

ANÁLISIS DE LA SITUACIÓN DEL SECTOR DE LAS TECNOLOGÍAS DEL LENGUAJE EN EUROPA

Plan de Impulso de las Tecnologías del Lenguaje (Plan TL)

Zabala Innovation Consulting, S.A.

Marzo 2018

INFORME FINAL



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 la Sociedad de la Información y la Agenda Digital y Red.es, que no comparten 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 las mismas.



Índice

1	Introducción	5
2	Resumen Ejecutivo	7
3	Mapa de instrumentos de financiación en Europa relacionados con las TL	9
3.1	Estrategias, políticas y programas de apoyo a las TL en 22 países europeos	10
3.2	Programas Europeos de financiación a las TLs	16
3.2.1	Visión general de programas europeos	16
3.2.2	Programas europeos que han financiado proyectos relacionados con las TL	19
3.3	Conclusión	31
4	Líneas de investigación, desarrollo e implantación de tecnologías innovadoras relacionadas con las TL financiadas en programas europeos	32
4.1	Líneas financiadas hasta 2017	32
4.2	Líneas a futuro	34
4.3	Conclusiones	36
5	Instrumentos de financiación europeos de interés para la ejecución del PLAN TL	38
5.1	Conclusiones	41
6	Herramientas, aplicativos y soluciones de tecnologías del lenguaje que se están utilizando sobre los que se está investigando en proyectos europeos	42
6.1	Conclusiones	45
7	Identificación de oportunidades dentro del marco de las estrategias europeas del Mercado Único Digital (DSM)	46
7.1	Concepto del Digital Single Market	46
7.2	¿Cómo apoya el DSM y sus correspondientes iniciativas/programas a las tecnologías del lenguaje?	48
7.2.1	Normas de Derecho de autor (copyright) modernas:	49
7.2.2	Economía de datos europea	51
7.2.3	Estándares e Interoperabilidad	53
7.2.4	Sociedad Digital	55



8	DIH (Digital Innovation Hubs) recomendaciones para el posicionamiento institucional de la SESIAD	57
8.1	DIH.....	58
8.1.1	Concepto	58
8.1.2	Miembros	59
8.1.3	Modelos de Negocio.....	60
8.1.4	Servicios.....	61
8.1.5	Indicadores	62
8.2	Listado de proyectos europeos y líneas del H2020 en los que aspectos relacionados con DIH han sido financiados o se pueden financiar	63
8.3	DIH en tecnologías del lenguaje	63
8.4	Los DIH como herramienta para el PLAN TL.....	65
8.5	Herramientas para apoyar la generación de DIH.....	68
9	Conclusiones.....	70
10	Glosario de siglas y acrónimos	79

1 Introducción

La Secretaría de Estado para la Sociedad de la Información y la Agenda Digital (en adelante SESIAD) como órgano superior del Ministerio de Energía, Turismo y Agenda Digital, es responsable de la ejecución y coordinación de la Agenda Digital para España (ADpE).

La Agenda Digital para España es la estrategia del Gobierno para desarrollar la economía y la sociedad digital en España. Para su puesta en marcha y ejecución, se han definido un conjunto de planes específicos entre los que se encuentra el Plan de Impulso a las Tecnologías del Lenguaje. Este Plan tiene como objetivo general desarrollar la industria del Procesamiento del Lenguaje Natural (PLN), Tecnologías del habla y sistemas conversacionales y la Traducción Automática (TA) en España, y especialmente en lengua española y lenguas cooficiales, y aprovechar estas novedosas capacidades para mejorar el servicio público. Para lograr sus objetivos, el Plan de Impulso de las Tecnologías del Lenguaje se ha estructurado en cuatro grandes ejes de actuación.

- Eje 1: Desarrollo de infraestructuras lingüísticas: Dirigido a aumentar el número, calidad y disponibilidad de las infraestructuras lingüísticas (recursos, procesadores y campañas de evaluación) de propósito general en español y lenguas cooficiales.
- Eje 2: Impulso de la industria de las Tecnologías del Lenguaje (TL): Para impulsar la Industria del lenguaje fomentando la transferencia de conocimiento entre el sector investigador y la industria. Incluye la ayuda a la internacionalización de las empresas e instituciones del sector y la mejora de la difusión de los proyectos en marchas.
- Eje 3: Las Administración Pública como impulsor de la industria del lenguaje: Dedicado a la mejora de la calidad y capacidad del servicio público incorporando las tecnologías de procesamiento de lenguaje natural y de la traducción automática, actuando, además, como tractor de la demanda. Apoya, también, la generación, estandarización y difusión de recursos lingüísticos creados en el contexto de la actividad de gestión pública propia de la Administración en el marco la política de Reutilización de la Información del Sector Público (RISP).
- Eje 4: Proyectos Faro: Se orienta a desarrollar nuevos servicios públicos, o a mejorar la capacidad y calidad de servicios públicos existentes y de procedimientos de las Administraciones Públicas mediante la aplicación de las tecnologías del lenguaje. Se trata de actuaciones en servicios públicos concretos de alto impacto social, que abarquen toda la cadena de valor y den lugar a productos y servicios acabados, con el fin de poner en valor las capacidades y beneficios de las tecnologías del procesamiento del lenguaje natural y la traducción automática. Estas actuaciones



servirán también para guiar el diseño y dar aplicación inmediata a las medidas horizontales del plan, como el desarrollo de infraestructuras lingüísticas, y, muy especialmente, a las plataformas comunes de procesamiento de lenguaje natural y traducción automática de las Administraciones Públicas.

En Europa, las tecnologías del lenguaje han sido objeto de interés y estudio dada la naturaleza multilingüe de la Unión Europea. Varias iniciativas han sido subvencionadas por fondos europeos con diferentes fines. Algunas iniciativas se han centrado en la labor de investigación y desarrollo de Tecnologías del Lenguaje o bien como finalidad, o bien como medio para otras aplicaciones y servicios. Otras iniciativas han tenido como objetivo coordinar esfuerzos y compartir recursos y buenas prácticas a través de diferentes plataformas y redes de excelencia. Es importante situar el ámbito de Tecnologías del Lenguaje español dentro del panorama europeo para identificar los agentes principales a nivel europeo, asociarse a nuevas iniciativas y conseguir una mayor participación española en este ámbito.

Para el desarrollo de parte de estos ejes, la Entidad Pública Empresarial Red.es y, en este caso, el ONTSI (Observatorio Nacional de las Telecomunicaciones y de la Sociedad de la Información) pone en marcha, dentro del desarrollo del Plan de Impulso a las Tecnologías del Lenguaje diversas actuaciones encaminadas a identificar y conocer mejor la situación del sector de las Tecnologías del Lenguaje.

Con este fin, se desarrolla el presente trabajo cuyo objetivo es realizar un estudio sobre la situación del sector de las tecnologías del lenguaje en Europa para identificar oportunidades en el ámbito del Plan de Impulso de las Tecnologías del Lenguaje (PLAN TL). Para ello, es necesario **comprender el estado actual del sector de las tecnologías del lenguaje en Europa: su desarrollo en los últimos años destacando las iniciativas y los mecanismos habilitados para su integración en la Agenda Digital Europea y el Mercado Único Digital.**

2 Resumen Ejecutivo

Con el objetivo de analizar la situación del sector de las tecnologías del lenguaje en Europa para identificar oportunidades en el ámbito del PLAN TL, se definieron una serie de actividades.

Primero, identificar la situación actual en cuanto a estrategias, políticas y programas de apoyo a las tecnologías del lenguaje en los países de la Unión Europea y desde la propia administración europea. Como consecuencia del análisis se identificaron por cada país los programas existentes y sus características. Asimismo, se identificaron los programas, las áreas y los proyectos de la Comisión Europea más relacionados con las Tecnologías del Lenguaje. De este análisis, se ha podido extraer conocimiento de la situación de las entidades españolas, y de España en el marco europeo, así como las entidades con más relevancia en los mismos y el tipo de instrumentos utilizados por la Comisión Europea, así como las líneas de investigación de más interés. Todo ello se recoge en este documento en las secciones: 3, 4, 5 y 6. Dichas secciones se apoyan en datos incluidos en el documento de apoyo.

Posteriormente, se pasa a analizar la identificación de oportunidades dentro del marco de las estrategias europeas del *Digital Single Market* (DSM) para el desarrollo del Plan TL así como estudiar la utilidad de los *Digital Innovation Hubs* (DIH) para el desarrollo del Plan TL. Dicho análisis se recoge en las secciones 7 y 8.

Finalmente, en la sección 10, con la información analizada, se presentan una serie de conclusiones agrupadas en los siguientes puntos:

- El peso de las TL dentro de los programas de Tecnologías de la Información y las Comunicaciones (TIC) europeos. Áreas financiadas y a financiar de las TL. Los programas más destacables son: 7PM, CIP-ICT-PSP, H2020 y CEF TELECOM.
- Importancia de las TL en la estrategia europea del Mercado Único Digital, principalmente en 4 de sus iniciativas: 1) Un marco de derechos de autor (Copyright) moderno y más europeo, 2) Abordar las barreras de la economía de datos europea, 3) Definir prioridades para los estándares y la interoperabilidad y 4) Apoyar una sociedad digital inclusiva.
- Instrumentos utilizados por los programas europeos, principalmente el H2020, para facilitar el apoyo a proyectos con resultados de impacto en el mercado, como el Instrumento PYMES, FTI y la financiación en cascada.
- Aplicación sectorial de los proyectos de TL.



- Interés de los Países Europeos en las TL: no existe actualmente ningún país en Europa con una estrategia tan clara como España, en relación a las TL.
- Situación de España en los Programas Europeos relacionados con las TL. La participación española en proyectos directamente relacionados con TL es bastante elevada tanto desde el punto de vista de entidades de investigación como por empresas.

3 Mapa de instrumentos de financiación en Europa relacionados con las TL

En esta sección, se presenta el resultado del análisis, realizado en 22 países europeos, de las estrategias, programas y oportunidades relacionados con las tecnologías del lenguaje (TL) y su desarrollo futuro.

El estudio comienza por un análisis de las oportunidades de financiación de proyectos relacionados con las TL a nivel nacional y regional. Posteriormente, se han analizado las iniciativas Eureka y Eurostars para después ampliar el ámbito del estudio a las convocatorias de la Comisión Europea.

Eureka: es una iniciativa intergubernamental de apoyo a la I+D+i cooperativa en el ámbito europeo, que tiene como objeto impulsar la competitividad de las empresas europeas mediante el fomento de la realización de proyectos tecnológicos, orientados al desarrollo de productos, procesos y/o servicios con claro interés comercial en el mercado internacional y basados en tecnologías innovadoras. Está dirigido a cualquier empresa o centro de investigación español capaz de realizar un proyecto de I+D+i de carácter aplicado en colaboración con, al menos, una empresa y/o centro de investigación de otro país de la red Eureka.

Cada país asume la financiación de sus entidades participantes. Eureka avala los proyectos aprobados mediante un "sello de calidad" que, además de ser un elemento promocional y de reconocimiento del nivel tecnológico de la compañía promotora, la hace acreedora de una financiación pública, que en el caso de España es especialmente favorable.

En EUREKA no existen líneas tecnológicas predeterminadas. Todas las tecnologías tienen cabida, siempre que tengan un carácter innovador.

Eurostars: es un programa de apoyo las PYMES intensivas en I+D en el desarrollo de proyectos transnacionales orientados al mercado. Está orientado a la generación de proyectos de I+D en consorcio liderados por PYMES intensivas en I+D. Dichos proyectos deben representar una ruptura con el estado del arte técnico y un reto comercial tales que permitan a estas empresas dar un salto cualitativo importante en su posición en el mercado. Las entidades participantes en los proyectos EUROSTARS reciben las ayudas a nivel nacional por medio de sus Agencias financiadoras. Los costes elegibles y los requisitos para su concesión son definidos en cada país por dichas Agencias.

A nivel regional y/o nacional resulta prácticamente imposible obtener información acerca de las actuaciones/proyectos concretos que hayan contado con algún tipo de apoyo financiero por parte de la



administración para su puesta en marcha. Sin embargo, a nivel de la administración europea sí que ha sido posible extraer datos. Así, el estudio incluye también un análisis de los resultados en los diferentes programas a nivel europeo (7º Programa Marco – FP7, Horizonte 2020 -H2020, CEF Telecom, INTERREG, EUROSTARS, EUREKA, etc.), incluyéndose unas tablas resumen comparativas.

De la información previa, se puede concluir que, en Europa, **las tecnologías del lenguaje han sido objeto de interés y estudio dada la naturaleza multilingüe de la Unión Europea**. Varias iniciativas han sido subvencionadas por fondos europeos con diferentes fines. De este modo, mientras que algunos proyectos se han centrado en la labor de investigación y desarrollo de Tecnologías del Lenguaje, en otros casos, el apoyo ha sido para la aplicación de las TL y desarrollo de nuevos servicios, aplicaciones o procesos. Igualmente, otras iniciativas han tenido como objetivo coordinar esfuerzos y compartir recursos y buenas prácticas a través de diferentes plataformas y redes de excelencia. No obstante, si bien es cierto que las TL son muy importantes y han recibido apoyo tanto a nivel europeo como nacional, **el análisis realizado refleja como todavía queda camino por recorrer**.

3.1 Estrategias, políticas y programas de apoyo a las TL en 22 países europeos

Tal y como se ha indicado anteriormente, se han analizado las estrategias, políticas y programas de apoyo a las TL en 22 países. Para ello, se ha completado por cada país y programa la siguiente ficha.

Nombre de la entidad responsable	
Dirección	
URL	
Nombre del programa/convocatoria	
Objetivos del programa/convocatoria	
Persona de contacto	
e-mail/teléfono	
Perfil de los solicitantes. Ej: PYMES, Grandes Empresas, entidades sin ánimo de lucro, etc.	
Tipo de proyecto: individual o cooperativo. (En caso de proyectos cooperativos, condiciones del consorcio)	
Ejes prioritarios o Temáticas	
Duración del proyecto	
Fecha límite presentación solicitudes	
Presupuesto proyecto	
Gastos elegibles	

Tipo de financiación	
Enlaces a manuales, documentación.	

TABLA 1. FICHA PARA DESCRIBIR PROGRAMAS RELACIONADOS CON LAS TL

Una lista de todas las fichas de programas identificados en los 22 países, se encuentra en el Anexo 1 del Documento de Anexos. Cabe destacar que esta información está en continuo cambio, por lo que hay que tener en cuenta la fecha de elaboración del documento (diciembre 2017).

De la información recogida en las fichas por países, lo que se puede concluir es que si bien **algunos países sí que definieron en el pasado estrategias para fomentar el impulso de las tecnologías del lenguaje** (Figura 1), éstas no se han traducido (salvo en casos excepcionales, como puede ser Estonia o Francia) en la puesta en marcha de iniciativas/convocatorias de ayuda específicas para impulsar el desarrollo de proyectos o actuaciones en el área. De hecho, **a fecha 21/12/2017 no hay ninguna convocatoria específica para TL abierta.**

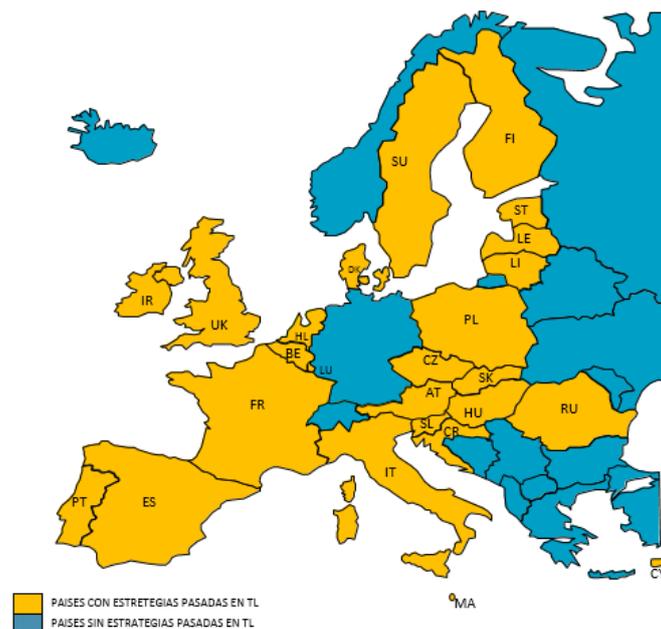


FIGURA 1 MAPA DE PAÍSES CON ESTRATEGIAS PASADAS EN TECNOLOGÍAS DEL LENGUAJE

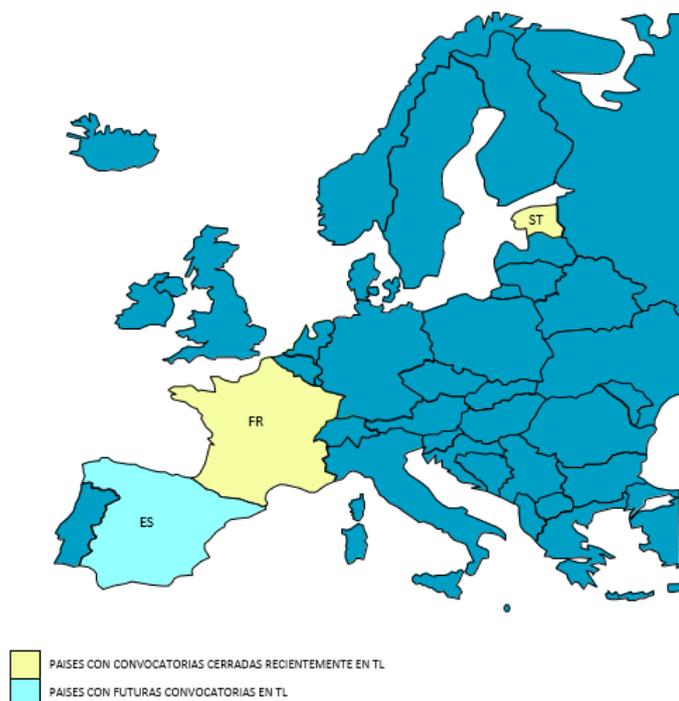


FIGURA 2 MAPA DE PAÍSES CON CONVOCATORIAS DE AYUDAS ACTUALES O RECIENTES EN TECNOLOGÍAS DEL LENGUAJE

Así pues, existen o han existido muy pocas convocatorias nacionales específicamente destinadas a financiar proyectos relacionados con las TL.

