa y el resto del mundo.

En el segundo grupo de enseñanzas de posgrado incluimos en la Tabla 6, una relación de los másteres en inteligencia artificial en España y países de habla hispana. La Tabla 6 (anexo 5) presenta algunos de los másteres en inteligencia artificial en Europa y resto del mundo. Finalmente, la Tabla 7 (anexo 6 y Tabla 8, anexo 7) recogen los másteres de “Artificial Intelligence”, “Big Data” y “Data Science” en las universidades americanas de más prestigio.

En lo que hace referencia a los estudios de doctorado también podemos distinguir entre los programas que incluyen algún área de especialidad relacionada con el Procesamiento del Lenguaje Natural, con aquellos programas de áreas afines, especialmente relacionadas con la informática y la inteligencia artificial, donde se pueden llevar a cabo tesis relacionadas con el Procesamiento del Lenguaje Natural. En la Tabla 9 (anexo 8) podemos observar programas de doctorado en España relacionados con esta área.

Denominación	Universidad	Idioma
Máster Universitario en I.A. Avanzada: Fundamentos, Métodos y Aplicaciones	UNED	Español
Máster en Inteligencia Artificial	UPC-UB-URV	Español
Máster en Inteligencia Artificial	UPM	Español
Máster Universitario en Investigación e Inteligencia Artificial	Universidad Internacional Menéndez Pelayo	Español
Máster en Inteligencia Artificial	Universitat Politècnica de Valencia	Español
Otros países de habla hispana		
Denominación	Universidad coordinadora	Idioma
Maestría en Ciencia de la Computación	Universidad de Oriente (UO), Santiago de Cuba, Cuba	Español
Maestría en Computación e Informática	Universidad de Costa Rica	Español
Maestría en Ingeniería de sistemas y computación	Pontificia Universidad Javeriana	Español
Magíster en Informática	Universidad de la República	Español

Tabla 6. Másteres en Inteligencia Artificial en España y otros países de habla hispana.



Denominación	Universidad	Idioma
MPhil in Advanced Computer Science	University of Cambridge	Inglés
Msc in Artificial Intelligence	University of Edinburgh	Inglés
Computer Science Master's Degree - Machine Learning	Columbia University in the City of New York	Inglés
Master of Science in Computer Science	Johns Hopkins	Inglés
MSc Advanced Computer Science	University of Sheffield	Inglés
MSc Cybersecurity and Artificial Intelligence	University of Sheffield	Inglés
MSc Artificial Intelligence	University of Manchester	Inglés
MSc Artificial Intelligence	Cardiff University	Inglés
MSc Artificial Intelligence	Imperial College London	Inglés
Computational Statistics and Machine Learning MSc	University College London	Inglés

Tabla 7. Másteres en Inteligencia Artificial en Europa y resto del mundo.

Denominación	Universidad coordinadora	Idioma
Master of Science in Data Science	Johns Hopkins	Inglés
Master of Science in Engineering in Data Science	University of Pennsylvania	Inglés
MSc Data Analytics	University of Sheffield	Inglés
MSc Data Science and Analytics	Cardiff University	Inglés
Data Science and Machine Learning MSc	University College London	Inglés

Tabla 8. Másteres en Big Data o Ciencia de Datos.



Denominación	Universidad	Idioma
Sistemas Inteligentes – Acceso a la información multilingüe y multimedia	Universidad Nacional de Educación a Distancia (UNED)	Español, inglés
Informática – Tecnologías del Lenguaje Humano	Universidad de Alicante (UA)	Español
Filosofía y Ciencias del Lenguaje – Lingüística Informática	Universidad Autónoma de Madrid (UAM)	Español
Ciencia Cognitiva y Lenguaje – Procesamiento, representación y adquisición del lenguaje	Universitat de Barcelona (UB)	Español y catalán
Ciencia y tecnología informática – Human Languages and Accessibility Technologies	Universidad Carlos II de Madrid (UC3M)	Español
Ingeniería Informática – Procesamiento del Lenguaje Natural	Universidad Complutense de Madrid (UCM)	Español
Traducción y Ciencias del Lenguaje – Lingüística Computacional e ingeniería del lenguaje	Universitat Pompeu Fabra (UPF)	Español
Análisis y procesamiento del lenguaje	Universidad del País Vasco (UPV/EHU)	Español
Informática – Machine Learning and Language Processing	Universidad Politécnica de Valencia (UPV)	Español

Tabla 9. Programas de doctorado en Procesamiento del Lenguaje Natural en España.

También es interesante destacar que las principales plataformas que ofrecen MOOCs disponen de cursos relacionados con el Procesamiento del Lenguaje Natural. En la Tabla 10 podemos observar la denominación del curso y la plataforma que lo ha ofrecido.

Denominación	Plataforma
moocTLH: Nuevos retos en las tecnologías del lenguaje humano	Miríada X
Natural Language Processing	Coursera
Natural Language Processing	Udacity
Natural Language Processing (NLP)	Edx
Deep Learning for Natural Language Processing	University of Oxford

Tabla 10. Cursos MOOC en Procesamiento del Lenguaje Natural.

5. OTRAS EXPERIENCIAS SIMILARES QUE HAYAN ABORDADO LA INCLUSIÓN DE NUEVAS TEMÁTICAS

Se han estudiado en detalle las iniciativas en el ámbito de la Ciberseguridad diseñadas por el Instituto Nacional de Ciberseguridad (INCIBE³), que pueden servir de referencia para su utilización en el campo de trabajo de las tecnologías lingüísticas. INCIBE plantea una serie de recomendaciones e iniciativas para promover la Ciberseguridad y contribuir al mercado digital para el uso seguro del ciberespacio en España. En el ámbito de la formación y la educación, promueve iniciativas para enseñanza primaria, FP y universitaria, con el objetivo de identificar y captar el talento para ampliar el número de profesionales del sector y hacer de España una referencia en Ciberseguridad.

Los estudios realizados en Ciberseguridad plantean acciones entre las que se incluyen jornadas escolares de motivación y captación de estudiantes, desarrollo de materiales on-line, formación y motivación de profesorado, becas de estudios, catálogos de formación y talleres, así como en el entorno empresarial, por medio de formación especializada, tanto presencial como MOOC.

Las experiencias recogidas por el Instituto Nacional de Ciberseguridad (INCIBE) en lo referente a su aproximación a la empresa, administración e iniciativas de investigación han sido la base sobre la que nos hemos apoyado para proponer la inclusión de temáticas en los sectores mencionados.

³ <https://www.incibe.es/>

6. RECOMENDACIONES DE ENTIDADES NACIONALES Y EUROPEAS SOBRE PLN E IA

Existen varios informes nacionales e internacionales⁴ recientes que contienen entre sus recomendaciones acciones formativas relacionadas con la IA y, por inclusión al PLN. En concreto, el informe [Artificial Intelligence: A European Perspective](#)⁵, publicado por la Comisión Europea, examina la situación de la IA en la Unión Europea y analiza profusamente el ámbito de la formación en IA. Además de presentar las necesidades y demandas de formación, pone de relieve la importancia de los retos que ésta permitirá enfrentar como posibles riesgos relacionados con la privacidad y aspectos éticos. A nivel nacional, la [Estrategia Española de I+D+i en Inteligencia Artificial](#) dispone entre sus prioridades desarrollar un sistema que fomente la formación transversal y profesional en Inteligencia Artificial. Recomienda incluir la IA en el sistema educativo como palanca de cambio tecnológico del país permitiendo así la incorporación del conocimiento y uso de la IA en el mercado laboral y fomento, recuperación y atracción del talento. Para ello subraya la importancia de crear de una plataforma y un ecosistema computacional compuesto por la universidad, los centros científicos y tecnológicos y las empresas. Propone fomentar la transferencia de conocimiento entre centros científicos y tecnológicos y empresas con mayor agilidad a través de consorcios de innovación tecnológica especializados en IA. La Big Data Value Association (BDA), en el informe elaborado en Junio 2019 [Strategic Research, Innovation and Deployment Agenda for an AI PPP](#)⁶ analiza la importancia de reconocer las competencias, tanto técnicas como no técnicas, a través de certificaciones avaladas por cursos universitarios, y vías de training profesional. Propone acciones complementarias como el desarrollo de cursos relacionados con la IA para formar a directivos de la industria y la administración pública sobre la potencialidad de la IA en sus modelos de negocio. El informe de la OCDE [Artificial Intelligence in Society](#)⁷ aborda el reto de preparar la sociedad para la transformación del mundo laboral. Se prevé que la IA y el PLN cambien la naturaleza del trabajo, complementando a los humanos en algunas tareas y generando a la vez nuevos tipos de trabajo. En cualquier caso, el impacto dependerá en gran parte de la velocidad de difusión en los diferentes

⁴ OECD (Organisation for Economic Co-operation and Development): *Identifying and Measuring Developments in Artificial Intelligence, Artificial Intelligence in Society*, WIPO (World Intellectual Property Organization): *Technology Trends 2019*.

⁵ <https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/artificial-intelligence-european-perspective>

⁶ <http://www.bdva.eu/node/1310>

⁷ https://www.oecd-ilibrary.org/science-and-technology/artificial-intelligence-in-society_eedfee77-en

sectores, lo que implica que su rápida adaptación al entorno educativo es crucial en el desarrollo de la sociedad. Las políticas clave para dirigir la transición a la IA deberán incluir salud, economía y, fundamentalmente, educación.

7. ESTUDIO DE LA DEMANDA DE LOS EMPLEADORES

Este estudio lo vamos a dividir en dos apartados:

1. Estudio de ofertas de empleo. Esto servirá para conocer las necesidades del mercado laboral en lo concerniente a tecnologías de TL.
2. Identificar las empresas que actualmente hacen ofertas de trabajo a titulados/as de esta área. Para identificar las empresas que actualmente ofertan puestos de trabajo en esta área, hemos acudido a los *cluster* de empresas relacionadas con las tecnologías lingüísticas, para a continuación realizar la encuesta a las personas responsables de recursos humanos de dichas empresas.

Los resultados se comentarán en las subsecciones 7.1 y 7.2.

7.1 ANÁLISIS DE OFERTAS DE EMPLEO EN TL

Inicialmente se han examinado los puestos/figuras requeridas en ofertas de empleo PLN (36 ofertas). Todas las ofertas de empleo se han recuperado con las cadenas “procesamiento del lenguaje natural” o “lenguaje natural”, por lo que todos requieren PLN. La Tabla 11 y la Figura 1 presentan distintos tipos de perfiles encontrados en esas ofertas.

Tipo	Palabras clave
CD (Científico de datos)	Data Scientist NLP, senior data governance, senior científico de datos, arquitecto big data, Data-scientist analytics, Científico de datos, Jefe de proyecto Big Data, <i>Senior Data Scientist, NLP Data Scientist, Sr Data Scientist (specialist in NLP), Data scientist Python&NLP, Senior Data Scientist-NLP Specialist, AI DATA SCIENTIST, Expert AI & Data Science</i>



LC (Lingüista computacional)	Lingüista computacional, <i>Data Researcher, ITA/ESP/ENG, Presales Engineer</i>
Voz	PLN+VOZ , investigador en tecnologías de procesamiento del habla, <i>Speech Analyst, Senior Machine Learning Engineer (voz)</i>
Desarrollador de aplicaciones	Desarrollador Java junior, Frontend developer, Full-Stack Developer - Inteligencia artificial
ML (Machine learning), incluye peticiones con inteligencia artificial	<i>Machine Intelligence/Natural Language generation Engineer, Tech Leader Artificial Intelligence & NLP, Machine Learning engineer, Graduado en informática, Applied scientist</i>
NS (no especifica)	Experto PLN, <i>Siri - Language Engineer (Norwegian), Senior NLP Engineer/NLP Engineer, Full stack developer</i>

Tabla 11. Tipología de perfiles requeridos de TL.

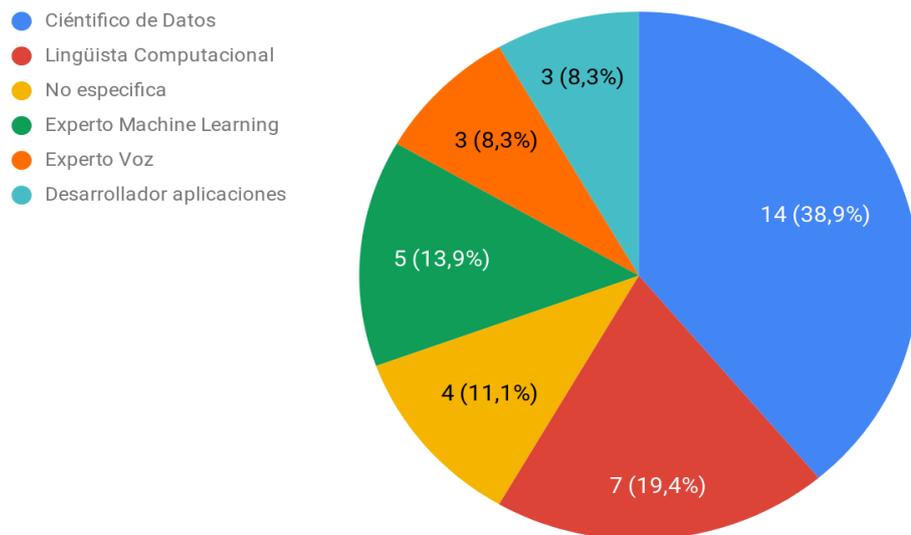


Figura 1. Distribución de perfiles de ofertas de empleo.

Referente a los estudios que se piden, la siguiente lista detalla diferentes tipos de perfiles demandados:

- FPII o Grado en Informática, Graduado en informática, Ingeniería informática o formación equivalente
- Grado o Máster en Ingeniería Informática o una rama del conocimiento relacionada (Telecomunicaciones, Matemáticas o Física)
- Científico de datos (7), Desarrollador (2), Machine learning (1), Lingüista computacional (2), Speech analyst (1), NS (1), Experto PLN (1)
- Licenciatura/Grado en Filología, Lenguas Aplicadas o Lingüística General o equivalente, o Lingüística computacional / Chatbots / Asistentes virtuales
- Lingüista computacional, *University background in Linguistic, Translating and Interpreting, Language Teaching, Philology, Natural Language Processing or related fields or equivalent practical experience.*
- Master/Doctorado (o equivalente) en Tecnologías del Habla o campo relacionado (6)

La Figura 2 resume los perfiles demandados. Se puede observar que un porcentaje importante de las peticiones no especifica el área de las TL demandada pero, teniendo eso en cuenta, también se puede apreciar que entre los perfiles que sí han sido especificados aparece un predominio de la formación STEM y también lingüística, observándose también una menor aunque relevante demanda de perfiles relativos al tratamiento de voz o habla.

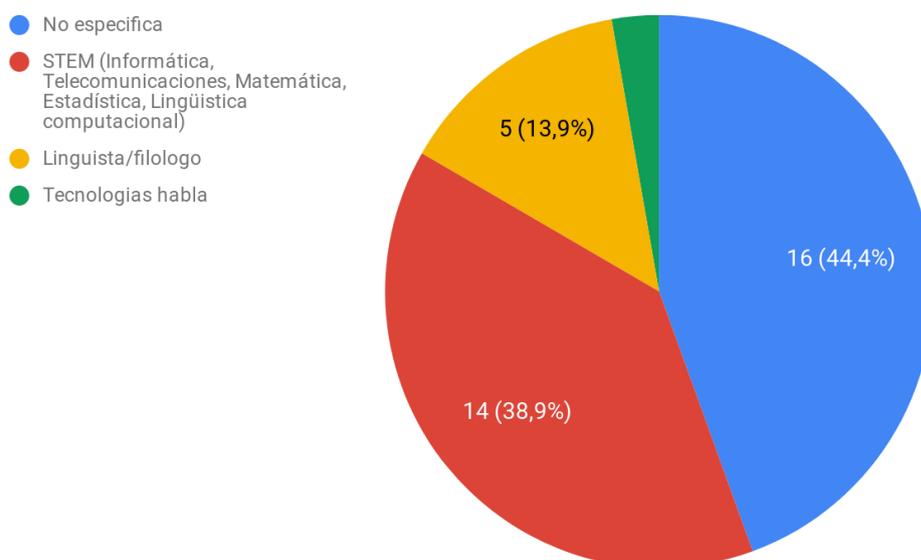


Figura 2. Perfiles de titulaciones de grado para expertos en TL.

Examinando el nivel de estudios requerido para las titulaciones de tipo STEM, la Figura 3 muestra que, aunque la mayoría de las peticiones se refieren a titulaciones de grado, un porcentaje importante de la demanda requiere estudios de nivel superior, correspondientes a máster o doctorado.

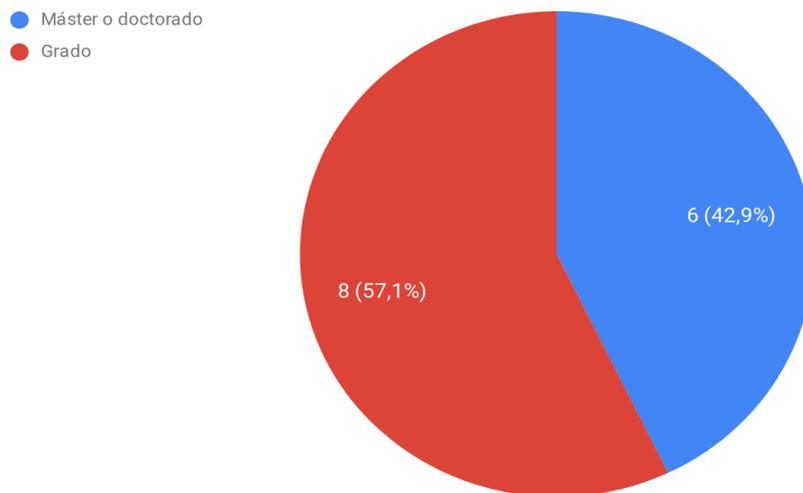


Figura 3. Categorías de estudios requeridas.

La Figura 4 presenta la distribución de los sectores que demandan profesionales con base en PLN, destacando las entidades que trabajan en Tecnologías de la Información, seguidas de consultorías y servicios.

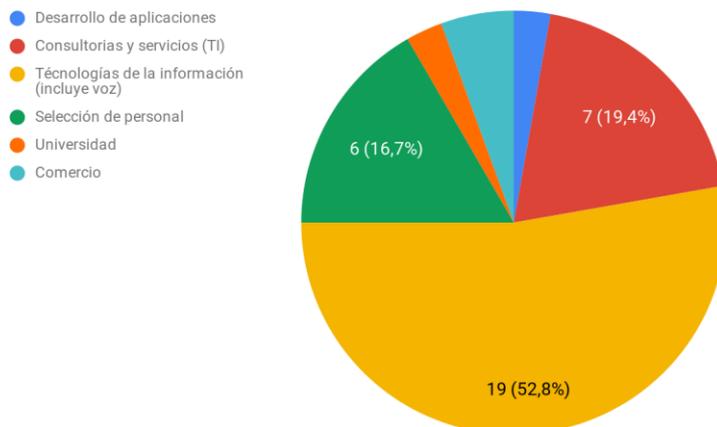


Figura 4. Sectores que demandan formación en TL.

7.2 ANÁLISIS DE ENCUESTAS SOBRE NECESIDADES DE EMPRESAS EN TL

Los datos presentados a continuación se han obtenido de una encuesta realizada a empresas públicas y privadas que trabajan en Tecnologías del Lenguaje. En total se recibieron 25 respuestas que dan una idea aproximada de la situación de los empleos en TL. De ellas, todas excepto una corresponden al sector de las TL (ver Figura 5).

Confirma que tu entidad se dedica al sector de las TL

25 respuestas

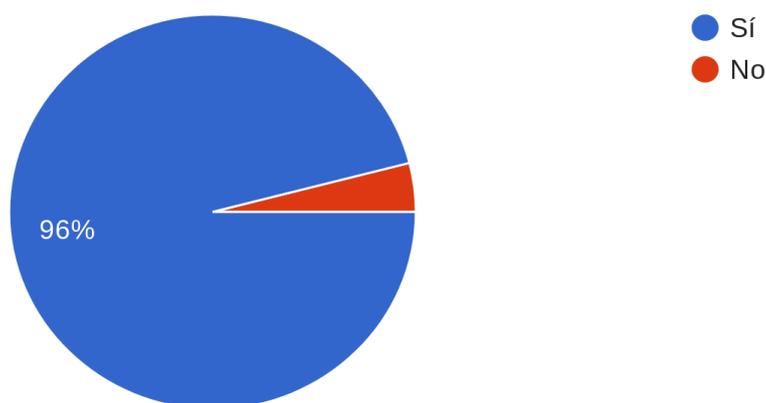


Figura 5. Proporción de respuestas de entidades de TL.

En cuanto a la distribución geográfica, las empresas se encuentran distribuidas en las diferentes comunidades autónomas (ver Figura 6). Madrid y Barcelona principalmente concentran la mayor parte de estas entidades, aunque también se encuentra un número relevante en otras comunidades.



¿En qué provincia se ubica la sede principal de tu entidad?

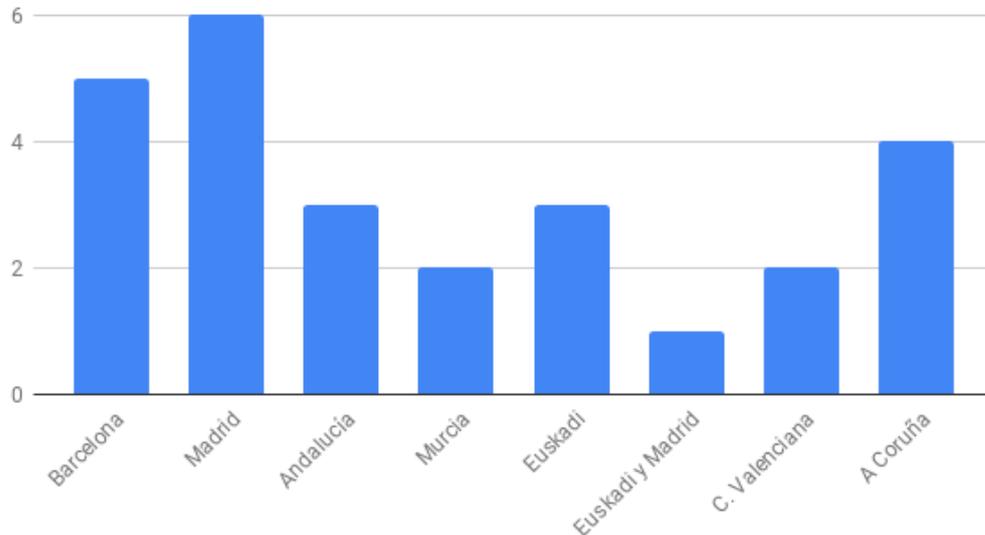


Figura 6. Distribución de entidades TL.

¿Cuántos empleados tiene tu entidad?

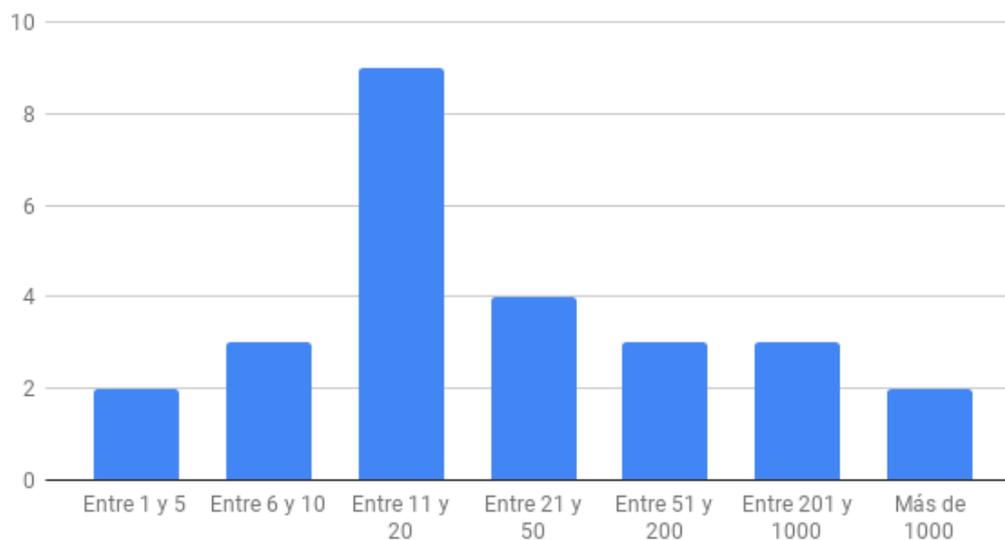


Figura 7. Tamaño de entidades de TL.

El número total de empleados de estas empresas se puede observar en la Figura 7, en la que se puede ver un predominio de empresas de tamaño medio y pequeño, aunque también son de gran importancia un número reducido de entidades de gran tamaño. La Figura 8 detalla el número de empleos de TL en esas empresas. En ella se puede ver cómo en la gran mayoría de los casos el

número de personas con formación en TL es reducido, entre 1 y 5 personas especialistas en la materia.

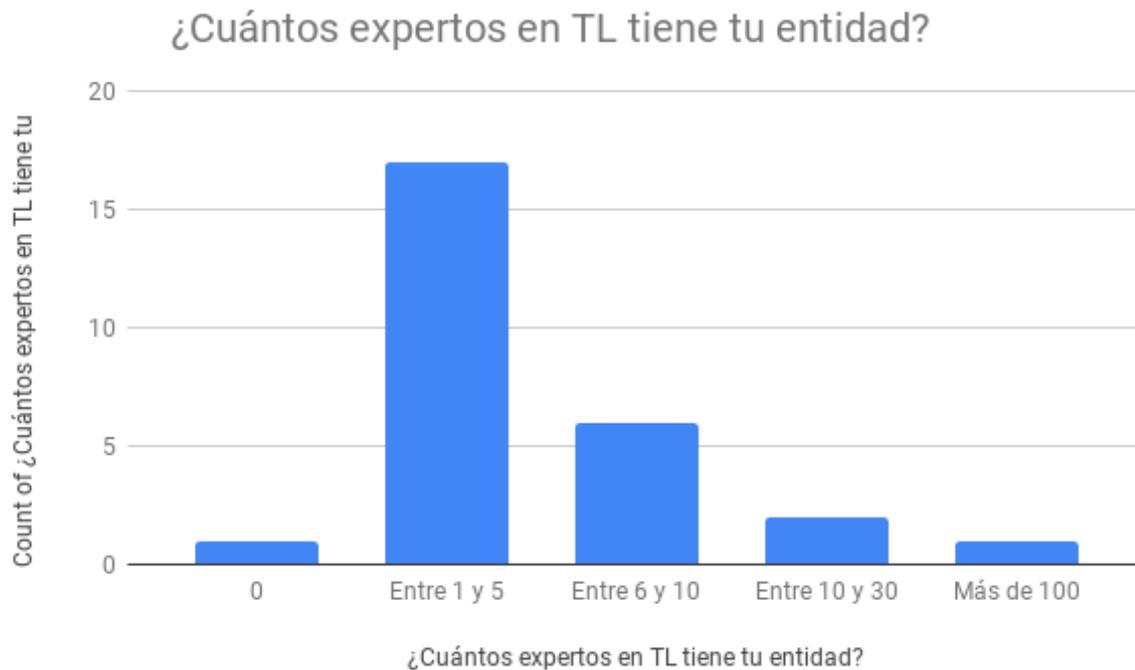


Figura 8. Distribución de expertos en TL.

Examinando los estudios de grado, máster o doctorado que han cursado los expertos en TL, se pueden destacar:

- Grado o equivalente en informática: 31
- Grado o equivalente en Telecomunicaciones, Física, Matemáticas: 9
- Grado en Filología, Traducción e Interpretación o Lingüística: 10
- Grado o equivalente en otras titulaciones: 4
- Máster (STEM): 20
- Máster (humanidades): 6
- Doctorado: 39

En relación al empleo de expertos en TL con formación alternativa a la universitaria, se destacan titulaciones de FP o cursos adicionales.

La

Figura 9 muestra el grado de dificultad percibido a la hora de encontrar expertos en TL. La mayoría de las respuestas destacaron la escasa opción de encontrar este tipo de profesionales, indicando una dificultad alta o muy alta.

Dificultades para encontrar los expertos en TL adecuados

25 respuestas

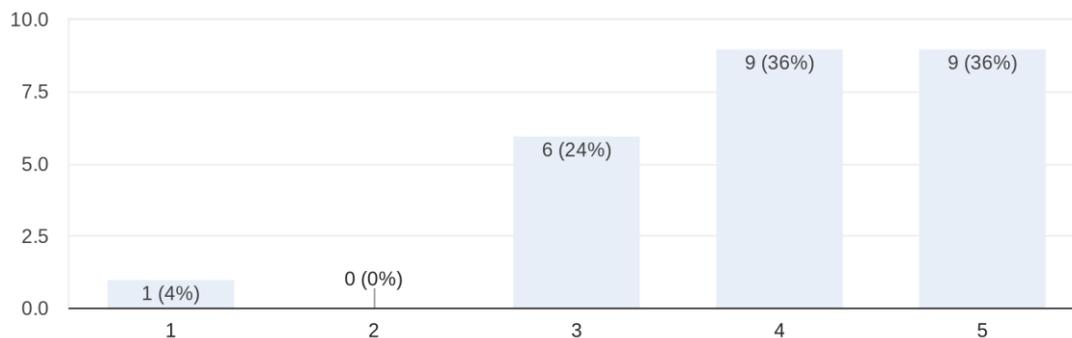


Figura 9. Dificultad para contratar expertos en TL (1: ninguna dificultad, 5: dificultad alta).

La Tabla 12 recoge las principales competencias que se requieren en los empleos de TL. Además de competencias más generales como *Computer Science* o *Artificial Intelligence*, se observa una fuerte demanda de tecnologías novedosas como *Big Data*, *PLN*, *Deep Learning*, *Machine Learning*, *Machine Translation*, *Speech Processing* o *Chatbots*.

Computer Science	2
Artificial Intelligence	1
Big Data	4
Natural Language Processing (NLP)	6
Deep Learning, Neural Nets	11

Machine Translation	6
Machine Learning	6
Speech Processing	5
Chatbots	3
Others: English, blockchain, problem solving, mathematics, graphs, ethical hacking, biomedical processing	

Tabla 12. Competencias requeridas en empleos de TL

Un aspecto interesante es la relación entre academia y mundo empresarial. La Figura 10 presenta el interés de las empresas en atraer talento investigador por medio de los doctorados industriales. Se aprecia cómo, aunque un 20% de las entidades encuestadas no consideran esta opción interesante, el 80% de las respuestas han utilizado o entienden esta posibilidad atractiva.

Contratos de formación de investigadores en empresas (doctorados industriales) y sus ayudas

25 respuestas

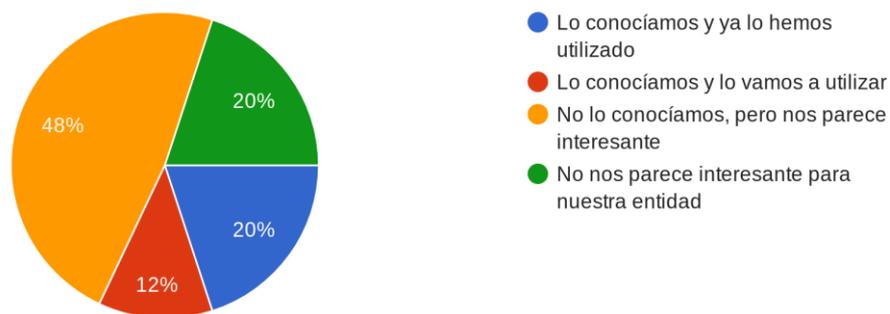


Figura 10. Ayudas a contratos de investigadores en empresas.

Como comentarios adicionales, la encuesta ha proporcionado la siguiente información:

- Se cree necesaria la organización de eventos anuales del sector TL para la dinamización de la oferta y demanda de empleos en TL (ver Figura 11).



- Se requiere la impartición de Máster o cursos especializados de tecnologías lingüísticas por perfiles.
- Además de estudios formales, se destaca la relevancia de vías alternativas (Coursera, Edx,...) a la educación formal, debido a su flexibilidad para su adaptación a las necesidades de las empresas.
- Respecto al doctorado industrial se ha destacado la exigencia de disponer de un doctor en la empresa, que limita el ámbito de aplicación, siendo más apropiado para empresas grandes o spin-off.
- Un aspecto importante es la falta de vocaciones estudiantiles en esta área, debido a la falta de grandes grupos de investigación. Por ello, se ha propuesto el aumentar la visibilidad a este sector de las TL en el ámbito estudiantil, mediante acciones formativas o participación en proyectos empresariales.
- En cuanto al ámbito de trabajo en TL, se destaca la importancia de aproximaciones multilingües, incluyendo, entre otros idiomas, a gallego/portugués, catalán, euskera, español e inglés.

¿Le parece interesante que en encuentros anuales del sector TL se cree un espacio para DINAMIZAR EL EMPLEO?