En el caso de estos países que han apoyado explícitamente las TL (Estonia y Francia), en la tabla siguiente se ofrece la información obtenida respecto a qué tipo de apoyo han dado, a qué tipo de TL y cuales han sido los proyectos seleccionados.



PAÍS	INSTRUMENTO NACIONAL	FINANCIA TL EXPLÍCITAMENTE
Estonia	Ministry of Education and Research/NPELT → <i>Programa Nacional de Tecnología del Idioma de Estonia (NPELT)</i>	TL en general, desde la compilación de recursos lingüísticos hasta la creación de prototipos de aplicaciones de tecnología del lenguaje.
Proyectos financiados	<ul style="list-style-type: none"> • UseMT: Usable Estonian Machine Translation • Toolkit for Acquiring Knowledge from Freetext Datasets • Tools for Lexical Resources • Speech Synthesis Improvements and Extension of Applications • Language hotline • Linguistic Knowledge in Estonian Machine Translation • EstNLTK - Open source tools for Estonian natural language processing • Statistical Models of the Emotionality of Speech and Written Text • Interfaces for Speech Synthesis • Modification of Lexicographer's Workbench • Prototype of Audiovisual Speech Synthesis • Speech Recognition • Speech Recognition 2 • New Resources in Machine Translation • Resources and Tools for Estonian Semantics • Template-based Mining of Facts from a Text Corpus • Analyser of Estonian Dialogue Pragmatics • Tools for Multi-layered Annotation of Text (applied to Estonian Reference Corpus) • Targeted Syntax for Corpora • Development of the Phonetic Corpus of Estonian Spontaneous Speech • Development of the Phonetic Corpus of Estonian Spontaneous Speech 2 • Compilation of Image-based Electronic Dictionary for Radiology Using an Estonian Corpus of Authentic Medical Texts • Compilation of the Audiovisual Discourse Corpus of Spoken Estonian and Development of a Query System • Võru and Seto Language Corpus • Võru and Seto Integrated Language Resources • Estonian Open Parallel Corpus • Estonian Open Parallel Corpus Follow-up Project 	



	<ul style="list-style-type: none"> • Extension of Estonian Wordnet • Extension of Estonian Wordnet 2 • Speech and Multimodal Corpora • Development of Speech Corpora • Estonian-French Parallel Corpus • Center of Estonian Language Resources • Application Suite for voicing and broadcasting subtitles on television • Framework for the Development of Estonian Dialogue Systems • Heliraamatute genereerija ja Digari helindamisliides • Addition of speech recognition tool to kõneravi.ee environment and the development of the templates to provide personalised speech therapy • Open Source Morphology Software 	
País	Instrumento nacional	Financia TL explícitamente
Francia	Agencia Nacional Francesa de Investigación - ANR → <i>Trans-Atlantic Platform Call for Proposals "Digging into Data"</i>	Ciencias de software y tecnología Datos, conocimientos y big data – contenido multimedia (Procesamiento del Lenguaje Natural)
Proyectos financiados	<ul style="list-style-type: none"> • ACLEW - Analyzing Child Language Experiences Around the World • OPILANGDATA - Understanding opinion and language dynamics using massive data • Machine Translation and Automated Analysis of Cuneiform Languages (MTAAC) • SPeech Across Dialects of English (SPADE): large-scale digital analysis of a spoken language across space and time • Cleaning, Organizing, and Uniting Linguistic Databases (the COULD project) • Digging into signs: Developing standard annotation practices for cross-linguistic quantitative analysis of sign language data • Trees and Tweets: Mining Billions to Understand Human Migration and Regional Linguistic Variation • Automating Data Extraction from Chinese Texts • Harvesting Speech Datasets for Linguistic Research on the Web • Mining a Year of Speech • Towards Dynamic Variorum Editions 	

No obstante, la financiación y el apoyo al desarrollo de TL en Europa, tanto a nivel nacional como regional, viene a través de convocatorias genéricas de desarrollo de tecnología, y no tanto de convocatorias específicas.

En este sentido, en el Anexo 1 se ofrece de manera detallada las características más relevantes de estas convocatorias en cuanto a entidades gestoras, contactos, tipología de ayuda, beneficiarios, prioridades temáticas, vigencia, enlaces de interés, etc. Con miras a que de una manera visual sea posible ver cuáles de estas convocatorias están en vigor a la fecha del informe, nuevamente se ha empleado un código de colores (verde = abierto, rojo = cerrado) en las tablas resumen de cada una de ellas.

Más concretamente, a nivel nacional en toda Europa se han analizado 67 convocatorias, mientras que a nivel regional 48 en toda Europa (a fecha diciembre de 2017).



FIGURA 3 ESTADO (ABIERTAS O CERRADAS) DE LAS CONVOCATORIAS NACIONALES

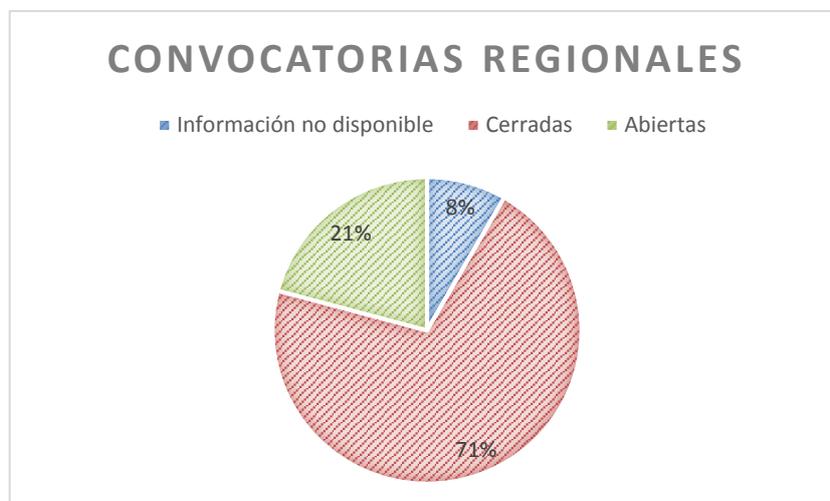


FIGURA 4 ESTADO (ABIERTAS O CERRADAS) DE LAS CONVOCATORIAS REGIONALES

3.2 Programas Europeos de financiación a las TLs

Antes de proceder a la descripción de los programas europeos y los proyectos financiados por dichos programas, vamos a hacer una breve introducción a los programas europeos y sus instrumentos de financiación. Describir cada programa en detalle sería muy amplio, por lo que se va a proporcionar una información genérica que ayude a entender mejor los siguientes puntos:

3.2.1 Visión general de programas europeos

De una forma muy sencilla, los proyectos europeos se podrían clasificar por:

- El tipo de proyectos que financian: más cercanos a la I+D o más cercanos al mercado.
- Por el grado de cooperación que requieren: proyectos individuales o en cooperación (varios socios).

Atendiendo a esta clasificación. Los programas europeos, más relacionados con este estudio se podrían clasificar de la siguiente manera:

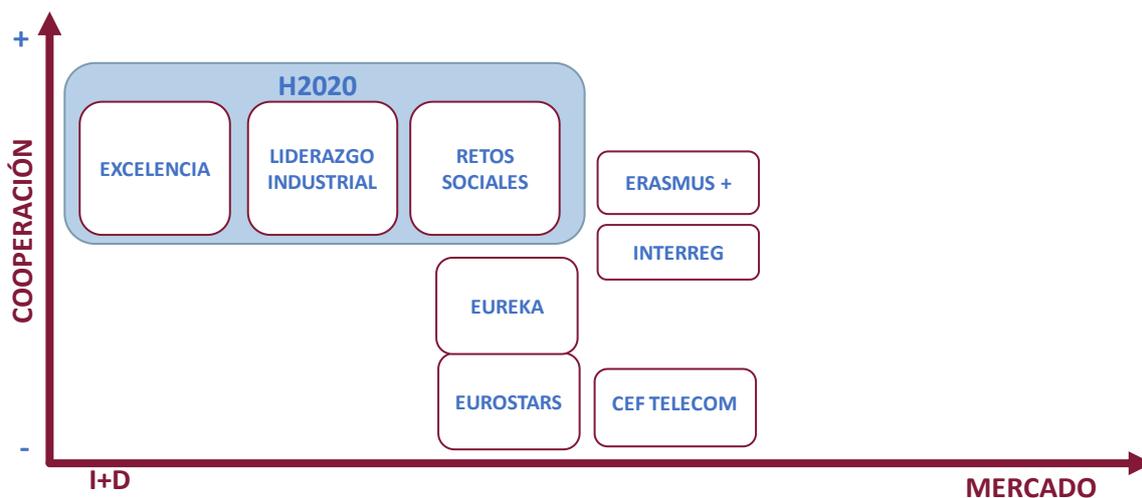


FIGURA 5. PROGRAMAS EUROPEOS DE FINANCIACIÓN

El **H2020** es el programa que toma el testigo al 7PM, y es el programa europeo de financiación a la I+D+I más grande. Este programa se divide en los siguientes bloques y sub-programas:



FIGURA 6. BLOQUES Y SUBPROGRAMAS DEL H2020

En el bloque de:

- **Ciencia Excelente:** se financian proyectos de investigación más básica, becas para investigadores, etc.
- **Retos sociales:** se financian proyectos en cooperación de investigación y desarrollo orientados a resolver retos sociales.
- **Liderazgo industrial:** se financian proyectos de investigación y desarrollo en tecnologías cuya aplicación es más transversal.

Financian diferentes tipos de proyectos en los bloques anteriores:

- **RIA (*Research and Innovation Action*):** proyectos de investigación y desarrollo en cooperación (más de 3 entidades de 3 países de la CE diferentes).
- **IA (*Innovation Actions*):** proyectos de desarrollo y validación de tecnologías novedosas en cooperación (más de 3 entidades de 3 países de la CE diferentes).
- **CSA (*Coordination and Support Actions*):** proyectos de coordinación o de soporte a las actividades de investigación y desarrollo.
- **Instrumento Pyme:** proyectos que apoyan el crecimiento de las empresas (exclusivamente PYMES) con ideas interesantes e innovadoras con potencial europeo o mundial, y que van a tener un impacto en la empresa en términos de crecimiento (ventas, plantilla, etc.) muy importante.
 - Para ello, el instrumento se estructura en 3 fases que abarcan las diferentes etapas del ciclo de innovación en la PYME (Fase I: concepto y evaluación de la viabilidad, Fase II I+D, demostración y replicación comercial, y Fase III Comercialización).



- **Fast track to Innovation:** son proyectos *bottom-up* y sin restricción temática y/o aplicación, pretende apoyar a las iniciativas de innovación cercanas al mercado. Para ello, y con el objetivo de fomentar también la participación entre el sector privado y las entidades investigadoras, se deben constituir consorcios de 3-5 socios, que tienen que contar como mínimo con un 60% de presupuesto de entidades industriales y pertenecer al menos a 3 estados miembro (o países asociados al H2020).

El H2020 no es un programa *bottom up*, es decir hay una serie de tópicos/líneas definidas a las que tienen que responder los proyectos. No se pueden presentar proyectos en cualquier temática.

EUREKA: Proyectos de cooperación internacional de Investigación y Desarrollo de cualquier sector o temática, de carácter aplicado para la creación y mejora significativa de un proceso productivo, producto o servicio. (Que genere un impacto en el mercado al finalizarlo)

EUROSTARS: Proyectos de I+D de cooperación internacional liderados por PYMES intensivas en I+D. Dichos proyectos deben representar una ruptura con el estado del arte técnico y un reto comercial, tales que permitan a estas empresas dar un salto cualitativo importante en su posición en el mercado.

ERASMUS + es el programa europeo en los ámbitos de la educación, la formación, la juventud y el deporte para el periodo 2014-2020. Erasmus+ es el programa único que trata de impulsar las perspectivas laborales y el desarrollo personal, además de ayudar a nuestros sistemas de educación, formación y juventud a proporcionar una enseñanza y un aprendizaje que doten a las personas de las capacidades necesarias para el mercado laboral y la sociedad actual y futura.

CEF TELECOM: es un programa orientado a facilitar la interacción transfronteriza entre administraciones públicas, empresas y ciudadanos, mediante el desarrollo de Infraestructuras de servicios digitales y redes de banda ancha. Apoya proyectos que contribuyan a la creación de un ecosistema europeo de servicios digitales interconectados e interoperables que sustenten el Mercado Único Digital.

INTERREG: programa de apoyo a la cooperación entre regiones europeas, con el fin implementar las políticas de cohesión social y mejorar la cohesión regional en Europa.

Adicionalmente en este documento se menciona el programa **CIP-ICT-PSP (2007 – 2013)**, el cual ya no está en marcha. The *ICT Policy Support Programme* (o CIP ICT PSP) tiene como objetivo, estimular la innovación y la competitividad a través de la puesta en marcha y el uso de las TICs, por los ciudadanos, administraciones y empresas. Fueron proyectos de aplicación y validación de tecnologías novedosas.

3.2.2 Programas europeos que han financiado proyectos relacionados con las TL

Los programas europeos que han financiado proyectos relacionados con las TL, están, normalmente, muy relacionados con las Tecnologías de la Información y las Comunicaciones (TICs). A modo de resumen, las siguientes tablas contienen los porcentajes de subvención concedida y de proyectos realizados.

Cabe destacar que para los programas Eurostars y Eureka no ha sido posible obtener los datos de financiación de los diferentes proyectos tras obtener el sello, por lo que en este caso se analizarán solamente el número de proyectos relacionados con las tecnologías del lenguaje, no las ayudas concedidas.

PROGRAMA	PRESUPUESTO DEL PROGRAMA (€)	AYUDA CONCEDIDA PROYECTOS (€)	% AYUDA FRENTE PRESUPUESTO	TL AL	Nº TOTAL PROYECTOS PROGRAMA	Nº PROYECTOS SUBVENCION TL	% PROYECTOS TL EN CADA PROGRAMA
FP7 - ICT	9.050.000.000	164.656.560	1,8%		2281	63	2,8%
H2020 - ICT	7.711.000.000	50.424.401	0,7%		1481	15	1,0%
CIP ICT PSP	3.621.000.000	28.494.024	0,8%		233	16	6,9%
CEF TELECOM	300.000.000	7.595.988	2,5%		256	10	3,9%

TABLA 2. PRESUPUESTO Y Nº DE PROYECTOS TL POR PROGRAMA

PROGRAMA	Nº TOTAL PROYECTOS APROBADOS	Nº PROYECTOS SUBVENCIONADOS TL	%PROYECTOS TL
EUROSTARS	1418	2	0,14%
EUREKA	3742	3	0,08%

TABLA 3. Nº DE PROYECTOS DE TL VS Nº PROYECTOS TOTAL EN LOS PROGRAMAS EUREKA Y EUROSTARS

Un listado de los proyectos identificados, tanto en el 7PM, CIP-ICT-PSP, H2020, CEF TELECOM, Eureka y Eurostars, se encuentran en el Anexo 2 (Documento de Apoyo).

A continuación, se van a presentar estos datos extrapolados a gráficos, para poder visualizar estas cifras con mayor facilidad y constatar de manera sencilla la evolución de las tecnologías del lenguaje a través de los diferentes programas estudiados. Los datos se han obtenido de la información publicada por el CORDIS (Community Research and Development Information Service) e INEA (Innovations and Network Executive Agency) hasta la fecha.

En primer lugar, se muestra el **programa FP7**, más concretamente en el Programa de Trabajo de ICT (*Information Communication and Technologies*), dónde se financiaron proyectos relacionados con las tecnologías del lenguaje por un valor del 1,8%, en el periodo de años desde el 2007 hasta el 2013.