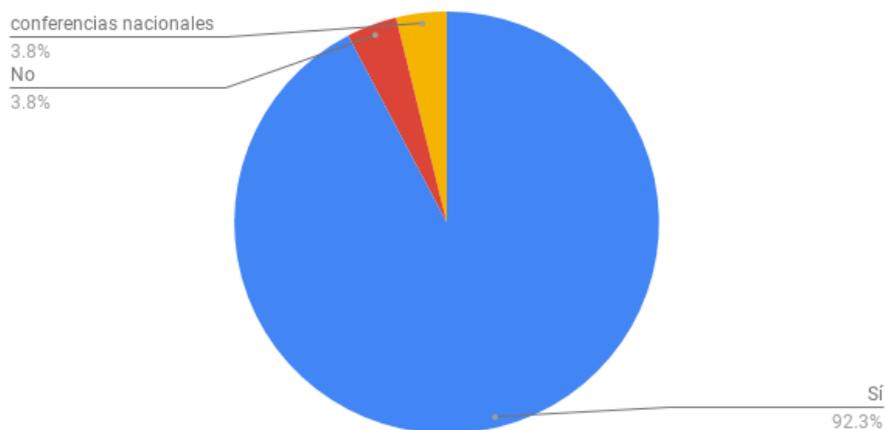


Figura 11. Interés en eventos de TL.

8. CONCLUSIONES

Este documento presenta el resultado del estudio de la oferta formativa actual en el ámbito universitario en el área de las tecnologías lingüísticas a nivel nacional, y europeo e internacional en países de referencia (EEUU, Reino Unido, Alemania, Países Bajos, Francia, Portugal, países latinoamericanos, etc.). Aunque esta relación no es exhaustiva, en particular en áreas muy relacionadas con las tecnologías lingüísticas como por ejemplo la inteligencia artificial o “Big Data” o “Data Science”, sí representa un mapa significativo de la oferta docente en estas materias a nivel mundial.

Este estudio nos ha permitido constatar que la enseñanza de Procesamiento del Lenguaje Natural se produce de forma muy dispar pero está presente en todos los niveles (grado, máster y doctorado) y también en cursos de tipo MOOC en diversas plataformas. De hecho, existen sólo algunos pocos grados, másteres, doctorados y cursos MOOC donde se imparten algunos contenidos propios de las tecnologías lingüísticas. Por ejemplo, la inmensa mayoría de los estudiantes de informática del país pueden terminar el grado sin las capacidades mínimas para poder procesar correctamente texto escrito u otras formas de datos digitales no estructurados. Debemos recordar que actualmente la información digital no estructurada (mayoritariamente texto en múltiples idiomas, audio y video) representa entre el 80% y 90% de la información digital disponible⁸. Es decir, tras terminar el grado la inmensa mayoría de estudiantes de informática desconocen cómo procesar adecuadamente la mayor parte de la información digital.

En lo que hace referencia a grado, esta disciplina se ofrece en España tanto en titulaciones relacionadas con la lingüística (filologías, traducción e interpretación, etc.) como en titulaciones relacionadas con la informática (o telecomunicaciones, ciencia de datos, inteligencia artificial, etc.).

Para estos grados recomendamos un fortalecimiento (incluyendo asignaturas obligatorias y optativas) y armonización de los contenidos sobre Procesamiento del Lenguaje Natural. Por ejemplo, todos los egresados de informática deberían estar capacitados para el procesamiento adecuado de datos no estructurados, en particular texto (multilingüe), audio y vídeo. Asimismo, los egresados de lingüística deberían estar capacitados para el procesamiento y análisis automático de textos. Los

⁸ White Paper IDC:

www.coveo.com/~/_/media/Files/WhitePapers/Coveo_IDC_Knowledge_Quotient_June2014.ashx

egresados de traducción e interpretación deberían saber utilizar sistemas de ayuda a la traducción, etc.

No existe en España ningún grado dedicado íntegramente al Procesamiento del Lenguaje Natural, a diferencia de otros países (por ejemplo EEUU) donde sí que existe esta oferta. Puede resultar interesante explorar el diseño de un grado universitario en Procesamiento del Lenguaje Natural (Lingüística Computacional o Tecnologías del Lenguaje), donde se estudien de una manera integrada tanto los aspectos más lingüísticos como los informáticos. Este tipo de grado podría tener una admisión de alumnos amplia e itinerarios algo diferenciados en los primeros cursos según la opción de especialidad de bachillerato de los alumnos.

Aunque no son muchos, a nivel de máster sí que encontramos en España algunos programas íntegramente dedicados al Procesamiento del Lenguaje Natural. Además, esta disciplina también está presente mediante asignaturas en programas de máster en el área de la informática y la inteligencia artificial. Fuera de España la situación es similar y se encuentran programas de máster íntegramente dedicados a las tecnologías del lenguaje y otros que contienen algunas asignaturas.

La situación de los estudios de doctorado en España es similar a los estudios de máster. Hay programas de doctorado con profesores especialistas en temas propios de las tecnologías lingüísticas.

Por último, cabe recalcar que el área del Procesamiento del Lenguaje Natural se encuentra presente también en diversos cursos tipo MOOC, lo que puede resultar especialmente interesante para la formación continua para profesionales de áreas relacionadas pero sin una formación específica.

Así pues, como conclusión de este apartado del estudio, se recomienda fortalecer y armonizar los contenidos sobre tecnologías lingüísticas en los grados relacionados (informática, ciencia de datos, inteligencia artificial, lingüística, traducción e interpretación, etc.), así como la creación de estudios de grado específicos en el área de las tecnologías del lenguaje.

9. REFERENCIAS

- [1] Plan de Impulso de las Tecnologías del Lenguaje.
<https://www.plantl.gob.es/Paginas/index.aspx>
- [2] US News and World Report, 2018. <https://www.usnews.com/best-graduate-schools/top-science-schools/computer-science-rankings>
- [3] LinkedIn. <https://www.linkedin.com/>
- [4] Instituto Nacional de Ciberseguridad (Incibe). <https://www.incibe.es/>
- [5] White Paper, International Data Consortium (IDC)
www.coveo.com/~/_media/Files/WhitePapers/Coveo_IDC_Knowledge_Quotient_June2014.ashx



10. GLOSARIO DE SIGLAS Y ACRÓNIMOS

AI	Artificial Intelligence
CPI	Compra Pública de Innovación
DL	Deep Learning
ML	Machine Learning
MT	Machine Translation
MOOC	Massive On-line Open course
NLP	Natural Language Processing
NN	Neural Network
SEAD	Secretaría de Estado para el Avance Digital
STEM	Science, Technology, engineering, Mathematics
TL	Tecnologías del Lenguaje



11. ANEXOS

Anexo 1. Grados con asignaturas en Procesamiento del Lenguaje Natural en España

Denominación	Lingüística y Lenguas Aplicadas
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universitat de Barcelona (UB)
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	240
Duración oficial	4 años
Idiomas en los que se imparte	Español y catalán
Orientación	Profesional y de investigación
Participación de empresas	
Página web	www.ub.edu/filologia/graus/linpld.htm

Denominación	Lingüística y Lenguas Aplicadas
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universidad de Cádiz (UCA)
Carácter	
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	240
Duración oficial	4 años
Idiomas en los que se imparte	Español
Orientación	Profesional y de investigación
Participación de empresas	
Página web	https://filosofia.uca.es/grado-en-linguistica-y-lenguas-aplicadas/

Denominación	Lingüística y lenguas aplicadas
Especialidad	Itinerario Lingüística Computacional
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universidad Complutense de Madrid (UCM)
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	240
Duración oficial	4 años
Idiomas en los que se imparte	Español y en los idiomas pertinentes en cada una de las lenguas estudiadas
Orientación	Profesional
Participación de empresas	
Página web	https://www.ucm.es/estudios/grado-linguisticaylenguasaplicadas



Denominación	Traducción, interpretación y lenguas aplicadas
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universitat Oberta de Catalunya (UOC)
Carácter	Nacional
Otros participantes	Universitat de Vic (UVic-UCC)
Tipo de enseñanza	Distancia
Nº créditos	240
Duración oficial	4 años
Idiomas en los que se imparte	Castellano y catalán
Orientación	Profesional
Participación de empresas	
Página web	https://estudios.uoc.edu/es/grados/traduccion-interpretacion-lenguas-aplicadas/presentacion

Denominación	Lengua y literatura catalanas
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universitat Oberta de Catalunya
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Distancia
Nº créditos	240
Duración oficial	4 años
Idiomas en los que se imparte	Catalán
Orientación	Profesional
Participación de empresas	
Página web	https://estudios.uoc.edu/es/grados/lengua-literatura-catalanas/presentacion

Denominación	Lenguas Aplicadas
Especialidad	Tecnologías del Lenguaje
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universitat Pompeu Fabra (UPF)
Carácter	Europeo
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	240
Duración oficial	4 años
Idiomas en los que se imparte	Castellano y catalán
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://www.upf.edu/web/traduccio/llengues-aplicades



Denominación	Grado en Ingeniería Informática
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universidad del País Vasco (UPV/EHU)
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	240
Duración oficial	4 años
Idiomas en los que se imparte	Castellano, inglés y euskera
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://www.ehu.eus/es/grado-ingenieria-informatica

Denominación	Grado en Ingeniería Informática
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universitat Pompeu Fabra (UPF)
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	240
Duración oficial	4 años
Idiomas en los que se imparte	Castellano, inglés y catalán
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://www.upf.edu/documents/1656590/3744563/PE_GEI_16.pdf/ed342710-bb4f-6d80-2b05-d1e8c51b8ffd

Denominación	Grado en Ingeniería Matemática en Ciencias de Datos
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universitat Pompeu Fabra (UPF)
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	240
Duración oficial	4 años
Idiomas en los que se imparte	Castellano, inglés y catalán
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://www.upf.edu/documents/1656590/3744563/PE_GEMCD_16.pdf/58ee71af-7b94-1f29-07ca-e4d8d20e167c

Denominación	Grado en Ingeniería Informática
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universidad Carlos III
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	240



Duración oficial	4 años
Idiomas en los que se imparte	Castellano e inglés
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://www.uc3m.es/ss/Satellite/Grado/es/Detalle/Estudio_C/1371212562160/1371212987094/Grado_en_Ingenieria_Informatica#programa
Denominación	Grado en Ingeniería Informática
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universidad de Alicante
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	240
Duración oficial	4 años
Idiomas en los que se imparte	Castellano
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://eps.ua.es/es/ingenieria-informatica/

Denominación	Grado en Ingeniería Informática
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universidad de Granada
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	240
Duración oficial	4 años
Idiomas en los que se imparte	Castellano
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://grados.ugr.es/informatica/pages/titulacion/titulacion/csi

Denominación	Grado en Ingeniería Informática
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universidad Politécnica de Cataluña
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	240
Duración oficial	4 años
Idiomas en los que se imparte	Castellano y catalán
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://eps.ua.es/es/ingenieria-informatica/



Anexo 2. Grados en Procesamiento del Lenguaje Natural en Estados Unidos

Denominación	Joint Major in CS+Linguistics
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad de Stanford
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://linguistics.stanford.edu/degree-programs/undergraduate-program/joint-major-cslinguistics

Denominación	NLP and CL
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Carnegie Mellon
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.lti.cs.cmu.edu/work/category/2717

Denominación	Cognitive Science
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Johns Hopkins
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://cogsci.jhu.edu/research-areas/linguistics/



Denominación	NLP & Speech
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Columbia University
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	http://www.upv.es/titulaciones/MULT/menu_1014516c.html

Denominación	Computational Linguistics
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	University of Pennsylvania
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.ling.upenn.edu/research/computational

Denominación	Computational Linguistics
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Cornell University
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://as.cornell.edu/linguistics
Denominación	Computational Linguistics
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	University of Washington
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	



Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://linguistics.washington.edu/fields/computational-linguistics

Denominación	Computational Linguistics
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	University of Southern California
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://thebestschools.org/online-colleges/university-southern-california-online/

Denominación	Linguistics
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	University of Illinois Urbana Champaign
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://linguistics.illinois.edu/academics/graduate-program

Denominación	Linguistics
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	University of California, Berkeley
Carácter	
Tipo de enseñanza	
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://lx.berkeley.edu/grad/program
Denominación	Linguistics



Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Massachusetts Institute of Technology
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	http://linguistics.mit.edu/undergraduate-program/

Anexo 3. Másteres en Procesamiento del Lenguaje Natural en España

Denominación	Lingüística Aplicada a las Tecnologías del Lenguaje y Gestión de Textos
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universidad de la Laguna (ULL)
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	60
Duración oficial	1 año
Idiomas en los que se imparte	Español e inglés
Orientación	Profesional e investigadora
Participación de empresas	
Página web	https://www.ull.es/view/master/tecnologiaLenguaje/Inicio/es

Denominación	Máster en Tecnologías de la Información y la Comunicación en la Enseñanza y el Tratamiento de Lenguas
Especialidad	2 itinerarios: enseñanza de lenguas y procesamiento del lenguaje
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universidad Nacional de Enseñanza a Distancia (UNED)
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Distancia
Nº créditos	60
Duración oficial	1 año
Idiomas en los que se imparte	Español, inglés e italiano
Orientación	Profesional e investigación
Participación de empresas	
Página web	http://portal.uned.es/portal/page?_pageid=93,22790448&_dad=portal&_schema=PORTAL

Denominación	Tecnologías del lenguaje
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universidad Nacional de Enseñanza a Distancia (UNED)
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Distancia
Nº créditos	60
Duración oficial	1 año
Idiomas en los que se imparte	Español
Orientación	Investigación
Participación de empresas	
Página web	https://nlpmaster.uned.es/



Denominación	Traducción y tecnologías
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universitat Oberta de Catalunya (UOC)
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Distancia
Nº créditos	60
Duración oficial	1 o 2 cursos
Idiomas en los que se imparte	Español y catalán
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://estudios.uoc.edu/es/masters-universitarios/traduccion-tecnologias/presentacion

Denominación	Lingüística Teórica y Aplicada
Especialidad	Lingüística Computacional
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universitat Pompeu Fabra (UPF)
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	60
Duración oficial	1 año
Idiomas en los que se imparte	Español, catalán e inglés
Orientación	Investigación
Participación de empresas	
Página web	https://www.upf.edu/web/masterlinguistica

Denominación	Lenguas y tecnología
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Universidad Politécnica de Valencia (UPV)
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	60
Duración oficial	1 año
Idiomas en los que se imparte	Español, valenciano, inglés
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://www.upv.es/titulaciones/MULT/

Anexo 4. Másteres en Procesamiento del Lenguaje Natural en Europa y el resto de mundo

Denominación	Language and Communication Technologies
Especialidad	
Tipo (oficial/título propio)	European Erasmus Mundus Master's programme
Universidad coordinadora	Saarbruecken University (Alemania)
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	120
Duración oficial	2 años
Idiomas en los que se imparte	Inglés
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://www.uni-saarland.de/en/master/study-programmes/internat/lct/description.html

Denominación	Language Science and Technology
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Saarland University (Alemania)
Carácter	Internacional
Otros participantes	U. Groningen, U. País Vasco/EHU
Tipo de enseñanza	Presencial
Nº créditos	120
Duración oficial	2 años
Idiomas en los que se imparte	Inglés
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://www.uni-saarland.de/en/master/study-programmes/lang-lit/lst/description.html

Denominación	Digital Humanities
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	KU Leuven (Bélgica)
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	60
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://onderwijsaanbod.kuleuven.be/opleidingen/e/CQ_52330579.htm#activetab=diploma_omschrijving
Denominación	Magíster en Lingüística Aplicada



Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Pontificia Universidad Católica de Valparaíso (Chile)
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	69
Duración oficial	2 años
Idiomas en los que se imparte	Español
Orientación	Investigación
Participación de empresas	NO
Página web	http://www.elv.cl/magister/

Denominación	IT and cognition
Especialidad	Language Processing
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Copenhaguen Univesity
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	120
Duración oficial	2 años
Idiomas en los que se imparte	Inglés
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://studies.ku.dk/masters/it-and-cognition/

Denominación	Computational Linguistics
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Brandeis University (Estados Unidos)
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	12 course credits at least + thesis or internship
Duración oficial	2 años
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	http://www.brandeis.edu/computer-science/computational-linguistics/graduate/masters/index.html



Denominación	Natural Language Processing
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Columbia University (Estados Unidos)
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	Almenos 30
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://cvn.columbia.edu/program/columbia-university-computer-science-masters-degree-natural-language-processing-masters

Denominación	Language Technologies
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Carnegie Mellon University (Estados Unidos)
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	Almenos 120
Duración oficial	2 años
Idiomas en los que se imparte	Inglés
Orientación	Investigación
Participación de empresas	
Página web	https://liti.cs.cmu.edu/intranet/education/msprogram

Denominación	International Masters in Natural Language Processing and Human Language Technology
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Franche-Comté (Francia)
Carácter	Erasmus Mundus
Otros participantes	U. Autónoma de Barcelona, U. do Algarve, U. Wolverhampton
Tipo de enseñanza	Presencial
Nº créditos	120
Duración oficial	2 años
Idiomas en los que se imparte	Inglés, francés, portugués, español
Orientación	
Participación de empresas	
Página web	http://mastermundusnlp-htl.univ-fcomte.fr/
Denominación	Computer Science with Speech and Language Processing
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	University of Sheffield (Reino Unido)
Carácter	Internacional
Otros participantes	



Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://www.sheffield.ac.uk/dcs/postgraduate-taught/slp
Denominación	Speech and Language Processing
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	University of Edinburgh (Reino Unido)
Carácter	
Otros participantes	
Tipo de enseñanza	
Nº créditos	120
Duración oficial	2 semesters + disertation
Idiomas en los que se imparte	
Orientación	
Participación de empresas	
Página web	https://www.ed.ac.uk/ppls/linguistics-and-english-language/prospective/postgraduate/msc/speech-language-processing

Denominación	Language Technology
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad de Göteborg (Suecia)
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	60 o 120
Duración oficial	1 o 2 años
Idiomas en los que se imparte	Inglés
Orientación	Profesional e investigación
Participación de empresas	
Página web	https://flov.gu.se/english/education/masters-second-cycle/mlt

Denominación	Language Technology
Especialidad	
Tipo (oficial/título propio)	Oficial
Universidad coordinadora	Uppsala University
Carácter	Internacional
Otros participantes	
Nº créditos	120
Duración oficial	2 años
Idiomas en los que se imparte	Inglés
Orientación	Profesional e investigación
Participación de empresas	
Página web	http://www.uu.se/en/admissions/master/selma/program/?pKod=HSP2M&pInr=&lasar=19%2F20



Anexo 5. Másteres en Inteligencia Artificial en España y otros países de habla hispana

Denominación	Máster Universitario en I.A. Avanzada: Fundamentos, Métodos y Aplicaciones
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	UNED
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	60
Duración oficial	1 año
Idiomas en los que se imparte	Español
Orientación	
Participación de empresas	
Página web	https://flov.gu.se/english/education/masters-second-cycle/mlt

Denominación	Máster en Inteligencia Artificial
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	UPC-UB-URV
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	90
Duración oficial	1 año y medio
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.fib.upc.edu/es/estudios/masteres/master-en-inteligencia-artificial

Denominación	Máster en Inteligencia Artificial
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	UPM
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	60
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	http://www.dia.fi.upm.es/masteria/



Denominación	Máster Universitario en Investigación e Inteligencia Artificial
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad Internacional Menéndez Pelayo
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Distancia
Nº créditos	60
Duración oficial	1 o 2 años
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	http://www.uimp.es/postgrado/estudios/fichaestudio.php?plan=P03S&any=2017-18&lan=es#

Denominación	Máster en Inteligencia Artificial
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universitat Politècnica de Valencia
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	60
Duración oficial	1 año
Idiomas en los que se imparte	Español
Orientación	
Participación de empresas	
Página web	http://www.upv.es/titulaciones/MUIARFID/index-es.html

Otros países de habla hispana

Denominación	Maestría en Ciencia de la Computación
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad de Oriente (UO), Santiago de Cuba, Cuba
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	70
Duración oficial	3 años
Idiomas en los que se imparte	
Orientación	
Participación de empresas	
Página web	email: silena@uo.edu.cu (coordinadora UO)

Denominación	Maestría en Computación e Informática
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad de Costa Rica
Carácter	Nacional
Otros participantes	



Tipo de enseñanza	Presencial
Nº créditos	60 o 62
Duración oficial	2 años
Idiomas en los que se imparte	
Orientación	
Participación de empresas	
Página web	http://www.pci.ucr.ac.cr/

Denominación	Maestría en Ingeniería de sistemas y computación
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Pontificia Universidad Javeriana
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	40 o 48
Duración oficial	2 años
Idiomas en los que se imparte	
Orientación	
Participación de empresas	
Página web	https://www.javeriana.edu.co/programas/maestria-ingenieria-sistemas-computacion

Denominación	Magíster en Informática
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad de la República
Carácter	Nacional
Otros participantes	
Tipo de enseñanza	Semipresencial
Nº créditos	160
Duración oficial	2 años
Idiomas en los que se imparte	
Orientación	
Participación de empresas	
Página web	http://www.pedeciba.edu.uy/informatica/indice.php

Anexo 6. Másteres en Inteligencia Artificial en Europa y resto del mundo

Denominación	MPhil in Advanced Computer Science
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	University of Cambridge
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.cst.cam.ac.uk/admissions/acs

Denominación	Msc in Artificial Intelligence
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	University of Edinburgh
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	180
Duración oficial	1 a 3 años
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.ed.ac.uk/studying/postgraduate/degrees?id=107&r=site/view

Denominación	Computer Science Master's Degree - Machine Learning
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Columbia University in the City of New York
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Online
Nº créditos	30
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://cvn.columbia.edu/program/columbia-university-computer-science-masters-degree-machine-learning-masters-science

Denominación	Master of Science in Computer Science
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Johns Hopkins
Carácter	Internacional
Otros participantes	



Tipo de enseñanza	Presencial y online
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://ep.jhu.edu/programs-and-courses/programs/computer-science

Denominación	MSc Advanced Computer Science
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	The University of Sheffield
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.sheffield.ac.uk/postgraduate/taught/courses/2019/advanced-computer-science-msc

Denominación	MSc Cybersecurity and Artificial Intelligence
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	The University of Sheffield
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.sheffield.ac.uk/postgraduate/taught/courses/2019/cyber-security-and-artificial-intelligence-msc

Denominación	MSc Artificial Intelligence
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	The University of Manchester
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	http://www.cs.manchester.ac.uk/study/taught-masters/courses/advanced-computer-science/ai/



Denominación	MSc Artificial Intelligence
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Cardiff University
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	180
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.cardiff.ac.uk/study/postgraduate/taught/courses/course/artificial-intelligence-msc

Denominación	MSc Artificial Intelligence
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Imperial College London
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	90
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.imperial.ac.uk/study/pg/computing/artificial-intelligence/

Denominación	Computational Statistics and Machine Learning MSc
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	University College London
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	180
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.ucl.ac.uk/prospective-students/graduate/taught-degrees/computational-statistics-machine-learning-msc



Anexo 7. Másteres en Big Data o Ciencia de Datos

Denominación	Master of Science in Data Science
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Johns Hopkins
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial y online
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://ep.jhu.edu/programs-and-courses/programs/data-science

Denominación	Master of Science in Engineering in Data Science
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	University of Pennsylvania
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://dats.seas.upenn.edu/

Denominación	MSc Data Analytics
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	The University of Sheffield
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.sheffield.ac.uk/postgraduate/taught/courses/2019/data-analytics-msc

Denominación	MSc Data Science and Analytics
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Cardiff University
Carácter	Internacional
Otros participantes	



Tipo de enseñanza	Presencial
Nº créditos	180
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.cardiff.ac.uk/study/postgraduate/taught/courses/course/data-science-and-analytics-msc

Denominación	Data Science and Machine Learning MSc
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	University College London
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	180
Duración oficial	1 año
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	https://www.ucl.ac.uk/prospective-students/graduate/taught-degrees/data-science-machine-learning-msc



Anexo 8. Programas de doctorado en Procesamiento del Lenguaje Natural en España

Denominación	Sistemas Inteligentes – Acceso a la información multilingüe y multimedia
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad Nacional de Educación a Distancia (UNED)
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	3 a 5 años
Idiomas en los que se imparte	Español, inglés
Orientación	
Participación de empresas	
Página web	http://portal.uned.es/portal/page?_pageid=93,62293503&_dad=portal&_schema=PORTAL&idtitulacion=9613#punto_432661539

Denominación	Informática – Tecnologías del Lenguaje Humano
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad de Alicante (UA)
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Español
Orientación	
Participación de empresas	
Página web	https://web.ua.es/es/phdinf/doctorado-en-informatica.html

Denominación	Filosofía y Ciencias del Lenguaje – Lingüística Informática
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad Autónoma de Madrid (UAM)
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	3 a 5 años
Idiomas en los que se imparte	Español, inglés y francés
Orientación	
Participación de empresas	
Página web	https://www.uam.es/EscuelaDoctorado/(es_ES)-Programa-Doctorado-en-Filosof%C3%ADa-y-Ciencias-de-Lenguaje/1429099818752.htm?language=es_ES&nDept=1&pid=1429099816822&pidDept=1429099817036



Denominación	Ciencia Cognitiva y Lenguaje – Procesamiento, representación y adquisición del lenguaje
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universitat de Barcelona (UB)
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Español y catalán
Orientación	
Participación de empresas	
Página web	http://www4.giga.ub.edu/acad/pops/pub0201.php?id=HDK08&lang=2

Denominación	Ciencia y tecnología informática – Human Languages and Accessibility Technologies
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad Carlos III de Madrid (UC3M)
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	3 a 5 años
Idiomas en los que se imparte	Español
Orientación	
Participación de empresas	
Página web	https://www.uc3m.es/ss/Satellite/Doctorado/es/Detalle/Estudio_C/1371210536455/1371210298470/Ciencia_y_Tecnologia_Informatica

Denominación	Ingeniería Informática – Procesamiento del Lenguaje Natural
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad Complutense de Madrid (UCM)
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Español
Orientación	
Participación de empresas	
Página web	https://informatica.ucm.es/doctorado-ing-inf-99

Denominación	Traducción y Ciencias del Lenguaje – Lingüística Computacional e ingeniería del lenguaje
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universitat Pompeu Fabra (UPF)
Carácter	Internacional
Otros participantes	



Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	3 a 5 años
Idiomas en los que se imparte	Español
Orientación	
Participación de empresas	
Página web	https://www.upf.edu/web/doctorats/traduccio-i-ciencies-del-llenguatge

Denominación	Análisis y procesamiento del lenguaje
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad del País Vasco (UPV/EHU)
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Español
Orientación	
Participación de empresas	
Página web	https://www.ehu.eus/es/web/doctoradohaplap/aurkezpena

Denominación	Informática – Machine Learning and Language Processing
Especialidad	
Tipo (oficial/título propio)	
Universidad coordinadora	Universidad Politécnica de Valencia (UPV)
Carácter	Internacional
Otros participantes	
Tipo de enseñanza	Presencial
Nº créditos	
Duración oficial	
Idiomas en los que se imparte	Inglés
Orientación	
Participación de empresas	
Página web	http://www.upv.es/entidades/EDOCTORADO/info/1004670normalc.html



Estrategia Formativa en TL

ET2: Directrices generales, propuesta de acciones formativas para la estrategia de formación en TL

Aitziber Atutxa (UPV/EHU), Nuria Bel (UPF), Arantza Díaz de Ilarraza (UPV/EHU), Patricio Martínez (UA), Raquel Martínez (UNED), Koldo Gojenola (UPV/EHU), Antoni Oliver (UOC), German Rigau (UPV/EHU), Alfonso Ureña (UJ)

Julio 2019

Este estudio ha sido realizado dentro del ámbito del Plan de Impulso de las Tecnologías del Lenguaje con financiación de la Secretaría de Estado para el Avance Digital, que no comparte necesariamente los contenidos expresados en el mismo. Dichos contenidos son responsabilidad exclusiva de sus autores.

Reservados todos los derechos. Se permite su copia y distribución por cualquier medio siempre que se mantenga el reconocimiento de sus autores, no se haga uso comercial de las obras y no se realice ninguna modificación de estas.

1. INTRODUCCIÓN	70
2. OBJETIVOS Y ORGANIZACIÓN DEL DOCUMENTO	70
3. METODOLOGÍA DE TRABAJO	71
4. DIRECTRICES GENERALES DE FORMACIÓN	71
5. PROPUESTA DE TEMÁTICAS.....	75
6. ACCIONES FORMATIVAS.....	76
7. CONCLUSIONES.....	80
8. REFERENCIAS	81
9. GLOSARIO DE SIGLAS Y ACRÓNIMOS	82

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

<i>TABLA 13. DIRECTRICES SOBRE LA FORMACIÓN DE LAS TL</i>	74
---	----

ET2: Directrices generales, propuesta de temáticas y acciones formativas para la estrategia de formación en TL

1. INTRODUCCIÓN

El Plan de Impulso de las Tecnologías del Lenguaje es una acción del Gobierno incluida en la Agenda Digital para España, coordinada por la Secretaría de Estado para la Sociedad de la Información y la Agenda Digital para España (SESIAD). El objetivo principal del plan es fomentar el desarrollo del procesamiento del lenguaje natural, la traducción automática y los sistemas conversacionales en lengua española y lenguas cooficiales. Uno de los subobjetivos del plan es el establecer las líneas maestras para un plan de formación de las tecnologías lingüísticas en el ámbito universitario con el fin de: i) maximizar la eficiencia en la gestión y asignación de los fondos de formación destinados a la formación continua en TIC, tanto de personal del sector privado como del sector público; ii) asignar parte de los recursos disponibles para la formación continua a la capacitación y adquisición de habilidades digitales de profesionales TIC; iii) reorientar la Formación Profesional relacionada con las TIC; iv) potenciar una mejora de la oferta universitaria destinada a la formación de profesionales TIC.

2. OBJETIVOS Y ORGANIZACIÓN DEL DOCUMENTO

El presente documento tiene como objetivo presentar una propuesta de temáticas, acciones formativas y directrices generales que contribuya al desarrollo de los objetivos estratégicos de la Agenda Digital para España y del Plan de Impulso de las Tecnologías del Lenguaje. Este documento es el resultado de la discusión e intercambio de ideas entre los participantes en la elaboración del plan de formación de TL y propone recomendaciones que permitan definir el Plan de Formación y Divulgación de TL.

Para ello:

1. Se presentarán las directrices generales para la estrategia de formación de TL.
2. Se presentará una propuesta de temáticas.
3. Se hará una exposición de las acciones formativas diseñadas según las directrices generales que recojan las temáticas identificadas.

La estructura del presente documento es la siguiente. Primero se justifica la necesidad de crear el Plan de Formación y Divulgación. A continuación, se describen los ámbitos (público, empresarial y académico) de actuación y los objetivos de formación. Estos objetivos se desarrollarán en una serie de actuaciones que incluyen la elaboración de cursos, materiales formativos, guías, eventos, premios o jornadas de divulgación. Todas estas actuaciones se describen en mayor detalle en los apartados siguientes.

3. METODOLOGÍA DE TRABAJO

La metodología de trabajo que se ha seguido para la realización del presente estudio se ha basado en:

1. La recogida de información sobre experiencias similares; para ello, se han estudiado en detalle las iniciativas en el ámbito de la Ciberseguridad diseñadas por el Instituto Nacional de Ciberseguridad (INCIBE)
2. La realización de entrevistas y contactos personales con expertos en la materia a nivel nacional, europeo e internacional.

Como resultado de esta recopilación de información, presentaremos en el apartado 3, las directrices generales de formación en TL.

4. DIRECTRICES GENERALES DE FORMACIÓN

Las directrices generales que se desarrollen han de satisfacer las necesidades de formación en tecnologías lingüísticas para un amplio público. Los ámbitos de actuación que hemos identificado incluyen el sector empresarial (Pymes u otro tipo de empresas), la administración pública, el ámbito académico y el público en general. Así podríamos distinguir los siguientes grupos objetivos: i) empresas y profesionales expertos en tecnologías lingüísticas; ii) empresas y profesionales que son regularmente usuarios de TL y son conocedores del valor añadido que las tecnologías lingüísticas proporcionan a su modelo de negocio; iii) empresas y profesionales en general, desconocedores del potencial de las tecnologías del lenguaje en sus áreas de negocio; iv) personas con orientación universitaria, es decir, personas que han decidido realizar una carrera universitaria o que, estando realizando sus estudios universitarios, deciden continuarlos cursando algún máster o programa de doctorado especializado. v) cuadros medios de la administración pública que, al igual que cualquier otro profesional pueden ser desconocedores del potencial de las tecnologías del lenguaje en sus

ámbitos de responsabilidad, tienen el deber añadido de no solo aplicar sino velar por la aplicación de criterios de igualdad de raza y género, y del Reglamento General de Protección de Datos vi) la sociedad en su conjunto.

1. **Empresas y profesionales expertos en tecnologías lingüísticas:** se trata de profesionales con amplios conocimientos en las TLs que realizan su labor en empresas para los que, considerando la velocidad de desarrollo de nuevas técnicas en el área, resulta crítico tener acceso a una formación especializada continua orientada a la mejora y actualización de su conocimiento. Esta formación ha de promoverse desde la academia, realizada por expertos investigadores en el área.
2. **Empresas y profesionales usuarios de TL mediante el que proporcionan un valor añadido a su modelo de negocio:** en este grupo incluimos a profesionales que hacen uso de TL en su tarea empresarial satisfactoriamente pero desconocen la influencia que herramientas de TL más avanzadas tendrían en su negocio. Para este grupo de profesionales se ha de proporcionar un apoyo formativo cuyas herramientas útiles serían: i) cursos on-line; ii) seminarios y talleres sobre herramientas avanzadas en TL; iii) cursos MOOC avanzados; iv) aulas Universidad-empresa para la resolución de casos prácticos; v) *hackatones* multidisciplinares que aporten nuevas ideas de trabajo.
3. **Empresas, cuadros medios de las administraciones públicas y profesionales en general:** para este grupo de profesionales se ha de hacer una labor de difusión de la potencialidad de las tecnologías lingüísticas en su empresa mediante la organización de: i) jornadas de presentación del área, orientadas a empresas, que incluyan la exposición de casos de uso realizados con éxito; ii) cursos on-line/MOOC introductorios; iii) *hackatones* multidisciplinarios; iv) aulas Universidad-empresa para la resolución de casos prácticos.
4. **Personas con orientación universitaria:** dada la cada vez mayor demanda de profesionales con conocimientos en TL por parte del tejido empresarial e industrial, se ha de proporcionar formación básica en cuanto a nivel de grados universitarios y formación avanzada a nivel de másteres y programas de doctorado. Además se debe impulsar que dichos másteres/programas de doctorado se realicen en el contexto de consorcios europeos y/o en colaboración con las empresas (doctorados industriales): i) Erasmus Mundus para másteres; ii) redes ITN para programas de doctorado.

5. **Jóvenes, educadores/as:** programas de difusión para potenciar el interés por la ciencia y tecnología en general y de TL en particular. Conviene que se realcen desde edades tempranas en colaboración con los/las educadores/as.

Pasamos a presentar las acciones formativas a tomar para los distintos ámbitos de actuación. La tabla que se incluye a continuación recoge un resumen de estas directrices y sus ámbitos de actuación.

Ámbitos de actuación	Directrices
Empresas y profesionales expertos/as en TL	Formación especializada continua sobre investigación puntera.
Empresas y profesionales usuarios/as de TL	Cursos on-line avanzados
	Seminarios y talleres sobre herramientas avanzadas en TL
	Cursos MOOC avanzados
	Hackatones multidisciplinares
	Aulas Universidad-empresa para la resolución de casos prácticos.
Empresas, cuadros medios de las administraciones públicas y profesionales en general	Cursos MOOC introductorios
	Jornadas orientadas a empresas de presentación del área con exposición de casos de uso exitosos
	Hackatones multidisciplinares
	Aulas Universidad-empresa para la resolución de casos prácticos.
Personas con orientación universitaria	Formación avanzada universitaria: grado, postgrados.
	Redes europeas de Másteres universitarios (Erasmus Mundus)
	Redes europeas de programas de doctorado (ITN)
Jóvenes, educadores	Programas de difusión

Tabla 13. Directrices sobre la formación de las TL

5. PROPUESTA DE TEMÁTICAS.

Pasamos a detallar las temáticas que proponemos para cada una de las directrices identificadas en el apartado anterior.

- **Formación especializada continua sobre investigación puntera.** Se tratarán temas avanzados relacionados con las TLs como, por ejemplo, *Chatbots*, aplicación de técnicas de *Deep Learning* para traducción, clasificación en diferentes áreas relacionadas con el procesamiento del lenguaje natural y/o los sistemas de diálogo. Dentro de esta directriz también podemos incluir los *summer school*, o cursos de verano donde se puede tratar de forma amplia un tema o incluir diferentes seminarios que tocan diversos temas. Ejemplos de este tipo de cursos son los cursos de verano organizados por diferentes universidades o los organizados por empresas como Vicomtech.
- **Cursos on-line.** Estos cursos pueden ser introductorios, en los que se tratan temas generales como: i) visión general del procesamiento del lenguaje natural, ii) definición y visión general de los sistemas de diálogo, iii) qué es la traducción automática y cómo se puede abordar. Estos cursos también pueden ser de carácter avanzado para tratar temas específicos como por ejemplo: i) procesamiento de texto mediante el uso de herramientas *unix*, ii) modelado del lenguaje iii) traducción automática utilizando técnicas de *Deep Learning*, iv) *machine learning* y sus aplicaciones en procesamiento del lenguaje natural.
- **Seminarios y talleres sobre herramientas avanzadas en TL.** Aquí se darán a conocer herramientas de utilidad para el usuario o experto de TLs. Normalmente estas herramientas no han sido desarrolladas con un objetivo comercial, sino más bien como software de libre distribución. Algunos ejemplos de herramientas a tratar, o mostrar, en estos seminarios y talleres son: i) analizadores morfo-sintácticos disponibles para castellano y lenguas cooficiales, ii) sistemas de traducción para las lenguas cooficiales iii) sistemas de diálogo para las lenguas oficiales y cooficiales. En la actualidad la SEPLN (Sociedad Española de Procesamiento del Lenguaje Natural) a lo largo de diferentes años está organizando seminarios o talleres de este tipo.
- **Hackatones multidisciplinares.** Se podrían organizar convenciones cuyo objetivo es crear software libre de forma colaborativa entre todos los participantes. Algo parecido a lo que se viene haciendo con algunas versiones del sistema operativo Linux pero aplicado a TL. Un

ejemplo de software a desarrollar sería: un sistema de traducción válido para cualquiera de las lenguas que oficialmente se pueden utilizar en el territorio español.

- **Aulas Universidad-empresa para la resolución de casos prácticos y jornadas orientadas a empresas de presentación del área con exposición de casos de uso exitosos.** El objetivo es que existan encuentros en los que profesionales de la empresa e investigadores de las universidades cambien impresiones mediante la exposición de casos prácticos. Un ejemplo es la herramienta Xuxen, corrector ortográfico para euskara, que comenzó siendo un tema de investigación dentro del grupo IXA y ha terminado siendo un producto creado por Elhuyar Hizkuntza eta Teknologia.
- **Formación universitaria avanzada.** Actualmente existen numerosas propuestas de formación universitaria avanzada a nivel nacional, europeo e internacional orientadas a la formación en TL. Estas incluyen máster y programas de doctorado aunque en España no existen estudios de grado. Algunos ejemplos son los másteres ofrecidos por la Universidad de Educación a Distancia (UNED) o el Máster en procesamiento del lenguaje natural de la Universidad del País Vasco (UPV/EHU).

6. ACCIONES FORMATIVAS

En este apartado haremos una exposición de las acciones formativas diseñadas según las directrices generales que recogen las temáticas identificadas.

FORMACIÓN CONTINUA SOBRE INVESTIGACIÓN PUNTERA

Cursos organizados por los grupos de investigación nacionales, europeos o internacionales que trabajan en el área. Para determinar qué cursos ofertar se seguirán las líneas de investigación puntera marcadas por congresos de prestigio nacional e internacional como SEPLN, ACL, COLING, etc. También se tendrán en cuenta las necesidades que vayan surgiendo desde el ámbito empresarial o las que marquen investigadores internacionales de prestigio.

SEMINARIOS Y TALLERES SOBRE HERRAMIENTAS AVANZADAS EN TL

Cursos organizados por expertos en TL que darán o conocer a otras personas que trabajan en TL herramientas de reciente desarrollo o creación. El objetivo es dar a conocer la utilidad de las mismas en el ámbito de las TLs. Estos cursos deben ser organizados tanto por las empresas como por ámbitos universitarios, de esta forma se establecerán sinergias entre ambos sectores facilitando su colaboración y desarrollo de futuros proyectos comunes.

CURSOS MOOC

MOOC es el acrónimo en inglés de **Massive Online Open Courses** (o cursos *online* masivos y abiertos). Todos los meses se ofrecen numerosos cursos MOOC (más de 100), ofrecidos por universidades de prestigio a nivel mundial. Actualmente hay más de 850 que lo están haciendo. En España la Universidad Politécnica de Madrid fue la primera en ofrecer este tipo de cursos, actualmente también lo está haciendo Telefónica y Universia (quien integra a 1.232 universidades de 23 países de Iberoamérica). Algunos ejemplos de este tipo de cursos relacionados con TL son: i) Introducción a la Minería de Datos ii) *Introduction to Deep Learning* iii) *Natural Language Processing*, etc. Estos cursos pueden ser generales relacionados con TL, donde se da una visión global de las tecnologías de la lengua o, pueden ser específicos donde se ahondan en conceptos o técnicas muy concretas; de esta forma se aborda la formación para personas en general o personas usuarias de TLs.

HACKATONES

Un *hackaton* es un encuentro de programadores cuyo objetivo es el desarrollo colaborativo de software libre. Esta idea se quiere llevar al campo de las TLs con objeto de desarrollar herramientas que cubran las necesidades que vayan surgiendo, tanto en empresas usuarias de estas tecnologías como en usuarios en general. De esta forma la puesta en común de problemáticas inherentes a las propias TLs hará que los participantes en los *hackatones* aprendan gracias a la resolución de problemáticas de forma colectiva. El software desarrollado será de libre uso una vez que haya sido debidamente testado por los propios participantes.

AULAS UNIVERSIDAD-EMPRESA PARA LA RESOLUCIÓN DE CASOS PRÁCTICOS

Las aulas universidad-empresa son aulas-laboratorio, situadas normalmente dentro de un ámbito universitario (campus universitario), que son creadas y financiadas por empresas con el objetivo de promover actividades de I+D+i. Son un instrumento de colaboración entre la universidad y las empresas que abarcan actividades relacionadas con: i) investigación, ii) desarrollo tecnológico, iii) innovación, iv) formación de nuevas personas que trabajen en la empresa o reciclaje de las que ya están trabajando. El objetivo sería orientar las aulas universidad-empresa a crear laboratorios relacionados con TL.

SEMINARIOS Y TALLERES PRESENTANDO CASOS DE USO REALIZADOS CON ÉXITO

Un taller o seminario es un evento en el cual los asistentes se forman sobre un determinado tema de manera intensiva. Estos talleres suelen contar con la participación de uno o más especialistas que se

encargan de dar una charla y de explicar conceptos teóricos. Después, bajo la supervisión de un moderador los asistentes realizan alguna actividad que permita llevar a la práctica esta teoría. En nuestro caso se pueden organizar seminarios relacionados con diferentes ámbitos de las TLs: i) procesamiento del lenguaje natural, ii) traducción automática iii) sistemas de diálogo.

CHARLAS DE PRESENTACIÓN DEL ÁREA

Serán charlas organizadas por expertos en TLs que deberán ser extensamente publicitadas para que jóvenes y educadores conozcan de su existencia y puedan asistir a las mismas. El objetivo es dar a conocer en qué consisten las TLs y transmitir una idea general de su utilidad para captar la atención de potenciales usuarios. De estas charlas también pueden surgir nuevas necesidades que pueden ser planteadas por los asistentes, y de esta forma se irá ampliando el espectro de necesidades a cubrir por las TLs.

FORMACIÓN AVANZADA UNIVERSITARIA: GRADO, POSTGRADOS

El objetivo es doble, por un lado, crear asignaturas optativas en los últimos cursos del grado que se especialicen en temas relacionados con TL ofreciendo una visión general. Por otro lado, a partir del nicho creado en las optativas de grado se pretende orientar a los alumnos hacia estudios de postgrado donde se ahonde en las TL. Tanto los cursos de grado como los de postgrado serán impartidos por profesores-investigadores universitarios que están desarrollando su labor investigadora en el área.

REDES EUROPEAS DE MÁSTERES UNIVERSITARIOS (ERASMUS MUNDUS)

El objetivo es crear estudios de postgrado en los que estén involucradas universidades de diferentes países. De esta forma los alumnos que cursen el máster recibirán clases en las distintas universidades percibiendo visiones diferentes de una misma temática. Un ejemplo es el máster EMLCT, organizado por el grupo de investigación en procesamiento del lenguaje natural IXA de la UPV/EHU.

REDES EUROPEAS DE PROGRAMAS DE DOCTORADO (ITN)

Una de las peculiaridades del Doctorado en Unión Europea es la participación de profesores universitarios de universidades de diferentes países. Los convenios de colaboración permiten la realización de tesis doctorales en régimen de co-tutela entre dos universidades y estancias de investigación en los centros involucrados. Actualmente la UNED, entre otras universidades españolas, dispone de programas de doctorado de este tipo.

JORNADAS DE DIFUSIÓN

Las jornadas de difusión pretenden dar a conocer a potenciales estudiantes de grado y postgrado las posibilidades de formación que tienen, en nuestro caso serían los tipos de formación explicados en los apartados 5.8 y 5.9. Actualmente esta labor se realiza en las diferentes universidades anualmente antes del periodo de matriculación. También es muy interesante, de cara a captar alumnos de grado, acudir a los centros de enseñanza secundaria para dar a conocer los posibles grados relacionados con TL que existen o se van a crear.

7. CONCLUSIONES

En este informe se han explicado las líneas maestras para establecer un plan de formación de las Tecnologías Lingüísticas (TL). Para ello se han expuesto las directrices generales a seguir para crear una estrategia de formación en TLs. También se han propuesto unas temáticas a tratar orientadas a establecer dicha formación. Se han expuesto las acciones formativas en función de los diferentes perfiles que se pueden requerir los potenciales expertos y profesionales en TLs.

Básicamente, para fomentar el desarrollo de las Tecnologías Lingüísticas, el estudio se ha dividido en tres áreas principales: Procesamiento del Lenguaje Natural, la Traducción Automática y los Sistemas Conversacionales en lengua española y lenguas cooficiales. Algunos de los ejemplos expuestos son directamente prácticos, es decir que ya están implantados, otros son propuestas.

Podemos concluir que en el campo de la formación en TLs hay muchos instrumentos formativos disponibles y muchas tareas formativas por emprender, algunas de ellas ya están comenzadas y otras se deben iniciar. Las primeras deben ser potenciadas y mejoradas. Por otro lado, las segundas requieren hacer un esfuerzo mayor aún. No debemos olvidar que en ambos casos es fundamental realizar una labor de difusión, más profunda de lo que se está haciendo hasta ahora, sobre todos los tipos de usuarios: la sociedad en general, expertos universitarios, formación continua para empresas y la administración, etc.

A nivel de formación académica avanzada se cree necesaria la creación de un grado orientado a capacitar expertos en el ámbito de las tecnologías lingüísticas. También se deben fortalecer y armonizar los contenidos sobre Tecnologías del Lenguaje en los grados relacionados: informática, ciencia de datos, inteligencia artificial, lingüística, Traducción e interpretación, etc. Y potenciar los másteres y doctorados industriales especializados en Tecnologías del Lenguaje. Hemos observado que es un área con amplia demanda tanto a nivel empresarial como científico (ver ET1, sección 6).

Además, al ser una tecnología transversal que puede aplicarse en todas las áreas de conocimiento humano (ya que todas usan el lenguaje para comunicarse), recomendamos también proporcionar ciertas actividades formativas a profesionales de ámbitos donde las Tecnologías Lingüísticas tienen un impacto potencial muy elevado (ciencias sociales y humanidades, periodistas, sociólogos, médicos, abogados, etc.) .

8. REFERENCIAS

- [1] ERASMUS MUNDUS LCT - European Masters Program in Language & Communication Technologies, <http://ixa2.si.ehu.es/erasmusmundus>
- [2] Curso de verano organizado por empresa <http://rtthss2019.vicomtech.org/>
- [3] Máster en procesamiento del lenguaje natural de la UNED http://portal.uned.es/portal/page?_pageid=93,61491281&_dad=portal&_schema=PORTAL&idAsignatura=24409117
- [4] Máster en procesamiento del lenguaje natural de la Universidad Politécnica de Cataluña <https://www.fib.upc.edu/es/investigacion/grupos-de-investigacion/gpln-grupo-de-procesamiento-del-lenguaje-natural>

9. GLOSARIO DE SIGLAS Y ACRÓNIMOS

AI	Artificial Intelligence
CPI	Compra Pública de Innovación
DL	Deep Learning
ML	Machine Learning
MT	Machine Translation
MOOC	Massive On-line Open course
NLP	Natural Language Processing
NN	Neural Network
SEAD	Secretaría de Estado para el Avance Digital
STEM	Science, Technology, engineering, Mathematics
TL	Tecnologías del Lenguaje

Estrategia Formativa en TL

ET3: Identificación de sectores para la estrategia de formación TL.

Aitziber Atutxa (UPV/EHU), Nuria Bel (UPF), Arantza Díaz de Ilarraza (UPV/EHU),
Patricio Martínez (UA), Raquel Martínez (UNED), Koldo Gojenola (UPV/EHU),
Antoni Oliver (UOC), German Rigau (UPV/EHU), Alfonso Ureña (UJ)

Julio 2019

Este estudio ha sido realizado dentro del ámbito del Plan de Impulso de las Tecnologías del Lenguaje con financiación de la Secretaría de Estado para el Avance Digital, que no comparte necesariamente los contenidos expresados en el mismo. Dichos contenidos son responsabilidad exclusiva de sus autores.

Reservados todos los derechos. Se permite su copia y distribución por cualquier medio siempre que se mantenga el reconocimiento de sus autores, no se haga uso comercial de las obras y no se realice ninguna modificación de estas.



1. OBJETIVOS Y ORGANIZACIÓN DEL DOCUMENTO	88
2. METODOLOGÍA DE TRABAJO	88
3. MARCOS NORMATIVOS.....	89
3.1 DISCIPLINAS FORMATIVAS UNIVERSITARIAS	89
3.2 FORMACIÓN CONTINUA	90
4. ÁREAS DE CONOCIMIENTO Y RECOMENDACIONES PARA LA FORMACIÓN EN TL	92
5. ÁMBITOS FORMATIVOS	95
5.1 INICIATIVAS NO UNIVERSITARIAS	95
5.2 INICIATIVAS UNIVERSITARIAS (GRADOS)	96
5.3 MÁSTER.....	103
7. REFERENCIAS	106
8. GLOSARIO DE SIGLAS Y ACRÓNIMOS	107

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS



ET3: Identificación de Sectores para la estrategia de formación TL. Directrices generales para abordar la estrategia de formación.

1. OBJETIVOS Y ORGANIZACIÓN DEL DOCUMENTO

La estrategia de formación en Tecnologías del Lenguaje (TL) debe tener una perspectiva amplia para responder al desafío al que se enfrenta Europa y en particular España. Es importante analizar la demanda y la oferta de las habilidades y conocimiento multidisciplinar que se requieren, así como la necesidad de formar en disciplinas no técnicas que pueden tener impacto en las aplicaciones de las TL y se beneficien de las mismas. La estrategia de formación en TL debe garantizar que existan planes de estudios adecuados para hacer frente a la demanda de conocimientos y para reconocer la necesidad del aprendizaje permanente y de formación de expertos profesionales. Otro aspecto importante es plantear la formación a la sociedad de manera que aumente el conocimiento de la ciudadanía sobre el uso adecuado de estas tecnologías, así como de sus riesgos.

El objetivo de este documento es presentar las áreas de conocimiento (STEM, humanidades, ciencias sociales y jurídicas, ciencias de la salud, economía, turismo, etc.) de mayor interés en las que convendría definir una estrategia de formación específica en las tecnologías del lenguaje. En este informe expondremos:

1. La metodología utilizada para la realización de este informe.
2. El marco normativo que debería aplicarse a cada ámbito formativo.
3. Las áreas de conocimiento donde debe centrarse la estrategia formativa en TL.
4. Los ámbitos formativos donde debe aplicarse la estrategia formativa en TL y centrándonos en particular en el ámbito universitario.

2. METODOLOGÍA DE TRABAJO

La metodología de trabajo que se ha seguido para la realización del presente estudio se ha basado en:

1. Estudio de marcos normativos para formación en las distintas disciplinas universitarias, incluyendo grados universitarios, másteres, doctorado y formación continua.



2. Prospección e identificación de diferentes casos de uso exitoso junto con la revisión de conclusiones de distintos informes y estudios que exponen las necesidades profesionales futuras en el ámbito de las TL para determinar las áreas de interés.
3. Estudio detallado de iniciativas similares. En particular, las iniciativas realizadas por INCIBE en el ámbito de la ciberseguridad.
4. Recomendaciones para formación en la agenda digital española y europea en temas de IA y PLN.

3. MARCOS NORMATIVOS

En esta sección se revisarán los distintos marcos normativos que regulan la definición y acreditación de formación de nivel universitario, incluyendo los grados, máster, doctorado y formación continua.

3.1 DISCIPLINAS FORMATIVAS UNIVERSITARIAS

Según el marco normativo actual, el objetivo primero de la formación universitaria es preparar para el ejercicio de la práctica profesional, aunque no exclusivamente. La educación basada en competencias en el marco del Espacio Europeo de Educación Superior (EEES) ha quedado definida en la normativa (Real Decreto 1393/2007 de 29 de octubre) como “el conjunto de conocimientos, destrezas y actitudes necesarias para desarrollar una ocupación”, implementando así el objetivo de primar una formación universitaria capacitadora para realizar una profesión. Así pues, la educación universitaria basada en competencias se ha de entender enmarcada en la necesidad social de formar profesionales, personas que tras conseguir un título universitario puedan desempeñar autónomamente las tareas asociadas a un empleo.

Máster

Teniendo en cuenta el Real Decreto 1393/2007 de 29 de octubre, las propuestas de las Universidades incluirán los contenidos generalistas y de formación básica en los planes de estudios de títulos de Grado y contenidos especializados en los planes de estudios de títulos de Máster. Estos estudios tienen como finalidad la adquisición "de una formación avanzada, de carácter especializado o multidisciplinar, orientada a la especialización académica o profesional, o bien a promover la iniciación en tareas investigadoras".

Ese documento también establece que los planes de estudios conducentes a la obtención del Máster tendrán entre 60 y 120 créditos, incluyendo "la formación teórica y práctica que el estudiante deba

adquirir: materias obligatorias, materias optativas, seminarios, prácticas externas, trabajos dirigidos, trabajo de fin de Máster, actividades de evaluación, y otras que resulten necesarias según las características propias de cada título".

Doctorado

El Real Decreto 99/2011, de 28 de enero, regula las enseñanzas oficiales de doctorado, indicando que la Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) o los órganos de evaluación de las comunidades autónomas inscritos en el Registro Europeo de Agencias de Calidad (*European Quality Assurance Register, EQAR*) establecerán conjuntamente los protocolos de evaluación necesarios para la verificación y acreditación de acuerdo con los mencionados criterios y directrices y conforme a lo dispuesto en este real decreto.

En ese decreto se añade que el doctorado debe jugar un papel fundamental como intersección entre el Espacio Europeo de Educación Superior (EEES) y el Espacio Europeo de Investigación (EEI) incluyendo, además de universidades y organismos públicos de investigación, a empresas, hospitales o fundaciones que han de participar en la formación doctoral y en la inclusión de los doctores en sus actuaciones cotidianas. Para garantizar la calidad del doctorado, la universidad deberá justificar la existencia de equipos investigadores solventes y experimentados en el ámbito correspondiente.

3.2 FORMACIÓN CONTINUA

La Ley Orgánica 6/2001, de 21 de diciembre, de Universidades, en su artículo primero establece que la Universidad realiza el servicio público de la educación superior mediante la investigación, la docencia y el estudio, y dentro de las funciones de la Universidad al servicio de la sociedad incluye explícitamente la difusión del conocimiento y la cultura a través de la extensión universitaria y la formación a lo largo de toda la vida. Por su parte, dicha ley establece en los artículos 2 y 34 que las universidades, en uso de su autonomía, podrán elaborar y aprobar planes de estudio de enseñanzas específicas de formación a lo largo de la vida, así como establecer enseñanzas conducentes a la obtención de diplomas y títulos propios.

Además, de acuerdo con la Disposición adicional undécima del Real Decreto 1393/2007, de 29 de octubre, por el que se establece la ordenación de las enseñanzas universitarias oficiales (Boletín Oficial del Estado de 30 de octubre de 2007), las universidades en uso de su autonomía podrán impartir otras enseñanzas conducentes a la obtención de otros títulos distintos a los de Grado, Máster y Doctorado, recogidos en el artículo 3.1 de este Real Decreto. La expedición de estos títulos se realizará del modo que determine la universidad, sin que ni su denominación ni el formato en que

se confeccionen los correspondientes títulos puedan inducir a confusión con los títulos oficiales anteriormente citados y regulados por los artículos 9, 10 y 11 de este Real Decreto.

Por otra parte, el Consejo de la Unión Europea realiza una serie de recomendaciones conducentes al espacio europeo de formación permanente (Resolución del Consejo de 27 de junio de 2002 sobre la educación permanente, DOCE 2002/C163/01), estableciendo así un marco de referencia común para los diferentes agentes implicados en la elaboración de los planes formativos.

De esta manera, cada universidad podrá establecer, de acuerdo a su propia normativa, qué títulos propios pueden expedirse y qué características tienen las enseñanzas conducentes a dichos títulos.

No obstante, desde la variedad que permite la legislación vigente, la mayoría de las universidades recogen los siguientes tipos de enseñanzas con características similares, generalmente con objetivo de ofrecer una formación multidisciplinar, adaptada a las demandas de la sociedad, y que favorezca la actualización de conocimientos y la especialización profesional:

1. Cursos / certificados de especialización: referidos a enseñanzas muy específicas con cargas típicas inferiores a 10 créditos ECTS.
2. Título propio de Experto Universitario/Profesional: cuyo objetivo es el perfeccionamiento y la especialización profesional, con cargas lectivas típicas entre 10 y 20 créditos ECTS.
3. Título propio de Especialista Universitario: que abordan la formación de alto nivel orientada a su aplicación en actividades profesionales, con cargas lectivas típicas entre 25 y 45 créditos ECTS
4. Título propio de Máster: que constituyen enseñanzas que tienen como finalidad la adquisición por el estudiante de una formación avanzada y complementaria de carácter especializada o multidisciplinar, orientada en general a la especialización académica y, sobre todo, profesional, con cargas lectivas entre 60 y 120 créditos ECTS.

4. ÁREAS DE CONOCIMIENTO Y RECOMENDACIONES PARA LA FORMACIÓN EN TL

Según el [Plan Coordinado en Inteligencia Artificial](#)⁹ [4] de la Comisión Europea, el rápido progreso tecnológico modificará las habilidades requeridas de los trabajadores, lo que significa que potencialmente un gran número de trabajadores necesitará mejorar sus habilidades, especialmente en las TICs. Casi todos los Estados miembros se enfrentan a la escasez de profesionales de la tecnología de la información y las comunicaciones, y actualmente hay más de 600.000 vacantes para expertos digitales¹⁰. Según este mismo informe, el progreso en IA abre la puerta a nuevas oportunidades en áreas como atención médica personalizada y de precisión, movilidad (conducción autónoma), tecnología financiera, economía circular y bio sostenible, detección e investigación mejoradas de actividades delictivas (por ejemplo, lavado de dinero, fraude fiscal), medios de comunicación, etc.

Existen actualmente claros ejemplos de casos de uso exitosos que emplean las TL en distintos ámbitos. Los siguientes ejemplos no forman una lista exhaustiva de dichos casos de uso, sino que simplemente pretenden ilustrar los múltiples ámbitos en los que ya se emplean herramientas que incorporan TL. En el sector de las telecomunicaciones, por ejemplo, Telefónica ha desarrollado *Aura*, un asistente virtual que incorpora procesamiento de lenguaje natural para inferir en tiempo real las intenciones de sus usuarios, relacionándolas con sus datos personales y ofreciéndoles propuestas individualizadas (por ejemplo, cuántos datos quedan por consumir y si podría convenir un plan distinto al contratado, o qué película ver). En el sector bancario, CaixaBank lanzó durante 2018 un asistente virtual que entiende el lenguaje natural y que da respuesta inmediata a más de 30.000 empleados en cuestiones complejas sobre normativa interna, legislación, productos bancarios, recursos humanos, tecnología, etc. Alrededor del 80% de las consultas habituales de los empleados pueden resolverse a través de este sistema. Casi todas las entidades bancarias cuentan con su *bot* para la atención al cliente. En el sector turístico, Telefónica en colaboración con Natoural Digital Solutions, Samsung, e historiadores, ha desarrollado una aplicación para *smartphone* o *tablet* y gafas de realidad mixta ofreciendo una nueva experiencia guiada de hacer turismo gracias a la realidad aumentada y la realidad mixta inmersiva. Estas iniciativas incorporarán guías virtuales con los cuales se podrá interactuar, planteándoles preguntas sobre la época, las costumbres, un personaje

⁹ https://ec.europa.eu/knowledge4policy/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en

¹⁰ https://www.pocbigdata.eu/monitorICTonlinevacancies/general_info/

determinado, etc. En el sector farmacéutico, Genpact, en colaboración con Bayer, ha creado la herramienta de farmacovigilancia PVAI que haciendo uso de herramientas de TL puede extraer y codificar Reacciones Adversas a Medicamentos que aparecen en documentos de origen diverso como por ejemplo redes sociales y foros de medicina no profesionales. En el sector médico, ASHO ha desarrollado *Ashoindex2* y *AshoCoode* basados en los últimos avances en *Machine Learning*, técnicas de procesamiento de lenguaje natural y métodos de búsqueda semántica para codificación de conceptos médicos a través de estándares como el CIE10 promovido por la Organización Mundial de la Salud. En el sector jurídico existen herramientas como *Everylaw*, que es una herramienta avanzada de recuperación de información de litigios, especialmente útil en sistemas jurídicos como el anglosajón, que se basa enormemente en la jurisprudencia. O el proyecto MIREL financiado por la unión europea, cuyo fin es definir un marco formal y desarrollar herramientas para extraer y razonar (*Mining and Reasoning with Legal texts*) sobre textos jurídicos, transformando estos textos jurídicos en representaciones sobre las que se pueden realizar consultas referentes a normativas, controles de conformidad, y soporte a las decisiones. En el sector de la comunicación se utilizan herramientas empleadas por los periodistas expertos en periodismo de rendición de cuentas (*watchdog journalism*) como *Document Cloud* que permite buscar, analizar y publicar fuentes primarias; *Timeline JS*, herramienta que crea líneas temporales visuales; *Ushahidi* que crea de forma colectiva mapas, etc. *Overview* (una plataforma más avanzada, que clasifica y ordena documentos en función de su contenido).

Como resultado de esta prospección se evidencia el carácter multidisciplinar de las aplicaciones de las TL. Nótese que las aplicaciones mencionadas han sido desarrolladas por multinacionales. Estas aplicaciones en un futuro cercano serán más y más comunes, no solo en las grandes empresas sino en todo el tejido empresarial español, europeo y latinoamericano. De hecho, el [Plan Coordinado en Inteligencia Artificial](#) subraya la importancia de fomentar la incorporación más amplia posible de IA (y por ende de las TL) en la economía, en particular por parte de *start-ups* y pequeñas y medianas empresas, a través del aumento de la conciencia pública y la divulgación de los últimos avances científicos y las tecnologías de vanguardia probadas y desarrolladas en Europa, para garantizar que todas las empresas, pequeñas o grandes, de alta tecnología o no, así como el sector público, puedan aprovechar estas oportunidades digitales.

En vista de estas previsiones de crecimiento del sector de Procesamiento de Lenguaje Natural, se considera conveniente ampliar la formación en esta tecnología en diversas áreas de conocimiento no solo restringidas a STEM, sino pertenecientes a las humanidades, ciencias de la salud, jurídicas, etc.