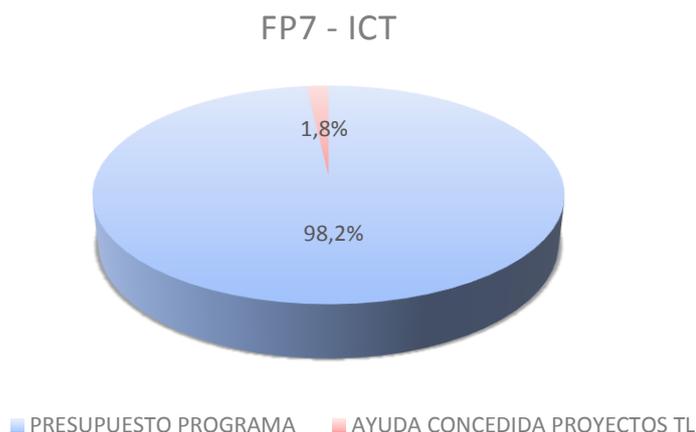


FIGURA 7 PORCENTAJE DE AYUDA TL FRENTE AL PRESUPUESTO DEL PROGRAMA - FP7

El **Horizon 2020** tomó el testigo del FP7. Cabe destacar que la información obtenida pertenece a las convocatorias de las que se dispone de información (publicada en CORDIS) hasta la fecha. El programa sigue vigente hasta 2020.

De esta manera, hasta la fecha, se han financiado proyectos relacionados con las tecnologías del lenguaje por un valor del 0,7%, una cifra bastante por debajo del 1,8% del anterior programa marco.

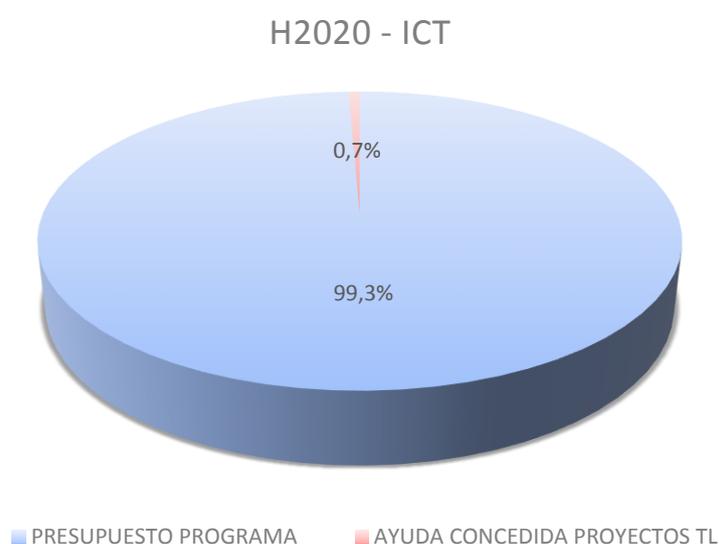


FIGURA 8 PORCENTAJE DE AYUDA TL FRENTE AL PRESUPUESTO DEL PROGRAMA - H2020 ICT

El programa **CIP PSP** (área de ICT) se desarrolló entre los años 2007 y 2013, paralelante al 7PM y contó con un presupuesto más reducido que los anteriores (3.621 M€). En este caso, los proyectos de tecnologías del lenguaje que se financiaron supusieron un 0,8% del presupuesto total del programa, una cifra que dista de los 1,8% conseguidos en el FP7 pero que queda muy emparejado con el conseguido en el programa H2020.

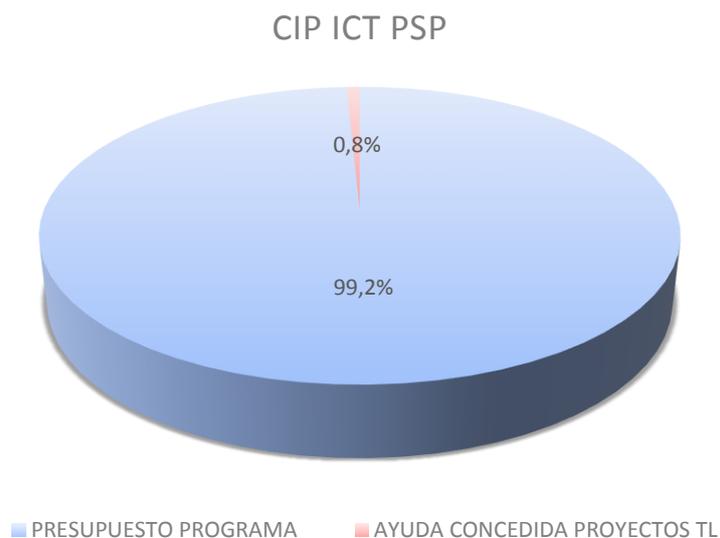


FIGURA 9 PORCENTAJE DE AYUDA TL FRENTE AL PRESUPUESTO DEL PROGRAMA - CIP ICT PSP

Finalmente, **el programa que más éxito ha tenido en términos de porcentaje de proyectos financiados de las tecnologías del lenguaje hasta la fecha ha sido el CEF Telecom, en el que los proyectos TL suponen un 2,5% del presupuesto del programa.** Si bien, cabe destacar que el presupuesto disponible para este programa está muy por debajo del resto de programas.

Aún y todo, el programa seguirá operativo hasta 2020, por lo que todavía esas cifras pueden variar, lo mismo que pasa con el programa H2020.

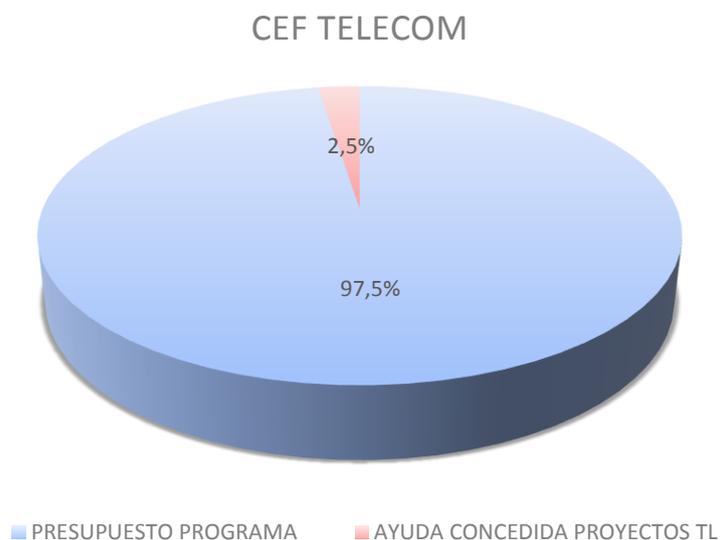


FIGURA 10 PORCENTAJE DE AYUDA TL FRENTE AL PRESUPUESTO DEL PROGRAMA - CEF TELECOM

7PM

Las siguientes imágenes representan la correlación de las entidades colaboradoras en los proyectos de tecnologías del lenguaje en el 7PM. Así pues, se distinguen a las **universidades e institutos tecnológicos y de investigación** como los principales valedores de este tipo de proyectos, siendo *la Deutsches Forschungszentrum fuer Kuenstliche Intelligenz GmbH* y *la Universidad de Sheffield*, los participantes con mayor número de proyectos (9), seguidos de *la Fondazione Bruno Kessler* (con 7 proyectos)

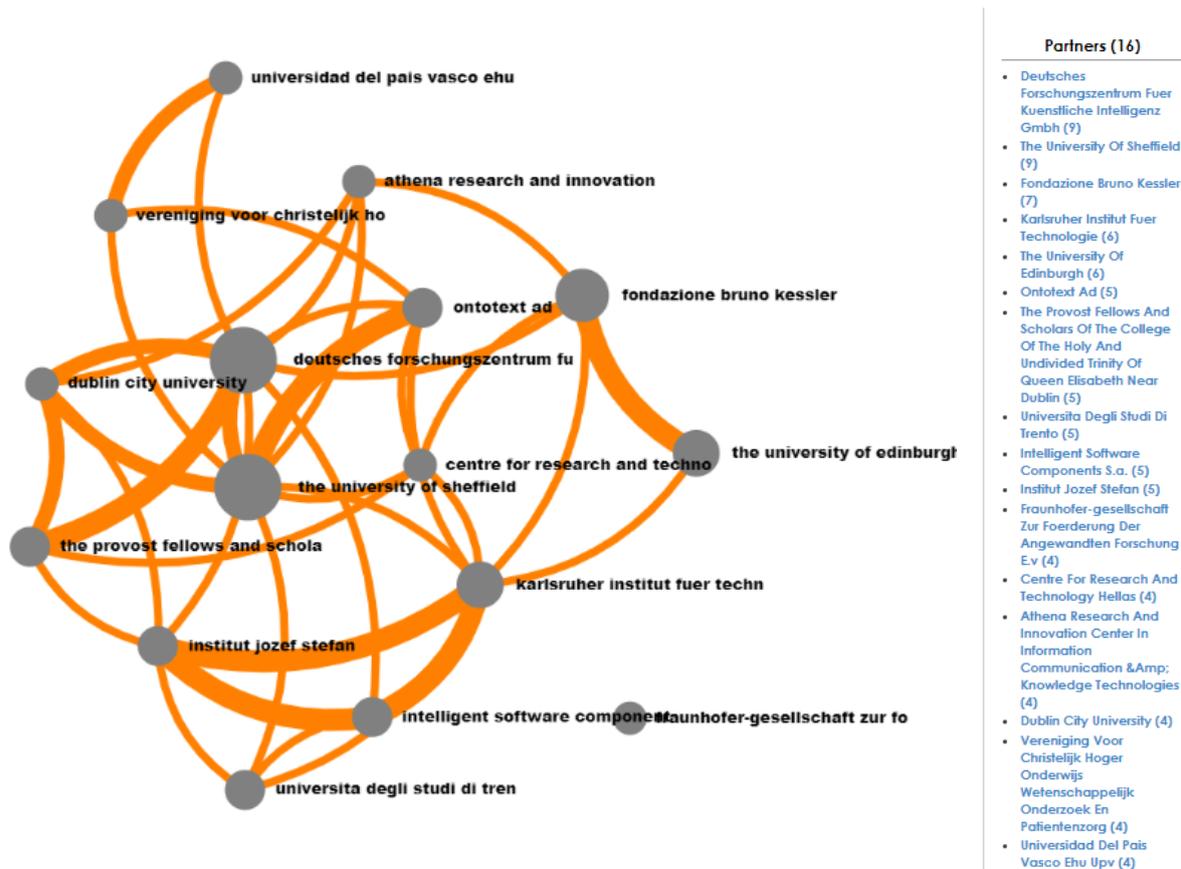


FIGURA 11 RELACIÓN ENTRE ENTIDADES PARTICIPANTES - FP7

Llegados a este punto, resulta de vital importancia conocer la procedencia de las diversas organizaciones que forman parte de los proyectos para saber **qué países apuestan por invertir más fuerte** en las **Tecnologías del Lenguaje**. Para ello, se va a diferenciar entre países que **participan como coordinadores** del proyecto y países que **participan como socios**, con el fin de obtener los países que impulsan con mayor fuerza el desarrollo y/o utilización de estas tecnologías.

La *intensidad de colaboración de cada país*, en forma de **número de socios** que aporta a los proyectos, se va a mostrar en el siguiente mapa con su correspondiente leyenda con código de colores. De esta manera, se puede resaltar cómo *Alemania* es el país que más socios aporta con más de 71, seguido de

España, que participa en el rango correspondiente a 51-60 socios. Países como *Italia* y *Reino Unido* están en tercer lugar con una cifra que va desde los 41 hasta los 50 socios.

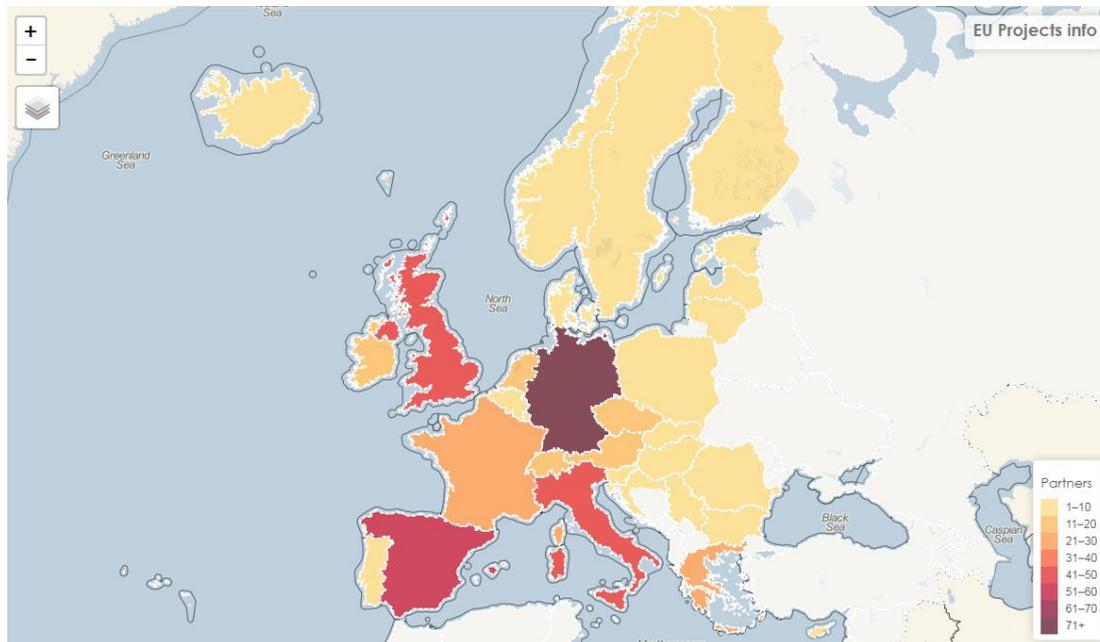


FIGURA 12 INTENSIDAD DE PARTICIPACIÓN DE CADA PAÍS EN FUNCIÓN DEL NÚMERO DE SOCIOS - FP7

En lo que concierne a las *entidades coordinadoras*, es decir, aquellos **países que traccionan los proyectos del área**, se tiene que *Alemania* y *España* son los países de procedencia que más veces actúan como líderes en los proyectos de Tecnologías del Lenguaje financiados por el programa FP7, con una participación de entre 11 y 12 proyectos, seguidos de *Italia* y *Reino Unido*, con de 7 a 8 proyectos.

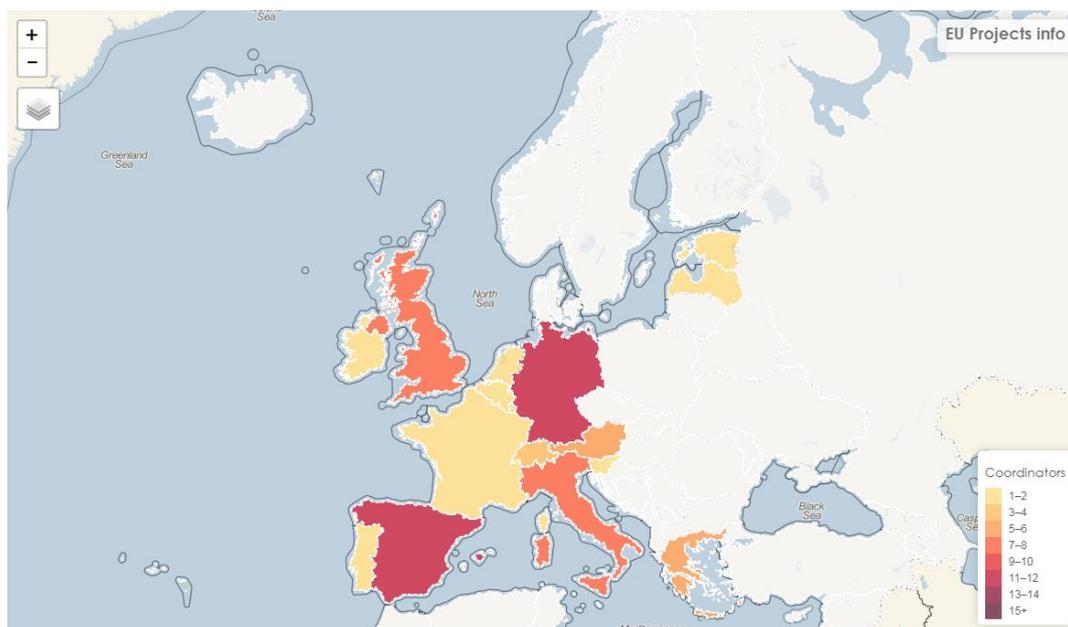


FIGURA 13 INTENSIDAD DE PARTICIPACIÓN DE CADA PAÍS EN FUNCIÓN DEL PAÍS COORDINADOR - FP7

H2020

El grafo siguiente muestra una relación entre las entidades que han participado en los proyectos financiados por el H2020 en TL, hasta diciembre de 2017. Se observa como las **universidades y centros de investigación** son aquellas organizaciones con una mayor participación en este tipo de proyectos, siendo la *Universidad de Edimburgo* la que participa en un mayor número de proyectos (6), seguida de la *Univerzita Karlova V Praze* y la *Deutches Forschungszentrum fuer Kienstliche Intelligenz GmbH* (con 4 proyectos cada una)

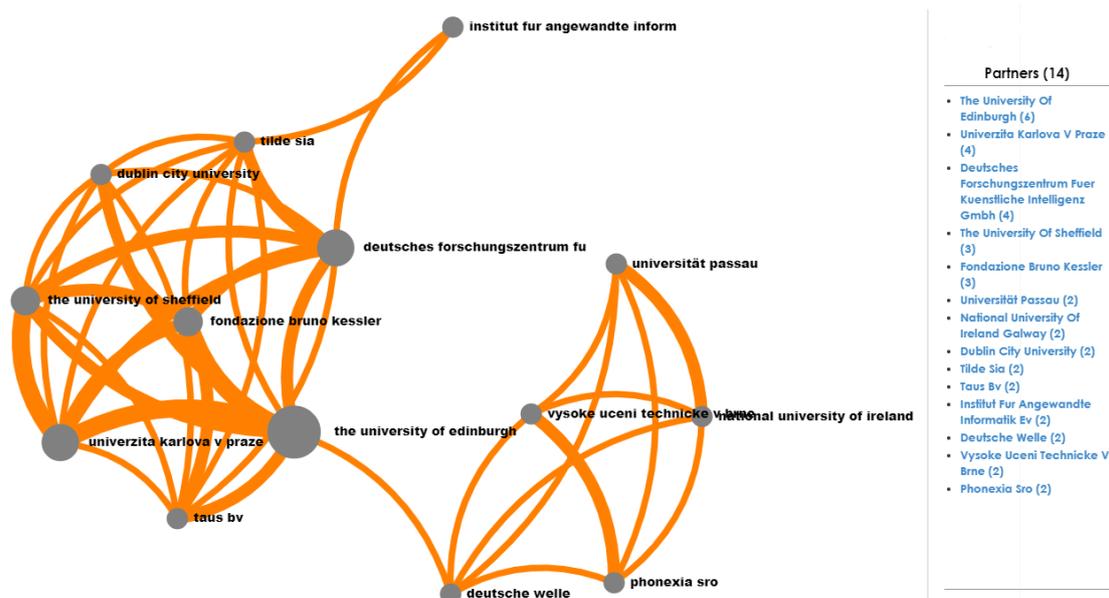


FIGURA 14 RELACIÓN ENTRE ENTIDADES PARTICIPANTES - H2020

En cuanto a los **países que apuestan más fuerte** por invertir en las **Tecnologías del Lenguaje**, al igual que las secciones anteriores, en el mapa siguiente se muestra (en base a un código de colores) la *intensidad de colaboración de cada país*, en forma de **número de socios** que aporta a los proyectos.

Así, se puede apreciar cómo el *Reino Unido* es el país que más socios aporta con más de 22, seguido de *Alemania*, que participa con de 16 a 18 socios. Países como *Italia*, *Grecia* o la *República Checa* completan el pódium con una cifra que va desde los 10 hasta los 12 socios.

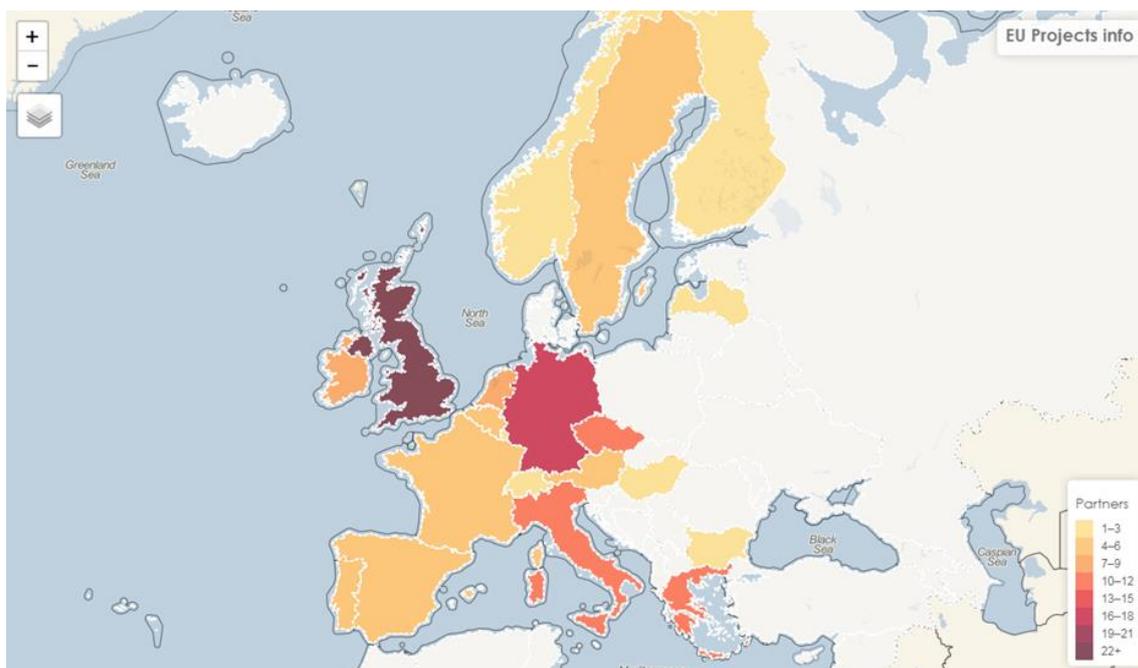


FIGURA 15 INTENSIDAD DE PARTICIPACIÓN DE CADA PAÍS EN FUNCIÓN DEL NÚMERO DE SOCIOS -H2020

Pensando en las *entidades coordinadoras*, es decir, aquellos **países que traccionan los proyectos del área**, se tiene que **Alemania** es el país de procedencia que más veces actúa como líder en los proyectos de Tecnologías del Lenguaje financiados por el Horizon 2020 hasta la fecha en hasta **5 ocasiones**.

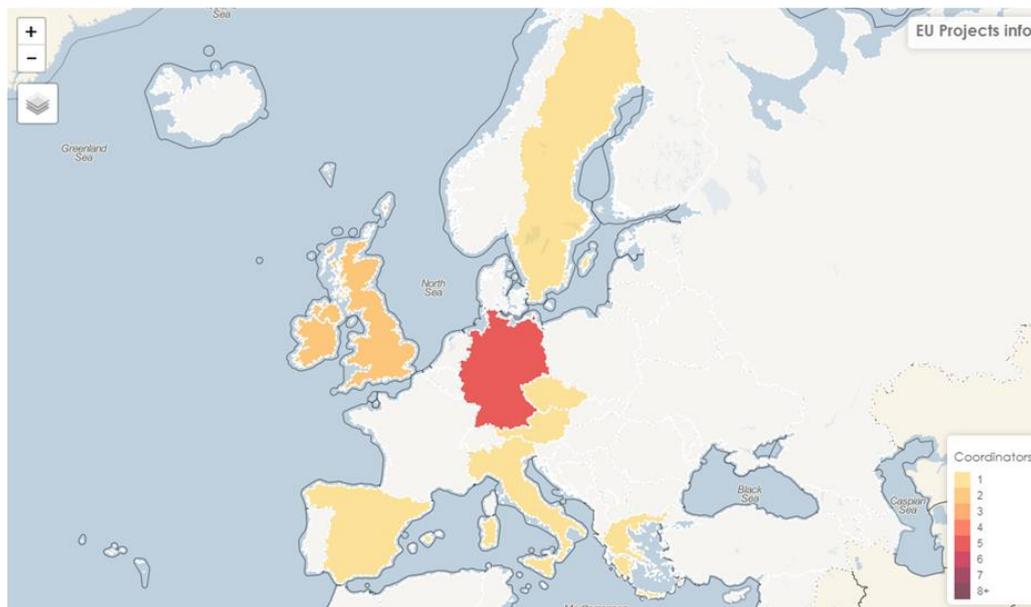


FIGURA 16 INTENSIDAD DE PARTICIPACIÓN DE CADA PAÍS EN FUNCIÓN DEL PAÍS COORDINADOR - H2020

Dentro del H2020, existe un instrumento, llamado **Instrumento Pyme**, que busca la financiación de PYMES que tienen un producto/servicio innovador con un *Technology Readiness Level* (TRL) o nivel de madurez de la tecnología alto, entre 6-7 (TRL 6: Modelo de sistema o subsistema o demostración de prototipo en un entorno relevante; TRL 7: Demostración de sistema o prototipo en un entorno real) y cuyo lanzamiento al mercado tiene previsto un altísimo impacto en el mercado internacional. Este



instrumento no existía en el 7PM. Otras de sus características, es que es abierto, “bottom up”, no se limita a temáticas concretas, sino a las siguientes grandes áreas, según su denominación en inglés:

- *SMEInst-02-2016-2017: Accelerating the uptake of nanotechnologies advanced materials or advanced manufacturing and processing technologies by SMEs*
- *SMEInst-02-2016-2017: Accelerating the uptake of nanotechnologies advanced materials or advanced manufacturing and processing technologies by SMEs*
- *SMEInst-03-2016-2017: Dedicated support to biotechnology SMEs closing the gap from lab to market*
- *SMEInst-04-2016-2017: Engaging SMEs in space research and development*
- *SMEInst-05-2016-2017: Supporting innovative SMEs in the healthcare biotechnology sector*
- *SMEInst-06-2016-2017: Accelerating market introduction of ICT solutions for Health, Well-Being and Ageing Well.*
- *SMEInst-07-2016-2017: Stimulating the innovation potential of SMEs for sustainable and competitive agriculture, forestry, agri-food and bio-based sectors*
- *SMEInst-08-2016-2017: Supporting SMEs efforts for the development - deployment and market replication of innovative solutions for blue growth*
- *SMEInst-09-2016-2017: Stimulating the innovation potential of SMEs for a low carbon and efficient energy system*
- *SMEInst-10-2016-2017: Small business innovation research for Transport and Smart Cities Mobility*
- *SMEInst-11-2016-2017: Boosting the potential of small businesses in the areas of climate action, environment, resource efficiency and raw materials*
- *SMEInst-12-2016-2017: New business models for inclusive, innovative and reflective societies*
- *SMEInst-13-2016-2017: Engaging SMEs in security research and development*

En las convocatorias del 2018 – 2020, estas áreas ya no existen y es totalmente abierto.

Adicionalmente, en el Instrumento Pyme, existen dos tipologías de proyectos.

- Los proyectos Fase 1, donde la Comisión Europea financia la elaboración de un estudio de viabilidad, principalmente, económico, del modelo de negocio, mercado, etc., donde a las empresas se les financia con una tasa fija de 50.000€.
- Los proyectos Fase 2, donde la Comisión Europea financia los últimos pasos para la obtención del producto/servicio final antes de lanzarlo al mercado. La financiación es un 70% en forma de subvención. Y la financiación está entre 0,5 – 2,5 millones de euros.

La siguiente gráfica muestra el número de proyectos Instrumentos PYMES relacionados con las tecnologías del lenguaje Fase 1 (11) y Fase 2 (4).

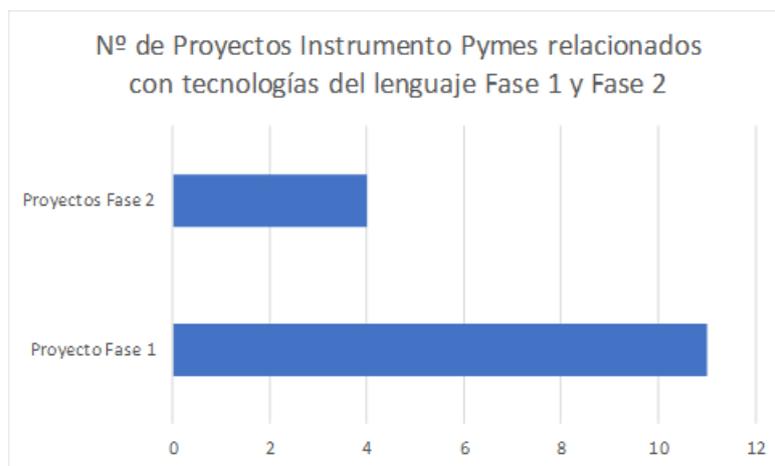


FIGURA 17. Nº DE PROYECTOS FASE 1 Y FASE 2 DE INSTRUMENTOS PYMES RELACIONADOS CON LAS TECNOLOGÍAS DEL LENGUAJE

El siguiente gráfico indica los países participantes. Destaca la participación de España con 3 proyectos Fase 1.

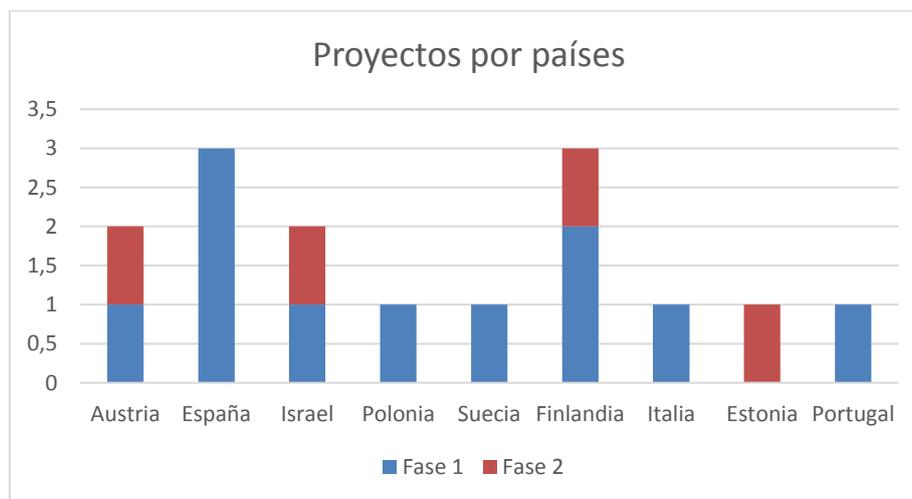


FIGURA 18. NÚMERO DE PROYECTOS DE INSTRUMENTO PYME POR PAÍS ORIGEN DEL SOLICITANTE

Las temáticas, la mayoría (11) están dirigidas al área de Tecnologías de la Información y las Comunicaciones. Las otras 4 temáticas restantes, están respectivamente dirigidas a las áreas de: transporte, seguridad, salud y modelos de negocio para PYMES. Como es habitual en este instrumento, el número de proyectos financiados en Fase 2 es menor que el de financiados en Fase 1, pero en porcentaje los proyectos financiados en Fase 2 relacionados con las TL es mayor que en Fase 1.

En el Anexo 2.3 se incluye la lista de proyectos europeos financiados dentro del subprograma ICT del H2020, relacionados con las TL, por si se quiere obtener más información de los mismos.

EUROSTARS

El siguiente grafo muestra la relación entre las entidades que han participado en los 2 proyectos relacionados con las TL. Se observa como al tratarse de tan pocos proyectos, no existe relación más allá de las que se dan en el mismo proyecto *per se*. En este caso se observa que el **tipo de organizaciones es muy variado** y que se trata de empresas colaborando con centros de investigación, universidades u otras empresas.

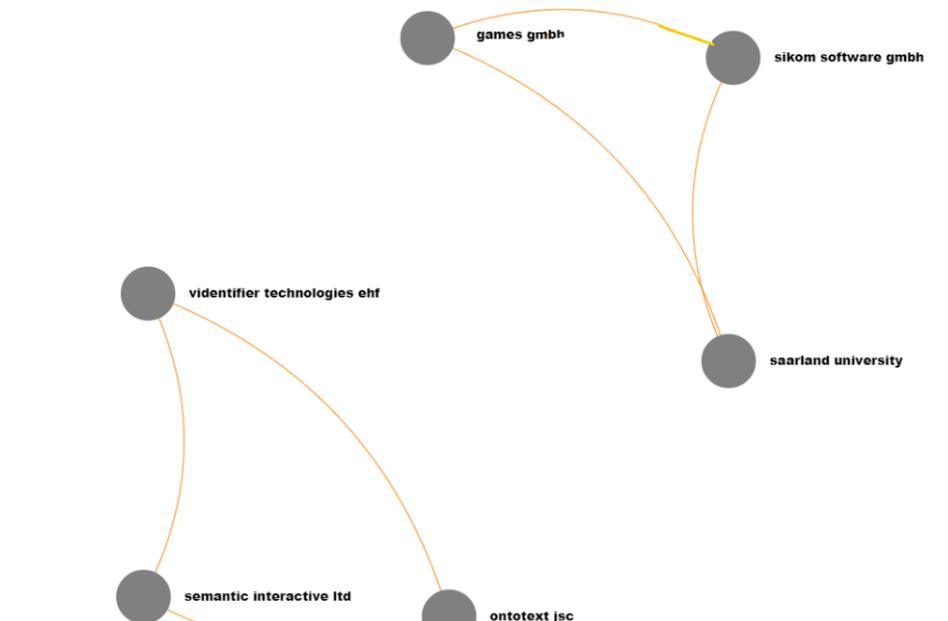


FIGURA 19 RELACIÓN ENTRE ENTIDADES PARTICIPANTES - EUROSTARS

Del mismo modo, se tiene el mapa con los países de procedencia de las diferentes organizaciones y se obtiene que los países que participan en uno o más de un proyecto en el ámbito de las Tecnologías del Lenguaje son **Islandia, Alemania, Austria y Bulgaria**.