Por otro lado, en el entregable ET2 se han identificado los sectores en los que es necesario incidir a la hora de hacer una propuesta de formación en el área de las tecnologías lingüísticas:

- La Administración pública: se prevé que la administración pública sea uno de los agentes impulsores principales en el ámbito de la industria del lenguaje ya que se generalizará el uso de plataformas que incorporen tecnologías de Procesamiento del Lenguaje para aumentar la capacidad, calidad y automatismo de los servicios de la propia administración así como de la atención al ciudadano por los canales telefónico e Internet. Por ejemplo, el uso de la traducción automática ampliará el alcance de comunicación y reducirá los costes.
- La empresa como demandante de avances en el área de las tecnologías lingüísticas: las empresas tienen que prepararse para los retos que se derivan de la internacionalización y reaccionar de la forma más adecuada y eficiente. Su modo de funcionar está sufriendo y sufrirá cambios importantes para los que tiene que estar preparada. Las tecnologías lingüísticas facilitan las tareas asociadas a la internacionalización pero no existen suficientes profesionales en las empresas para poder aplicarlas adecuadamente de manera que permita obtener de ellas un rendimiento adecuado. Es imprescindible para mejorar las oportunidades de las empresas españolas, reclutar, retener y formar (y desde las propias empresas participar y facilitar de esta formación) el talento adecuado con las habilidades necesarias para las tecnologías lingüísticas. Por lo tanto, a medida que los sectores industriales tradicionales experimentan una transformación, también deben hacerlo sus fuerzas de trabajo. Además, los expertos en TL necesitan una comprensión de las consecuencias éticas planteadas como consecuencia de las aplicaciones que desarrollan, necesitan una buena comprensión del panorama legal y normativo como, por ejemplo, del Reglamento General de Protección de Datos (GDPR en inglés). Es decir, la necesidad de desarrollar e integrar la fiabilidad, la seguridad y privacidad a través del desarrollo de tecnología, productos y servicios adecuados.
- Centros de investigación y transferencia: se han realizado grandes avances en el ámbito de las tecnologías del lenguaje, pero el área está sufriendo una revolución con la incorporación de nuevos métodos y algoritmos. La investigación en España en el área de TL ha alcanzado un nivel de madurez remarcable, sus logros son reconocidos a nivel internacional y es importante continuar en la misma línea. Por ello, es crucial formar investigadores jóvenes que impulsen la investigación avanzada que se realiza en los grupos de investigación.

- Universidades y organizaciones de formación profesional que representan la oferta; así como estudiantes, profesionales y personas que participan como consumidores directos de educación, capacitación y otros servicios profesionales, como certificación y búsqueda de empleo.
- La Sociedad en general: En el contexto de inclusión las Tecnologías del Lenguaje (TL) y, por extensión, las derivadas de las aplicaciones en Inteligencia Artificial afectarán las habilidades que necesita la sociedad en general. Normalmente, los usuarios de sistemas basados en TL serán personas sin experiencia en estadística, matemáticas o ciencias de la computación. Para que las tecnologías lingüísticas sean aceptadas por la sociedad, debemos asegurarnos de que los usuarios no expertos tengan una comprensión y conocimiento básicos de los sistemas con capacidades de procesamiento lingüístico y de cómo operan. Esto es necesario para evitar el uso indebido y garantizar que las personas puedan aceptar y confiar en las soluciones basadas en el procesamiento del lenguaje hablado o escrito.

5. ÁMBITOS FORMATIVOS

5.1 INICIATIVAS NO UNIVERSITARIAS

Podemos distinguir las iniciativas dirigidas a menores, familias y educadores, a los cuales es indispensable dirigirse, por un lado para informarles sobre las aportaciones de la TL y su transferencia a la sociedad, por otro, para concienciarles sobre los riesgos del empleo no ético de las TL. Por ejemplo, la capacidad de ciertas herramientas (*deepfake*) de fabricar de forma automática noticias falsas (*fakenews*) difícilmente distinguibles de las reales o que conozcan el alcance e implicaciones derivadas de compartir información personal, opiniones en las redes, etc (ver el caso [Cambridge Analytics](#)).

Para ello sería preciso implementar acciones similares a las realizadas por is4K ([Internet Segura for Kids](#)) y la ESA ([European Space Agency](#)) que tienen como objetivo explicar las actividades realizadas en el ámbito de la ciberseguridad la primera y el espacio la segunda, adaptadas a un público infantil y juvenil. Este tipo de iniciativas buscarían mostrar a estos colectivos sin conocimientos técnicos las TLs de una manera entretenida. También sería interesante organizar Jornadas escolares y crear materiales didácticos adaptados.

Además, deberían existir iniciativas similares dirigidas a Bachilleres, FP, Universidades y profesores de estos cursos. Estas iniciativas no solo perseguirían informar, sino formar al mismo tiempo que se busca y capta el talento en TL pensando en el medio-largo plazo. Para ello sería también

imprescindible convocar becas de estudios de especialización en TL para, por ejemplo, realizar campus de verano en las TL.

El Instituto Nacional de Administración Pública (INAP) debería configurar cursos técnicos y divulgativos específicos para el personal de toda la administración pública.

En cuanto a las iniciativas de orientación profesional, estas se deben orientar sobre los cursos de grados y posgrados que obtienen la capacitación para convertirse en un profesional de las TL.

No hay que olvidar las iniciativas dirigidas a profesionales y emprendedores de las TL: cursos de formación en formato MOOC, organización de Jornadas profesionales, convocatorias de programas de incorporación de talento y formación, etc.

Para finalizar, los recursos formativos para empresas en general, no específicamente de las TL, incluirían MOOC, títulos propios y programas de formación continua en tecnologías de la lengua, que permitirían formar transversalmente a profesionales en el campo de las TL.

5.2 INICIATIVAS UNIVERSITARIAS (GRADOS)

Seguidamente, realizaremos un breve repaso a las distintas titulaciones de grado que deberían incorporar conocimientos adecuados a cada perfil, empezando por las titulaciones de tipo STEM y continuando por las de humanidades, ciencias sociales y jurídicas, y ciencias de la salud.

Telecomunicaciones

En los grados en Ingeniería de Tecnologías y Servicios de Telecomunicación, en Ingeniería de Sistemas de Telecomunicación y otros grados afines encontramos en su mayoría una asignatura genérica de tratamiento o procesamiento digital de señales. Sin embargo, no en todos los grados se ofertan asignaturas más específicas de voz o habla y relacionadas con su procesamiento desde el punto de vista del tratamiento del lenguaje natural. Destaca el Grado en Ingeniería en Tecnologías de Telecomunicación de la UC3M que oferta dos asignaturas optativas de este perfil: [Procesamiento del lenguaje natural](#) y Análisis de Datos Audiovisuales. En la mayoría de este tipo de grados se ofertan únicamente asignaturas centradas en audio y voz en itinerarios de intensificación en sonido e imagen. Ejemplos de este tipo de asignaturas son: Tratamiento de señales de voz y audio y Tecnologías de audio.



Informática

El [libro blanco de Informática](#)¹¹ (pp. 163) que recoge las funciones propias a desarrollar por un Ingeniero en Informática destaca las siguientes: inteligencia artificial y nuevas tecnologías; optimización de métodos y medios de comunicación con el computador y los usuarios. Más concretamente, en la Categoría 2, Contenidos Específicos de la Ingeniería en Informática incluye en la subcategoría 2.2 (Ingeniería del Software, Sistemas de Información y Sistemas Inteligentes) la Interacción Persona-Computadora e Inteligencia Artificial. Así, el Bloque temático 7 (Sistemas Inteligentes) incluye entre sus contenidos formativos la Inteligencia Artificial, Sistemas basados en el conocimiento, Minería de datos, y Control inteligente.

Con ello, el perfil de Ingeniería en Informática definido anteriormente le capacita para el ejercicio de algunas Actividades Profesionales entre las que se incluye el de Ingeniería del Conocimiento y que comprende entre otros la minería de datos, redes neuronales, redes bayesianas, aprendizaje automático, lenguaje natural, etc.

Como se ha visto en la sección 3 del ET1, en la mayoría de los grados en informática que se pueden cursar en España se imparten asignaturas genéricas relacionadas con las TL como pueden ser Minería de Datos, Aprendizaje Automático, o Técnicas de Inteligencia Artificial. Sin embargo, solo unos pocos grados de informática ofertan materias sobre Procesamiento del Lenguaje Natural y, en los que se hace, estas materias se limitan muchas veces a ser optativas de libre elección, con alguna excepción como en el caso de la ya citada UC3M.

En los grados de informática es a través de las especialidades y las menciones en las que el alumnado adquiere cierta especialización. Algunas de estas especializaciones o menciones contienen más carga lectiva en esas asignaturas genéricas que acabamos de nombrar. Por ejemplo, en especializaciones como Computación, la carga lectiva en asignaturas genéricas de IA relacionadas con TL suele ser superior a la carga en especializaciones como Sistemas o Ingeniería de Computadores. En cualquier caso y sea cual sea la especialización no existe una oferta coherente y suficiente en asignaturas de Procesamiento del Lenguaje Natural, ni en general de tratamiento de datos no estructurados. Recordemos que según algunas estimaciones, los datos digitales no estructurados (texto, audio, video, etc.) engloban el 80% de los datos digitales totales.

¹¹ http://www.aneca.es/media/150388/libroblanco_jun05_informatica.pdf



Como resumen podemos destacar el hecho de que, al no tener el PLN una gran presencia en titulaciones técnicas como informática que en teoría podrían presentar mayor afinidad, es previsible que en otras titulaciones de la rama STEM su ocurrencia sea menor todavía.

Matemáticas

Las asignaturas de los grados en matemáticas están centradas en las materias propias de esa disciplina y no ofertan, en general, asignaturas de campos de aplicación de esas materias. En el primer curso de estos grados se oferta una asignatura de programación o informática, en algunos casos hasta dos asignaturas, y algunas universidades ofertan itinerarios de computación, con asignaturas más avanzadas en esta materia, pero no se ha encontrado ningún itinerario en el que aparezcan asignaturas de procesamiento de lenguaje natural.

Es habitual encontrar oferta de dobles titulaciones, como el doble grado en Informática y Matemáticas, y recientemente se están empezando a ofertar grados de matemáticas más centrados en computación, como por ejemplo el Grado en Matemática Aplicada y Computación de la UC3M, que incluye asignaturas de inteligencia artificial, o el Grado en Matemática Computacional de la UNIR. Pero en estos casos tampoco se incluyen asignaturas específicas de procesamiento de lenguaje natural.

Ciencia e Ingeniería de datos

En las propuestas de grados de Ciencias e Ingeniería de datos se ofertan asignaturas genéricas como Minería de Datos Masivos, Tratamiento de los Datos, Visualización de Datos, Recuperación de la Información y Análisis de la Web o incluso Aspectos Legales de Datos. En lo referente a las asignaturas relacionadas con las TL la situación es heterogénea. Por ejemplo, en el Grado en Ciencia de Datos de la Universidad de Valencia se oferta la asignatura obligatoria Procesado del Lenguaje Natural. En otros grados similares, como por ejemplo el Grado en Ingeniería Matemática en Ciencias de Datos de la Universidad Pompeu Fabra existe una asignatura denominada Técnicas de Procesamiento de Lenguaje Natural para Aplicaciones Cotidianas pero que es de carácter optativo. Finalmente, en la gran mayoría de estos grados, no existe ninguna asignatura de PLN.

Física

Al igual que en los grados en matemáticas, los grados en física se centran en materias propias de esa disciplina. En algunos casos en primer curso se oferta una asignatura de programación o informática, pero no en todos los planes de estudios. Como era de esperar, no se ha

encontrado ningún itinerario en el que aparezcan asignaturas de procesamiento de lenguaje natural.

Nuevos grados de Inteligencia Artificial

Algunos de los nuevos grados de Inteligencia Artificial que actualmente se están diseñando y que se espera implantar en breve en varias facultades de informática, incorporan contenidos de asignaturas obligatorias y optativas de Procesamiento del Lenguaje Natural. Por ejemplo, el diseño del nuevo grado de Inteligencia Artificial de la UPV/EHU¹² (actualmente en trámite de acreditación por UNIBASQ) incorpora dos asignaturas obligatorias y una optativa (18 créditos totales) con contenidos de Tecnologías del Lenguaje: Procesamiento del Lenguaje Natural, Minería de Datos Textuales y Técnicas Avanzadas de PLN. Sin embargo, los detalles concretos de sus planes de estudio aún no son públicos. Esperamos que dichos planes de estudio generalicen la implantación de contenidos y asignaturas de Tecnologías del Lenguaje.

Lingüística

En cuanto a la formación en grados de Lingüística, Lenguas Aplicadas y/o Filología en España se observa que se están incorporando las materias de TL en algunos planes de estudio. Por ejemplo, la Universidad Complutense de Madrid imparte desde el curso 2011-2012 el Grado en Lingüística y Lenguas Aplicadas con un itinerario específico en Lingüística Computacional de 36 ECTS, con asignaturas básicas de Introducción a la Informática, Introducción a la Estadística, y asignaturas de especialización: (i) Lingüística Computacional, (ii) Lingüística del Corpus y Herramientas de Análisis Textual, (iii) Procesamiento del Lenguaje Natural, y (iv) Programación para el Procesamiento del Lenguaje Natural. Es el programa más completo que hemos visto en la materia de TL.

En el Grado en Lenguas Aplicadas de la Universidad Pompeu Fabra se impartirán 31 ECTS a partir del curso 2019-2020, en seis asignaturas, dos de introducción y cuatro que incluyen la formación en competencias específicas de un perfil profesional de Tecnologías del Lenguaje que es optativo: Marcaje de textos, Procesamiento Automático de Textos, Procesamiento del Habla e Industrias de la Lengua. Estas asignaturas se complementan con una introducción a la informática y Métodos Empíricos para el Estudio del Lenguaje.

¹² <https://www.ehu.eus/ehusfera/ifbloga/2019/07/15/la-upv-ehu-ha-aceptado-el-grado-en-inteligencia-artificial/>



La Facultad de Filosofía y Letras de la Universidad de Cádiz imparte desde 2010 un Grado en Lingüística y Lenguas Aplicadas que incluye dos asignaturas, de 6 ECTS cada una, Lingüística Computacional e Ingeniería del Lenguaje.

La Facultad de Filología de la UB, en su Grado en Lingüística ofrece asignaturas de la materia: Lingüística Computacional y Tecnologías de la Información y la Comunicación. En los demás grados en Lenguas Aplicadas en España, hay una cobertura muy modesta de la materia. Por ejemplo, la Universidad de Lérida ha creado recientemente (curso 2017-18) un Grado en Lingüística Aplicada y Traducción que ofrece únicamente una asignatura de TL.

En cuanto a centros europeos, la especialización en tecnologías del lenguaje en la formación de lingüistas se está configurando como una materia que requiere de más formación específica y ligada a materias que no son tradicionalmente de lo que llamamos letras. El Itinerario de Tecnologías del Lenguaje en el grado de Lingüística de la Universidad de Helsinki (2017), por ejemplo, ofrece para el Grado en Lingüística una especialización en Tecnologías del Lenguaje con 9 asignaturas de 5 ECTS: Introducción a la Lingüística Computacional, Construcción de aplicaciones de Procesamiento del Lenguaje Natural, Morfología computacional, Sintaxis computacional, Semántica Computacional, Modelos y algoritmos en aplicaciones de Procesamiento del Lenguaje Natural, y las que destacan por aportar los conocimientos básicos: Programación para lingüistas, Aprendizaje automático para lingüistas y Matemáticas para lingüistas. El Grado en Lingüística Computacional de la Universidad de Saarland también incluye asignaturas de fundamentos matemáticos, de lingüística, de lingüística computacional y seminarios de proyectos de software para la formación de técnicos especialistas en esta materia.

Traducción e Interpretación

En España la mayoría de grados en traducción no incluyen asignaturas de tecnologías del lenguaje en general y solamente algunas incluyen formación específica en tecnologías de la traducción. Por ejemplo, el Grado en Traducción, interpretación y lenguas aplicadas de la Universidad de Vic y Universitat Oberta de Catalunya que ofrece dos asignaturas de tecnología: una orientada a traductores e intérpretes en la que se tratan con profundidad las herramientas de traducción asistida por ordenador y otra en forma de asignatura optativa que es una introducción al procesamiento del lenguaje natural con programación en Python y uso de la plataforma *Natural Language Toolkit*.

La Universidad de Granada ofrece como asignatura obligatoria de su Grado en Traducción la asignatura obligatoria Herramientas Informáticas para Traductores e Intérpretes en la que



estudian únicamente herramientas básicas de traducción asistida por ordenador (TAO): memorias de traducción y gestión terminológica. El Grado de Traducción de la Universidad de Alicante también ofrece una única asignatura de Tecnologías de la traducción.

La Universidad Pompeu Fabra ofrece también una asignatura obligatoria de Tecnologías de la Traducción en segundo curso en la que se presentan las herramientas de TAO y la traducción automática. Sin embargo, ofrece en cursos posteriores la posibilidad de cursar asignaturas optativas relacionadas con la tecnología: Localización, Traducción Automática y Gestión de terminología.

Periodismo

Existen unas 105 ofertas de distintos grados de periodismo, relacionados o afines a este en el estado español entre las universidades públicas y privadas, que van desde el grado en Periodismo, grado en Gestión de la Información y Contenidos Digitales, a dobles grados como doble grado en Periodismo y Comunicación Audiovisual, Periodismo y Humanidades, Periodismo e Información y Documentación, Periodismo y Redes Sociales o Ciencias Políticas y Periodismo. En ninguno de estos grados se ha encontrado ninguna asignatura relacionada con las TL y esto contrasta con las nuevas tendencias que comienzan a asomar como el periodismo computacional, *robot journalism*s, periodismo de datos, etc. Un informe reciente de *Reuters Institute* [6], trata sobre la incorporación de este tipo de tecnologías por parte de las grandes agencias de noticias, por ejemplo para la generación automática de noticias empleando herramientas de Generación Automática de Lenguaje (*Natural Language Generation* NLG) para generar noticias relativas al tiempo o a los deportes o de herramientas de traducción automática de vídeos. También nos habla de las cuestiones éticas que suscita el uso de este tipo de herramientas.

Ciencias Sociales y Ciencias Políticas

En los grados de Ciencias Sociales y Ciencias Políticas no se imparten asignaturas relacionadas con las TL. Sin embargo, varios trabajos recientes versan sobre la transformación de la investigación en ciencias políticas que se está iniciando a través de las TL como consecuencia de la disponibilidad de explorar grandes cantidades de datos empleando herramientas sofisticadas de TL. Esta nueva vía está derivando en importantes avances en los métodos de investigación, abre nuevas interrogantes y vías de investigación, y modifica las actuales formas de entender. Los autores de estos trabajos argumentan que no



es tanto la necesidad de informáticos, sino de formar a los politólogos y sociólogos para obtener la capacitación de aplicar las TL [7,8,9].

Administración y dirección de empresas

En los grados de ADE no existe oferta alguna de asignaturas que incorporen las TL. En grados relacionados, como *Business Analytics*, impartido en la Universidad Francisco de Vitoria, existe la posibilidad de cursar la especialidad o mención en ADE, con varias asignaturas de Minería de Datos (I, II, III) pero no existe una específica de Procesamiento del Lenguaje Natural o similar. En la Universidad Pública de Navarra existe un doble grado en ADE y Ciencias de Datos, pero en este, al igual que en el anterior, se imparten asignaturas generales de IA, relacionadas con las TL como puedan ser Aprendizaje Automático, Preparación de Datos, Aprendizaje Profundo, pero no hay ninguna asignatura de Procesamiento del Lenguaje Natural. También se considera una carencia que los alumnos de este tipo de grados no tengan la posibilidad de cursar ninguna asignatura que verse sobre las TL, dado que cada vez se están empleando más en el sector financiero este tipo de herramientas para por ejemplo mejorar la interacción con los clientes, medir el impacto y conocimiento que estos tienen sobre nuestro negocio, conocer la situación de nuestros competidores, monitorizar las actividades de estos últimos para por ejemplo recibir avisos de nuevos productos que hayan desarrollado etc.

Medicina y Farmacia

En los grados de medicina y farmacia impartidos en España no se han encontrado ofertas de asignaturas relacionadas con las TL. Sin embargo, en el grado de Ingeniería Biomédica impartido en la Universidad Politécnica de Madrid, en el Itinerario en Informática Biomédica, se oferta la asignatura PLN y recuperación de información como optativa. Sería interesante que los alumnos de estos grados conociesen la existencia de ontologías médicas como *Snomed*, *UMLS (Universal Medical Language System)*, o estándares como el CIE-10 (Clasificación Internacional de Enfermedades), que se podrían mostrar en asignaturas de las TL junto con herramientas de recuperación y extracción de información que les permitan estar continuamente actualizados en cuanto a los logros más recientes en su especialidad, nuevas medicaciones, efectos adversos, etc.

Derecho

La oferta de asignaturas en TL en los grados de derecho es inexistente a día de hoy, excepto en un grado relacionados como es el previamente mencionado *Business Analytics* de la Universidad Francisco de Vitoria en la especialidad o mención en derecho empresarial.

Otros grados

Las Tecnologías del Lenguaje son transversales y pueden ser aplicadas en múltiples ámbitos, disciplinas y dominios. En esta sección hemos mencionado algunos grados universitarios donde el conocimiento de las Tecnologías del Lenguaje desde un punto de vista técnico (informática, lingüística, telecomunicaciones, traducción e interpretación, ciencia de datos, inteligencia artificial, etc.) o de dominio de aplicación (derecho, ciencias sociales, medicina, etc.) puede ser interesante incluir conocimientos más o menos profundos sobre la materia. Puede haber otros grados donde el conocimiento de las capacidades actuales de las Tecnologías del Lenguaje puede ser relevante (seguridad, audiovisuales, psicología, etc.)

5.3 MÁSTER

En el entregable ET1, “Estudio realizado sobre la oferta formativa en tecnologías lingüísticas a nivel nacional, europeo e internacional en países de referencia”, se ha visto que en España ya existe un número relativamente importante de másteres sobre PLN e IA (Tablas 4 a 8). A este respecto se puede afirmar que, aunque todavía haya una distancia significativa respecto a países de mayor tradición científica y tecnológica como EEUU y varios países europeos, la demanda de formación a nivel de máster en PLN e IA queda cubierta en mayor grado, en comparación a otros niveles educativos, al menos en algunas universidades. Sin embargo, la situación dista de ofrecer una especialización en PLN en la mayor parte de las universidades españolas.

6. CONCLUSIONES

De todo lo expuesto se deduce que es importante desarrollar la formación transversal y profesional en tecnologías lingüísticas. La investigación ha de ser la responsable en participar en la transformación educativa y responder a la necesidad de formaciones transversales y específicas siguiendo los últimos desarrollos de las tecnologías de TL. Esta adaptación de la formación debe extenderse a las metodologías de capacitación profesional en un contexto de aprendizaje permanente y accesible.

Uno de los mayores problemas a los que se enfrenta el desarrollo de la TL en Europa es la escasez de expertos. La irrupción de la TL requiere de una educación adaptada a la sociedad digital en la que nos encontramos, así como la necesaria formación en competencias digitales en nuestro sistema educativo y laboral. Esta transformación educativa implica la necesidad de formaciones transversales y específicas en temáticas relacionadas con las TL tanto para el desarrollo del talento en profesionales del área como para la necesaria actualización de conocimientos de la sociedad en general. Es fundamental para que los ciudadanos puedan ser conscientes de la disrupción provocada por la transformación digital y la aplicación masiva de las tecnologías del lenguaje en su vida cotidiana. Para conseguir que los ciudadanos comprendan las implicaciones del uso de las TL y maximicen las oportunidades que conlleva, es necesario que dispongan de competencias para relacionarse con las tecnologías lingüísticas y sus distintos ámbitos de actuación.

De la misma manera que se recoge en el informe Estrategia Española de I+D+i en Inteligencia Artificial para la enseñanza de la IA, vemos que en el caso de las tecnologías lingüísticas es necesario también adaptar el modelo educativo, contemplando la hibridación formativa de ciencias, tecnologías y humanidades. Estos conceptos deberían estar presentes en un gran número de grados universitarios, en especial, pero no únicamente en los vinculados con la informática, las ciencias, las tecnologías, las ingenierías y las matemáticas, sino también en los vinculados a las humanidades, especialmente a la lingüística. Las grandes empresas tecnológicas ya están aplicando este modelo, creando equipos multidisciplinares que han supuesto un aumento en las contrataciones de profesionales con perfil humanístico. Parece obligado actualizar los planes de educación para que sean acordes al presente (y futuro) tecnológico en que vivimos, en lo que a nosotros nos concierne, desde la formación de expertos en la Universidad o la Formación Profesional, hasta alcanzar su especialización en el mercado laboral y manteniendo una actualización mediante programas de aprendizaje permanente. Esto necesariamente requiere, también, formar a los formadores.

Deberemos fomentar la formación continua a lo largo de la vida profesional. Profesionales de áreas como economía, abogacía, medicina, o administración entre otras, deberán ser capaces de entender y emplear estas tecnologías, en la medida de lo necesario. El papel de la I+D+i será imprescindible en el proceso de reflexionar, reformular e impulsar programas educativos de distintas tipologías como grados y postgrados específicos pero también cursos de especialización y cursos MOOC (cursos online de carácter abierto) de alta calidad que permitan dar respuesta a las competencias que requieran las futuras profesiones.

Los planes de estudio deberían ser más flexibles para poder adaptarse a los contextos científicos, tecnológicos, económicos y sociales. En cuanto a niveles de posgrado, y especialmente en el caso de los Programas de Doctorado, sería recomendable impulsarlos, en el marco adecuado, mediante instrumentos como las becas FPI con perfiles de TL, así como en el fomento de tesis industriales que permitan transferencia entre las Universidades, los centros de investigación y las empresas públicas y privadas.

Sería deseable disponer de un programa de fomento de las vocaciones en TL, de acuerdo a la demanda del mercado laboral, con especial atención a la igualdad de género que en este momento supone una brecha importante. Cabe también una mención especial a la atracción, recuperación y retención del talento para que nuestros/as investigadores e ingenieros/as más brillantes formen parte de este proyecto de transferencia de conocimiento y transición hacia una nueva sociedad y economía.

7. REFERENCIAS

- [1] Plan de Impulso de las Tecnologías del Lenguaje.
<https://www.plantl.gob.es/Paginas/index.aspx>
- [2] US News and World Report, 2018. <https://www.usnews.com/best-graduate-schools/top-science-schools/computer-science-rankings>
- [3] LinkedIn. <https://www.linkedin.com/>
- [4] Instituto Nacional de Ciberseguridad (Incibe). <https://www.incibe.es/>
- [5] https://ec.europa.eu/knowledge4policy/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en
- [6] <https://reutersinstitute.politics.ox.ac.uk/our-research/putting-europes-robots-map-automated-journalism-news-agencies>
- [7] John Wilkerson y Andreu Casas. Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges May 2017 Annual Review of Political Science 20(1):529-544 DOI: 10.1146/annurev-polisci-052615-025542
- [8] Sendhil Mullainathan and Jann Spiess. Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives—Volume 31, Number 2—Spring 2017—Pages 87–106
- [9] Justin Grimme. We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together. Political Science and Politics 48(01):80-83 DOI: 10.1017/S1049096514001784



8. GLOSARIO DE SIGLAS Y ACRÓNIMOS

AI	Artificial Intelligence
CPI	Compra Pública de Innovación
DL	Deep Learning
ML	Machine Learning
MT	Machine Translation
MOOC	Massive On-line Open course
NLP	Natural Language Processing
NN	Neural Network
SEAD	Secretaría de Estado para el Avance Digital
STEM	Science, Technology, engineering, Mathematics



Plan TL

B. Diseño de la estrategia de formación

**ET4: Diseño de la estrategia de formación en TL específica
para el sistema universitario español**

Aitziber Atutxa (UPV/EHU), Nuria Bel (UPF), Arantza Díaz de Ilarraza (UPV/EHU),
Patricio Martínez (UA), Raquel Martínez (UNED), Koldo Gojenola (UPV/EHU),
Antoni Oliver (UOC), German Rigau (UPV/EHU), Alfonso Ureña (UJ)

Diciembre 2019

Este estudio ha sido realizado dentro del ámbito del Plan de Impulso de las Tecnologías del Lenguaje con financiación de la Secretaría de Estado para el Avance Digital, que no comparte necesariamente los contenidos expresados en el mismo. Dichos contenidos son responsabilidad exclusiva de sus autores.

Reservados todos los derechos. Se permite su copia y distribución por cualquier medio siempre que se mantenga el reconocimiento de sus autores, no se haga uso comercial de las obras y no se realice ninguna modificación de estas.



1. OBJETIVOS Y ORGANIZACIÓN DEL DOCUMENTO	113
2. METODOLOGÍA DE TRABAJO	113
3. ESTADO ACTUAL DE NECESIDADES FORMATIVAS EN TL	114
3.1 CONCLUSIONES EXTRAÍDAS DEL ESTUDIO DE CARACTERIZACIÓN	114
3.2 CONCLUSIONES DERIVADAS DE DISTINTOS INFORMES	115
4. PREVISIÓN FUTURA DE NECESIDADES EN EL SECTOR	120
5. ANÁLISIS DE COMPETENCIAS DEMANDADAS E IDENTIFICACIÓN DE PERFILES	123
6. PROPUESTAS FORMATIVAS POR NIVELES	124
6.1 PROPUESTA FORMATIVA NO UNIVERSITARIA	124
6.2 PROPUESTA FORMATIVA UNIVERSITARIA	126
7. CONCLUSIONES	162
8. ANEXO. COMPETENCIAS TÉCNICAS GENERALES	163
9. ANEXO. GRADOS DE LINGÜÍSTICA COMPUTACIONAL Y TECNOLOGÍAS DEL LENGUAJE	173
10. REFERENCIAS	174
11. GLOSARIO DE SIGLAS Y ACRÓNIMOS	175

ÍNDICE DE FIGURAS

FIGURA 12. PERFILES REQUERIDOS EN LA IA.....	117
FIGURA 13. CAPACIDADES INCLUIDAS EN AL MENOS UNA FUNCIÓN DE LA EMPRESA.....	117
FIGURA 14. STARTUPS DE IA EN LA UNIÓN EUROPEA QUE HAN RECIBIDO FINANCIACIÓN EN EL ÚLTIMO AÑO.	118
FIGURA 15. PORCENTAJE DE AUTORAS DE ARTÍCULOS CIENTÍFICOS EN EL ÁREA DE LA IA.	119

ÍNDICE DE TABLAS

ET4: Diseño de la estrategia de formación en TL específica para el sistema universitario español

1. OBJETIVOS Y ORGANIZACIÓN DEL DOCUMENTO

En el presente documento se detallarán las propuestas para desarrollar una estrategia de formación en PLN y TL. Esta estrategia abarca varios niveles de formación, incluyendo, la formación pre-universitaria, formación continua, formación dirigida a las Administraciones Públicas, y MOOC, y principalmente la formación de nivel universitario, incluyendo grado, máster y doctorado.

La propuesta se basa en la información recogida en los entregables previos: ET1 (estudio sobre la presente oferta formativa en tecnologías lingüísticas), ET2 (directrices generales) y ET3 (identificación de sectores).

Este documento se organiza de la siguiente forma:

- Descripción de la metodología empleada.
- Análisis del estado actual de necesidades formativas en TL.
- Previsión futura de necesidades en el sector.
- Análisis de competencias basado en la demanda identificada en el ET1.
- Propuestas formativas.

2. METODOLOGÍA DE TRABAJO

La metodología de trabajo que se ha seguido para la realización de este estudio se ha basado en:

1. Valorar las recomendaciones formativas de informes recientes nacionales e internacionales en el ámbito de la Inteligencia Artificial y las Tecnologías del Lenguaje.
2. Identificar necesidades, perfiles y sus competencias asociadas a partir de los estudios anteriores (ET1, ET2, ET3).
3. Desarrollar la propuesta formativa completa para alcanzar esas competencias en los distintos ámbitos de actuación, y en particular el ámbito universitario.
4. Diseñar la estrategia de implantación.

3. ESTADO ACTUAL DE NECESIDADES FORMATIVAS EN TL

Para definir las necesidades formativas en TL, primeramente, examinaremos las conclusiones del estudio de caracterización de TL en España, y seguidamente pasaremos a examinar distintos informes sobre las necesidades de la IA y las TL en distintos ámbitos (España, Europa, países de la OCDE, ...).

3.1 CONCLUSIONES EXTRAÍDAS DEL ESTUDIO DE CARACTERIZACIÓN

En el ESTUDIO DE CARACTERIZACIÓN DEL SECTOR DE TECNOLOGÍAS DEL LENGUAJE EN ESPAÑA¹³, elaborado dentro del Plan de Impulso de las TL, se examinan en detalle las necesidades del sector español de las TL:

1. En general, se trata de un sector de actividad concentrado en microempresas y pymes, tan solo el 17,1% de las empresas consultadas son medianas y alrededor del 8% son grandes empresas. El 70,9% de los empleados del sector de los agentes consultados tienen titulación superior, ingeniería o doctorado; 14,6% ingenierías técnicas o titulaciones medias; 9,2% de nivel directivo; 4,8% administrativo (Encuesta del ONTSI Caracterización del sector de tecnologías del lenguaje 2017. Base Total de 60 entidades, con 35 Empresas y 25 Centros de investigación). De aquí se deduce que claramente es un sector vinculado a la educación superior universitaria.
2. Según el estudio, el 75% de los agentes consultados afirmó haber contratado personal durante el año 2017 y dentro de estos, un 82,2% afirma haber contratado personal especializado en TL: procesamiento del lenguaje natural, traducción automática y sistemas conversacionales. Se estima que existirá una gran demanda de especialistas en previsión de crecimiento del área.
3. De todos los agentes del sector encuestados, un 28,3% de los agentes expresaron haber pensado en contratar personal y haber tenido **dificultad** para encontrar los técnicos adecuados. Según este informe, la mayoría de las empresas considera que existe un déficit de profesionales adecuados puesto que no existe formación con perfiles mixtos que aúnen la necesaria combinación de conocimientos tecnológicos con lingüísticos e idiomáticos. Son finalmente las empresas las que deben asumir la tarea de formar al especialista técnico con competencias lingüísticas, o al especialista lingüista con competencias técnicas. Todo ello implica una insuficiente oferta de titulados especialistas con formación mixta.