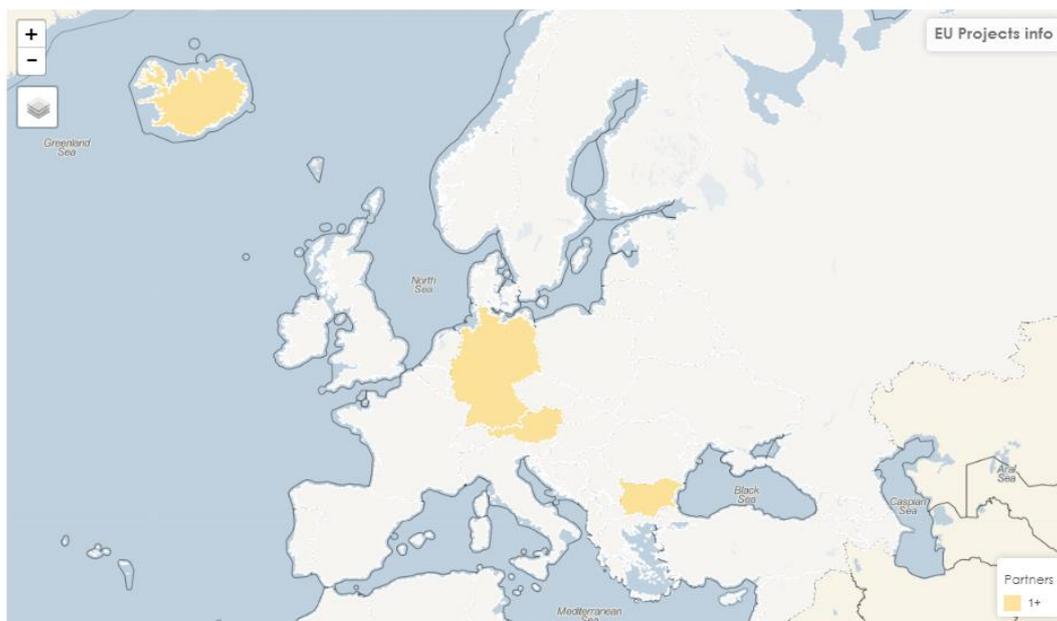


FIGURA 20 INTENSIDAD DE PARTICIPACIÓN DE CADA PAÍS EN FUNCIÓN DEL NÚMERO DE SOCIOS – EUROSTARS

EUREKA

En la siguiente imagen se analiza la relación entre las entidades que han participado en los proyectos relacionados con TL. Al igual que pasa con los analizados en EUROSTARS, se observa que, al tratarse de tan pocos proyectos, no existe relación más allá de las que se dan en los mismos proyectos. Se demuestra que el **tipo de entidades es muy heterogéneo** y que se trata de empresas colaborando con centros de investigación, universidades u otras empresas.

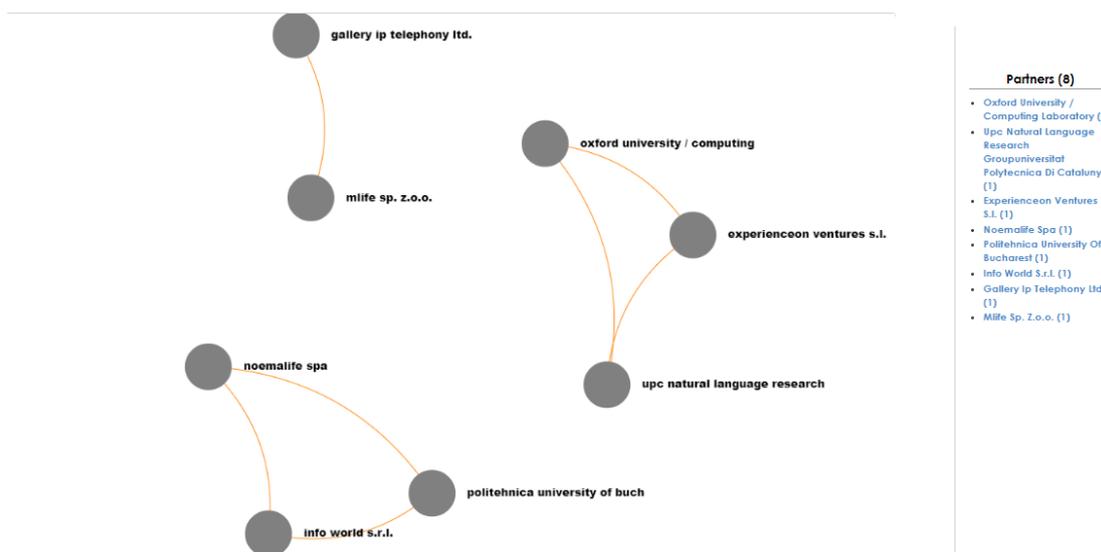


FIGURA 21 RELACIÓN ENTRE ENTIDADES PARTICIPANTES - EUREKA

Se ha obtenido el mapa donde es posible ver los países de procedencia de los participantes de los proyectos que han sido analizados anteriormente. De esta manera, y al tratarse de tan pocos

proyectos, los países que participan en uno o más de un proyecto del área de las Tecnologías del Lenguaje como socios son los siguientes: **Reino Unido, España, Italia, Polonia, Rumanía e Israel.**



FIGURA 22 INTENSIDAD DE PARTICIPACIÓN DE CADA PAÍS EN FUNCIÓN DEL NÚMERO DE SOCIOS - EUREKA

En lo concerniente a las *entidades coordinadoras*, se obtiene que *España, Polonia y Rumanía* son los países de procedencia que más veces actúan como líderes en los proyectos de Tecnologías del Lenguaje, ya que han participado en una o más de una ocasión.

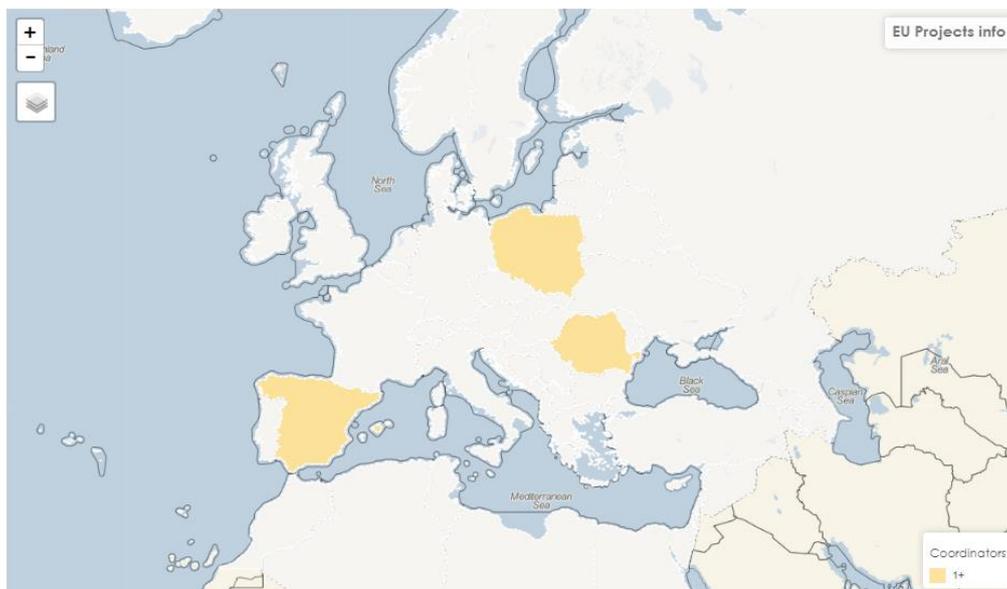


FIGURA 23 INTENSIDAD DE PARTICIPACIÓN DE CADA PAÍS EN FUNCIÓN DEL PAÍS COORDINADOR – EUREKA



3.3 Conclusión

En Europa las tecnologías del lenguaje han sido objeto de interés y estudio dada la naturaleza multilingüe de la Unión Europea.

A nivel nacional, existen algunos países que definieron en el pasado estrategias para fomentar el impulso de las tecnologías del lenguaje. Sin embargo, éstas no se han traducido (salvo en casos excepcionales, como puede ser Estonia o Francia) en la puesta en marcha de iniciativas/convocatorias de ayuda específicas para impulsar el desarrollo de proyectos o actuaciones en el área. De hecho, **a fecha de 21/12/2017 no hay ninguna convocatoria específica para TL abierta**. De manera general, se puede afirmar que ***existen o han existido muy pocas convocatorias nacionales específicamente destinadas a financiar proyectos relacionados con las TL***. La financiación en estos casos para el desarrollo de las TL se obtiene a través de convocatorias genéricas de I+D, no de programas específicos.

A nivel europeo, sí que las TL han tenido un mayor apoyo habiendo conseguido movilizar un número importante de fondos a través de proyectos en diferentes programas: FP7, H2020, CEF Telecom, CIP ICT PSP, EUREKA o EUROSTARS. No obstante, es importante destacar que también en este caso se está viendo como ***cada vez hay menos líneas/tópicos específicos para las TL***.

Así pues, si bien es cierto que las TL son muy importantes y han recibido apoyo tanto a nivel europeo como nacional, ***el análisis realizado refleja como el apoyo recibido se ha conseguido tradicionalmente a través de iniciativas de apoyo a la I+D, y no tanto de líneas específicas***.

4 Líneas de investigación, desarrollo e implantación de tecnologías innovadoras relacionadas con las TL financiadas en programas europeos

En este punto vamos a mostrar cuáles han sido las líneas de investigación, desarrollo e implantación de tecnologías innovadoras en TL (en adelante líneas) apoyadas en los últimos años a través de los programas de financiación pública, así como las líneas a futuro, de acuerdo con la información disponible, para finalmente, establecer unas conclusiones, de qué líneas podrían ser de interés para el Plan TL.

4.1 Líneas financiadas hasta 2017

De acuerdo, al análisis de los programas, la Comisión lleva desde el comienzo de 2007 apoyando proyectos de tecnologías del lenguaje.

En el **7 PM** (2007 – 2013) las líneas que se apoyaron fueron:

- *ICT-2013.4.1 Content analytics and language technologies*
- *ICT-2013.4.3 SME initiative on analytics*
- *ICT-2011.4.1 SME initiative on Digital Content and Languages*
- *ICT-2011.4.2 Language Technologies*
- *ICT-2007.4.2 Intelligent content and semantics*

De dichos tópicos se extrae que las **líneas** objeto de los proyectos fueron: *Semantic technologies, knowledge management technologies, integration of language-enabled content technologies (combination of content authoring, management, translation and publishing tasks and tools within typical production processes and translation/localisation workflows, in real-life multilingual settings), advancing machine translations, natural spoken interaction, Information access and mining, Cross-media content analytics*

De forma paralela al 7PM, durante el programa **CIP ICT PSP (2009-2010)**, las líneas directamente relacionadas con TL financiadas se indican a continuación:

- *CIP-ICT-PSP.2009.5.1 Multilingual Web: Machine translation for the multilingual web*
- *CIP-ICT-PSP.2009.5.2 Multilingual Web: Multilingual Web content management: standards and best practices*
- *CIP-ICT-PSP.2009.5.3 Multilingual Web: Multilingual Web content management: methods, tools and processes*
- *CIP-ICT-PSP.2010.6.2 Multilingual online services*
- *CIP-ICT-PSP.2010.6.1 Open linguistic infrastructure*



A través de dichos proyectos, se apoyaron proyectos de **integración de tecnologías**, más que de investigación y se dieron lugar a nuevos productos que facilitarían el multilingüismo en Internet y en los servicios ofrecidos por Internet. Fueron proyectos donde la validación con usuarios finales tenía un papel relevante.

Sin embargo, en el **H2020**, sólo en el año 2014 ha habido un tópico propio de tecnologías del lenguaje, muy orientado a Traducción Automática, pero después hasta el año 2017, las tecnologías del lenguaje pierden fuerza como área propia, y se incluyen dentro del área de Big Data. **Las tecnologías del lenguaje se consideran como un elemento necesario en Big Data, pero no como un área de investigación propia.** Los tópicos financiados hasta 2017 fueron:

- *ICT-16-2015 Big data - research*
- *ICT-15-2014 Big data and Open Data Innovation and take-up*
- *ICT-17-2014 Cracking the language barrier*

Y las **líneas** apoyadas fueron: *“language understanding technologies for carrying out data analytics, data quality assessment and improvement, prediction and visualization tasks at extremely large scale and with diverse structured and unstructured data, advanced machine translation, re-use, sharing of language resources”*

Paralelamente, el H2020, quiere apoyar a las PYMES con posibilidades de convertirse en empresas de gran volumen de negocio, empleo e impacto internacional, a través del **Instrumento Pyme**. Como se ha mostrado anteriormente, hay un número de proyectos instrumentos PYMES, relacionados con tecnologías del lenguaje, pero en ellos no se financia la investigación, sino el lanzamiento de un producto/servicio al mercado novedoso, que no tiene por qué implicar nuevas tecnologías.

A través del programa **CEF TELECOM**, tampoco se ha financiado la investigación sino más bien el apoyo a crear la infraestructura digital base para el desarrollo de servicios públicos multilingües. Se ha apoyado sobre todo la recopilación de recursos lingüísticos en diferentes países y el desarrollo de algunos proyectos de introducción de traducción automática en algunos servicios públicos digitales. La Comisión considera el CEF TELECOM como un programa base, para desarrollar la infraestructura necesaria para el desarrollo de las tecnologías del lenguaje.

Otros programas como **Eureka o Eurostars** no tienen incidencia en la estrategia de investigación e innovación en las tecnologías del lenguaje.



4.2 Líneas a futuro

En el programa de trabajo del **H2020 para los años 2018 – 2020**. Aparece un único tópico ([ICT-29-2018: A multilingual Next Generation Internet](#)) relacionado con las tecnologías del lenguaje, que busca:

- Facilitar el acceso de los investigadores y de la industria de tecnologías del lenguaje a recursos del lenguaje y herramientas de calidad.
- Aumentar la calidad y el alcance de las soluciones multilingües por parte de las empresas en los sectores relevantes para el Mercado Único Digital.
- Aumentar el uso de tecnologías del lenguaje en varios sectores.
- Ahorrar coste a los sectores públicos y privados que utilizan soluciones de tecnologías del lenguaje.

A continuación, se indica el detalle de lo que se busca financiar ya que se considera de interés:

ICT-29-2018: A multilingual Next Generation Internet

Innovation Action: A European Language Grid

- *Architecture and components for a public, open and interoperable grid connecting resources and tools to support effective development and deployment of language technologies (software and services) across Europe*
 - *ecosystem for accessing tools, services and data resources*
 - *enable effective pooling and usage of existing but often "dispersed" tools, services and data resources for EU languages*
 - *lowering barrier of entry/facilitating growth for SMEs active in LT*
 - *not only technology integration issues but also organisational aspects and obstacles, e.g. maintenance responsibility; connection modalities; services type and scope; billing models if you combine tools/services from several providers; confidentiality requirements for language resources data sets etc.*
 - *end users, i.e. developers of multilingual solutions and services based on LT must be tightly involved in the design, testing and appreciation of the grid.*
- *Coordinate the work of European Language grid...*
 - *relevant results of small-scale projects and RIA actions become part of the grid (interoperability)*
 - *structured dialogue with relevant stakeholders and activities*
 - *... leading to an improved approach underlying the grid (legal, organisational, technological aspects)*
- *Establish competence centres / nodes in Member States*
 - *to facilitate coordination with national and regional activities*
 - *building on previous EC-funded actions within the FP7, H2020 and CEF and existing networks/federations Member States (CLARIN, EFNIL, CEF Automated Translation National Anchor Points...)*
- *Pilot the European Language Grid ... through small scale demonstrators (mini-projects for third parties)*
 - *cascading grants instrument (Annex K of the general Annexes)*
 - *best-practice examples: [H2020 ODINE](#) and [DATA PITCH](#) projects*

Research and Innovation Action: Domain-specific/challenge-oriented Human Language Technology

- *Researchers and industrial users of language technologies working together on R&I challenges:*
 - *in a specific sector of high commercial and/or societal impact, or*
 - *a technological challenge common/relevant to several sectors*
- *Challenges defined by industrial users*
- *Aim of sectorial projects: create a sustainable ecosystem of multilingual applications and services tailored for the specific needs of the addressed sector*

Detailed analysis of the expected advances in LT-related research



En **CEF TELECOM**, la información disponible es sólo la correspondiente a la convocatoria del 2018. En este sentido, las líneas a financiar serán continuistas con respecto a años anteriores. Concretamente la línea que se abrirá en la segunda convocatoria de CEF TELECOM 2018 es:

Convocatoria	Fecha de inicio	Fecha de cierre
CEF-TC-2018-2 Automated Translation (Presupuesto: €5 million)	2 Mayo 2018	18 Septiembre 2018

Específicamente lo que van a buscar en la nueva convocatoria es:

The main objectives of the actions planned in 2018 are:

- *to continue operating the core service platform and to complete and operate the automated adaptation facility based on the results of the previous work programmes. The core service platform will be further enhanced by the introduction of new processing tools and services.*
- *to continue language resource collection for specific domains and to stimulate the porting and integration of the eTranslation service to online systems requiring multilingual functionality via Generic Services.*

En la [Strategic Agenda for Multilingual Digital Single Market](#), documento elaborado principalmente en los proyectos europeos financiados por el H2020: *CRACKER* y *LT_Observatory*, con el fin de identificar cuáles serán las áreas de investigación e innovación en las tecnologías del lenguaje en los próximos años, se identifican **4 áreas de investigación** que habitualmente son una información muy importante para la Comisión en la definición de futuros programas:

- *Cross-lingual Big Data Language Analytics*
- *High-Quality Machine Translation*
- *Meaning, Semantics, Knowledge*
- *Conversational Technologies*

Y 6 áreas horizontales:

- *Standardisation and Interoperability*
- *Business Models and Ecosystems*
- *Language Policies and Public Procurement*
- *Copyright and Data Protection*
- *Open Source*
- *Related Areas, Applications and Societal Challenges*

El detalle de cada una de ellas se puede encontrar en la [SRIA](#) (Strategic Research and Innovation Agenda)

4.3 Conclusiones

El Plan de Impulso de las Tecnologías del Lenguaje tiene como objetivo fomentar el desarrollo del procesamiento del lenguaje natural, la traducción automática y los sistemas conversacionales en lengua española y lenguas cooficiales

Estos objetivos de apoyo al avance en procesamiento del lenguaje natural, la traducción automática y los sistemas conversacionales se han apoyado durante el 7PM, CIP ICT PSPS y durante el H2020. **No se ha detectado una línea de investigación diferente en los programas europeos a la del PLAN TL.**

También es destacable la adopción por parte de la CE de la visión de LT-Innovate (Asociación Europea de las Tecnologías del lenguaje) de cómo debe ser la **Plataforma Europea del lenguaje**, que responde a este esquema:

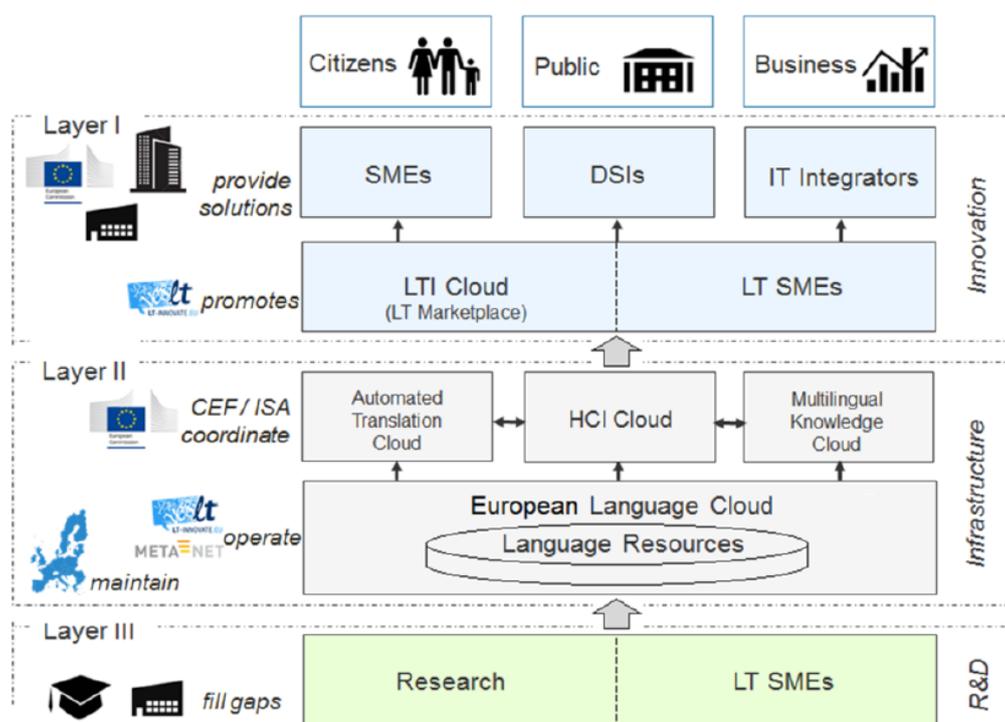


FIGURA 24. PLATAFORMA EUROPEA PARA UN MERCADO DIGITAL EUROPEO MULTILINGÜE. FUENTE. EUROPEAN PLATFORM FOR THE MULTILINGUAL DIGITAL SINGLE MARKET: CONCEPTUAL PROPOSAL

Analizando la próxima convocatoria del H2020, consideramos, que ésta va a apoyar proyectos de Investigación, que pertenecerían a la *Layer III* a través de los proyectos RIA (Research and Innovation Actions), pero también quiere establecer la base para la *Layer I* a través del proyecto de IA (Innovation Actions).

Este esquema y lo que significa, también puede ser un buen input, de cara a las líneas de investigación a tener en cuenta en el PLAN TL.



Adicionalmente, para el Plan TL, consideramos que podría ser de interés para puntualizar sus próximas líneas de investigación la **SRIA**, con sus 4 líneas de investigación identificadas en la misma, así como las 6 horizontales. En dicho documento se nombra la iniciativa de **Human Language Project**, que desde el Centro de Investigación DFKI se está intentando lanzar, a través de una propuesta que se presentó a la convocatoria [FETFLAG-01-2018](#) del H2020 y que si surgiera, sería posiblemente el marco en el que se gestionara la financiación europea pública orientada a las tecnologías del lenguaje en los próximos años. Esta iniciativa, sería similar a la actual iniciativa "[Human Brain Project](#)", también financiada por la Comisión.

Las "Flagships" son iniciativas visionarias, de investigación a largo plazo, orientadas a grandes retos científicos y tecnológicos. Son iniciativas a largo plazo, que agrupan a equipos multidisciplinares de investigación excelente, que comparten un único objetivo y una ambiciosa hoja de ruta de investigación para conseguirlo.

Con las FET Flagships, se quiere establecer a Europa como líder en la temática correspondiente, a la vez que un polo de atracción para la cooperación internacional. Se nutrirá de la creatividad, el talento, atrayendo a las mejores mentes de todo el mundo y creando los investigadores multidisciplinares mejores en Europa. Las FET FLAGSHIP son un nuevo modelo de consorcio para la investigación colaborativa europea a largo plazo.

A través de las FET FLAGSHIP, se conseguirá:

- Convertir los avances científicos en oportunidades de innovación concretas, crecimiento y empleo.
- Contribuir a hacer frente a los mayores retos sociales de Europa.

La duración de las FET FLAGSHIP, se espera que sea de 10 años, con un presupuesto alrededor de 1000 millones de euros. En los consorcios participarán entidades de investigación, grandes empresas y pymes. La financiación no proviene solo de las CE sino también de los países y las regiones.

Por ahora las dos FET FLAGSHIP financiadas son: GRAPHENE y "The Human Brain Project".

5 Instrumentos de financiación europeos de interés para la ejecución del PLAN TL

Con el fin de identificar instrumentos de financiación a nivel europeo que pudieran ser de interés para apoyar la ejecución del PLAN TL, a continuación, se indica una tabla con los ejes del PLAN TL y con instrumentos de financiación que consideramos podrían ser de interés. Estos instrumentos, son los mecanismos de financiación utilizados por los programas europeos, principalmente H2020.

Ejes del PLAN TL	Instrumentos de financiación de programas europeos
<p>Eje I: Apoyo al desarrollo de infraestructuras lingüísticas</p> <p>Este eje pretende facilitar el desarrollo de procesadores (reconocedores de entidades, desambiguadores, calculadores de proximidad semántica, etc.) y recursos lingüísticos (corpus paralelos, diccionarios, taxonomías, etc.) que sirvan de combustible para el desarrollo de la industria española de procesamiento de lenguaje natural y de traducción automática.</p>	<p>Instrumentos tradicionales de financiación a la investigación y al desarrollo del H2020. Es decir:</p> <ul style="list-style-type: none"> • RIA: Research and Innovation Actions para proyectos de investigación. Suelen ser proyectos donde el resultado está lejano al mercado. Los consorcios suelen estar formados por entidades de investigación y alguna industria, pero como validador de tecnología, no tanto para validar un producto. El nivel de maduración de la tecnología está entre 3 y 5. (TRL3: prueba de concepto; TRL5: validación a nivel de componentes en un entorno relevante) • IA: Innovation Actions, para proyectos de innovación. En estos proyectos, el objetivo suele ser la obtención de prototipos de productos/servicios comercializables en un plazo corto, unos 2 años. El nivel de maduración de la tecnología final suele ser 5 – 6 (TRL5: validación a nivel de componentes en un entorno relevante; TRL6: validación de sistema o subsistema en un entorno relevante). En los consorcios hay entidades de investigación, pero también industria que explotaría el producto/servicio resultante e incluso usuarios finales que validan la solución en un entorno real. <p>Proyectos CEF TELECOM: para mejorar y aumentar la recolección de recursos lingüísticos pertenecientes a distintos dominios, así como la mejora del servicio de la Comisión Europea MT@EC (servicio de traducción</p>

Ejes del PLAN TL	Instrumentos de financiación de programas europeos
	automática europea, desarrollado por la CE para administraciones públicas).
<p>Eje II: Impulso de la Industria de las Tecnologías del Lenguaje</p> <p>El segundo de los ejes responde al objetivo de apoyar la transferencia de conocimiento entre el sector investigador y la industria, así como la internacionalización de las empresas e instituciones que componen el sector.</p>	<p>En este eje, consideramos que pueden ser de interés los nuevos instrumentos de financiación que la Comisión Europea está utilizando para favorecer el origen de nuevas empresas y el crecimiento de las PYMES/empresas existentes. Estos nuevos instrumentos son:</p> <ul style="list-style-type: none"> • Instrumento PYMES: se trata de una tipología de proyecto dentro del H2020. No es una herramienta de apoyo a la creación de empresas, sino que apoya el crecimiento de las empresas (exclusivamente PYMES) con ideas interesantes e innovadoras con potencial europeo o mundial, y que van a tener un impacto en la empresa en términos de crecimiento (ventas, plantilla, etc.) muy importante. <ul style="list-style-type: none"> ○ Para ello, el instrumento se estructura en 3 fases que abarcan las diferentes etapas del ciclo de innovación en la PYME (Fase I: concepto y evaluación de la viabilidad, Fase II I+D, demostración y replicación comercial, y Fase III Comercialización). ○ Las PYME beneficiarias (pueden ser proyectos en cooperación entre PYMES) se benefician de medidas de apoyo como redes, formación, asesoramiento y consejos. Asimismo, esta parte puede enlazarse con medidas para promover la contratación precomercial y la contratación de soluciones innovadoras. • Fast track to Innovation (FTI): se trata de un tipo de iniciativa dentro del H2020 que, mediante un enfoque <i>bottom-up</i> y sin restricción temática y/o aplicación, pretende apoyar a las iniciativas de innovación cercanas al mercado. Para ello, y con el objetivo de fomentar también la participación entre el sector privado y las entidades

Ejes del PLAN TL	Instrumentos de financiación de programas europeos
	<p>investigadoras, se deben constituir consorcios de 3-5 socios, que tienen que contar como mínimo un 60% de presupuesto de entidades industriales y pertenecer al menos a 3 estados miembro (o países asociados al H2020).</p> <p>Adicionalmente existe otro modelo también interesante denominado Cascade funding. Se trata de una tipología de financiación contemplada dentro del H2020 que consiste en distribuir a terceros (startups y PYMES) fondos a través de consorcios de proyectos europeos financiados.</p> <p>Estos proyectos se centran en impulsar la adopción de nuevas tecnologías por sus usuarios finales y se implantan como financiación para realizar experimentos con ellas o extensiones para mejorarlas.</p>
<p>Eje III: La Administración Pública como impulsor de la Industria del Lenguaje</p> <p>Este tercer eje busca, por un lado, crear plataformas comunes de procesamiento de lenguaje natural y de traducción automática de las AAPP (mejorar calidad, dotar de herramientas comunes ...) y por otro, impulsar la reutilización de la información del sector público.</p>	<p>En cuanto al papel que la administración puede desempeñar, se ven fundamentalmente dos líneas de actuación:</p> <ul style="list-style-type: none"> • Compra Pública de Tecnología Innovadora (CPTI): instrumento de contratación pública para fomentar la innovación y potenciar el desarrollo de mercados innovadores desde el lado de la demanda. Compra Pública Precomercial (CPP): instrumento de contratación pública que, a diferencia de la CPTI, se circunscribe exclusivamente a la compra de servicios de I+D+i dirigidos a conseguir unos resultados innovadores respecto a lo existente en ese momento en el mercado
<p>Eje IV: Proyectos faro de tecnologías de procesamiento de lenguaje natural.</p> <p>Estos proyectos de aplicación de las tecnologías del lenguaje en sectores estratégicos que pretenden ser</p>	<p>Instrumentos empleados por la Comisión Europea en el Programa H2020 con este fin son los grandes proyectos como:</p> <ul style="list-style-type: none"> • <i>Large scale projects</i> • <i>Lighthouse projects</i> <p>En los que se llevan a cabo pilotos en entornos reales para la incorporación de tecnologías pero que a su vez sirvan de</p>

Ejes del PLAN TL	Instrumentos de financiación de programas europeos
demostradores, generar industria y recursos reutilizables.	<p>proyectos tractores para el impulso de dichas tecnologías. La Comisión los identifica así, porque cuando describe el “Tópico correspondiente” indica que es un <i>Lighthouse</i> o si no por el gran tamaño de los mismos, normalmente más de 15 millones de euros de financiación, y por el requisito de llevar a cabo varios pilotos en entornos reales. La duración suele ser entre 3-4 años.</p> <p>Proyectos de este tipo se han financiado en el H2020 en el área de Smart Cities (ver ejemplos en el siguiente link https://smartcities-infosystem.eu/scc-lighthouse-projects) o en el área de IoT -<i>Internet of Things</i> (ver ejemplos en el siguiente link: https://iot-epi.eu/projects/).</p> <p>Este tipo de proyectos no son lo mismo que los FET FLAGSHIP que son iniciativas de mucho mayor alcance en todo: investigación, duración, financiación, entidades cooperando, etc.</p>

TABLA 4. EJES VS INSTRUMENTOS DE FINANCIACIÓN

De los proyectos financiados hasta el momento en TL, por programas europeos, se han identificado proyectos tipo RIA e IA, así como Proyectos Instrumento PYME, pero no se han identificado proyectos tipo, “*Fast Track to Innovation*”, “*Large scale projects*” o “*lighthouse projects*”, ni proyectos CPTI y CPP.

5.1 Conclusiones

Los proyectos colaborativos entre empresas y entidades de investigación, pueden ser un buen vehículo para avanzar en el sector de las tecnologías del lenguaje desde un punto de vista de desarrollo de tecnología y de nuevos productos y servicios a medio – largo plazo, pero herramientas tipo el *Fast Track to Innovation*, el instrumento PYME, o la financiación en cascada pueden ayudar a activar el sector apoyando a pequeñas empresas, start ups, etc. a generar nuevos productos y servicios, de gran impacto en el mercado. Los proyectos de grandes pilotos, pueden ayudar a implantar en mayor escala, tecnologías/productos, novedosos pero que ya estén prácticamente en el mercado.

6 Herramientas, aplicativos y soluciones de tecnologías del lenguaje que se están utilizando sobre los que se está investigando en proyectos europeos

En base a proyectos financiados en el H2020 relacionados con las TL, se han identificado proyectos cuyos resultados pudieran ser de interés. Para ello se ha utilizado un análisis hecho por la Comisión Europea sobre los principales proyectos en el área ([link donde se incluye una diapositiva al respecto dentro de una presentación de la CE](#)). Cabe destacar que, en muchos casos, los resultados no son públicos porque se van a explotar en soluciones comerciales por lo que la información que se presenta es aquella disponible y accesible pero no se puede afirmar que sea todo lo existente:

Machine translation:

- **MMT**: *Adaptive domain-sensitive MT infrastructure*. Dispone de una versión descargable en GitHub
- **QT21**: *Quality translation*. En el Proyecto van a desarrollar:
 - *substantially improved statistical and machine-learning based translation models for challenging languages and resource scenarios,*
 - *improved evaluation and continuous learning from mistakes, guided by a systematic analysis of quality barriers, informed by human translators,*
 - *all with a strong focus on scalability, to ensure that learning and decoding with these models is efficient and that reliance on data (annotated or not) is minimised.* Varios de los deliverables están accesibles por la Web.
- **TraMOOC**: *Translation for Massive Open Online Courses. The main expected outcome of the project is a high quality machine translation service for all types of educational textual data available on a MOOC platform. The service will support 11 target languages, 9 European (German, Italian, Portuguese, Dutch, Bulgarian, Greek, Polish, Czech, Croatian) and 2 BRIC (Brazil, Russia, India, China).* Tiene publicados “papers”, posters,... pero no el acceso a la herramienta, que probablemente será una herramienta comercial y no de acceso libre.
- **HimL**: *Automated translation of public health information*. En la web están disponibles entregables con resultados de la validación de la herramienta pero no los módulos desarrollados que indican que se van a incluir en productos de los socios como: [NHS 24](#) and [Cochrane](#) y por tanto no son de acceso libre.

Proyectos de **Big data** con una orientación de tecnologías del lenguaje, pero de aplicación en sectores como: el médico, la agricultura o medios.