¹³ <https://www.plantl.gob.es/tecnologias-lenguaje/actividades/estudios/Paginas/tecnologias-del-lenguaje-en-Espana.aspx>

La mayoría de los centros de investigación entrevistados afirmó que las universidades no cuentan con titulaciones técnicas especializadas en tecnologías del lenguaje con ese perfil mixto buscado. El informe recomienda “la creación de una titulación técnica que integre los conocimientos técnicos y de programación, y los conocimientos lingüísticos necesarios para desarrollar la actividad de tecnologías NECESIDADES DEMANDADAS Y COMPETENCIAS (DEL ET1) del lenguaje”. Esta recomendación junto con las ofertas de trabajo analizadas en el entregable ET1 del presente informe apuntan claramente la necesidad de generar un nuevo perfil multidisciplinar.

3.2 CONCLUSIONES DERIVADAS DE DISTINTOS INFORMES

Como ya hemos comentado en el entregable ET1, existen varios informes nacionales e internacionales¹⁴ recientes que contienen entre sus recomendaciones acciones formativas relacionadas con la IA y, por inclusión, al PLN. En concreto, el informe [Artificial Intelligence: A European Perspective](#)¹⁵, publicado por la Comisión Europea, examina la situación de la IA en la Unión Europea y analiza profusamente el ámbito de la formación en IA. Además de presentar las necesidades y demandas de formación, pone de relieve la importancia de los retos que ésta permitirá enfrentar como posibles riesgos relacionados con la privacidad y aspectos éticos. A nivel nacional, la [Estrategia Española de I+D+i en Inteligencia Artificial](#)¹⁶ dispone entre sus prioridades desarrollar un sistema que fomente la formación transversal y profesional en Inteligencia Artificial. Recomienda incluir la IA en el sistema educativo como palanca de cambio tecnológico del país permitiendo así la incorporación del conocimiento y uso de la IA en el mercado laboral y fomento, recuperación y atracción del talento. Para ello subraya la importancia de crear una plataforma y un ecosistema computacional compuesto por la universidad, los centros científicos y tecnológicos y las empresas. Propone fomentar la transferencia de conocimiento entre centros científicos y tecnológicos y empresas con mayor agilidad a través de consorcios de innovación tecnológica especializados en IA. La Big Data Value Association (BDA), en el informe elaborado en Junio 2019 [Strategic Research, Innovation and Deployment Agenda for an AI PPP](#)¹⁷ analiza la importancia de reconocer las competencias, tanto técnicas como no técnicas, a través de certificaciones avaladas por

¹⁴ OECD (Organisation for Economic Co-operation and Development): *Identifying and Measuring Developments in Artificial Intelligence, Artificial Intelligence in Society*, WIPO (World Intellectual Property Organization): *Technology Trends 2019*.

¹⁵ <https://ec.europa.eu/jrc/en/publication/artificial-intelligence-european-perspective>

¹⁶

<http://www.ciencia.gob.es/portal/site/MICINN/menuitem.26172fcf4eb029fa6ec7da6901432ea0/?vgnnextoid=70fcdb77ec929610VqnVCM1000001d04140aRCRD>

¹⁷ <http://www.bdva.eu/node/1310>

cursos universitarios, y vías de *training* profesional. Propone acciones complementarias como el desarrollo de cursos relacionados con la IA para formar a directivos de la industria y la administración pública sobre la potencialidad de la IA en sus modelos de negocio. El informe de la OCDE [Artificial Intelligence in Society](#)¹⁸ aborda el reto de preparar la sociedad para la transformación del mundo laboral. Se prevé que la IA y el PLN cambien la naturaleza del trabajo, complementando a los humanos en algunas tareas y generando a la vez nuevos tipos de trabajo. En cualquier caso, el impacto dependerá en gran parte de la velocidad de difusión en los diferentes sectores, lo que implica que su rápida adaptación al entorno educativo es crucial en el desarrollo de la sociedad. Las políticas clave para dirigir la transición a la IA deberán incluir salud, economía y, fundamentalmente, educación.

Además, el 26 de junio de 2019, el Grupo de expertos de alto nivel de la Comisión Europea sobre IA (AI HLEG) publicó su segundo informe principal, sobre [Recomendaciones de política e inversión para IA confiable](#)¹⁹. El apartado F del capítulo 2 (página 31) está dedicado a “Generando habilidades y educación apropiadas para la IA”. En particular, las recomendaciones de este informe se centran en cuatro apartados: a) rediseñar los sistemas educativos desde la educación preescolar hasta la superior, b) desarrollar y retener talento en los sistemas europeos de educación superior, c) aumentar la proporción de mujeres en ciencia y tecnología, d) mejorar y renovar las capacidades de la fuerza laboral actual, d) concienciar a los interesados y apoyar la toma de decisiones en las políticas de capacitación.

También queremos mencionar el detallado informe que se elabora anualmente [índice de la IA](#)²⁰ sobre el estado actual del área de la Inteligencia Artificial. La Figura 1 muestra el número de vacantes de empleo por año y las habilidades en IA requeridas. Las habilidades de IA no son mutuamente excluyentes. Es más que evidente el interés creciente en perfiles del ámbito de las Tecnologías del Lenguaje (Procesamiento del Lenguaje en Natural y Reconocimiento del Habla) y las técnicas asociadas (Aprendizaje Automático y Profundo).

¹⁸ https://www.oecd-ilibrary.org/science-and-technology/artificial-intelligence-in-society_eedfee77-en

¹⁹ https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60343

²⁰ <https://hai.stanford.edu/ai-index/previous-reports/2018>

Job openings by AI skills required (2015 – 2017)

Source: Monster.com

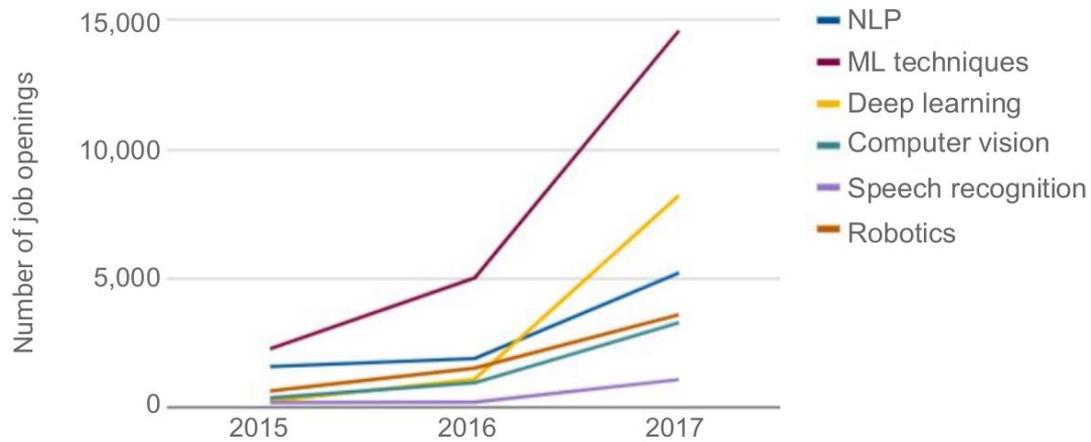
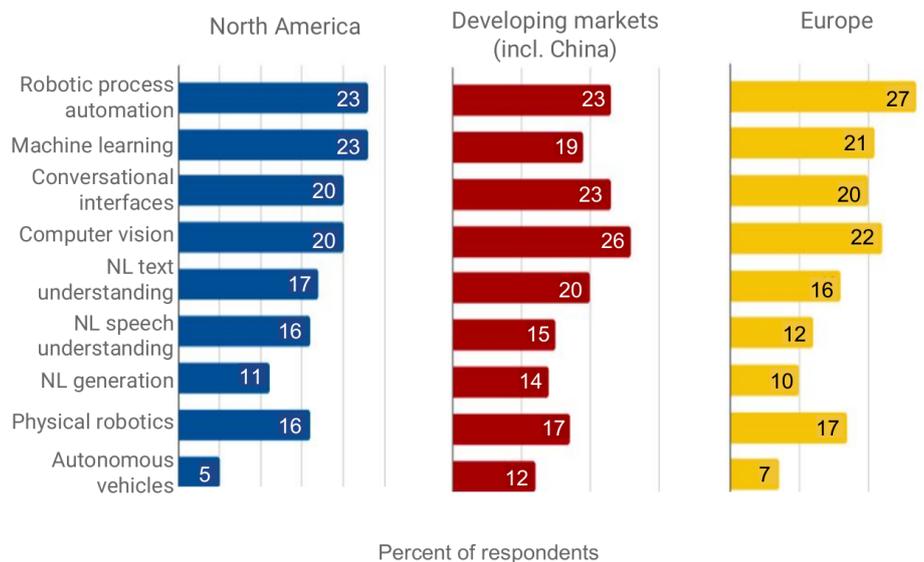


Figura 12. Perfiles requeridos en la IA.

La Figura 12 muestra el porcentaje de encuestados cuyas organizaciones tienen capacidades integradas de inteligencia artificial en al menos una función o unidad de negocio. Los encuestados pueden seleccionar múltiples capacidades de IA.

Capabilities embedded in at least one company function (2018)

Source: McKinsey & Company



Note: The size of each bar is relative to the capabilities within each region; North America: N = 479; Developing markets (incl. China): N = 189 (China N = 35); Europe: N = 803

Figura 13. Capacidades incluidas en al menos una función de la empresa.

Como vemos, las capacidades en Tecnologías del Lenguaje son mayoritarias (Interfaces Conversacionales, Comprensión del Lenguaje Natural, Comprensión del Habla, Generación de

Lenguaje Natural) y las técnicas asociadas (Aprendizaje Automático) y superan con mucho al resto (Automatización de Procesos Robóticos, Visión por Computador, Robótica Física o Vehículos Autónomos).

En el [informe](#) de este mismo año, también se aprecian tendencias muy parecidas a las de los últimos años. La Figura 14 muestra las empresas *startups* de IA en la Unión Europea que han recibido financiación en el último año, por área (julio de 2018-julio de 2019). Entre las áreas que han recibido mayor financiación destacan varias integradas en las Tecnologías del Lenguaje, como el Análisis Textual (4.5%) y también Herramientas de Datos (4.0%), *Chatbots* (3.7%) o Reconocimiento del Habla (2.7%).

AI startups in the European Union: Companies that have received any funding within the last year, by focus area (July 2018-July 2019)

Source: CAPIQ, Crunchbase, Quid, 2019.

Fig. 4.2.6b.

Notes: Network highlighting 993 AI startups in Europe that received investment between July 2018 and July 2019. Colored by focus area with top five labeled.

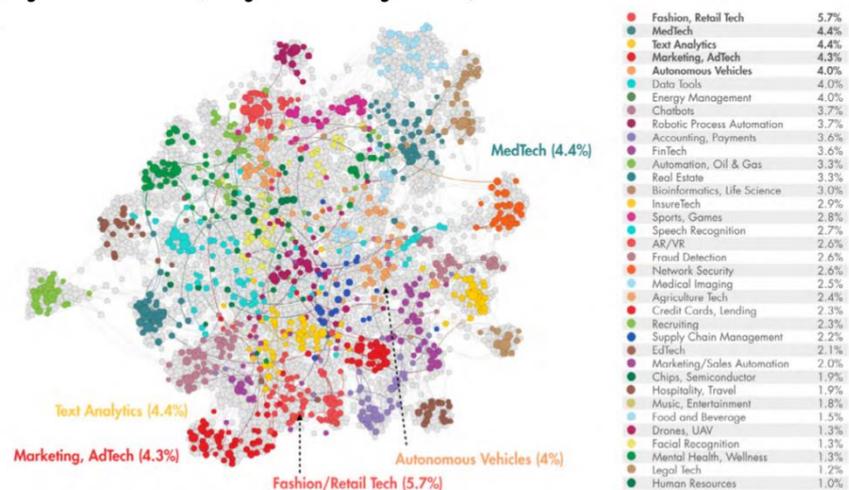


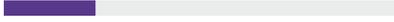
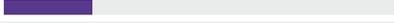
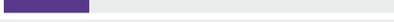
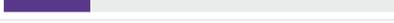
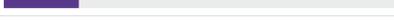
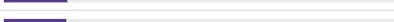
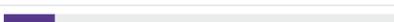
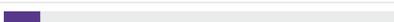
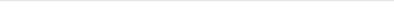
Figura 14. Startups de IA en la Unión Europea que han recibido financiación en el último año.

El informe sobre [talento en IA global](#)²¹ también coincide en manifestar que existe una fuerte evidencia de que la oferta de talento de IA de primer nivel no satisface la demanda. La Figura 4 muestra el porcentaje de autoras de artículos científicos en el área de la IA. España, aunque no llega al 30% del total se sitúa entre los países con mayor porcentaje de autoras.

²¹ <https://ifgaqne.ai/talent-2019/>

Women represented 18% of the conference authors overall

Spain, Taiwan and Singapore had the highest percentage of female conference authors, but the U.S. accounted for almost half in absolute numbers.

Country	Talent by Gender	Total	Ratio 2019
Spain		170	26%
Taiwan		150	23%
Singapore		310	23%
Australia		550	22%
China		2295	22%
United States		9835	20%
Switzerland		260	19%
Italy		475	18%
United Kingdom		1420	18%
India		580	17%
France		795	16%
Germany		1025	16%
Israel		330	14%
Canada		970	14%
Netherlands		230	13%
South Korea		345	12%
Japan		755	9%

Source: 2019 Global Talent Report • jfgagne.ai/talent

Figura 15. Porcentaje de autoras de artículos científicos en el área de la IA.

En resumen, todos los informes coinciden en la demanda creciente de perfiles de Inteligencia Artificial y de Tecnologías del Lenguaje en particular. También coinciden en la escasez de perfiles con capacidades técnicas apropiadas para el área IA y por inclusión de las Tecnologías del Lenguaje. Estos perfiles también deberían incluir capacidades transversales, así como aspectos relacionados con la privacidad y aspectos éticos o legales. También coinciden en el sesgo de género en toda el área STEM. Es por ello que todos los informes también coinciden en las mismas recomendaciones:

- a) rediseñar los sistemas educativos desde la educación preescolar hasta la superior
- b) desarrollar y retener talento de educación superior
- c) aumentar la proporción de mujeres en ciencia y tecnología
- d) mejorar y renovar las capacidades de la fuerza laboral actual
- e) concienciar a los interesados y a la sociedad en general de las oportunidades y riesgos de la IA, y por extensión de las Tecnologías del Lenguaje.

4. PREVISIÓN FUTURA DE NECESIDADES EN EL SECTOR

En lo referente a la previsión futura de las necesidades en el sector de las TL para anticiparse y controlar el impacto en el empleo de los cambios tecnológicos que se avecinan, varios informes nacionales e internacionales llegan a las mismas conclusiones; El informe [Artificial Intelligence: A European Perspective](#)²² dedica la sección 8 (página 71) al tema de la formación. Según dicho informe, en la UE, las estimaciones del número real de personas con conocimientos de inteligencia artificial varían significativamente, de decenas de miles a cien mil. Se estima que hay alrededor de 22.000 estudiantes de doctorado en AI y alrededor de 5.000 personas que han escrito artículos académicos, o realizado presentaciones en conferencias de AI en los últimos años. Se espera que la alta visibilidad de la IA y la demanda actual dirijan el talento de forma relativamente rápida a esta área.

La comunicación final que recoge el [Plan Coordinado en Inteligencia Artificial](#)²³ de la Comisión Europea, identifica la existencia en Europa de una brecha significativa y persistente en las habilidades de TIC. Según este informe, esta brecha es particularmente aguda en áreas emergentes como la IA donde el problema se está acrecentando a medida que la oferta se queda atrás con respecto a la demanda del mercado. Así el informe hace especial hincapié en el hecho de que casi todos los Estados miembros se enfrentan a la escasez de profesionales en el área de AI. La oferta actual de programas especializados de educación superior es limitada y no está igualmente disponible en todos los Estados miembros.

En el [Estudio de Caracterización del Sector de Tecnologías del Lenguaje en España](#)²⁴ la escasez de expertos en el sector se señala como uno de los mayores problemas a los que se enfrenta el desarrollo de la IA en Europa. Es por esto que la sexta recomendación de dicho informe insta a incluir la IA en el sistema educativo como palanca de cambio tecnológico del país. En el informe [For a Meaningful Artificial Intelligence Towards a French and European Strategy](#)²⁵ solicitado por el parlamento francés en 2018, se prevé la necesidad de multiplicar por tres el número de profesionales en IA (y entre ellos especialistas en TL) en los próximos tres años a nivel de grado, máster y doctorado, subrayándose especialmente la creciente necesidad de profesionales calificados de IA

²² <https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/artificial-intelligence-european-perspective>

²³ https://ec.europa.eu/knowledge4policy/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en

²⁴ http://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf

²⁵ https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

que puedan graduarse con el equivalente de un título de licenciatura (es decir, tres o cuatro años de educación superior). Entre los expertos en IA se menciona especialmente a los expertos en Procesamiento del Lenguaje Natural (traducción automática, implicación discursiva, comprensión, generación de texto), que según el informe es probablemente dentro de la IA, el área que más se ha visto beneficiada por los avances en *deep learning*. El objetivo, según este último informe, debería ser no solo capacitar un tipo de profesional que disponga de conocimiento profundo y multidisciplinar, más o menos experto en IA (y en este caso de las TL) y/o otras áreas (jurídica, sanitaria, en ciencias sociales, etc.) a través de estudios de postgrado, sino capacitar un perfil dirigido a las ocupaciones mucho más técnicas de IA, en las que el conocimiento de la IA sería considerado un activo por las empresas. Algunos ejemplos incluyen la industrialización de las técnicas de IA, visualización de datos y producción de análisis o la integración y adaptación de componentes de IA.

Ambos informes también expresan como fundamental adaptar los modelos educativos respectivos dirigiéndolos hacia la multidisciplinariedad formativa en ciencias, tecnologías y humanidades. Ambos concluyen que no solo los grados vinculados con la programación, las ciencias, las tecnologías (los conocidos como STEM por sus siglas en inglés) deberían capacitar a profesionales con habilidades en TL, obviamente a distintos niveles. En concreto, en el informe francés se identifican, por ejemplo, las siguientes áreas: ciencias sociales, económicas, jurídicas, física y química, biología y salud, ecología y desarrollo sostenible. Todo ello sin olvidar que todos los informes y estudios mencionados destacan la necesidad de formar también en lo referente a los posibles riesgos relacionados con la inteligencia artificial, especialmente los de privacidad y ética.

Para finalizar, el [Plan Coordinado en Inteligencia Artificial](#)²⁶ de la Comisión Europea ya mencionado invita a los estados miembros a llevar a cabo el necesario compromiso político para tomar medidas siguiendo las presentes recomendaciones:

1. Incluir la dimensión de habilidades en las estrategias nacionales de IA (a mediados de 2019). Hacer coincidir la oferta educativa nacional, las necesidades de habilidades (la IA también debe ser parte de integrarse en otras disciplinas como la jurídica, ciencias humanas, medio ambiente, salud) y las prioridades de capacitación para IA, prestando especial atención a la inclusión y a la atracción de más mujeres a los estudios de IA (para fines de 2020).
2. Las estrategias deben abordar todo el ciclo de la educación formal, la formación profesional, la educación superior y los posdoctorados. Al mismo tiempo, se debe prestar más atención al

²⁶ https://ec.europa.eu/knowledge4policy/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en

aprendizaje a lo largo de la vida, para permitir que las personas que ya están empleadas adquieran y mejoren las habilidades relacionadas con la IA.

3. Explorar cómo la IA podría incorporarse a los planes de estudio de los programas de educación secundaria y terciaria, incluida la formación profesional.

En lo referente a la Comisión, esta emitirá un informe con el apoyo de los Estados miembros al respecto a principios de 2020 y apoyará acciones modelo en regiones seleccionadas. La Comisión incluirá un componente de programas de doctorado comunes y programas post-doctorado en la convocatoria para fortalecer los centros de excelencia en investigación de IA en 2020 con un enfoque en desafíos industriales. La ambición es establecer una marca única y reconocida mundialmente para un programa europeo de doctorados en IA orientados a la industria y mantener a los investigadores en Europa después de completar sus doctorados. Las acciones Marie Skłodowska-Curie contribuirán a este objetivo.

4. Explorar vías para apoyar la inclusión de módulos de IA en programas de maestría multidisciplinares (por ejemplo, en salud electrónica, tecnología financiera, gobierno electrónico) y en programas de capacitación para adultos con enfoque en personas con educación superior y experiencia laboral.
5. Los Estados miembros y la Comisión trabajarán juntos y desarrollarán material para ser utilizado en campañas de sensibilización sobre los beneficios de la IA. A partir de 2020, la Comisión propone que la Unión proporcione un total de 700 millones de euros para apoyar las habilidades avanzadas (para IA, HPC y ciberseguridad) como parte del Programa Europa Digital a través de Másteres.
6. Proponer prácticas en el trabajo para jóvenes y profesionales que necesitan adquirir experiencia, entrenamiento y capacitaciones a corto plazo para que la fuerza laboral se familiarice con la inteligencia artificial.
7. Incorporar los principios éticos desarrollados por la Unión en los esquemas y programas de capacitación propuestos anteriormente.

5. ANÁLISIS DE COMPETENCIAS DEMANDADAS E IDENTIFICACIÓN DE PERFILES

Como se ha concluido en el ET1, se puede hablar de una incipiente presencia a todos los niveles (grado, máster y doctorado) de formación en Procesamiento del Lenguaje Natural y también en cursos de tipo MOOC en diversas plataformas. Desgraciadamente, esta formación se produce de forma muy heterogénea, y aún son muy escasos los grados, másteres, doctorados y cursos MOOC que imparten contenidos propios de las TL. Por ejemplo, la inmensa mayoría de los estudiantes de informática del país finalizan el grado careciendo de las capacidades mínimas para poder procesar correctamente texto escrito u otras formas de datos digitales no estructurados. Por no hablar de estudiantes de otras disciplinas, en las que como hemos comprobado existe una clara carencia de formación en TL.

El estudio realizado en el informe ET1 junto con las conclusiones contenidas en los distintos informes nacionales e internacionales consultados, reflejan la demanda proveniente de las empresas asociada a distintos grupos de competencias, así como la necesidad de formar profesionales multidisciplinares. Básicamente se identifican 2 grandes perfiles. Por un lado, los profesionales de las TL cuya actividad principal consistirá en desarrollar herramientas de PLN destinadas a terceros. Por otro, los profesionales de distintas áreas de conocimiento o dominios (ciencias sociales, económicas, jurídicas, física y química, biología y salud, ecología y desarrollo sostenible) que identificarán formas de aportar valor añadido en sus propias áreas a través de la inclusión y empleo de herramientas de TL desarrolladas por los primeros.

En el Anexo 1 se pueden encontrar las competencias identificadas. Estas se pueden dividir básicamente en dos grandes grupos:

- Competencias técnicas generales
- Competencias asociadas a utilidades y funcionalidades concretas

En cada uno de estos 2 grupos el número de competencias es grande y es por ello que se ha procedido a agruparlas en distintas tablas. También se incluyen competencias asociadas a otros profesionales aunque su identificación no proviene del estudio presentado en el ET1.

6. PROPUESTAS FORMATIVAS POR NIVELES

6.1 PROPUESTA FORMATIVA NO UNIVERSITARIA

i. Formación pre-universitaria

Se propone crear contenidos formativos atractivos para secundaria y bachillerato que promuevan el conocimiento en las TL. Sería interesante añadir en los cursos de tecnología y programación conocimiento de herramientas básicas de TL. Para los centros que emplean aprendizaje por proyectos diseñar proyectos y talleres atractivos (por ejemplo, de *webscraping*, descarga y análisis y tratamiento de *tweets*, o construir un *chatbot* básico). Una idea a desarrollar será impulsar competiciones de proyectos entre centros a nivel nacional.

Se propone también jornadas para profesorado de enseñanzas medias que tengan como objetivo la capacitación como formadores en TL de dichos profesores, de manera que adquieran las habilidades necesarias para impartir los citados talleres a su alumnado o guiarles en la consecución de un determinado proyecto TL. En dichas jornadas sólo se tratarán conocimientos de TL, y se presupondrá que los asistentes tienen suficientes conocimientos teóricos y prácticos de nuevas tecnologías y programación. Deberán ser completamente gratuitos y aportar al formador algún tipo de acreditación. La duración aproximada sería de 10 horas como máximo. Se facilitarán materiales de apoyo que los profesores podrán utilizar en sus clases, y se realizarán los principales ejercicios de manera tutelada. Finalizada la jornada el profesorado habrá adquirido los conocimientos necesarios para poder impartir adecuadamente talleres de TL a sus alumnos.

ii. Formación continua

Se propone como objetivo desarrollar la cultura de la importancia del dato en las organizaciones, que es uno de los objetivos de la Estrategia Digital de la Comisión Europea. Esto requiere una planificación de acciones formativas continuadas en el tiempo. Más aun cuando los avances en tecnologías se suceden de forma continua e incesante. Las características que deben poseer dichas acciones formativas son las siguientes:

- Guardar un buen equilibrio entre aportar una visión global, pero centrándose en las especificidades propias de cada negocio.
- Ser modulares. Diseñadas a varios niveles. Por ejemplo, a modo de *nanogrados* que se componen de distintos módulos de distinta complejidad. Desde los más básicos a los más avanzados.

- Revelar la importancia de los datos y la soberanía de los mismos, analizando cómo transformarlos en información. Impulsar el conocimiento del beneficio que TL pueden aportar como ventaja competitiva para el negocio en cuestión, etc.

Para ello se proponen distintos tipos de actividades formativas dirigidas a profesionales:

- Itinerarios sectoriales: Cursos formativos que muestren casos de uso exitosos de aplicación de TL en el sector o actividad solicitante de la formación.
- Talleres presenciales: Talleres específicos dirigidos a formar a empresas no proveedoras de TL en lo referente a una acción concreta (atención al cliente a través de *chatbots*, customización de la oferta, etc.). Talleres específicos dirigidos a formar a empresas proveedoras de TL en lo referente a últimas tecnologías para el desarrollo de productos concretos.
- *Nanogrados* de menos de 60 créditos: Presenciales o no, sería deseable que permitan obtener una certificación. Podrán estar dirigidos a profesionales de las TL o a empresas consumidoras de TL. Sería conveniente que estuvieran divididos en módulos de pocos créditos (por ejemplo 3 o menos). La formación cubrirá desde módulos de formación básica (por ejemplo, utilidades de las TL, aplicaciones de técnicas básicas de aprendizaje basadas en *Machine Learning* para soluciones PLN. Herramientas ML: TensorFlow, PyTorch, scikit-learn... recursos hardware y software necesarios, instalación de los mismos) hasta módulos más avanzados (escalabilidad, diseños de arquitecturas hardware y software, clusters, Apache Spark, etc).

iii. Formación a la Administración Pública

Como ya se ha mencionado, hoy en día uno de los principales activos de la sociedad y del tejido productivo son los datos recolectados y transformados en información. Es importante que existan políticas públicas que se encarguen de su correcta gestión y explotación, así como de generar sinergias y potenciar el uso de esta para asegurar y expandir la ventaja competitiva que se haya conseguido, por ejemplo, a través de planes como el Industria 4.0. Para poder lograrlo, uno de los principales retos es la formación a distintos niveles de los profesionales de las administraciones públicas. Se propone al INAP la organización de cursos sobre TL con un enfoque adaptado a las necesidades de estos profesionales. Mandos de nivel superior que deben conocer las herramientas de TL y posibles aplicaciones de estas, riesgos derivados de un uso no reglado, para planificar las

futuras estrategias nacionales. Mandos de nivel intermedio que deben implementar dichas estrategias, generando el ecosistema necesario. Para ello, también estos deberán conocer las posibilidades que brindan las herramientas de TL, casos de uso exitosos, etc. Y por último, cualquier profesional de la administración pública que, como usuario de dichas tecnologías, deberá conocerlas y explotarlas. Para todos ellos se desplegarán distintas convocatorias y como en otras iniciativas se sugiere que se lleven a cabo a través de las plataformas de formación del INAP. Los cursos podrán tener asociados certificados de superación expedidos por dichas plataformas de formación.

6.2 PROPUESTA FORMATIVA UNIVERSITARIA

La presente propuesta se centra en los estudios universitarios y quiere potenciar una mejora de su oferta con el objetivo de formar profesionales de las Tecnologías del Lenguaje que puedan cubrir las necesidades del mercado laboral en los próximos diez años. Constatado el interés que tiene (y tendrán) las aplicaciones de las tecnologías del lenguaje, se tiene que velar por diferentes acciones que corresponden a la estrategia de implantación de los contenidos necesarios para cada uno de los perfiles universitarios.

A grandes rasgos los contenidos específicos sobre Tecnologías del Lenguaje (TL) a incluir en los grados universitarios pueden clasificarse en:

1. Lingüística para las TL:
 - a. Comunicación, fonología, morfología, sintaxis, semántica, pragmática, etc.
2. Técnicas para las TL:
 - a. Basadas en conocimiento, en corpus, estadísticas, aprendizaje automático, *deep learning*, etc.
3. Aplicaciones funcionales de las TL:
 - a. Extracción de Información, Recuperación de Información, Traducción Automática, Síntesis de Voz, Reconocimiento del lenguaje hablado, Sistemas de Diálogo, *chatbots*, etc.
4. Campos de aplicación de las TL:
 - a. Medicina, legal, sociología, periodismo, educación, etc.

Así, cada ámbito académico (expertos, grado, máster, etc.) deberá incorporar en su currículum los contenidos específicos sobre Tecnologías del Lenguaje apropiados a su perfil. Por ejemplo, sería conveniente que las carreras relacionadas con la salud (medicina, farmacia, etc.) conocieran las distintas ontologías médicas como UMLS, SNOMED-CT, y sistemas de codificación como ICD-10 que se aplican en el ámbito sanitario, y las capacidades actuales que ofrecen las aplicaciones funcionales de diagnóstico, etc. de las Tecnologías Lingüísticas y de la Inteligencia Artificial en general.

Teniendo en cuenta estos factores, esta es la propuesta que presentamos para el ámbito universitario:

1. Propuesta de título específico de grado en Lingüística Computacional, a imagen de los que ya existen a nivel europeo.
2. Propuesta de título específico de grado en Inteligencia Artificial, a imagen de los que ya existen a nivel europeo, con intensificaciones en Tecnologías del Lenguaje.
3. Inclusión de asignaturas obligatorias en Tecnologías del Lenguaje y en general de procesamiento de datos no estructurados en el currículum académico de los grados actuales en Ingeniería en Informática (y títulos relacionados Ciencia de Datos, etc.) e Ingeniería de Telecomunicaciones, así como el resto de titulaciones STEM.
4. Inclusión de asignaturas obligatorias en Tecnologías del Lenguaje en los grados actuales en Lingüística (y títulos relacionados: Lenguas Aplicadas, Filología y Lingüística, Traducción e Interpretación, etc.)
5. Inclusión de contenidos en Tecnologías del Lenguaje en grados actuales relacionados con los campos de aplicación de la TL como medicina, sociología, periodismo, derecho, administración y dirección de empresas, etc.
6. Propuesta de títulos de máster específicos de TL repartidos geográficamente de forma equilibrada, para visibilizar su importancia y dar cabida a la demanda prevista. Apoyo efectivo con becas específicas.
7. Apoyo efectivo a la creación de programas de doctorado (incluidos programas de doctorado industrial) y títulos de postgrado (no-oficiales) con becas especializadas (también de fondos de formación destinados a la formación continua en materia de TIC).
8. Cursos de expertos para formación continua.
9. MOOCs (*Massive Online Open courses*).

iv. Grados

Seguidamente, realizaremos las distintas propuestas para las titulaciones de grado, comenzando por las dos propuestas de nuevos grados Lingüística Computacional e Inteligencia Artificial con línea

curricular en Tecnologías del Lenguaje, para continuar con la propuesta en los grados tipo STEM y finalizar por las de humanidades, ciencias sociales y jurídicas, y ciencias de la salud.

Lingüística computacional

Se propone como titulación un grado de Lingüística Computacional dado que el perfil abarca competencias de dos ramas tradicionalmente separadas, de forma que los contenidos técnicos y lingüísticos necesarios no se pueden obtener de forma simultánea. Como ha quedado patente en el ET1, la formación universitaria en lingüística apenas ofrece contenidos de matemática o programación. Las competencias que necesitan adquirir los estudiantes de lingüística no pueden adquirirse en un máster de un año o dos. Esta no es una característica únicamente del panorama universitario en España, en otros países se encontraron en la misma situación, pero hace años que ya ofrecen grados en procesamiento del lenguaje natural como se ha visto en el ET1 (y en el Anexo 2 del presente documento).