- **FREME**: *Open Framework of e-Services for Multilingual and Semantic Enrichment of Digital Content*. En la web presenta 4 casos de negocio (eBooks, traducción y localización, acceso a



datos abiertos de agricultura y alimenticios, recomendaciones personalizadas del contenido de las web para incrementar la participación de los usuarios) reales en los que aplica las innovaciones del proyecto. También hay un video en el que muestra como FREME hace más eficiente la creación de contenido. (<https://vimeo.com/143390983>)

- **Kconnect:** *ML search for medical information*: la tecnología puede utilizar una gran variedad de recursos de información médica incluyendo: historia clínica, publicaciones médicas, mejores prácticas y pautas de tratamiento, revisiones sistemáticas, páginas web indexadas, etc. La tecnología está disponible para empresas, proveedores del sector médico e investigadores.

En cuanto a resultados, y entregables del proyecto se tienen:

- Toolkits de anotación semántica (D1.5)
- Toolkit e informe acerca de la adaptación a nuevas lenguas (D1.7)
- *Cloud market plataform* (datos médicos y análisis de texto como servicio) (D2.2) Se puede registrar una cuenta <https://kconnect-ui.s4.ontotext.com/#/>
- **SUMMA:** *Scalable Understanding of Multilingual Media*. El proyecto todavía está en curso hasta finales de 2018. El resultado va a ser una plataforma que va a permitir el análisis automatizado de los flujos de los medios en varios idiomas. Así, se van a agregar y analizar diferentes aspectos (texto, audio, video). La plataforma consiste en una serie de módulos para el procesamiento del lenguaje natural, que se desarrollan de manera independiente:
 - Reconocimiento de voz (ASR),
 - Machine Translation (MT), Entity Linking (NEL),
 - Knowledge Data Base (KDB),
 - Análisis de topics,
 - Análisis de sentimiento,
 - Story telling...

Proyectos de **ámbito más político y estratégico**:

- **LT Observatory** (*Observatory for Language Resources and Machine Translation in Europe*). El resultado del proyecto es un observatorio (**LT-Observe**) disponible en la web de LT-Innovate y donde hay disponible:
 - Observatorio de políticas públicas
 - Catálogo de los recursos del lenguaje
 - Observatorio del mercado.
 - *MT EcoGuide*



Además, como resultado del proyecto se establecieron las bases para la creación de una Agenda Estratégica para las Tecnologías del Lenguaje y Machine Learning (MT): SRIA en cooperación con el otro proyecto financiado en el mismo tópico (proyecto CRACKER)

En la web se pueden encontrar los entregables públicos del proyecto.

- **Cracker** (*Cracking the Language Barrier*) El proyecto finalizó en diciembre de 2017 por lo que puede ser que todavía no haya tenido lugar la reunión de cierre con la EC, y por ello, la información que aparece en la web del proyecto en cuanto a los resultados es muy limitada. Lo que se ha encontrado ha sido a través de la web de la Comisión (https://cordis.europa.eu/project/rcn/194311_en.html) Así, se encuentra disponible una nueva [versión del software META-SHARE](#).

Otro resultado del proyecto ha sido una versión más avanzada de la SRIA (Agenda Estratégica para las Tecnologías del Lenguaje y Machine Learning (MT))

Machine translations							
Acrónimo	Título	Fecha de Inicio	Fecha de Fin	Prespto. €	Subven. €	Instr.	Tópico
MMT	MMT will deliver a language independent commercial online translation service based on a new open-source machine translation distributed architecture	01/01/2015	01/01/2018	3.695.200	2.994.700	IA	ICT-17-2014
QT21	QT21: Quality Translation 21	01/02/2015	01/02/2018	3.997.428	3.977.428	RIA	ICT-17-2014
TraMOOC	Translation for Massive Open Online Courses	01/02/2015	01/02/2018	3.270.710	3.081.147	IA	ICT-17-2014
HimL	Health in my Language	01/02/2015	01/02/2018	3.051.383	2.949.571	IA	ICT-17-2014
Big Data							
Acrónimo	Título	Fecha de Inicio	Fecha de Fin	Prespto.	Subven.	Instr.	Tópico
FREME	Open Framework of E-Services for Multilingual and Semantic Enrichment of Digital Content	01/02/2015	01/02/2017	3.606.751	3.212.626	IA	ICT-15-2014
KConnect	Khresmoi Multilingual Medical Text	01/02/2015	01/08/2017	3.889.842	3.083.083	IA	ICT-15-2014



	Analysis, Search and Machine Translation Connected in a Thriving Data-Value Chain						
SUMMA	Scalable Understanding of Multilingual Media	01/02/2016	31/01/2019	7.963.951	6.193.361	RIA	ICT-16-2015
Políticas y estrategia							
Acrónimo	Título	Fecha de Inicio	Fecha de Fin	Prespto.	Subven.	Instr.	Tópico
LT Observatory	LT_OBSERVATORY - OBSERVATORY FOR LR and MT in EUROPE	01/01/2015	01/01/2017	982.565	982.563	CSA	ICT-17-2014
CRACKER	Cracking the Language Barrier: Coordination, Evaluation and Resources for European MT Research	01/01/2015	01/01/2018	999.995	999.995	CSA	ICT-17-2014

TABLA 5. LISTADO DE PROYECTOS DEL H2020 CON RESULTADOS DE INTERÉS

Existe más información sobre los proyectos en los hipervínculos a las webs de los proyectos.

6.1 Conclusiones

De los proyectos europeos, pueden surgir nuevas tecnologías, productos y servicios de gran interés. Pero la realidad es que la publicidad de estos depende de los socios y en muchos casos no es adecuada. A este hecho se une que normalmente hay muchos resultados parciales, que pueden ser de interés pero que no se difunden adecuadamente o que están en un estado lejano al mercado, y tampoco se publicitan. Desde la Comisión se está intentado mejorar este hecho y cada vez más, se obliga a mejores difusiones y a que los resultados publicables de las investigaciones financiadas por programas europeos se publiquen obligatoriamente de forma abierta. A medida que se avance en este sentido, se irá mejorando el acceso a los resultados de proyectos.

Actualmente la mejor manera de acceder a los resultados, es a través de las páginas web de los proyectos, donde se suelen incluir los entregables públicos del proyecto o contactando directamente con los responsables del proyecto.

7 Identificación de oportunidades dentro del marco de las estrategias europeas del Mercado Único Digital (DSM)

En el estudio se ha analizado la Estrategia Europea, de Mercado Único Digital (Digital Single Market – DSM) y las oportunidades que ofrece para el desarrollo de las tecnologías del lenguaje en España. Para ello, primeramente, se ha identificado en qué consiste la estrategia, qué iniciativas comprende, y el estado de éstas.

Una vez reflejado en qué consisten la estrategia DSM, se han identificado 4 iniciativas que afectan en mayor medida al sector de las tecnologías del lenguaje:

- Un marco de derechos de autor (Copyright) moderno y más europeo.
- Abordar las barreras de la economía de datos europea.
- Definir prioridades para los estándares y la interoperabilidad.
- Apoyar una sociedad digital inclusiva.

Dichas actuaciones se han analizado en detalle para identificar las acciones que se están llevando a cabo, grupos de interés trabajando en ellas y resultados.

Finalmente, por cada una de las iniciativas detectadas, se han definido unas recomendaciones sobre las actuaciones que se podría llevar a cabo, desde la Secretaría de Estado, para aprovechar dichas actuaciones. Como las actividades que incluyen las iniciativas, son de alto nivel, es decir, más de definición de políticas, regulaciones, las implementaciones más concretas se llevan a cabo a través de licitaciones por parte de la DG Connect, principalmente, y de los programas H2020, CEF Telecom y Erasmus +. El seguimiento de dichas iniciativas por parte de la Secretaría de Estado, y hasta su participación, permitiría fomentar:

- La divulgación de la información relevante al sector de las tecnologías del lenguaje en España.
- Fomentar la participación de otras entidades del sector de las TL en programas europeos como el H2020, CEF Telecom y Erasmus +.

7.1 Concepto del Digital Single Market

En Europa nos encontramos que las barreras reales al mercado único se extienden, también, al mundo digital virtual. Por ejemplo, los mercados virtuales aún son principalmente nacionales en lo que a servicios en línea se refiere: tan solo el 7% de las pequeñas y medianas empresas de la UE venden en otros países. Esta situación puede cambiar con el establecimiento del mercado único en línea.

El propósito de la Comisión Juncker es crear un **mercado único digital** en el que esté garantizada la libre circulación de mercancías, personas, servicios y capitales y en el que ciudadanos y empresas

tengan un acceso en línea ininterrumpido y equitativo a bienes y servicios, independientemente de su nacionalidad o de su lugar de residencia.



FIGURA 25. ESTRATEGIA – MERCADO ÚNICO DIGITAL

El estado actual de las iniciativas del DSM, se reflejan en la siguiente gráfica:

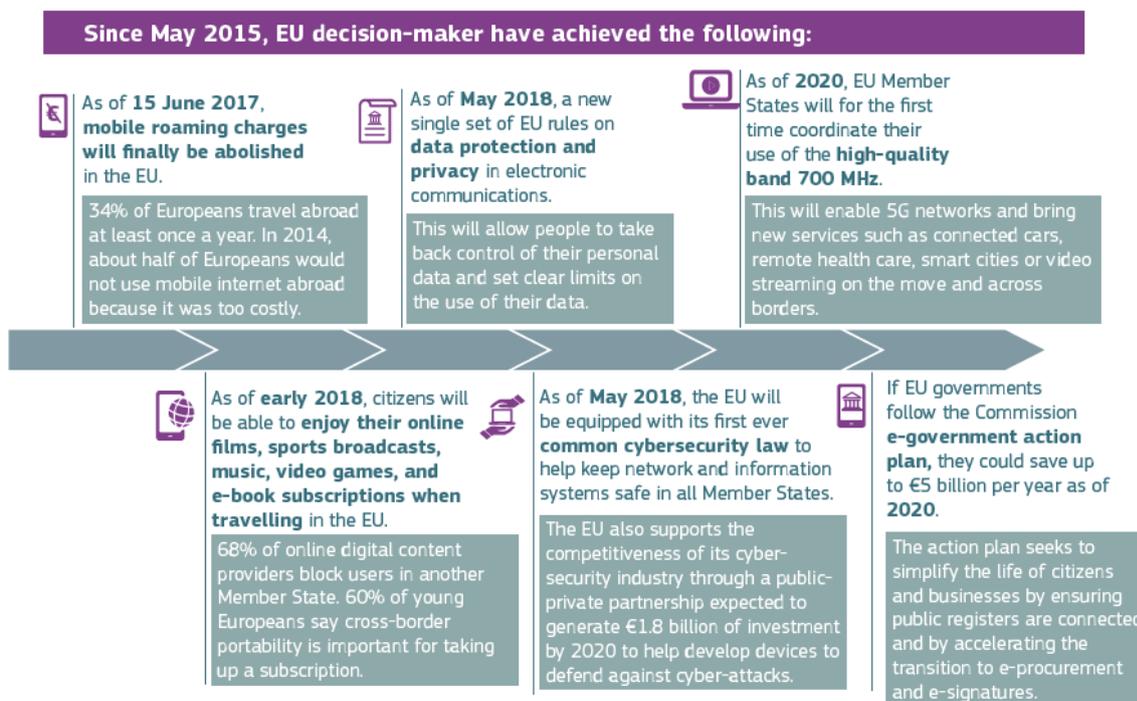


FIGURA 26. ESTADO ACTUAL DE LAS INICIATIVAS LANZADAS POR EL DSM

Para la implementación de las iniciativas anteriormente citadas, la Comisión Europea, está utilizando los siguientes instrumentos:

- Licitaciones de la DG CONNECT principalmente:
 - Suministro de mercado de datos en contenidos en línea, servicios y aplicaciones disponibles en plataformas de banda ancha fijas y móviles.
 - Nuevas oportunidades para las start ups y PYMES proveedoras de soluciones digitales para la salud.
- H2020 a través de sus diferentes subprogramas: ICT e Infraestructuras principalmente.
- CEF TELECOM.

7.2 ¿Cómo apoya el DSM y sus correspondientes iniciativas/programas a las tecnologías del lenguaje?

Atendiendo a todos los aspectos que engloba la estrategia europea del Mercado Único Digital, a continuación, se identifican los objetivos y las iniciativas que, hemos considerado, afectan más directamente a las Tecnologías del Lenguaje, y las que desde la Secretaría de Estado deberían tenerse en cuenta:

Objetivos del DSM que afectan más directamente a las Tecnologías del Lenguaje:

- Impulsar el comercio electrónico en la UE combatiendo el bloqueo geográfico y haciendo así más asequible y eficiente la entrega transfronteriza de paquetes.
- Liberar el potencial de una economía de datos europea con un marco para la libre circulación de datos no personales en la UE.
- Adaptar las normas de privacidad electrónica al nuevo entorno digital.
- Ayudar a las pequeñas y grandes empresas, a los investigadores, a la ciudadanía y a las autoridades públicas a sacar el máximo provecho de las nuevas tecnologías garantizando que todo el mundo tenga las competencias digitales necesarias y financiando la investigación de la UE en sanidad e informática de alto rendimiento.

Iniciativas del DSM que afectan más directamente a las tecnologías del lenguaje. A continuación, se detalla la información sobre las iniciativas que más afectan a las tecnologías del lenguaje y que pueden ser de interés para la SESIAD:

- Mejor acceso de los consumidores y las empresas a los bienes en línea.
 - Un marco de derechos de autor (Copyright) moderno y más europeo.
- El sector digital como motor de crecimiento.

- Abordar las barreras de la economía de datos europea.
- Definir prioridades para los estándares y la interoperabilidad.
- Apoyar una sociedad digital inclusiva.

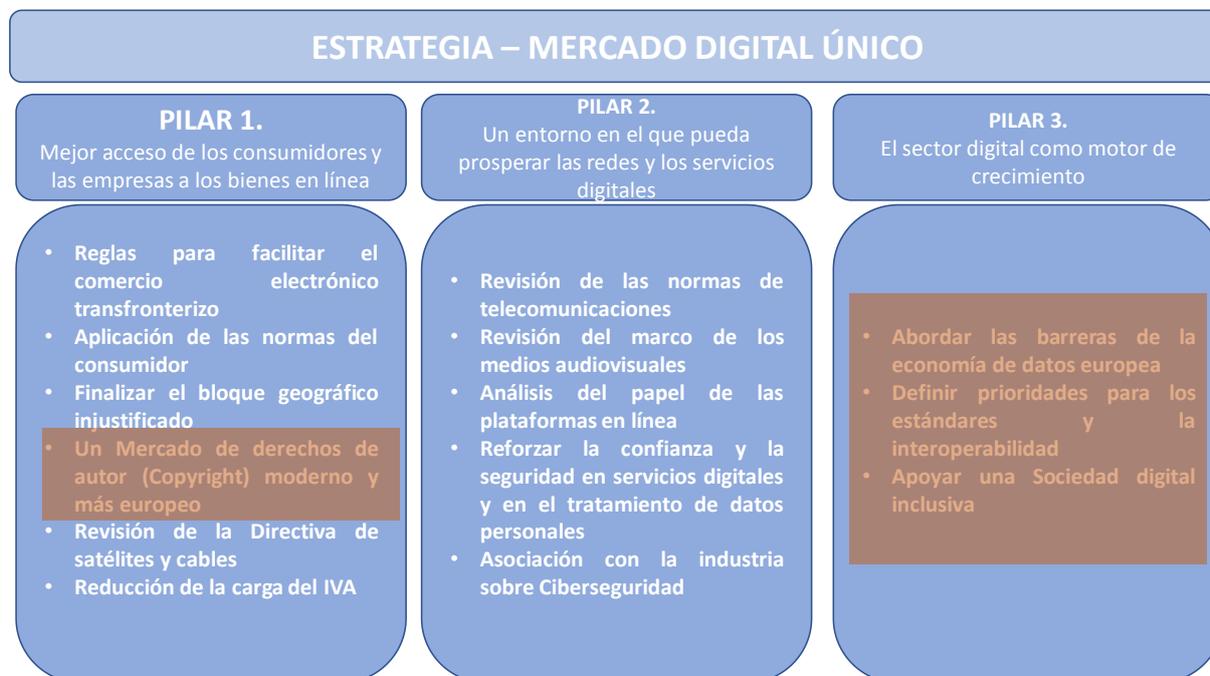


FIGURA 27. INICIATIVAS QUE AFECTAN A LAS TECNOLOGÍAS DEL LENGUAJE MÁS DIRECTAMENTE

En los siguientes puntos, se profundizará en las actuaciones que la Comisión Europea está llevando a cabo en cada una de las iniciativas, para reflejar como la Secretaría de Estado se puede beneficiar y/o contribuir a las mismas:

7.2.1 [Normas de Derecho de autor \(copyright\) modernas:](#)

La UE está llevando a cabo una iniciativa con el fin de mejorar las reglas de Derechos de Autor (Copyright) tanto en educación, investigación, herencia cultural e inclusión de personas con discapacidad. Dicha revisión consiste en una [regulación](#) y una [directiva de “Copyright”](#) en un Mercado Único Digital, así como una [Directiva y un Reglamento para implementar el tratado de Marrakech en la legislación de la UE](#).

Los Objetivos de la modernización son:

- Más acceso transfronterizo a contenido en línea;
- Ampliar las oportunidades de utilizar materiales con derechos de autor en educación, investigación y patrimonio cultural;
- Un mercado de derechos de autor que funcione mejor.