Es un grado enfocado a crear una formación profesionalizadora para los estudiantes de lingüística y lenguas aplicadas que suelen tener pocas salidas profesionales en las que ejerzan como lingüistas, ya que los estudiantes de ingeniería informática (y afines) disponen de muchas salidas profesionales y no suelen especializarse en el grado. También puede resultar interesante para los estudiantes que ahora escogen carreras científicas en la rama de ciencias sociales como psicología y que no desean seguir la vertiente clínica. La propuesta de este grado quiere hacer la misma apuesta que han hecho otras disciplinas como Biotecnología, Tecnología de los alimentos, o Diseño de Videojuegos, que se imparte como un grado a cursar de forma simultánea con Ingeniería informática (por ejemplo, en la Universidad Juan Carlos I).

Así proponemos un grado de 180 créditos ECTS como los grados europeos que hemos mencionado en el ET1 que pueda entrar en un esquema de doble grado o estudios simultáneos con alguna filología, traducción, lenguas aplicadas o alguna ingeniería. De esta forma los estudiantes que quieran proseguir con sus estudios pueden optar a un máster de especialización más técnica con el contenido básico requerido.



Descripción general del grado

DATOS BÁSICOS	<p>Nivel: Grado</p> <p>Denominación corta: Lingüística Computacional y Tecnologías del Lenguaje</p> <p>Denominación específica: Grado en Lingüística Computacional y Tecnologías del Lenguaje</p> <p>Rama: Artes y Humanidades</p> <p>Habilita para profesión regulada: No</p> <p>Profesión regulada: No</p> <p>Condición de acceso para título profesional: No</p> <p>Título profesional: No</p> <p>Especialidades: No</p>
DISTRIBUCIÓN DE CRÉDITOS	<p>Créditos formación básica: 48</p> <p>Créditos obligatorios: 84</p> <p>Créditos optativos: 33</p> <p>Créditos prácticas externas: no se contemplan</p> <p>Créditos trabajo de fin de grado o máster: 15</p> <p>Créditos ECTS (total): 180</p>

Programa Marco

En este documento no se propone un plan de estudios completo, sino que se realiza un inventario de los contenidos específicos que debería incluir un grado que responda a las necesidades que hemos analizado. Presentaremos asignaturas posibles agrupadas por módulos que estructuran los estudios según las contribuciones de las diferentes ramas que se integran en estos estudios de marcado carácter transversal a las humanidades y los estudios de informática e inteligencia artificial.



Nombre del módulo: Fundamentos	
Asignaturas que la conforman	<p>Matemática discreta</p> <p>Introducción a la programación</p> <p>Estructuras de datos y algoritmos</p> <p>Estadística y Probabilidad</p> <p>Lingüística</p> <p>Industrias de la lengua</p>
Contenido	
<p>Esta materia contiene la formación y conocimientos de matemática, programación y lingüística que han de permitir al estudiante manejarse bien en las materias específicas que aparecen en el grado. Para ello, una vez dispone de los conocimientos básicos de álgebra, cálculo, lógica, conjuntos, grafos y teoría de la información, por un lado y de estadística y probabilidad por otro se puede avanzar en contenidos directamente relacionados con la lingüística y los conceptos de esta disciplina que se utilizan en el procesamiento del lenguaje natural, en particular sobre niveles de análisis y problemas de la comprensión del lenguaje natural.</p>	
Observaciones	
<p>Es importante que las diferentes asignaturas técnicas utilicen ejemplos de procesamiento del lenguaje natural.</p>	
Competencias asociadas	<p>STEM01: Capacidad para la resolución de los problemas matemáticos básicos. Aptitud para aplicar los conocimientos sobre: álgebra lineal, cálculo diferencial e integral, métodos numéricos, algorítmica numérica, estadística y optimización</p> <p>STEM02: Capacidad para comprender los conceptos básicos de matemática discreta, lógica y algorítmica y su aplicaciones para la resolución de problemas propios de la ingeniería</p> <p>PRG01: Programación en lenguajes enfocados al desarrollo PLN</p> <p>PRG02: Ser competente programando en lenguajes enfocados al análisis estadístico</p> <p>PRG03: Conocer y saber emplear expresiones regulares y construcción de patrones de búsqueda</p>



	<p>PRG06: Conocer y saber emplear plataformas de GV</p> <p>PRG07: Ser competente programando en orientación a objetos</p> <p>PRG08: Saber programar APIs</p> <p>PRG10: Disponer de conocimientos multientorno y capacidad de adaptación a cambios</p> <p>PRG11: Conocer y saber realizar desarrollos en plataformas cloud computing</p> <p>SOP01: Linux</p>
Resultados de aprendizaje	<p>RA1: Aplica conocimientos de álgebra lineal para resolver problemas que puedan plantearse en el procesamiento del lenguaje natural</p> <p>RA2: Demuestra conocimientos de cálculo diferencial e integral para entender problemas que puedan plantearse en el modelado de información lingüística</p> <p>RA3: Emplea conocimientos de métodos numéricos para resolver problemas que puedan plantearse en el modelado de información lingüística</p> <p>RA4: Emplea conocimientos de estadística para resolver problemas que puedan plantearse en el modelado de información lingüística</p> <p>RA5: Aplica conocimientos de matemática discreta para resolver problemas que puedan plantearse en el modelado de información lingüística</p> <p>RA6: Aplica conocimientos de programación para realizar análisis de datos</p> <p>RA7: Aplica conocimientos de programación para realizar preproceso de datos</p>

Nombre del módulo: Tecnología	
Asignaturas que la conforman	<p>Inteligencia Artificial</p> <p>Aprendizaje automático y redes neuronales</p> <p>Recuperación de la información</p>



Contenido

En este módulo se introducen las redes complejas y se desarrollan los conocimientos que permiten trabajar con datos masivos, heterogéneos y distribuidos, formalizar los problemas que se plantean, conocer los métodos y las aproximaciones que aporta la estadística computacional para resolver y visualizar estos problemas. Por último, se muestra cómo aplicar estos contenidos en los sistemas de recuperación y predicción más exitosos.

Por un lado, existe una línea de estadística computacional y aprendizaje automático que incluye las asignaturas de aprendizaje automático, minería de datos masivos y modelos gráficos probabilísticos, donde se introducen los paradigmas de aprendizaje automático (supervisado, no supervisado y aprendizaje por refuerzo). En ellos se profundiza en las técnicas de inferencia Bayesiana usando grandes volúmenes de datos, métodos de reducción de dimensionalidad y métodos de optimización no lineal, incluyendo enfoques variacionales y estocásticos.

Por otro lado, las asignaturas de Recuperación de la información y análisis de la web y Procesamiento del sonido avanzan en los contenidos de introducción a la programación paralela distribuida y de bases de datos. Se introduce el diseño y arquitectura de bases de datos distribuidas y el procesamiento de consultas y búsquedas para datos heterogéneos y distribuidos.

Observaciones

Competencias asociadas

PLN03: Conocer y saber emplear librerías y sistemas abiertos para PLN

PLN04: Conocer y saber emplear técnicas de aprendizaje basadas en *Machine Learning* para soluciones PLN

PLN05: Conocer y saber emplear técnicas de aprendizaje profundo basadas en *Deep Learning* para soluciones PLN. Herramientas DL

PLN013: Ser competente en la aplicación de métodos estadísticos al desarrollo PLN. Formalización matemática de las relaciones entre los datos textuales

PLN027: Ser competente en la aplicación de técnicas basadas en reglas para soluciones PLN

PLN029: Saber evaluar y validar herramientas y aplicaciones para tareas de PLN en textos



	<p>PLN030: Saber evaluar y validar herramientas y aplicaciones para tareas de PLN en sistemas conversacionales</p> <p>PLN032: Grafos de conocimiento</p> <p>PLN14: Conocimiento de Sistemas de diálogo colaborativo. Estrategias de diálogo colaborativo</p> <p>PLN20: Conocimiento de las Tecnologías de <i>chatbot</i>. Sistemas y motores cognitivos</p> <p>PLN029: Evaluar y validar herramientas y aplicaciones para tareas del procesamiento del lenguaje natural en textos</p> <p>PLN030: Evaluar y validar herramientas y aplicaciones para tareas del procesamiento del lenguaje natural en sistemas conversacionales</p>
<p>Resultados de aprendizaje</p>	<p>RA1: Conoce los aspectos estadísticos fundamentales de la ciencia de las redes</p> <p>RA2: Reconoce las técnicas estadísticas aplicadas a la minería de datos</p> <p>RA3: Modela problemas reales de clasificación mediante paradigmas computacionales</p> <p>RA4: Reconoce algoritmos de optimización comúnmente usados en tareas de ingeniería</p> <p>RA5: Reconoce las características, estructuras y técnicas asociadas a los sistemas distribuidos</p> <p>RA6: Aplica modelos gráficos para la realización de cálculos probabilísticos</p> <p>RA7: Reconoce y aplica técnicas de minería de datos</p> <p>RA8: Utiliza paquetes software existentes para realizar cálculo multivariable y álgebra lineal aplicado a técnicas de aprendizaje automático</p> <p>RA9: Utiliza modelos y métodos visuales para realizar análisis de datos</p> <p>RA10: Reconoce y utiliza infraestructuras para la visualización de datos</p> <p>RA11: Crea elementos básicos de visualización, como gráficos o animaciones interactivas</p> <p>RA12: Reconoce y utiliza técnicas de recuperación de la información, incluyendo lenguajes de petición (<i>query</i>), indexado y búsqueda</p> <p>RA13: Aplica técnicas de análisis de la web, incluyendo textos estructurados e información multimedia y multilingüe</p>



RA14: Elabora proyectos que tengan por objeto, de acuerdo con los conocimientos adquiridos, la recogida o la explotación de datos textuales y aplicaciones del procesamiento del lenguaje natural

Nombre del módulo: Lenguaje, idiomas y generación de recursos

Asignaturas que conforman la materia

La estructura de las lenguas
 El significado de las lenguas
 Métodos empíricos en la investigación lingüística
 Inglés: comprensión y uso
 Textos especializados y terminología
 Generación de recursos

Contenido

“Estructura de las Lenguas” y “El Significado en las Lenguas” introducen respectivamente el estudio de la estructura fonológica, morfológica y sintáctica de las lenguas, y los procesos semánticos y pragmáticos de construcción del significado. Por su parte, “Métodos Empíricos para el Estudio del Lenguaje” introduce al estudiante en el uso de los métodos de investigación propios de las ciencias humanas y sociales, para los ámbitos de aplicación de las lenguas en contextos diversos. Debería abarcar los siguientes temas:

La naturaleza del lenguaje. El lenguaje como fenómeno dinámico. El lenguaje como hecho social y como hecho psicológico. Diversidad tipológica; similitudes y diferencias entre las lenguas. Lenguas de distinta modalidad: oral y visual. El estudio de los aspectos estructurales del lenguaje en sus distintos componentes o niveles de análisis: fonética y fonología, morfología, sintaxis, semántica y pragmática. El significado en las lenguas. Significado convencional e inferencia. Significado léxico y composicionalidad. La dinamicidad del significado. Semántica y pragmática. La investigación mediante corpus lingüísticos.

“Inglés” y “Textos especializados” se enfocan con dos objetivos, dar a conocer las características singulares del lenguaje especializado y familiarizar al alumno con la redacción y comprensión de textos de alta especialización en inglés.



<p>Competencias asociadas</p>	<p>LIN01: Ser capaces de formalizar el lenguaje natural</p> <p>LIN02: Ser capaces de describir, analizar e interpretar datos lingüísticos de todos los niveles gramaticales de cualquier lengua mediante la aplicación de diversas teorías y métodos formales para descomponer estructuras lingüísticas en sus elementos constituyentes e identificar patrones estructurales</p> <p>LIN03: Ser capaces de identificar patrones léxicos/morfológicos/sintácticos para modelado estadístico</p> <p>LIN04: Poseer conocimientos de IPA (<i>International Phonetic Alphabet</i>) y SAMPA (<i>Speech Assessment Methods Phonetic Alphabet</i>)</p> <p>LIN05: Ser capaces de leer espectrogramas</p> <p>REC01: Crear recursos lingüísticos: diccionarios, léxicos, corpus.</p> <p>REC02: Anotación de corpus con información lingüística utilizando las herramientas adecuadas</p> <p>REC03: Capacidad de elaboración de recursos para asistentes conversacionales</p> <p>REC03A: Diseño de estrategias de diálogo y diálogos en función de la personalidad definida para el agente conversacional</p> <p>REC03B: Creación de diálogos</p> <p>REC03C: Conocimiento de las técnicas de creación de diálogos</p>
<p>Resultados de aprendizaje</p>	<p>RA1. Demuestra la comprensión de una variedad de construcciones y discursos lingüísticos</p> <p>RA2. Maneja funcionalmente las estructuras y estrategias lingüísticas pertinentes para la producción oral y escrita en inglés</p> <p>RA3. Describe lingüísticamente elementos estructurales y de significado de los textos</p> <p>RA4. Conoce las diferencias entre el lenguaje natural y otros sistemas semióticos</p> <p>RA5. Conoce las nociones analíticas necesarias para el análisis lingüístico en sus diferentes dimensiones</p> <p>RA6. Describe estructuras complejas en las diferentes dimensiones del lenguaje (fonológica, sintáctica o semántica) usando los conceptos analíticos apropiados</p>



	<p>RA7. Aplica principios teóricos para explicar fenómenos del lenguaje natural en sus diferentes dimensiones</p> <p>RA8. Explica diferencias entre las lenguas en los distintos niveles de análisis del lenguaje natural</p> <p>RA9. Aplica con rigor los distintos métodos de investigación empírica para el estudio de cuestiones lingüísticas variadas</p>
--	---

Nombre del módulo: Procesamiento del lenguaje natural

Asignaturas que la conforman

Introducción al procesamiento del lenguaje natural
 Procesamiento automático de textos
 Marcaje de textos
 Procesamiento del habla
 Traducción automática

Contenido

En esta materia se plantean los temas relacionados con la industria del Procesamiento del Lenguaje Natural y la Inteligencia Artificial. En la introducción se presenta la clasificación de documentos como ejemplo que permita conocer todos los pasos de una aplicación industrial: Localización, preparación y normalización de textos para su procesamiento. Usabilidad de textos: restricciones legales. Además, los estudiantes se introducen en las técnicas estándar del procesamiento: Manipulación del texto. Representación vectorial de la información lingüística (bolsa de palabras y modelos distribucionales). Distancias como similitud lingüística de palabras y de frases. Aplicaciones en semántica léxica. Clasificación y aprendizaje automático para el análisis de sentimiento. Métodos de evaluación de clasificación de textos. Conjuntos de datos de prueba, *gold-standards*, equilibrio, selección y estadísticos relacionados. Validación cruzada y comparación con referencias. Análisis de errores (tabla de contingencia y análisis cualitativo). Finalmente se introduce la creación de modelos vectoriales con Python y word2vec y el cálculo de distancias.

En Procesamiento automático de textos se profundiza en la programación en Python para el procesamiento automático de textos: variables, manipulación de datos, funciones, estructuras de control, estructuras de datos (caracteres, cadenas, tuplas, listas) y la representación de textos. Segmentación del texto en frases y palabras, la búsqueda de patrones con expresiones regulares con Python, sobre texto plano y texto etiquetado, la manipulación de texto plano con ejemplos de corrección



ortográfica con distancia de edición y evaluación del sistema creado y análisis de errores. También se trabaja la manipulación de texto etiquetado morfosintácticamente y la aplicación a e-learning / iCALL (software de aprendizaje de idiomas) con búsqueda de patrones y diagnóstico de errores.

En Marcaje de textos se profundiza en la programación con Python de estructuras de datos complejas (diccionarios para el cálculo de frecuencias) y recursividad. La anotación lingüística automática: métodos de anotación por reglas y de aprendizaje automático. Evaluación mediante precisión, cobertura y F1. Análisis de errores (tabla de contingencia, análisis cualitativo). Finalmente se estudian tareas concretas: Etiquetado morfosintáctico (POS *tagging*), Etiquetado sintáctico (dependencias) y Etiquetado de entidades con nombre propio (NER). Métodos de anotación: por reglas y estadísticos. Evaluación mediante precisión, cobertura y F1. Se utilizarán para la formación los diferentes ámbitos como BioNLP y textos clínicos, Análisis de Sentimientos en marketing, etc.

En Procesamiento del habla se estudian: Conceptos básicos del procesamiento de la voz. La señal. Modelo acústico. Métodos básicos de representación de la señal de voz. Identificación de los sonidos del habla. Correlatos acústicos. Procesamiento digital del habla. Señales analógica y digital y su conversión. Conversión texto-habla. Aplicaciones. Tecnología de síntesis. Desarrollo y evaluación. Reconocimiento de voz. Aplicaciones. Tecnología de reconocimiento. Desarrollo y evaluación. Los sistemas de diálogo. Aplicaciones. Tecnologías, análisis semántico, gestión del diálogo y generación de texto. Desarrollo y evaluación.

Finalmente, la asignatura de Traducción automática contempla: a) Tipo de métodos y sistemas de TA. Los recursos lingüísticos para la TA: memorias de traducción y corpus paralelo, diccionarios y léxico adicional. b) Sistema de TA basado en transferencia sintáctica superficial (Apertium). Modificación de diccionarios. c) Entrenamiento y traducción con un sistema de traducción automático estadístico (Moses) d) Entrenamiento y traducción con un sistema de traducción automática neuronal (ModernMT). La postedición de TA. Evaluación de la calidad de la TA y de los textos posteditados: tipología de los errores más comunes, métricas automáticas, criterios para la evaluación manual. El mercado de la TA y la postedición.

Competencias asociadas

PLN14: Conocimiento de Sistemas de diálogo colaborativo. Estrategias de diálogo colaborativo

PLN20: Conocimiento de las Tecnologías de chatbot. Sistemas y motores cognitivos (Google DialogFlow, IBM



	<p>Watson, Amazon Alexa, Facebook messenger, Whatsapp Business), y asistentes (Google Home ...)</p> <p>PLN01: Conocimientos generales de PLN. Conceptos, fases de análisis, técnicas principales, sistemas de PLN</p> <p>PLN02: Conocimiento sobre datos textuales (lingüísticos, estructurados, digitales, textuales, multimedia). Análisis e interpretación de datos lingüísticos en todos los niveles gramaticales mediante la aplicación de teorías y métodos formales, descomponer las estructuras en sus constituyentes e identificar patrones estructurales</p> <p>PLN31: Conocimientos en adquisición, análisis, transformación y utilización de datos no estructurados y multilingües</p> <p>PLN17: Conocimiento de la Ingeniería ontológica: Modelado conceptual ontológico. Ontologías y mecanismos de inferencia. Metodologías, lenguajes y herramientas para la representación conceptual en dominios: OWL, Protégé, ...</p> <p>PLN26: Conocimiento sobre Procesos de comprensión del lenguaje (NLU). Inferencia y extracción de conocimiento</p> <p>PLN18: Conocimiento de Sistemas de clasificación textual y <i>clustering</i></p> <p>PLN21: Conocimiento sobre Modelado de <i>topics</i>: reconocimiento del tema</p> <p>PLN19: Conocimiento de Sistemas de recomendación, técnicas y motores (ThinkAnalytics, BoxFish, Kannuu, Jinni, DigitalSmiths, Watchwith, etc.)</p> <p>PLN22: Conocimiento de Sistemas de generación del Lenguaje Humano. Arquitectura para NLG. Macro y micro planificación y realización lingüística</p> <p>PLN23: Adquisición y generación de estructuras de conocimiento para NLG</p> <p>PLN24: Conocimiento de Sistemas de traducción automática</p> <p>PLN25: Conocimiento de Sistemas de búsqueda de respuestas. <i>Question-Answering</i></p> <p>PLN9: Construcción de recursos del lenguaje: diccionarios, lexicones, corpus, <i>datasets</i>,...)</p> <p>PLN08: Ser capaces de construir herramientas para el Análisis morfológico automático: <i>taggers</i>, lematización, <i>stemming</i></p> <p>PLN28: Ser capaces de construir herramientas para el Análisis sintáctico automático: <i>chunkers</i>, análisis de dependencias</p>
--	---



	<p>PLN10: Ser capaces de construir herramientas para el Análisis semántico: Representación y composición del conocimiento, etiquetado de roles semánticos. Técnicas de similitud semántica y modelos de representación del conocimiento (<i>word representation models: word embeddings, word2vec...</i>)</p> <p>PLN011: Ser capaces de construir herramientas para el Análisis discursivo y de la pragmática: análisis del contexto, marcas contextuales, etc.</p> <p>PLN12: Saber construir sistemas de análisis de sentimientos: minería de opiniones y reconocimiento de emociones</p> <p>PLN15: Saber diseñar y desarrollar procesos de extracción de información</p> <p>PLN16: Saber diseñar y desarrollar procesos de recuperación de información</p>
<p>Resultados de aprendizaje</p>	<p>RA1: Diseña e implementa cadenas de procesamiento de textos usando plataformas de desarrollo</p> <p>RA2: Diseña y lleva a cabo un proyecto de servicio lingüístico con análisis automático de textos, y uso de herramientas y métodos informáticos, con especial énfasis en el control de calidad</p> <p>RA3: Implementa un sistema de etiquetado automático y lo evalúa con métodos formales</p> <p>RA4: Aplica algoritmos y medidas de distancia con representaciones vectoriales de textos para diferentes tareas de clasificación</p> <p>RA5: Escribe expresiones regulares eficientes para la extracción de datos del corpus y la preparación de vectores</p> <p>RA6: Construye un clasificador supervisado para análisis de sentimiento y lo evalúa con métricas estándar</p> <p>RA7: Lleva a cabo evaluaciones formales de aplicaciones relacionadas con el habla y la comunicación persona-máquina (basadas en un marco de evaluación y con criterios objetivos y cuantificables)</p> <p>RA8: Implementa cadenas de procesamiento para la confección de un asistente (sistema de diálogo)</p> <p>RA9: Pone en marcha un sistema de TA utilizando programas de acceso abierto para un dominio concreto entre dos lenguas de ámbito europeo.</p>

Inteligencia Artificial en las Tecnologías del Lenguaje



Se propone como titulación un grado de Inteligencia Artificial en las Tecnologías del Lenguaje que se debería adscribir a la rama de conocimiento de Ingeniería y Arquitectura y no habilitaría para el ejercicio de profesiones reguladas.

El grado en Ingeniería en Inteligencia Artificial que aquí se propone está enfocado a obtener un perfil técnico especializado en Procesamiento del Lenguaje Natural/Tecnologías del lenguaje (Grado IA-PLN) formaría profesionales que necesitan responder a las demandas crecientes en el análisis de información digital que requieren de técnicas de IA de vanguardia. En él se deberían proporcionar los fundamentos y las habilidades avanzadas en los principios y tecnologías que subyacen a la IA, incluyendo la lógica, la representación del conocimiento, los modelos probabilísticos y el aprendizaje automático. Dentro de las diferentes áreas de la inteligencia artificial, los estudiantes profundizarán en procesamiento de lenguaje natural, una de las principales áreas y actualmente con un alto impacto tanto en la industria como en la administración. Así, el grado incluiría la hibridación formativa a la que se refiere la Estrategia Española de I+D+i en Inteligencia Artificial²⁷ incluyendo en este caso materias de ciencias, tecnologías y humanidades.

Para que el plan de estudios de un Grado IA-PLN esté de acuerdo a las recomendaciones que establece la Conferencia de Directores y Decanos de Ingeniería Informática (CODDII) para una titulación oficial de ingeniería informática, deberá incluir, como mínimo, los siguientes módulos:

DATOS BÁSICOS	<p>Nivel: Grado en Inteligencia Artificial en las Tecnologías del Lenguaje</p> <p>Denominación corta: Grado en IA en las Tecnologías del Lenguaje</p> <p>Denominación específica: Grado en Inteligencia Artificial en las Tecnologías del Lenguaje</p> <p>Rama: Ingeniería y Arquitectura</p> <p>Habilita para profesión regulada: No</p>
----------------------	--



	Profesión regulada: No Condición de acceso para título profesional: No Título profesional: No Especialidades: No
DISTRIBUCIÓN DE CRÉDITOS	Créditos formación básica: 60 Créditos obligatorios: 108 <ul style="list-style-type: none"> • <i>Común a la rama de Informática (60 créditos).</i> • <i>Tecnología específica (48 créditos).</i> Créditos prácticas externas: no se contemplan Créditos trabajo de fin de grado: 12 mínimo Créditos ECTS (total): 180

Programa Marco

Como se ha comentado previamente, no se propondrá un plan de estudios completo. Se realiza un inventario de los contenidos específicos que debería incluir un grado que responda a las necesidades que hemos analizado. Presentaremos asignaturas posibles agrupadas por materias.

Nombre del módulo: Formación Básica	
Asignaturas que la conforman	Álgebra Matemática discreta Cálculo Probabilidades y Estadística Fundamentos de programación Aspectos Avanzados de la Programación Ingeniería de Software Bases de Datos Estructuras de datos y algoritmos Fundamentos de sistemas digitales



Contenido

Esta materia contiene la formación y conocimientos fundamentales en cualquier STEM que comprenden matemática, programación, algoritmos y estructuras de datos. Y algunas más específicas de desarrolladores e ingenieros de software.

Observaciones

Competencias asociadas

STEM01: Capacidad para la resolución de los problemas matemáticos básicos. Aptitud para aplicar los conocimientos sobre: álgebra lineal, cálculo diferencial e integral, métodos numéricos, algorítmica numérica, estadística y optimización

STEM02: Capacidad para comprender los conceptos básicos de matemática discreta, lógica y algorítmica y su aplicaciones para la resolución de problemas propios de la ingeniería

PRG01: Programación en lenguajes enfocados al desarrollo PLN

PRG02: Ser competente programando en lenguajes enfocados al análisis estadístico

PRG03: Conocer y saber emplear expresiones regulares y construcción de patrones de búsqueda

PRG04: Ser capaz de desarrollar aplicaciones en plataformas

PRG05: Ser capaz de desarrollar pruebas en plataformas de pruebas de servicios API REST

PRG06: Conocer y saber emplear plataformas de GV

PRG07: Ser competente programando en OO

PRG08: Saber programar APIs

PRG09: Conocer y saber emplear plataformas para desarrollo web

PRG010: Disponer de conocimientos multientorno y capacidad de adaptación a cambios

PRG011: Conocer y saber realizar desarrollos en plataformas Cloud Computing

BDA01: Ser competente en el acceso a Bases de Datos (BD) para la gestión de flujos de trabajo: *data-pipelines* y *data-stores*

BDA02: Conocer sistemas relacionales de BD

BDA03: Conocer bases de datos NOSQL



	<p>ISW01: Conocer y aplicar metodologías y la correcta perspectiva ingenieril del desarrollo del software</p> <p>ISW02: Saber diseñar y desarrollar productos de software siguiendo los principios de la ingeniería del Software</p> <p>ISW03: Saber diseñar y desarrollar productos de software aplicando metodologías ágiles</p> <p>SOP01: Linux</p>
Resultados de aprendizaje	<p>RA1: Aplica conocimientos de álgebra lineal para resolver problemas de ingeniería</p> <p>RA2: Demuestra conocimientos de cálculo diferencial e integral para resolver problemas de ingeniería</p> <p>RA3: Emplea conocimientos de métodos numéricos para resolver problemas de ingeniería (o de modelado lingüístico)</p> <p>RA4: Emplea conocimientos de estadística para resolver problemas de ingeniería</p> <p>RA5: Aplica conocimientos de matemática discreta para resolver problemas de ingeniería</p> <p>RA6: Aplica conocimientos de programación de ingeniería</p> <p>RA7: Aplica conocimientos de programación para realizar preproceso de datos</p> <p>RA8: Demuestra saber programar y probar aplicaciones empleando el mejor algoritmo, paradigma de programación e identifica y emplea las estructuras de datos adecuadas para resolver los problemas de ingeniería de una forma eficiente, robusta y reutilizable</p> <p>RA9: Demuestra saber modificar y mantener programas previamente creados</p>

Nombre del módulo: Inteligencia Artificial	
Asignaturas la que conforman	<p>Fundamentos de Inteligencia Artificial</p> <p>Aprendizaje Automático</p> <p>Ampliación Inteligencia Artificial</p>
Contenido	
A continuación, se presentan los contenidos que configurarían el núcleo de la	



especialización de un grado de ingeniería informática en la materia de Inteligencia Artificial con intensificación en procesamiento de lenguaje natural. Aspectos conceptuales de la inteligencia artificial y la ingeniería del conocimiento. Resolución de problemas mediante técnicas de búsqueda

Conocimiento y razonamiento. Conocimiento incierto y razonamiento. Introducción al aprendizaje automático. Aprendizaje supervisado. Métodos lineales por regresión, Métodos lineales por clasificación, Métodos jerárquicos, Redes neuronales feed-forward superficiales, Redes neuronales recurrentes superficiales, Métodos basados en *kernels*, Métodos de conjuntos, Aprendizaje no supervisado: *Clustering*, Aprendizaje por refuerzo, Redes neuronales y aprendizaje profundo, Mejoras de las redes neuronales (ajuste de hiperparámetros, regularización y optimización), Redes neuronales convolucionales, Agentes inteligentes, Sistemas multiagente, Computación evolutiva, Conjuntos borrosos

Observaciones

Competencias asociadas

IA01: Conocer técnicas y conceptos básicos de IA: Búsquedas en espacios de estados, representación del conocimiento, agentes, etc.

IA02: Conocer las técnicas y algoritmos de IA apropiados para resolver problemas de aprendizaje supervisado y no supervisado. Algoritmos genéticos, *Deep Learning*, *embeddings*, KNN etc.

IA03: Conocer, ser capaces de analizar, entender la formalización y saber aplicar las técnicas apropiadas y modificar aplicaciones de IA en problemas reales

PLN034: Saber programar herramientas de PLN aplicando técnicas de aprendizaje profundo basadas en *Deep Learning* para soluciones PLN. Saber programar en TensorFlow, PyTorch, Keras, etc.

Resultados de aprendizaje

RA10: Demuestra haber adquirido una perspectiva global de qué es la inteligencia artificial, su historia, áreas de aplicación y métodos genéricos aplicables

RA11: Conoce las técnicas posibles y sabe distinguir las más adecuadas para la adquisición, gestión, representación y ejecución del conocimiento (cierto o difuso) junto con la generación rápida de prototipos y la aplicación de métodos de ingeniería para su integración en productos

RA12: Comprende la necesidad de integrar conocimiento existente y conocimiento aprendido automáticamente a partir de datos



	<p>RA13: Conoce los paradigmas más importantes de aprendizaje</p> <p>RA14: Demuestra ser capaz de identificar qué tipo de tareas están implicadas en un problema de aprendizaje y saber elegir la técnica de aprendizaje y algoritmo más adecuados en función de las características de cada tarea</p> <p>RA15: Demuestra ser capaz de evaluar e informar sobre la calidad de lo aprendido por un sistema de aprendizaje</p> <p>RA16: Conoce y maneja adecuadamente plataformas software de uso generalizado que implementan algoritmos de aprendizaje automático</p> <p>RA17: Posee la capacidad de diseñar e implementar un sistema basado en agentes sobre herramientas de uso generalizado</p> <p>RA18: Demuestra ser capaz de evaluar e informar sobre sistemas basados en agentes de acuerdo a sus características</p> <p>RA19: Conoce el paradigma evolutivo y los conjuntos borrosos</p> <p>RA20: Aplica y evalúa algoritmos avanzados para la resolución de problemas de búsqueda</p>
--	--

Nombre del módulo: Materia de Procesamiento del Lenguaje Natural / Tecnologías del lenguaje

Asignaturas que conforman la materia

Introducción al Procesamiento del lenguaje natural
 Representaciones y modelos
 Métodos, técnicas y aplicaciones de PLN

Contenido

Introducción al PLN: historia, aplicaciones, enfoques. Niveles lingüísticos para PLN.

Análisis morfológico-léxico y análisis morfosintáctico. Análisis sintáctico. Gramáticas independientes del contexto. Análisis de dependencias. Semántica. Semántica léxica. Bases de conocimiento léxico-semánticas. Reconocimiento de entidades nombradas, desambiguación del sentido de las palabras. Semántica distribucional. Pragmática. Interpretación contextual y relativa al mundo. Generación de lenguaje; Etapas y aplicaciones. Herramientas y recursos lingüísticos disponibles. Modelos de lenguaje. Modelos del espacio vectorial y funciones de pesado. Representaciones distribuidas. *Word embeddings*. Análisis de semántica latente (LSA-*Latent Semantic Analysis*) y análisis de semántica latente probabilístico (PLSI) Modelos de temas (*Topic*



Modelling). Herramientas y recursos disponibles. Aprendizaje automático aplicado al PLN: Técnicas genéricas: clasificación y *clustering* y aplicación a los analizadores de los niveles lingüísticos. Métodos y técnicas aplicadas a sistemas de pregunta-respuesta. Métodos y técnicas aplicadas al análisis de sentimientos y minería de opinión. Métodos y técnicas aplicadas a los sistemas de diálogo y *chatbots*. Métodos y técnicas aplicadas a la recuperación y extracción de información. Métodos y técnicas aplicadas a la generación de resúmenes y simplificación de textos. Métodos y técnicas aplicadas a la traducción automática

Competencias asociadas

PLN01: Disponer de conocimientos generales de PLN. Conceptos, fases de análisis, técnicas principales, sistemas de PLN

PLN03: Conocer y saber emplear librerías y sistemas abiertos para PLN

PLN04: Conocer y saber aplicar técnicas de aprendizaje basadas en *Machine Learning* para soluciones PLN. Herramientas ML

PLN05: Conocer y saber aplicar técnicas de aprendizaje profundo basadas en *Deep Learning* para soluciones PLN. Herramientas DL

PLN08: Ser capaz de construir herramientas para el Análisis morfológico automático: *taggers*, lematización, *stemming*

PLN10: Ser capaces de construir herramientas para el Análisis semántico: Representación y composición del conocimiento, etiquetado de roles semánticos. Técnicas de similitud semántica y modelos de representación del conocimiento (*word representation models: word embeddings, word2vec...*)

PLN11: Ser capaces de construir herramientas para el Análisis discursivo y de la pragmática: análisis del contexto, marcas contextuales, etc.

PLN12: Saber construir sistemas de análisis de sentimientos: minería de opiniones y reconocimiento de emociones

PLN013: Ser competente en la aplicación de métodos estadísticos al desarrollo PLN. Ser competente en la formalización matemática de las relaciones entre los datos textuales

PLN15: Saber diseñar y desarrollar procesos de extracción de información

PLN16: Saber diseñar y desarrollar procesos de recuperación



	<p>de información</p> <p>PLN28: Ser capaces de construir herramientas para el Análisis sintáctico automático: <i>chunkers</i>, análisis de dependencias</p> <p>PLN032: Saber representar información lingüística a través de Grafos de conocimiento</p>
<p>Resultados de aprendizaje</p>	<p>RA21: Demuestra conocer la historia del PLN, sus aplicaciones principales y los enfoques con que se aborda</p> <p>RA22: Conoce los niveles lingüísticos relevantes en PLN, sus fundamentos y características principales</p> <p>RA23: Conoce y aplica las herramientas y recursos disponibles para los diversos niveles de análisis lingüístico</p> <p>RA24: Demuestra conocer la secuencia de analizadores (<i>pipeline</i>) necesarios para alcanzar un determinado nivel de análisis</p> <p>RA25: Conoce los fundamentos de la generación de lenguaje natural</p> <p>RA26: Conocer los fundamentos de los modelos de representación algebraica y su aplicación en la herramientas PLN</p> <p>RA27: Conoce las funciones de pesado y sus características</p> <p>RA28: Sabe extraer y representar el conocimiento necesario para poder construir una aplicación de PLN</p> <p>RA29: Demuestra conocer y aplicar las herramientas disponibles para representación de textos</p> <p>RA30: Sabe analizar un problema e identificar los métodos y técnicas apropiadas de PLN para abordarlo</p> <p>RA31: Conoce cómo integrar componentes ya existentes de PLN (tanto recursos lingüísticos como herramientas) para resolver aplicaciones del mundo real</p> <p>RA32: Conoce en detalle los principales campos de aplicación del PLN</p> <p>RA33: Demuestra capacidad de seleccionar y utilizar las herramientas necesarias para resolver una aplicación final de PLN</p>



STEM

Se estima conveniente que los estudiantes de estos grados tengan la opción de conocer conceptos y técnicas básicas del tratamiento de textos, módulos y librerías de tratamiento de texto, representación vectorial de información y conocimientos básicos de técnicas que permiten la explotación básica de la misma, aprendizaje automático. Para ello se propone una asignatura optativa para grados STEM.

Nombre de la asignatura: Métodos de Inteligencia Artificial para el Procesamiento del Lenguaje Natural	
Observaciones	
Prerrequisitos: conocimientos de programación, matemática discreta, álgebra lineal, cálculo matricial, probabilidades y estadística.	
Competencias asociadas	<p>PLN01: Disponer de conocimientos generales de PLN. Fases, análisis, principales técnicas, sistemas pre-existentes de PLN</p> <p>PLN03: Conocer y saber emplear librerías y sistemas abiertos para PLN</p> <p>PLN04: Conocer y saber emplear técnicas de aprendizaje basadas en <i>Machine Learning</i> para soluciones PLN</p> <p>PLN05: Conocer y saber emplear técnicas de aprendizaje profundo basadas en <i>Deep Learning</i> para soluciones PLN. Herramientas DL</p> <p>PLN034: Saber programar herramientas de PLN aplicando técnicas de aprendizaje profundo basadas en <i>Deep Learning</i> para soluciones PLN. Saber programar en TensorFlow, PyTorch, Keras, etc.</p> <p>PLN033: Ser capaces de emplear y entender las representación y composición del conocimiento, modelos de representación del conocimiento (word representation models: word embeddings, word2vec...)</p> <p>PLN029: Saber evaluar y validar herramientas y aplicaciones para tareas de PLN en textos</p>
Resultados de aprendizaje	<p>RA01: Conocer los fundamentos del PLN, los enfoques con que se aborda y los niveles lingüísticos relevantes en PLN, sus fundamentos y características principales</p> <p>RA02: Conocer técnicas de aprendizaje automático aplicado a tareas de PLN</p>



	<p>RA03: Conocer los fundamentos de los modelos de representación</p> <p>RA04: Conocer la secuencia de analizadores (pipeline) necesarios para alcanzar un determinado nivel de análisis</p> <p>RA05: Saber analizar un problema e identificar los métodos y técnicas apropiadas de PLN para abordarlo</p> <p>RA06: Conocer cómo integrar componentes ya existentes de PLN (tanto recursos lingüísticos como herramientas) para resolver aplicaciones del mundo real</p>
--	--

Lingüística, Filologías y Traducción e Interpretación

Se estima conveniente que los estudiantes de estos grados tengan la opción de conocer conceptos y técnicas básicas del tratamiento de textos, como puedan ser el preprocesado de textos, representación vectorial de información y conocimientos básicos de técnicas que permiten la explotación básica de la misma. Para ello se propone una asignatura optativa para grados de Lingüística, Filología, Traducción, Lenguas Aplicadas, etc.

Nombre de la asignatura: Métodos de Inteligencia Artificial para el Procesamiento del Lenguaje Natural	
Contenido	
Preparación y normalización de textos para su procesamiento.	
Representación vectorial de la información lingüística (bolsa de palabras y modelos distribucionales). Distancia como similitud (distancia coseno). Aplicación a semántica léxica.	
Clasificación y aprendizaje automático para tareas de PLN (árboles de decisión, regresión logística, máquinas de vectores, agrupación con k-vecinos más próximos).	
Métodos de evaluación, Precisión y cobertura, F1. Conjuntos de datos de prueba, <i>gold-standard</i> , equilibrio, selección, estadísticos relacionados. Validación cruzada y comparación con referencias (<i>Baselines</i>). Análisis de errores (tabla de contingencia y análisis cualitativo).	
Observaciones	
Competencias asociadas	<p>PLN01: Disponer de conocimientos generales de PLN. Fases, análisis, principales técnicas, sistemas pre-existentes de PLN</p> <p>PLN03: Conocer y saber emplear librerías y sistemas abiertos para PLN</p>



	<p>PLN04: Conocer y saber emplear técnicas de aprendizaje basadas en <i>Machine Learning</i> para soluciones PLN</p> <p>PLN05: Conocer y saber emplear técnicas de aprendizaje profundo basadas en <i>Deep Learning</i> para soluciones PLN. Herramientas DL</p> <p>PLN33: Ser capaces de emplear y entender la representación y composición del conocimiento, modelos de representación del conocimiento (<i>word representation models: word embeddings, word2vec...</i>)</p> <p>PLN013: Ser competente en la aplicación de métodos estadísticos al desarrollo PLN. Formalización matemática de las relaciones entre los datos textuales</p> <p>PLN027: Ser competente en la aplicación de técnicas basadas en reglas para soluciones PLN</p> <p>PLN029: Saber evaluar y validar herramientas y aplicaciones para tareas de PLN en textos</p>
<p>Resultados de aprendizaje</p>	<p>RA1. Analiza las tareas que realizan las aplicaciones informáticas que tienen por objeto textos (orales o escritos) y su aplicación en el ámbito de la investigación y de la industria</p> <p>RA2. Describe y diseña un proyecto de servicio lingüístico con análisis automático de textos, y uso de herramientas y métodos informáticos</p> <p>RA3. Implementa cadenas de procesamiento de textos utilizando plataformas de desarrollo y herramientas simples de programación</p> <p>RA3.1. Recoge un corpus de datos textuales para entrenamiento y lo prepara para su procesamiento</p> <p>RA3.2. Diseña y construye aplicaciones informáticas para preparar corpus de textos para su procesamiento, para la extracción de muestras específicas y para el cálculo y análisis de datos</p> <p>RA3.3. Aplica algoritmos y medidas de distancia con representaciones vectoriales de textos para diferentes tareas de clasificación</p> <p>RA3.4. Escribe expresiones regulares eficientes para la extracción de datos del corpus y la preparación de vectores</p> <p>RA3.5. Construye un clasificador supervisado para análisis de sentimiento y la evalúa con métricas estándar</p>



Ciencias de la información

Se propone añadir la materia optativa Retos frente al Tratamiento, procesamiento y análisis de enormes cantidades de información (Big Data) que tendrá como objetivo preparar a los futuros profesionales evitando que estén abrumados por la cantidad de datos y que puedan enfrentarse a los futuros retos planteando estrategias y empleando técnicas para mejorar el análisis de datos y su aplicación para usos editoriales y periodísticos.

Nombre de la materia: Retos frente al Tratamiento, procesamiento y análisis de enormes cantidades de información (Big Data)	
Asignaturas que la conforman	Retos frente al tratamiento, procesamiento y análisis de enormes cantidades de información (Big Data)
Contenido	
A continuación, se presentan los contenidos que configurarían el núcleo de la materia <i>Retos frente al tratamiento, procesamiento y análisis de enormes cantidades de información (Big Data)</i>	



Observaciones	
Competencias asociadas	<p>PER01: Conocer APIs y herramientas de Recuperación de Información y <i>scraping</i> tanto de internet como de redes sociales</p> <p>PER02: Conocer herramientas y librerías para el pre-procesado de texto (<i>stemming, stop words, etc.</i>)</p> <p>PER03: Conocer herramientas y librerías de extracción de información de dominio</p> <p>PER04: Conocer herramientas y librerías para la identificación automática de <i>fake-news</i></p> <p>PER05: Conocer herramientas y librerías para realizar un análisis de modelado de temas</p>
Resultados de aprendizaje	<p>RA1: Demuestra haber adquirido una perspectiva global de qué es la inteligencia artificial y sus aplicaciones actuales y futuras en el ámbito del periodismo</p> <p>RA2: Conoce herramientas para poder obtener información a partir de la red y de redes sociales</p> <p>RA3: Conoce herramientas para poder procesar grandes cantidades de información</p> <p>RA4: Conoce herramientas que permiten obtener distintas características como las fuentes originarias de la información, veracidad de la información, etc.</p>

Ciencias Sociales y Ciencias Políticas

Se propone añadir la materia optativa de *Técnicas de Procesamiento del Lenguaje Natural aplicadas a las ciencias sociales*. Esta materia estaría conformada por una asignatura que permitiría conocer a los alumnos técnicas de recuperación y análisis de la información recogida de distintas fuentes: web, redes sociales, media, etc.



Nombre de la materia: Técnicas de Procesamiento del Lenguaje Natural aplicadas a las ciencias sociales	
Asignaturas la que conforman	Técnicas de Procesamiento del Lenguaje Natural aplicadas a las ciencias sociales
Contenido	
A continuación, se presentan los contenidos que configurarían el núcleo de la materia <i>Técnicas de Procesamiento del Lenguaje Natural aplicadas a las ciencias sociales</i>	
Observaciones	
Competencias asociadas	<p>POL01: Conocer APIs y herramientas de Recuperación de Información y <i>scraping</i></p> <p>POL02: Conocer herramientas y librerías para el pre-procesado de texto (<i>stemming, stop words, etc.</i>)</p> <p>POL03: Conocer herramientas y librerías para realizar un análisis de sentimientos</p> <p>PLO04: Conocer herramientas y librerías para realizar un análisis de modelado de temas</p> <p>PLO05: Conocer herramientas y librerías para realizar un análisis discursivo</p>
Resultados de aprendizaje	<p>RA1: Demuestra haber adquirido una perspectiva global de qué es la inteligencia artificial y sus aplicaciones actuales y futuras en el ámbito de las ciencias sociales</p> <p>RA2: Conoce herramientas para poder recuperar información de distintas fuentes, y analizar dicha información a distintos niveles (análisis de sentimientos, contrastación de información, modelado de temas, análisis discursivo)</p> <p>RA3: Demuestra ser capaz de identificar y proponer potenciales aplicaciones y usos de las técnicas y herramientas estudiadas en las ciencias sociales</p>

Administración y dirección de empresas

Se propone añadir la materia optativa de *Tecnologías de la Lengua aplicadas al ámbito empresarial*. Esta materia estaría conformada por una asignatura que permitiría a los alumnos de administración de empresa familiarizarse con distintas herramientas que permitirían obtener una ventaja competitiva a su empresa. Herramientas de análisis de



sentimientos para obtener retroalimentación sobre la forma es que su empresa es percibida por los usuarios, herramientas de análisis de textos para poder tramitar y analizar automáticamente quejas de los usuarios, *chatbots* para poder interactuar directamente con los usuarios, etc.

Nombre de la materia: Tecnologías de la Lengua aplicadas al ámbito empresarial	
Asignaturas que la conforman	Tecnologías de la Lengua aplicadas al ámbito empresarial.
Contenido	
A continuación, se presentan los contenidos que configurarían el núcleo de la materia <i>Tecnologías de la Lengua aplicadas al ámbito empresarial</i> .	
Observaciones	
Competencias asociadas	<p>ADM01: Conocer las posibilidades y limitaciones de herramientas de dialogo (<i>chatbots</i>)</p> <p>ADM02: Conocer herramientas de extracción de información</p> <p>ADM03: Conocer herramientas y librerías para realizar un análisis de modelado de temas</p> <p>ADM04: Conocer herramientas y librerías de <i>sentiment analysis</i></p>
Resultados de aprendizaje	<p>RA1: Demuestra haber adquirido una perspectiva global de qué es la inteligencia artificial y sus aplicaciones actuales y futuras en el ámbito empresarial</p> <p>RA2: Demuestra conocer herramientas que permitirían obtener una ventaja competitiva a su empresa. Herramientas de análisis de sentimientos para obtener <i>feedback</i> sobre como su empresa es percibida por los usuarios, herramientas de análisis de textos para poder tramitar y analizar automáticamente quejas de los usuarios, <i>chatbots</i> para poder interactuar directamente con los usuarios, etc.</p> <p>RA3: Demuestra ser capaz de identificar y proponer potenciales aplicaciones y usos de las técnicas y herramientas estudiadas</p>



Medicina y Farmacia

Se propone añadir en ambas titulaciones la materia optativa de *Tecnologías de la Lengua aplicadas al ámbito sanitario* (o *Técnicas de IA aplicadas al ámbito sanitario*, si se considerase más adecuado como nombre). Esta materia estaría conformada por dos asignaturas. Una asociada al conocimiento de las distintas ontologías médicas y formas de explotar la información en ellas almacenada. Y por último otra asignatura que cubra la necesidad de tener conocimiento sobre las herramientas, aplicaciones y casos de uso de las TL al ámbito de la medicina. Con estas asignaturas la brecha entre los profesionales técnicos en TL y los profesionales del ámbito sanitario quedaría reducida potenciando que se generen sinergias entre ambos profesionales, surgiendo nuevo conocimiento y nuevas necesidades.

Nombre de la materia: Tecnologías de la Lengua aplicadas al ámbito sanitario	
Asignaturas la que conforman	Ontologías médicas, usos y aplicaciones Aplicaciones de la IA en el ámbito sanitario
Contenido	
A continuación, se presentan los contenidos que configurarían el núcleo de la materia <i>Tecnologías de la Lengua aplicadas al ámbito sanitario</i>	
Observaciones	
Competencias asociadas	<p>MED01: Conocer herramientas de extracción de información</p> <p>MED02: Conocer ontologías médicas y sus aplicaciones</p> <p>MED03: Conocer técnicas y herramientas de procesamiento de información del dominio médico</p> <p>MED04: Conocer casos de uso exitosos en el ámbito de la medicina</p>
Resultados de aprendizaje	<p>RA1: Demuestra haber adquirido una perspectiva global de qué es la inteligencia artificial y sus aplicaciones actuales y futuras en el ámbito de la sanidad</p> <p>RA2: Conoce las ontologías médicas, la importancia de su aplicación para la estandarización, intercambio y explotación de la información médica</p> <p>RA3: Demuestra ser capaz de identificar y proponer potenciales aplicaciones y usos de las técnicas y herramientas estudiadas</p>



Derecho

Se propone añadir la materia optativa de *Retos Jurídicos derivados del uso masivo de la información*. Esta materia estaría conformada por dos asignaturas. Una asociada al conocimiento de las distintas técnicas y herramientas de TL aplicadas al ámbito jurídico. Y otras relacionadas con los retos éticos y jurídicos asociados al uso masivo de la información. Igual que sucede con la propuesta dirigida al ámbito sanitario, con estas asignaturas se pretende reducir la brecha entre los profesionales técnicos en TL y los profesionales del ámbito jurídico, preparando a estos profesionales ante los nuevos retos de una sociedad cada vez más informatizada.

Nombre de la materia: Retos Jurídicos derivados del uso masivo de la información	
Asignaturas que la conforman	Retos éticos y jurídicos derivados del uso masivo de la información Aplicaciones de la IA en el ámbito jurídico
Contenido	
A continuación, se presentan los contenidos que configurarían el núcleo de la materia <i>Retos Jurídicos derivados del uso masivo de la información</i>	
Observaciones	



Competencias asociadas	<p>DER01: Conocer herramientas de extracción de información</p> <p>DER02: <i>Conocer aspectos jurídicos asociados al uso masivo de datos</i></p> <p>DER03: <i>Conocer retos éticos asociados a la inteligencia artificial</i></p> <p>DER04: <i>Conocer técnicas y herramientas de procesamiento de información del dominio jurídico</i></p> <p>DER05: Conocer casos de uso exitosos en el ámbito de la justicia</p>
Resultados de aprendizaje	<p>RA1: Demuestra haber adquirido una perspectiva global de qué es la inteligencia artificial y sus aplicaciones actuales y futuras en el ámbito de la justicia</p> <p>RA2: Conoce los nuevos retos éticos ante los que hay que regular</p> <p>RA3: Conoce los riesgos y beneficios derivados del uso masivo de la información</p> <p>RA4: Demuestra ser capaz de identificar y proponer potenciales aplicaciones y usos de las técnicas y herramientas estudiadas en el ámbito jurídico</p>

v. Máster

Los estudios de máster dan opción a la especialización en áreas de conocimiento, completando la formación obtenida en el grado, y deben satisfacer las demandas de la industria y la investigación en un área de rápido crecimiento. El informe ET1, en las Tablas 5, 6, 7 y 8, ha presentado las propuestas existentes actualmente en TL e IA, tanto en España como en el resto del mundo, mostrando que actualmente ya existen distintos másteres que abarcan las áreas de tecnologías del lenguaje y la comunicación. Como ejemplo, podemos citar los másteres ofrecidos en la Universidad del País Vasco sobre procesamiento del lenguaje: el máster local a nivel nacional “Análisis y Procesamiento del Lenguaje” (HAP/LAP) y el máster a nivel europeo “Máster Universitario Erasmus Mundus en Tecnologías del Lenguaje y la Comunicación” (LCT), en colaboración con distintas universidades europeas.

Examinando esas propuestas, vemos que esos másteres están organizados de manera modular, al combinar aspectos de la lingüística con la informática e inteligencia artificial, dado su carácter multidisciplinar. Por ejemplo, el máster HAP/LAP incluye personal investigador y profesorado de los departamentos de Didáctica de la Lengua y la Literatura, Lenguajes y Sistemas Informáticos,

Arquitectura y Tecnología de Computadores, Ciencia de la Computación e Inteligencia Artificial, Lengua Vasca y Comunicación e Ingeniería de Comunicaciones, así como de distintos centros tecnológicos. En concreto, el máster LCT incluye, en consonancia con otros másteres europeos, los siguientes módulos:

- Theoretical linguistics
- Natural language processing
- Phonetics and phonology
- Cognitive models for speech language processing
- Speech signal processing
- Pattern recognition
- Language engineering applications
- Students also need certain background programming skills
- Digital Humanities
- Data Science

Se puede apreciar cómo el grupo de módulos propuesto abarca múltiples áreas y permite un gran número de combinaciones o especializaciones de acuerdo al *background* o interés de cada estudiante, de cara a adquirir o reforzar los aspectos menos desarrollados en sus respectivos grados. Por ejemplo, los estudiantes que provienen de grados en Filología dispondrían de la posibilidad de elegir una especialización en *Digital Humanities*, incluyendo competencias de programación, o bien escoger una orientación técnica, abarcando módulos de programación, *Data Science* y *Pattern Recognition*.

Teniendo estos factores en cuenta, y viendo que ya existe una oferta definida de másteres en TL e IA según esas líneas, nuestra propuesta es que se deberá incidir en los siguientes aspectos:

- Refuerzo de los actuales másteres en TL e IA, con posibilidad de becas y ayudas, y promoviendo la cooperación a nivel estatal e internacional entre distintos agentes investigadores (departamentos, universidades y centros tecnológicos)
- Ampliación de la oferta, promoviendo la creación de másteres allí donde exista una importante base investigadora en TL y una demanda adecuada, así como la posibilidad de másteres online o semipresenciales para cubrir toda la demanda prevista.

Por otra parte, además de los másteres en TL e IA, la oferta de contenidos en estas materias también debería producirse en los másteres del resto de disciplinas, tanto de tipo STEM como humanísticas o artísticas. La incorporación de estos contenidos se produciría en la siguiente línea:

- Grupo STEM. Como ejemplo, los másteres en ingeniería deberían incluir los módulos de Natural Language Processing y Language Engineering Applications, mientras que un máster de Física incluiría el módulo de Natural Language Processing y Speech Signal Processing.
- Inclusión de módulos de Tecnologías del Lenguaje en los másteres en Lingüística, como Digital Humanities, Natural Language Processing y Language Engineering Applications.
- Inclusión de módulos de Tecnologías del Lenguaje en másteres relacionados con los campos de aplicación de la TL como medicina, sociología, periodismo, derecho, administración y dirección de empresas, etc.

vi. Doctorado

Los estudios de doctorado suponen el más alto grado académico y de especialización del conocimiento. Consecuentemente, de su adecuada planificación dependerá en gran medida el éxito de la implantación de la IA y las TL en la sociedad, ya que de ella se derivan la creación y consolidación de expertos que trabajarán en la creación de nuevo conocimiento y supondrán uno de los ejes fundamentales del progreso en los ámbitos empresarial, industrial y académico. Adicionalmente, este núcleo de expertos de alto nivel también se incorporarán al proceso formativo de grados, másteres, doctorados y títulos de especialización en IA y TL, por lo que el establecimiento de una base sólida en este apartado supone un elemento crucial en esta estrategia, que de manera circular y en sucesivos ciclos podrá consolidar un liderazgo en el área.

En el informe ET1 se han analizado las diferentes propuestas para programas de doctorado que tienen como ámbito las TL en España. El informe ha mostrado cómo ya existe una importante base empresarial, universitaria y de centros tecnológicos que realizan investigación puntera a nivel mundial en IA y TL. Sin embargo, esa fuerza no se traduce directamente en un conjunto amplio de programas de doctorado, que se encuentran dispersos y además focalizados en un número reducido de universidades (ver ET1, Tabla 9), o como subgrupos minoritarios de otros programas de doctorado en otras universidades. Desde esa perspectiva, creemos que es de suma importancia el reforzar y renovar el conjunto de propuestas de doctorado actuales y proponemos las siguientes líneas de actuación:



- Consolidación de los programas actuales de doctorado en TL. Esto implica el aumento de financiación, siempre sujeto a una evaluación objetiva basada en resultados, para los siguientes apartados:
 - Movilidad del profesorado, permitiendo a los investigadores y profesores de los programas de doctorado la estancia en centros de amplia trayectoria investigadora.
 - Igualmente, se deberá posibilitar la incorporación, mediante estancias de corta duración o contratos, de investigadores de prestigio desde otros centros de investigación de referencia, con el objetivo de impartir cursos o establecer propuestas de colaboración conjuntas.
 - Formación y actualización del profesorado, mediante la organización o asistencia a cursos de especialización en áreas de investigación punteras.
- Creación de programas de doctorado en TL e IA en entornos donde haya un ecosistema de investigación en TL competitivo y relevante. Actualmente existen núcleos importantes de investigación y producción científica que se encuentran englobados en programas de doctorado más generalistas (como Lingüística, Informática o Telecomunicaciones), que tienen potencial de iniciar una especialización mediante un programa de doctorado propio.
- Doctorados industriales. Un aspecto fundamental de nuestra propuesta consiste en fomentar la colaboración entre la universidad, los centros tecnológicos y la empresa, de cara a disminuir la brecha existente actualmente entre investigación básica y aplicada.
- Alumnado. En este apartado, aunque existe actualmente un gran interés en los estudios en IA y TL, existen diversas barreras como la escasez de plazas y becas o ayudas pre y postdoctorales. En concreto planteamos:
 - Potenciación de becas predoctorales y contratos postdoctorales. Para ampliar el área y cubrir la demanda en TL, es fundamental ampliar el número de becas y contratos, así como incrementar su atractivo, económicamente y profesionalmente, de cara a disponer de un personal que responda al perfil de excelencia requerido.
 - Potenciación de la movilidad en el ámbito doctoral, tanto predoctoral como postdoctoral, mediante la promoción de estancias en universidades y centros de investigación españoles y extranjeros de reconocido prestigio, así como la recepción

de investigadores externos (estancias o contratos) en los centros responsables de los estudios de doctorado.

vii. MOOC

Los cursos MOOC suponen un mecanismo de democratización de la enseñanza de contenidos online, permitiendo una llegada simultánea y gratuita a una gran comunidad de estudiantes. De esta forma los cursos MOOC se suelen plantear como introductorios para la especialización en sectores específicos, así como una manera ágil de actualizar el currículum del profesional.

En el caso de las tecnologías del lenguaje, los cursos MOOC se plantean como un mecanismo básico y atómico para la adquisición de conocimientos, que posteriormente podrían dar lugar a una especialización de mayor nivel mediante otros tipos de cursos más avanzados.

Los destinatarios de estos cursos podrían ser el personal novel de reciente incorporación a una entidad que trabaje en tecnologías del lenguaje y que necesita iniciarse en alguna de las técnicas propias de la disciplina, personal que necesita reciclarse en su puesto de trabajo para adaptarse al uso de estas nuevas técnicas, o bien estudiantes que máster o doctorado que necesitan recibir un curso introductorio acerca de dichas tecnologías para el desarrollo de sus competencias.

De esta forma, los cursos MOOC deben plantearse como una secuencia -no necesariamente ordenada- de unidades básicas de conocimiento que permitan al estudiante configurar de manera ágil su propio itinerario formativo, profundizando en aquellas áreas que sean de su interés.

Los contenidos de los cursos MOOC pueden abarcar todas las competencias analizadas anteriormente en diferentes grados de especialización. No obstante, teniendo en cuenta que los mayores grados de especialización se encontrarán en propuestas de cursos y títulos más intensivos que el MOOC, se propone al menos la creación de cursos MOOC en los siguientes ámbitos:

- Cursos MOOC de carácter generalista sobre las Tecnologías del Lenguaje Humano, incluyendo competencias sobre análisis de lenguaje, uso de enfoques y recursos básicos, así como aplicaciones fundamentales de las Tecnologías del Lenguaje.
- Cursos MOOC sobre las TLH aplicadas a diferentes entornos de trabajo, con enseñanzas específicas sobre aplicaciones y técnicas útiles específicamente en el área de negocio objeto de estudio.
- Cursos MOOC para la introducción de nuevas técnicas, tendencias, y aplicaciones, que por su inmediatez aún no hayan podido incorporarse a otros niveles de enseñanza.

7. CONCLUSIONES

En el presente documento se ha detallado una estrategia de formación en PLN y TL. Se ha realizado una propuesta para los distintos niveles de formación, incluyendo la formación pre-universitaria, formación continua, formación dirigida a las Administraciones Públicas, y MOOC, y principalmente la formación de nivel universitario, incluyendo grado, máster y doctorado.

Para ello, se ha llevado a cabo inicialmente un análisis del estado actual de necesidades formativas en TL, la previsión futura de necesidades en el sector, el análisis de competencias basado en la demanda identificada y por último, se han detallado las propuestas formativas, en las que podemos destacar dos grados: 1) Grado en Lingüística Computacional y 2) Grado en Inteligencia Artificial en las Tecnologías del Lenguaje. Además, se han propuesto distinta optatividad para otros grados no relacionados directamente con las tecnologías del lenguaje, pero cuyos profesionales se beneficiarán enormemente de los conceptos en ellas adquiridos, incluyendo grados humanísticos, jurídicos, sanitarios y ciencias sociales.

8. ANEXO. COMPETENCIAS TÉCNICAS GENERALES

Grupo de competencias asociadas al desarrollo de recursos lingüísticos (REC)

Este grupo aúna las competencias relacionadas con la generación de recursos. Para este grupo, hemos seleccionado las competencias descritas en la Tabla 1, basadas en competencias relacionadas solicitadas en las ofertas de trabajo.

REC01	Disponer de capacidad para crear recursos lingüísticos: diccionarios, léxicos, corpus
REC02	Estar cualificado para anotar corpus con información lingüística utilizando las herramientas adecuadas
REC03	Disponer de capacidad de elaboración de recursos para asistentes conversacionales
REC03A	Poder diseñar adecuadamente estrategias de diálogo y diálogos en función de la personalidad definida para el agente conversacional
REC03B	Ser capaz de crear diálogos
REC03C	Disponer de conocimiento de las técnicas de creación de diálogos

Tabla 1. Competencias para el perfil Recursos.

Grupo de competencias asociadas a la programación (PRG)

Se trata de competencias asociadas con la Programación, en diferentes lenguajes y entornos relacionados con los desarrollos de PLN. La Tabla 2 muestra las competencias de este grupo.

PRG01	Ser competente programando en lenguajes enfocados al desarrollo PLN: Java, Javascript, Python, PHP
PRG02	Ser competente programando en lenguajes enfocados al análisis estadístico: R
PRG03	Conocer y saber emplear expresiones regulares y construcción de patrones de búsqueda
PRG04	Ser capaz de desarrollar aplicaciones en plataformas (Eclipse, IBM Bluemix, Vue.js, React)
PRG05	Ser capaz de desarrollar pruebas en plataformas de pruebas de servicios API REST (Postman)
PRG06	Conocer y saber emplear plataformas de gestión y control de versiones. GIT
PRG07	Ser competente programando en orientación a objetos: Java, Scala
PRG08	Saber programar APIs
PRG09	Conocer y saber emplear plataformas para desarrollo específico web (Laravel, Symphony)
PRG10	Disponer de conocimientos multientorno y capacidad de adaptación a cambios
PRG11	Conocer y saber realizar desarrollos en plataformas cloud computing (Cloudera, Hortonworks , Amazon AWS, Heroku)

Tabla 2. Competencias para el perfil de Programación.



Este grupo, se listan varias competencias básicas generales solicitadas para los profesionales de las TL.

PLN01	Disponer de conocimientos generales de PLN. Conceptos, fases de análisis, técnicas principales, sistemas de PLN
PLN02	Disponer de conocimiento sobre datos textuales (lingüísticos, estructurados, digitales, textuales, multimedia). Análisis e interpretación de datos lingüísticos en todos los niveles gramaticales mediante la aplicación de teorías y métodos formales, descomponer las estructuras en sus constituyentes e identificar patrones estructurales
PLN31	Disponer de conocimientos en adquisición, análisis, transformación y utilización de datos no estructurados y multilingües
PLN03	Conocer y saber emplear librerías y sistemas abiertos para PLN: NLTK, Gate, Freeling, Gensim, GlounNLP, Beautiful-Soup, spaCy, Apache OpenNLP, Stanford's CoreNLP, Apache UIMA
PLN04	Conocer y saber aplicar técnicas de aprendizaje basadas en Machine Learning para soluciones PLN. Librerías y entornos ML: TensorFlow, scikit-learn
PLN05	Conocer y saber aplicar técnicas de aprendizaje profundo basadas en Deep Learning para soluciones PLN. Herramientas DL
PLN013	Ser competente en la aplicación de métodos estadísticos al desarrollo PLN. Ser competente en la Formalización matemática de las relaciones entre los datos textuales
PLN027	Ser competente en la aplicación de técnicas basadas en reglas para soluciones PLN
PLN029	Saber evaluar y validar herramientas y aplicaciones para tareas del procesamiento del lenguaje natural en textos
PLN030	Saber evaluar y validar herramientas y aplicaciones para tareas del procesamiento del lenguaje natural en sistemas conversacionales
PLN032	Saber representar información lingüística a través de Grafos de conocimiento
PLN034	Saber programar herramientas de PLN aplicando técnicas de aprendizaje profundo basadas en Deep Learning para soluciones PLN. Saber programar en TensorFlow, PyTorch, Keras, etc.

Tabla 3. Competencias para el grupo de PLN General

Grupo Lingüística: (LIN)

En este grupo, se listan varias competencias solicitadas para los profesionales de TL, relacionadas con habilidades lingüísticas.

LIN01	Ser capaces de formalizar el lenguaje natural
LIN02	Ser capaces de describir, analizar e interpretar datos lingüísticos de todos los niveles gramaticales de cualquier lengua mediante la aplicación de diversas teorías y métodos formales para descomponer estructuras lingüísticas en sus elementos constituyentes e identificar patrones estructurales
LIN03	Ser capaces de identificar patrones léxicos/morfológicos/sintácticos para modelado estadístico
LIN04	Disponer de conocimientos de la IPA (International Phonetic Alphabet) y SAMPA (Speech Assessment Methods Phonetic Alphabet)
LIN05	Ser capaces de leer de espectrogramas utilizando Pratt/Wave Surfer

Tabla 4. Competencias para el grupo Lingüistas

Grupo: Ingeniería del software: ISW

Bajo este grupo, se listan varias competencias solicitadas para los profesionales de las TL que al crear productos software deben poseer competencias propias de cualquier desarrollador software.

ISW01	Conocer y aplicar metodologías y la correcta perspectiva ingenieril del desarrollo del software
ISW02	Saber diseñar y desarrollar productos de software siguiendo los principios de la ingeniería del Software
ISW03	Saber diseñar y desarrollar productos de software aplicando metodologías ágiles

Tabla 5. Competencias para el grupo de Ingeniería del Software

Grupo: General (GEN)

Bajo este grupo, se listan competencias generales de comunicación, trabajo en equipo y aprendizaje, y de idiomas.

GEN01	Disponer de conocimiento funcional y técnico de la cadena End to End (E2E) del proyecto en cuestión
GEN02	Ser capaces de desarrollar trabajo en equipo con habilidades interpersonales excepcionales y una actitud orientada a la solución
GEN03	Disponer de capacidad de trabajar bajo plazos (dado por horas) y en varios proyectos simultáneamente
GEN04	Ser competente como usuario de los editores de texto más comunes (Microsoft Office y documentos Google)
GEN05	Disponer de orientación hacia la excelencia, atención a los detalles
GEN06	Disponer de habilidad de comunicar de forma efectiva. Interacción con clientes, especialmente cara a cara en conversaciones técnicas
GEN07	Disponer de conocimiento demostrable en metodologías de investigación
GEN08	Saber preparar y dirigir presentaciones ante una amplia variedad de público



GEN09	Disponer de habilidad de comunicar contenidos técnicos a una audiencia no técnica
GEN10	Ser capaz de captar, entender y aprender rápidamente nuevas habilidades técnicas

Tabla 6. Competencias Generales

Grupo: STEM (STEM)

Bajo este grupo, se listan varias competencias básicas de las STEMs.

STEM01	Poseer capacidad para la resolución de los problemas matemáticos que puedan plantearse en la ingeniería. Aptitud para aplicar los conocimientos sobre: álgebra lineal, cálculo diferencial e integral, métodos numéricos, algorítmica numérica, estadística y optimización
STEM02	Disponer de capacidad para comprender y dominar los conceptos básicos de matemática discreta, lógica, algorítmica y complejidad computacional, y sus aplicaciones para la resolución de problemas propios de la ingeniería

Tabla 7. Competencias Generales STEM

Grupo: Inteligencia Artificial (IA)

Bajo este grupo, se listan varias competencias básicas de Inteligencia Artificial.

IA01	Conocer técnicas y conceptos básicos de IA: Búsquedas en espacios de estados, representación del conocimiento, agentes, etc.
IA02	Conocer las técnicas y algoritmos de IA apropiados para resolver problemas de aprendizaje supervisado y no supervisado. Algoritmos genéticos, <i>Deep Learning</i> , <i>embeddings</i> , KNN etc.
IA03	Conocer, ser capaces de analizar, entender la formalización y saber aplicar las técnicas apropiadas y modificar aplicaciones de IA en problemas reales existentes

Tabla 8. Competencias Generales IA

Grupo: Idiomas (ID)

Bajo este grupo, se listan varias competencias relacionadas con el dominio o conocimiento de una o varias lenguas.

ID01	Ser competente en Español (nativo) / inglés / francés
ID02	Ser competente en Español (nativo) con conocimientos de español de América
ID03	Ser competente en habla y escritura fluida e idiomática en inglés (nativo o conocimiento avanzado)
ID04	Ser competente en italiano y español en los negocios
ID05	Conocer de forma fluida y experta de una de las siguientes lenguas: alemán, francés canadiense, noruego, sueco y holandés

Tabla 9. Competencias de Idiomas

Grupo: Bases de Datos (BDA)

Bajo este grupo, se listan varias competencias relacionadas con bases de datos

BDA01	Ser competente en el acceso a Bases de Datos (BD) para la gestión de flujos de trabajo: data-pipelines y data-stores. MySQL, DynamoDB, Kafka
BDA02	Conocer sistemas relacionales de BD: MySQL, MariaDB, Oracle, SQL Server
BDA03	Conocer bases de datos NOSQL: Cassandra, MongoDB

Tabla 10. Competencias para el grupo de Bases de Datos

Grupo: Data Science - Data Analysis (DAS)

Bajo este grupo, se listan varias competencias de Análisis de Datos.

DAS01	Principios y métodos de análisis de datos y ciencia de datos
DAS02	Procesos de limpieza y calidad de los datos
DAS03	Herramientas y frameworks de gobierno del dato: Collibra, IBM Infosphere, Cloudera, Talencl, Experian, Hortonworks, Mapr, AWS, Azure
DAS04	Arquitectura y plataformas BigData: Hadoop, Spark
DAS05	Modelado de sistemas descriptivos y predictivos
DAS06	Sistemas de visualización de datos.

Tabla 11. Competencias para el grupo de Data Science

Grupo: Sistemas Operativos (SOP)

Bajo este grupo, se listan varias competencias básicas relacionadas con el empleo de Sistemas Operativos

SOP01	Ser competente en el empleo de Sistemas operativos basados en Unix (Linux). Manipulación del texto con línea de comandos, bash, shell, awk, egrep y tr. Introducción a la programación con herramientas de línea de comandos (variables, tipos de datos, funciones).
--------------	--

Tabla 12. Competencias para el grupo de Sistemas Operativos

Grupo de competencias asociadas a diferentes tareas y utilidades de las herramientas del Procesamiento del Lenguaje Natural (PLN)

SISTEMAS DE DIÁLOGO Y CHATBOTS

Este grupo, se listan varias competencias solicitadas para los profesionales de las TL que desarrollen chatbots.

PLN14	Disponer de conocimiento de Sistemas de diálogo colaborativo. Estrategias de diálogo colaborativo
PLN20	Disponer de conocimiento de las Tecnologías de chatbot. Sistemas y motores cognitivos (Google DialogFlow, IBM Watson, Amazon Alexa, Facebook messenger, Whatsapp Business), y asistentes (Google Home ...)

Tabla 13. Competencias para el grupo de Sistemas de Diálogo y Chatbots

SISTEMAS DE INGENIERÍA ONTOLÓGICA Y COMPRENSIÓN DEL LENGUAJE (NLU)

Este grupo, se listan varias competencias solicitadas para los profesionales de las TL que desarrollen sistemas que requieran módulos de comprensión del lenguaje.

PLN17	Disponer de conocimiento de la Ingeniería ontológica: Modelado conceptual ontológico. Ontologías y mecanismos de inferencia. Metodologías, lenguajes y herramientas para la representación conceptual en dominios: OWL, Protegee...
PLN26	Disponer de conocimiento sobre Procesos de comprensión del lenguaje (NLU). Inferencia y extracción de conocimiento

Tabla 14. Competencias para el grupo de PLN Sistemas de Ingeniería Ontológica y Comprensión del Lenguaje

SISTEMAS DE CLASIFICACIÓN TEXTUAL

Este grupo, se listan varias competencias solicitadas para los profesionales de las TL que desarrollen sistemas que requieran módulos de clasificación de documentos.

PLN18	Disponer de conocimiento de Sistemas de clasificación textual y clustering (agrupamiento)
PLN21	Disponer de conocimiento sobre Modelado de tópicos: reconocimiento del tema

Tabla 15. Competencias para el grupo de PLN Sistemas de Clasificación Textual

SISTEMAS DE RECOMENDACIÓN

Este grupo, se listan varias competencias solicitadas para los profesionales de las TL que desarrollen sistemas de recomendación.

PLN19	Disponer de conocimiento de Sistemas de recomendación, técnicas y motores (ThinkAnalytics, BoxFish, Kannuu, Jinni, Digitalsmiths, Watchwith, etc)
--------------	---

Tabla 16. Competencias para el grupo de PLN Sistemas de Recomendación

GENERACIÓN DEL LENGUAJE HUMANO NLG

Este grupo, se listan varias competencias solicitadas para los profesionales de las TL que desarrollen sistemas que requieran módulos de generación del lenguaje.

PLN22	Disponer de conocimiento de Sistemas de generación del Lenguaje Humano. Arquitectura para la Generación de Lenguaje Natural (NLG). Macro y micro planificación y realización lingüística
PLN23	Ser competente en la adquisición y generación de estructuras de conocimiento para NLG

Tabla 17. Competencias para el grupo de PLN Sistemas de Recomendación

TRADUCCIÓN AUTOMÁTICA

Este grupo, se listan varias competencias solicitadas para los profesionales de las TL que desarrollen sistemas de traducción automática.

PLN24	Conocer los Sistemas de traducción automática
--------------	---

Tabla 18. Competencias para el grupo de PLN Traducción Automática

SISTEMAS DE QA

Este grupo, se listan varias competencias solicitadas para los profesionales de las TL que desarrollen sistemas de Question-Answering.

PLN25	Conocer los Sistemas de búsqueda de respuestas. Question-Answering (QA)
--------------	---

Tabla 19. Competencias para el grupo de PLN Sistemas de Question Answering

ANÁLISIS LINGÜÍSTICO

Este grupo, se listan varias competencias solicitadas para los profesionales de las TL que desarrollen módulos de análisis lingüístico.

PLN08	Ser capaces de construir herramientas para el Análisis morfológico automático: <i>taggers, lematización, stemming</i>
PLN28	Ser capaces de construir herramientas para el Análisis sintáctico automático: <i>chunkers, análisis de dependencias</i>
PLN10	Ser capaces de construir herramientas para el Análisis semántico: Representación y composición del conocimiento, etiquetado de roles semánticos. Técnicas de similitud semántica y modelos de representación del conocimiento (<i>word representation models: word embeddings, word2vec...</i>)
PLN33	Ser capaces de emplear y entender las representación y composición del conocimiento, modelos de representación del conocimiento (<i>word representation models: word embeddings, word2vec...</i>)
PLN011	Ser capaces de construir herramientas para el Análisis discursivo y de la pragmática: análisis del contexto, marcas contextuales, etc.

Tabla 20. Competencias para el grupo de PLN Análisis Lingüístico

SISTEMAS DE ANÁLISIS DE SENTIMIENTOS

Este grupo, se listan varias competencias solicitadas para los profesionales de las TL que desarrollen sistemas de análisis de sentimientos.

PLN12	Ser competentes en la construcción de sistemas de análisis de sentimientos: minería de opiniones y reconocimiento de emociones
--------------	--

Tabla 21. Competencias para el grupo de PLN Sistemas de Sentimientos

SISTEMAS DE RECUPERACIÓN Y EXTRACCIÓN DE INFORMACIÓN

En este grupo, se listan varias competencias solicitadas para los profesionales de las TL que desarrollen sistemas de Recuperación y Extracción de Información.

PLN15	Saber diseñar y desarrollar procesos de extracción de información
PLN16	Saber diseñar y desarrollar procesos de recuperación de información

Tabla 22. Competencias para el grupo de PLN Sistemas de Recuperación y Extracción de Información

Competencias asociadas a profesiones no relacionadas con las TL

A continuación pasamos a listar competencias en TL asociadas a distintos profesionales en áreas ajenas a las TL y que aparecen citadas en distintos informes y/o masters existentes a día de hoy.

Competencias relevantes para el perfil de un periodista

PER01	Conocer APIs y herramientas de Recuperación de Información y <i>scraping</i>
PER02	Conocer herramientas y librerías para el pre-procesado de texto (<i>stemming, stop words, etc.</i>)
PER03	Conocer herramientas y librerías de extracción de información de dominio
PER04	Conocer herramientas y librerías para realizar un análisis de modelado de temas

Tabla 23. Competencias para el perfil Sociólogo/Politólogo Computacional.

Competencias relevantes para el perfil de sociólogo o politólogo computacional

POL01	Conocer APIs y herramientas de Recuperación de Información y <i>scraping</i>
POL02	Conocer herramientas y librerías para el pre-procesado de texto (<i>stemming, stop words, etc.</i>)
POL03	Conocer herramientas y librerías para realizar un análisis de sentimientos
POL04	Conocer herramientas y librerías para realizar un análisis de modelado de temas
POL05	Conocer herramientas y librerías para realizar un análisis discursivo

Tabla 24. Competencias para el perfil Sociólogo/Politólogo Computacional.

Competencias relevantes para el perfil de economista

ECO01	Conocer APIs y herramientas de Recuperación de Información y <i>scraping</i>
ECO02	Conocer herramientas y librerías para el pre-procesado de texto (<i>stemming, stop words, etc.</i>)
ECO03	Conocer herramientas y librerías para realizar un análisis de sentimientos
ECO04	Conocer herramientas y librerías para realizar un análisis de modelado de temas
ECO05	Conocer herramientas y librerías para la generación de resúmenes

Tabla 25. Competencias para el perfil Economista.

Competencias relevantes para el perfil en Administración de empresas



ADM01	Conocer las posibilidades y limitaciones de los <i>chatbots</i> .
ADM02	Conocer herramientas de extracción de información
ADM03	Conocer herramientas y librerías para realizar un análisis de modelado de temas
ADM04	Conocer herramientas y librerías de <i>sentiment analysis</i>

Tabla 26. Competencias para el perfil Administración de empresa.

Competencias relevantes para el perfil biosanitario

MED01	Conocer herramientas de extracción de información
MED02	Conocer ontologías médicas
MED03	Conocer técnicas y herramientas de procesamiento de información del dominio médico

Tabla 27. Competencias para el perfil biosanitario.

Competencias relevantes para el perfil jurídico

DER01	Conocer herramientas de extracción de información
DER02	Conocer aspectos jurídicos asociados al uso masivo de datos
DER03	Conocer retos éticos asociados a la inteligencia artificial
DER04	Conocer técnicas y herramientas de procesamiento de información del dominio jurídico
DER05	Conocer casos de uso exitosos en el ámbito de la justicia

Tabla 28. Competencias para el perfil jurídico

Otros

OTR01	Conocer las posibilidades y limitaciones de los <i>chatbots</i>
OTR 02	Conocer herramientas de extracción de información

Tabla 29. Competencias para el perfil Administración de empresa.

9. ANEXO. GRADOS DE LINGÜÍSTICA COMPUTACIONAL Y TECNOLOGÍAS DEL LENGUAJE

EEUU

1. California State University, Fresno. Linguistics - [Computational Linguistics Option, B.A.](#)
2. Indiana University Bloomington. [Computational Linguistics B.S.](#)
3. Stanford University. [Joint Major in Computer Science + Linguistics](#)
4. Queens College. [Minor in Computational Linguistics](#)
5. University of Illinois. [Computer Science + Linguistics \(BA\)](#)
6. University of Southern California. [Computational Linguistics \(BS\)](#)

Europa

Alemania

1. Friedrich-Alexander Universität. Erlangen-Nüremberg. [Computational Linguistics \(BA\)](#)
2. University of Heidelberg. [Computational Linguistics Bachelor](#)
3. Universität Postdam Studium. [Computerlinguistik Single Subject Bachelor](#)
4. Universität des Saarlandes. [Computational Linguistics \(BA\)](#)
5. Universität Stuttgart. [Bachelor of Science program "Maschinelle Sprachverarbeitung"](#).
6. Universität Trier. [Bachelor- Studiengang Computerlinguistik](#)
7. Universität Tübingen. [International Studies in Computational Linguistics](#)

Irlanda

1. Trinity College Dublin. [BA Computer Science and Language](#)

Italia

1. Università di Pisa. [Lingüística Computazionale](#)

República Checa

1. Masaryk University. [Computational Linguistics \(BA\)](#)

Suiza

1. University of Zurich. [Computational Linguistics and Language Technology](#)

Rusia

1. St. Petersburg University. Applied, [Computer and Computational Linguistics \(English\)](#)

10. REFERENCIAS

11. GLOSARIO DE SIGLAS Y ACRÓNIMOS

AI	Artificial Intelligence
API	Interfaz de Programación de Aplicaciones (Application Programming Interface)
BD	Bases de Datos
DL	Aprendizaje Profundo (Deep Learning)
GV	Gestión de Versiones
ML	Aprendizaje Automático (Machine Learning)
MT	Traducción Automática (Machine Translation)
MOOC	Massive On-line Open course
NLP	Procesamiento del Lenguaje Natural (Natural Language Processing)
NN	Redes Neuronales (Neural Networks)
OO	Orientación a Objetos (Object Oriented)
SEAD	Secretaría de Estado para el Avance Digital
STEM	Ciencia, Tecnología, Ingeniería y Matemáticas (Science, Technology, Engineering, Mathematics)
TL	Tecnologías del Lenguaje
ITM	Redes de formación innovadora (Innovative Training Networks)
NLG	Generación de Lenguaje Natural (Natural Language Generation)
QA	Obtención de respuestas a preguntas (Question-Answering)
IPA	Alfabeto fonético internacional (International Phonetic Alphabet)
SAMPA	Speech Assessment Methods Phonetic Alphabet



Plan TL

C. Plan de acción en formación

**ET5: Plan de acción a corto y medio plazo
para la formación en TL en el sector universitario español**

**Aitziber Atutxa (UPV/EHU), Nuria Bel (UPF), Arantza Díaz de Ilarraza (UPV/EHU),
Patricio Martínez (UA), Raquel Martínez (UNED), Koldo Gojenola (UPV/EHU),
Antoni Oliver (UOC), German Rigau (UPV/EHU), Alfonso Ureña (UJ)**

Diciembre 2019

Este estudio ha sido realizado dentro del ámbito del Plan de Impulso de las Tecnologías del Lenguaje con financiación de la Secretaría de Estado para el Avance Digital, que no comparte necesariamente los contenidos expresados en el mismo. Dichos contenidos son responsabilidad exclusiva de sus autores.

Reservados todos los derechos. Se permite su copia y distribución por cualquier medio siempre que se mantenga el reconocimiento de sus autores, no se haga uso comercial de las obras y no se realice ninguna modificación de estas.

1. OBJETIVOS Y ORGANIZACIÓN DEL DOCUMENTO.....	181
2. METODOLOGÍA DE TRABAJO	181
3. ESTRATEGIA DE IMPLANTACION	181
4. CONCLUSIONES.....	185
5. REFERENCIAS	186
6. GLOSARIO DE SIGLAS Y ACRÓNIMOS	187

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

TABLA 14. EJEMPLO DE CALENDARIO DE PROPUESTA Y VERIFICACIÓN DE UN TÍTULO DE GRADO.....	183
TABLA 15. EJEMPLO DE CALENDARIO DE PROPUESTA Y VERIFICACIÓN DE UN TÍTULO DE MÁSTER.	184

ET5. Implantación de la estrategia de formación en TL específica para el sistema universitario español

1. OBJETIVOS Y ORGANIZACIÓN DEL DOCUMENTO

En el presente documento se detallarán en lo posible los pasos a seguir para implantar la propuesta de formación universitaria en PLN y TL presentada en el ET4. La propuesta consiste en dos grados: uno denominado Grado en Lingüística Computacional y otro denominado Grado en Inteligencia Artificial en las Tecnologías del Lenguaje. El grado en Lingüística Computacional es un grado enfocado a crear una formación profesionalizadora en las TLs para los estudiantes provenientes de un bachillerato de Humanidades y Ciencias Sociales, perfil lingüístico especializado en Tecnologías del lenguaje. El grado en Inteligencia Artificial en las Tecnologías del Lenguaje es un grado que está enfocado a obtener un perfil en ingeniería en Inteligencia Artificial especializado en Tecnologías del lenguaje (Grado IA-PLN).

Con respecto a los estudios de postgrado, el protocolo a seguir es similar al del grado.

2. METODOLOGÍA DE TRABAJO

La metodología de trabajo que se ha seguido para la realización del presente estudio se ha basado en:

1. Recopilación de la información recogida en varios Reales Decretos referentes a la verificación de títulos de grado y postgrado.
2. Presentación de un calendario orientativo que recoge los principales hitos del protocolo de verificación necesario para implantar los grados propuestos.

3. ESTRATEGIA DE IMPLANTACION

En la transición al EEE, se hizo un cambio normativo en el que se pasó de una lista cerrada de títulos de Licenciatura o Ingeniería determinada por el ministerio de referencia al marco normativo vigente, en el que, para la implantación de un nuevo título, es cada universidad (o agrupación de universidades) la que, por iniciativa propia, define el título y, en coherencia, una vez verificado y autorizado por la comunidad autónoma de referencia, se aprueba. Los títulos se inventarían en el Registro de Universidades, Centros y Títulos (RUCT).

Las normativas que se aplican son las siguientes:

- RD 1393/2007, de 29 de octubre, por el que se establece la ordenación de las enseñanzas universitarias oficiales.
- Decreto 274/2017, de 19 de diciembre, de implantación y supresión de las enseñanzas universitarias oficiales conducentes a la obtención de los títulos de Grado, Máster y Doctorado.
- Normativa de implantación de titulaciones de grado, máster o doctorado, según el caso propio de cada universidad.

La legislación española que regula la ordenación de las enseñanzas universitarias en España atribuye a ANECA y agencias autonómicas equivalentes²⁸ la competencia de elevar informes al ministerio competente en materia universitaria y al Consejo de Universidades sobre la evaluación, certificación y acreditación de títulos oficiales. Así, los estudios conducentes a un título oficial deben someterse a un proceso de evaluación externa en diferentes momentos: una verificación; un seguimiento tras su implantación y una acreditación ex post. Los criterios de evaluación son los que se definieron en la Conferencia de Ministros de la Unión Europea de Ereván (2015), "[Standards and Guidelines for Quality Assurance in the European Higher Education Area](#)²⁹" (ESG) y la acreditación por parte de una de las agencias evaluadoras oficiales significa la homologación del título en todos los países de la Unión Europea. El calendario de acreditación es oficial y depende de la aprobación por parte de la universidad y la agencia acreditadora que puede pedir más de una corrección.

El calendario de acreditación es oficial pero depende de la aprobación por parte de la universidad, la agencia acreditadora y de la comunidad autónoma, y de la cantidad de correcciones que se soliciten por parte de cada estamento. En cualquier caso el presente documento muestra un ejemplo de calendario que puede ser orientativo para quien se sumerja en dicho proceso. En cuanto a los plazos, es imposible establecerlos de forma precisa dado que como ya se ha comentado la tramitación de la propuesta depende de la comunidad autónoma en la que se tramite y los diversos estamentos de la universidad por los que deba ser aprobado, pero conviene que se inicie aproximadamente dos años antes del curso académico previsto para su implantación (siempre teniendo en cuenta lo anteriormente mencionado, es decir que dependiendo de la comunidad autónoma los plazos pueden variar en meses).

El calendario orientativo que aquí se presenta está diseñado asumiendo que los nuevos grados quisieran comenzar a impartirse en el año 2022 y partiendo de que existe consenso con el equipo rectoral.

²⁸ Ver lista completa en <http://www.aneca.es/Agencias-de-las-Comunidades-Autonomas>

²⁹ https://enqa.eu/wp-content/uploads/2015/11/ESG_2015.pdf

Actuaciones para el inicio de un nuevo título de máster universitario para el curso 2022-2023	Previsión fechas
PROCEDIMIENTOS INTERNOS	
Análisis de viabilidad: estudio que suele realizar el Vicerrectorado correspondiente para medir el impacto en la docencia de los departamentos afectados y, en su caso, acuerdos con el centro promotor. El informe final debe ser favorable.	Octubre y Noviembre
Redactar un informe para la Comisión de Grado/Postgrado. Y realizar las tareas administrativas que ello conlleve en cada universidad.	Diciembre y Enero
Revisión de la propuesta en el Vicerrectorado correspondiente y, en el caso de máster, estudio por parte de una Comisión evaluadora nombrada al efecto.	Marzo
Redactar el informe de la Comisión de Grado/Máster. En caso de modificaciones se devolvería al órgano promotor para incluir las modificaciones a realizar.	Abril
Aprobación de la propuesta por el Consejo de Gobierno de la Universidad.	Mayo
Grabación de la propuesta en el aplicativo de la sede electrónica del Ministerio y envío a la Agencia Evaluadora.	Junio
PROCEDIMIENTOS EXTERNOS	
Evaluación del título por parte de la agencia evaluadora.	Julio-Enero
Resolución del Consejo de Universidades (del Consejo de Universidades se tramita al Ministerio para la aprobación del título).	Febrero
Tramitación del expediente para solicitar el informe preceptivo del Consejo Social que se eleva al Departamento de Educación del Gobierno Vasco.	Marzo
Autorización de implantación del título por parte de la administración correspondiente Gobierno Autónomo.	Junio
Se publica la autorización de las enseñanzas en el Boletín Oficial de la Autonomía correspondiente	Julio
Aprobación del título por el Ministerio (BOE) y publicación del plan de estudios en el BOE y Boletín Oficial Autónomo correspondiente.	Septiembre
PROCEDIMIENTOS INTERNOS DE CADA UNIVERSIDAD PARA LA PUESTA EN MARCHA	
Preparación de la información pública (web, folletos informativos, información centros de Bachillerato, impresos de preinscripción, ...etc.)	Octubre
Distintas tareas de gestión interna de cada universidad (procesos de preinscripción, selección de candidatos etc.)	Noviembre-Julio
Inicio de las enseñanzas.	Septiembre CURSO 2022-2022

Tabla 14. Ejemplo de calendario de propuesta y verificación de un título de grado.

Actuaciones para el inicio de un nuevo título de máster universitario para el curso 2022-2023	Previsión fechas
Redacción de la ficha de programación	15 noviembre 2020
Aprobación de la ficha de programación definitiva y petición de programación del título por parte del centro	15 noviembre 2020
Aprobación de la programación del curso 2019-2020 por la Comisión de Postgrado y Doctorado de la Universidad	27 noviembre 2020
Aprobación de la programación del curso 2019-2020 por el Consejo de Dirección de la Universidad	13 diciembre 2020
Fase 1: Introducción de las solicitudes de programación de estudios	31 diciembre 2020
Propuesta de aprobación de la programación universitaria y autorización para la verificación por parte de la Consejería de universidades	Enero 2021
Redacción de la memoria de verificación	Primera versión del documento el 8 enero 2021
Aprobación de la programación universitaria y autorización para la verificación por parte de la Consejería de universidades	Febrero 2021
Aprobación de la memoria de verificación definitiva por parte del Centro y petición de elevar el documento a los órganos de gobierno de la Universidad	7 Febrero 2021
Aprobación de la memoria de verificación definitiva por la Comisión de Postgrado y Doctorado de la Universidad	12 Febrero 2021
Aprobación de la memoria de verificación definitiva por el Consejo de Dirección de la Universidad	21 Febrero 2021
Fecha límite carga de la memoria en el aplicativo VERIFICA	15 Marzo 2021
Documentación fase 2, redacción	15 Mayo 2021
Fase 2: Introducción de los datos de implantación de los estudios	15 Septiembre 2021 a 30 Octubre 2021
Data límite Resolución favorable de verificación Másteres, del Consejo de Universidades.	Febrero 2022
Resolución de programación universitaria, Consejería de universidades	Abril 2022
Publicación en el Diario oficial	Mayo 2022
Inicio de la primera edición del máster	Septiembre 2022 CURSO 2022-2022

Tabla 15. Ejemplo de calendario de propuesta y verificación de un título de máster.

Cabe mencionar los títulos propios y diplomas de posgrado que a diferencia de un máster oficial que es el único que da acceso al tercer ciclo o estudios de doctorado, pueden ser ofertados también por las universidades y no han de ser aprobados por las agencias de verificación. Un máster propio tiene,

como el oficial, entre 60 y 120 créditos ECTS, y se puede cursar en uno o dos años. El Diploma de Postgrado es también un título propio para estudios profesionalizadores con un programa de entre 30 a 60 ECTS.

Los Diploma de posgrado son cursos de un elevado grado de especialización y/o actualización en una disciplina o ámbito muy concreto, más que el ámbito de un máster, y normalmente con una clara orientación profesionalizadora.

Normalmente estos últimos, además, pueden cursarlos estudiantes sin título de grado universitario (o equivalente) a condición de que en lugar recibir el título de diploma de postgrado reciban un certificado, que puede estar firmado por el rector o rectora.

Tanto los másteres propios como los diplomas de posgrado suelen ser interuniversitarios o también co-organizados con entidades especializadas.

La aprobación de los títulos propios tiene un reglamento en cada universidad, y normalmente se propone desde el centro a una comisión de doctorado y postgrado de la universidad en cuestión. Para los títulos propios, el precio también lo estipula la universidad (sujeto a aprobación del Consejo Social de la entidad) y en algunas universidades se exige que los estudios propios se autofinancien, satisfaciendo además el canon de servicios generales que se impone a las actividades que tienen lugar usando las infraestructuras universitarias.

4. CONCLUSIONES

En este documento se han detallado los pasos a seguir para implantar la propuesta de formación universitaria de grado y máster en PLN y TL presentada en el ET4. Los grados que se han propuesto en el ET4 deben someterse a un proceso de evaluación externa en diferentes momentos: una verificación; un seguimiento tras su implantación y una acreditación años después. El calendario de acreditación es oficial pero depende de la aprobación por parte de cada universidad, agencia acreditadora correspondiente y Comunidad Autónoma, y de la cantidad de correcciones que se soliciten por parte de cada estamento. Como existen varias agencias de acreditación (cada Comunidad Autónoma posee su propia agencia de acreditación) además de la ANECA (Agencia Nacional), los calendarios de implantación pueden variar en cada caso. Es por esto que a pesar de que hemos presentado unos calendarios para que sirvan de soporte a quien quisiera seguir el proceso de implantación de los grados propuestos en el ET4, estos deberán de ser ajustados en cada caso. Asimismo, otros títulos propios, de expertos, cursos de formación continua, o doctorados siguen también procesos de acreditación y aprobación similares a los ya detallados.

5. REFERENCIAS

- [1] ERASMUS MUNDUS LCT - European Masters Program in Language & Communication Technologies, <http://ixa2.si.ehu.es/erasmusmundus>
- [2] Curso de verano organizado por empresa <http://rtthss2019.vicomtech.org/>
- [3] Máster en procesamiento del lenguaje natural de la UNED http://portal.uned.es/portal/page?_pageid=93,61491281&_dad=portal&_schema=PORTAL&idAsignatura=24409117
- [4] Máster en procesamiento del lenguaje natural de la Universidad Politécnica de Cataluña <https://www.fib.upc.edu/es/investigacion/grupos-de-investigacion/gpln-grupo-de-procesamiento-del-lenguaje-natural>

6. GLOSARIO DE SIGLAS Y ACRÓNIMOS

AI	Artificial Intelligence
API	Interfaz de Programación de Aplicaciones (Application Programming Interfaze)
BD	Bases de Datos
DL	Aprendizaje Profundo (Deep Learning)
GV	Gestión de Versiones
ML	Aprendizaje Automático (Machine Learning)
MT	Traducción Automática (Machine Translation)
MOOC	Massive On-line Open course
NLP	Procesamiento del Lenguaje Natural (Natural Language Processing)
NN	Redes Neuronales (Neural Networks)
OO	Orientación a Objetos (Object Oriented)
SEAD	Secretaría de Estado para el Avance Digital
STEM	Ciencia, Tecnología, Ingeniería y Matemáticas (Science, Technology, Engineering, Mathematics)
TL	Tecnologías del Lenguaje
ITM	Redes de formación innovadora (Innovative Training Networks)
NLG	Generación de Lenguaje Natural (Natural Language Generation)
QA	Obtención de respuestas a preguntas (Question-Answering)
IPA	Alfabeto fonético internacional (International Phonetic Alphabet)
SAMPA	Speech Assessment Methods Phonetic Alphabet