Las normas que más relación tienen con tecnologías del lenguaje son las que se indican a continuación, por la importancia que tienen en la obtención de corpus y textos que permitan alimentar los sistemas de traducción automática. Es importante poder reutilizar material existente, que de otra manera sería muy complejo de elaborar, sobre todo en idiomas que se hablan en menor medida. Por ello es importante tener en cuenta la legislación que regula el “copyright” y los contenidos en internet, así como la posibilidad de su uso. Concretamente las propuestas de la Comisión han sido:

- **Ampliar las oportunidades de utilizar materiales con derechos de autor en educación, investigación y patrimonio cultural.**

La mayoría de las excepciones de derechos de autor en la legislación de la UE son actualmente opcionales y no se aplican a través de las fronteras. Además, algunos de ellos deben ser reevaluados a la luz de las realidades tecnológicas actuales. El objetivo de la directiva propuesta es modernizar las normas de la UE aplicables a las principales excepciones y limitaciones en los ámbitos de la enseñanza, la investigación y la conservación del patrimonio cultural, centrándose en particular en los usos digitales y transfronterizos. Las excepciones obligatorias que anuncia la directiva propuesta están relacionadas con:

- Actividades de enseñanza
 - Minería de texto y datos
 - Preservación del patrimonio cultural
- **Un mercado de derechos de autor que funcione mejor.**

El “Copyright” impulsa la creatividad y la industria de la cultura en Europa. Los límites al acceso transfronterizo a contenido protegido por derechos de autor son comunes. Esto a menudo significa que no puede llevar su propio contenido al extranjero y que no puede acceder a muchos programas de radio y televisión de otros Estados miembros. La innovación y la investigación también se ven obstaculizadas por la falta de un marco jurídico claro en toda la UE. También es necesario asegurarse de que el valor generado por algunas de las nuevas formas de distribución de contenido en línea sea bastante compartido. La Comisión ha propuesto normas de copyright modernizadas para facilitar una mayor disponibilidad en línea de contenido en toda la UE, para modernizar el marco de excepciones y limitaciones y para lograr un mercado de derechos de autor que funcione correctamente.

Recomendación: seguimiento de la [Regulación](#) y la [directiva de “Copyright”](#) en un Mercado Único Digital y cómo ésta afecta a la utilización de contenidos en Internet así como a resultados de investigación, para su utilización en investigación o en productos/servicios comercializables, con el fin de poder divulgar dicha información al sector de las Tecnologías del Lenguaje español.

7.2.2 Economía de datos europea

Dentro de esta iniciativa las actividades que más afectan a las tecnologías del lenguaje, son aquellas que tienen relación con la Iniciativa “European Cloud”. Esta iniciativa se centra en:

- La European Open Science Cloud ([EOSC](#)): un entorno confiable y abierto para almacenar, compartir y reutilizar datos científicos y resultados.
- La European Data Infrastructure ([EUDAT](#)): una infraestructura digital de clase mundial para acceder, mover, compartir y procesar datos de manera segura en Europa.

La **iniciativa European Cloud** lanzada en abril de 2016 se orientó a facilitar servicios de certificación claros y fiables que permitan a los usuarios beneficiarse de servicios en la nube, de alta calidad, seguros y fiables. En base a esta iniciativa, están previstas las siguientes actuaciones:

- A partir de 2016, desarrollar una nube europea de ciencia abierta para investigadores europeos y sus colaboradores científicos globales mediante la integración y consolidación de plataformas de infraestructura electrónica, federación de nubes científicas e infraestructuras de investigación existentes y apoyo al desarrollo de servicios basados en la nube.

Evolución:

- En su resolución de 19 de enero de 2016 “*Towards a Digital Single Market Act*”, el Parlamento pidió a la Comisión que estableciera una nube de ciencia abierta europea para finales de 2016, que debería integrar sin problemas las redes existentes, los datos y los sistemas informáticos de alto rendimiento. -servicios de infraestructura en todos los campos científicos en el marco de políticas, estándares e inversiones compartidos.
- El 27 de marzo de 2017, ministros de siete países europeos (Francia, Alemania, Italia, Luxemburgo, Países Bajos, Portugal y España) firmaron una declaración conjunta en Roma para establecer el proyecto EuroHPC para adquirir y desplegar una infraestructura informática integrada de alto rendimiento de clase mundial e impulsar las capacidades científicas y la competitividad industrial de Europa. Desde entonces, Bélgica y Eslovenia se han unido a la declaración del EuroHPC en junio y julio de 2017, convirtiéndose en el octavo y noveno país en unirse.
- En la revisión de la mitad de período de la estrategia del mercado único digital, la Comisión anunció su intención de proponer, antes de finales de 2017, un instrumento jurídico para la aplicación de la iniciativa EuroHPC. La Comisión se está preparando actualmente para una evaluación de impacto de esta iniciativa, basada en colecciones de datos existentes y futuras y un análisis de los comentarios recibidos por los



interesados, con el objetivo de definir el instrumento, su estructura de gobierno y presupuesto y así poder poner el EuroHPC en funcionamiento para 2019.

- 2017: apertura por defecto de todos los datos científicos producidos por proyectos futuros en el marco del programa de investigación e innovación Horizonte 2020 de 77.000 millones de euros, para garantizar que la comunidad científica pueda reutilizar la gran cantidad de datos generados.
- 2018: lanzamiento de una iniciativa emblemática para acelerar el desarrollo incipiente de la tecnología cuántica, que es la base de la próxima generación de superordenadores.
- Para el año 2022, desarrollar e implementar una infraestructura europea de computación, almacenamiento de datos y redes de alto rendimiento a gran escala, que incluya la adquisición de dos prototipos de supercomputadoras de última generación, una de las tres mejores del mundo, que establezca un centro de datos grande europeo y la actualización de la red troncal de investigación e innovación (GEANT).

EOSC (European Open Science Cloud)

En octubre de 2017, la Comisión Europea puso a disposición de todas las partes interesadas científicas la Declaración EOSC, para su aprobación y compromiso con la realización del EOSC para 2020. La Declaración surgió de la Cumbre EOSC del 12 de junio, donde más de 80 partes interesadas científicas clave mostraron un fuerte sentido de compromiso, dedicación y rigor intelectual con respecto a la implementación de la EOSC. En general, la respuesta ha sido muy positiva. Hasta hoy, 70 actores científicos han respaldado los principios de la Declaración.

EUDAT (European Data Infrastructure)

El EUDAT - CDI es esencialmente una infraestructura electrónica europea de servicios de datos integrados y recursos para apoyar la investigación. Esta infraestructura y sus servicios se han desarrollado en colaboración con otras instituciones de investigación. El establecimiento de EUDAT CDI apoya la implementación inminente de la European Open Science Cloud, que tiene como objetivo ofrecer servicios abiertos y sin fisuras para el almacenamiento, la gestión, el análisis y la reutilización de la investigación, a través de las fronteras y las disciplinas científicas.

Los investigadores pueden confiar en los servicios de datos innovadores para respaldar su colaboración de investigación y gestión de datos. Se benefician de un marco de gestión de servicios común entregado por los proveedores de servicios CDI y la conexión entre los sitios. Estos proveedores de servicios son líderes europeos en el despliegue de servicios de gestión de datos e infraestructura de

datos. Al unirse al CDI de EUDAT, pueden prestar un mejor servicio a sus clientes locales a través de las fronteras y respaldar las colaboraciones internacionales.

Como ejemplo de un caso de uso de EUDAT se muestra [CLARIN](#).

El siguiente gráfico resume cómo están relacionadas el DSM y la Iniciativa European Cloud.

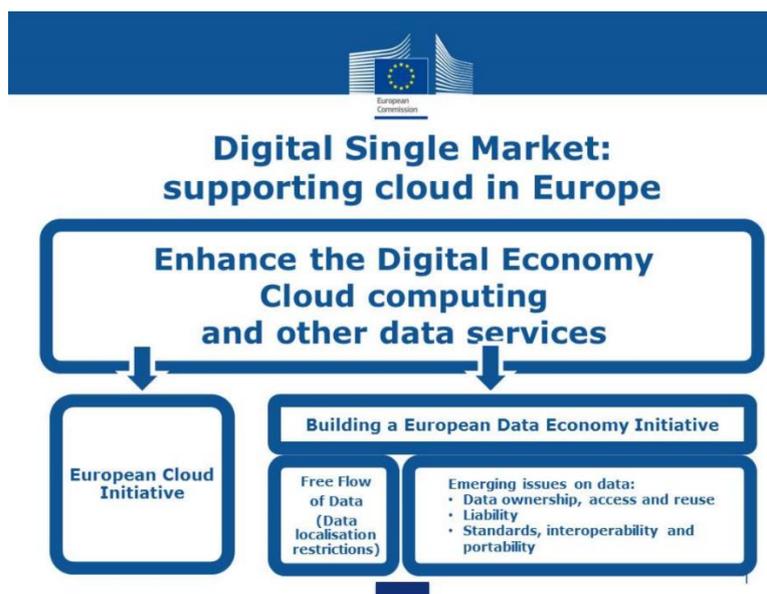


FIGURA 28. EL DSM Y LA INICIATIVA EUROPEAN CLOUD

Estas iniciativas se ven apoyadas para su **financiación** a través del **H2020**, del **Programa Infraestructuras**, concretamente **e-Infraestructuras** en sus convocatorias:

- *Implementing the European Open Science Cloud.*
- *Integrating and opening research infrastructures of European interest.*
- *European Data Infrastructure.*
- *Demonstrating the role of Research Infrastructures in the translation of Open Science into Open Innovation.*

Recomendación: Dada la importancia que tiene para el desarrollo de las tecnologías del lenguaje, principalmente de la traducción automática, el compartir recursos lingüísticos, resultantes de la investigación, desde la Secretaría de Estado se podría fomentar la participación de entidades de investigación españolas, tanto en las iniciativas EOSC como EUDAT, así como en las convocatorias de e-infraestructuras, para que las necesidades propias del área de tecnologías del lenguaje, estén recogidas dentro de estas grandes iniciativas.

7.2.3 Estándares e Interoperabilidad



La estandarización es un componente esencial en la competitividad industrial. El [Reglamento 1025/2012](#) sobre normalización europea establece el marco jurídico en el que operan los actores de la estandarización (la Comisión Europea, las organizaciones europeas de estandarización, la industria, las PYME y las partes interesadas de la sociedad).

La Comisión soporta financieramente el trabajo de 3 Organizaciones de estandarización europeas:

- [ETSI](#) – *the European Telecommunications Standards Institute*.
- [CEN](#) – *the European Committee for Standardization*.
- [CENELEC](#) – *the European Committee for Electrotechnical Standardization*.

Con la comunicación “[Communication on ICT Standardisation Priorities](#)”, la CE propuso focalizar las actividades de estandarización en 5 áreas prioritarias: El 5G, el Internet de las cosas, la computación en la nube, la ciberseguridad y las **tecnologías de datos**. La estandarización en estos ámbitos acelerará la digitalización en áreas como: eSalud, sistemas de transporte inteligente, vehículos conectados y automáticos, ciudades y hogares inteligentes y fabricación avanzada. Esta comunicación complementa a la [European Multi Stakeholder Platform on ICT Standardisation \(MSP\)](#), que es un grupo asesor de expertos en todos los asuntos relacionados con la estandarización europea de las TIC. Sus principales actividades son:

- Detectar las futuras y potenciales necesidades de estandarización de las TIC en apoyo de la legislación, las políticas y la contratación pública europeas;
- Definir las especificaciones técnicas para la contratación pública, desarrolladas por organizaciones mundiales de desarrollo de estándares TIC;
- Establecer la cooperación entre las organizaciones que establecen estándares TIC, Definir el Plan continuo ([Rolling Plan](#)), que proporciona una visión general plurianual de las necesidades de actividades de normalización de las TIC preliminares o complementarias en la UE.

El MSP está compuesto por representantes de autoridades nacionales de los estados miembros y países de la “Asociación Europea de libre comercio – EFTA”, autoridades internacionales y por organizaciones de partes interesadas que representan a la industria, las pequeñas y medianas empresas y los consumidores. Está copresidido por las Direcciones Generales: [Internal Market, Industry, Entrepreneurship and SME](#) and [CONNECT](#). Se reúne cuatro veces al año.

Recomendación: en el desarrollo de las tecnologías del lenguaje, la estandarización, al igual que en otras áreas TIC, es esencial para el desarrollo de productos y servicios que lleguen al mercado. Propiamente, las tecnologías del lenguaje no aparecen como tal en las prioridades, pero sí las

tecnologías de datos, que tienen relación. Por tanto, sería recomendable conocer los avances por parte de estos organismos y darlos a conocer, al sector de tecnologías del lenguaje.

7.2.4 Sociedad Digital

La Comisión Europea quiere una sociedad digital inclusiva que se beneficie del Mercado Único Digital. Por tanto, la construcción de ciudades más inteligentes, los servicios de eSalud y las habilidades digitales permitirán una sociedad europea verdaderamente digital. La sociedad digital europea se está desarrollando aprovechando los beneficios que la estrategia del Mercado Único Digital provee. La Comisión Europea, está llevando a cabo acciones concretas para el desarrollo de servicios públicos digitales transfronterizos y quiere asegurar el uso de herramientas y sistemas digitales que provean mejores y más modernos servicios de eGobierno a las empresas y a los ciudadanos. Para ello la Comisión Europea está llevando a cabo las siguientes iniciativas:

- La política de [digital skills](#) con iniciativas como la *Skills agenda for europe*.
- En eSalud: [eHealth Action Plan](#) and Telemedicine.
- Apoyando tecnologías digitales inteligentes como: [self-driving cars](#) o consumo inteligente de la energía en hogares.
- Establecimiento de servicios armonizados y accesibles para todos los ciudadanos de la Unión Europea, por ejemplo: [emergency and support lines](#) como [the 112 emergency number](#) and [116 helpline](#).

De estas iniciativas, la que más afectaría a las tecnologías del lenguaje, sería la política de Digital Skills con la iniciativa de **“Skills Agenda for Europe”**

El 10 de junio de 2016, la Comisión Europea publicó “The Skills Agend for Europe”, con el fin de fortalecer el capital humano, la empleabilidad y la competitividad. Dicha agenda, presenta una serie de acciones e iniciativas con la ambición de abordar el déficit de habilidades digitales en Europa. Dentro de esta Agenda, la [Digital Skills and Jobs Coalition](#) es el buque insignia, entre otras iniciativas que se presentaron. La Coalición está formado por los Estados Miembros, las empresas, los interlocutores sociales, las organizaciones sin fines de lucro y los proveedores de educación, que toman medidas para hacer frente a la falta de habilidades digitales en Europa. Todas las organizaciones que toman medidas para impulsar las habilidades digitales en Europa pueden convertirse en miembros de la Coalición respaldando los objetivos y principios de la Coalición. Las actividades pueden ser muy variadas, desde formación para gente desempleada, MOOCS (*Massive Open Online Course*) orientados a profesores para dar clases de informática a niños, hasta formación para especialistas TICs.

Para financiar las actividades relacionadas, en algunos casos hacen uso de convocatorias dentro del H2020, relacionadas con Digital Skills, Tenders de la Comisión Europea, Premios, etc.



El **programa Europeo Erasmus Plus**, es el programa encargado de financiar nuevas metodologías de formación. Dentro de su programa de trabajo 2018, existe un área dentro de su Acción Clave 2 llamada: “Sector Skill Alliances”, donde proyectos relacionados con identificación de necesidades en el sector de tecnologías del lenguaje, nuevas metodologías etc., podrían tener lugar. La convocatoria terminó el 28 de febrero. Para más información ver: <http://www.erasmusplus.gob.es/documentos.html>

Recomendación: dentro de la estrategia del DSM se da importancia a la capacitación de la sociedad en habilidades digitales, desde los ciudadanos a los profesionales. En el sector de las tecnologías del lenguaje, también hace falta una especialización de profesionales. Los instrumentos más directos para apoyar a esta especialización, serían conocer y divulgar las iniciativas y seguir las iniciativas identificadas dentro del [Digital Skills and Jobs Coalition](#) para darlas a conocer a los agentes interesados (colegios, universidades, etc.) y en algún caso participar. Otra opción interesante es participar en proyectos europeos dentro del programa Erasmus +, específicamente en los proyectos tipo “Sector Skill Alliances”.

8 DIH (Digital Innovation Hubs) recomendaciones para el posicionamiento institucional de la SESIAD

Los Digital Innovation Hubs (DIHs) son “one-stop-shops” que ayudan a las empresas a ser más competitivas en sus procesos de negocio/producción, productos o servicios utilizando tecnologías digitales, que se crean sobre infraestructuras tecnológicas (Centros de Competencia) y proveen acceso a los conocimientos, experiencia y tecnología más avanzados para apoyar a sus clientes a través de pilotos, pruebas y experimentación con innovaciones digitales. Además, **ofrecen soporte de negocio y financiación** para llevar a cabo dichas innovaciones, si es necesario, a través de la cadena de valor.

Así, un DIH es una **cooperación regional entre diferentes organizaciones** (entidades de investigación universidades, centros tecnológicos, asociaciones industriales, cámaras de comercio, incubadoras/aceleradoras, agencias de desarrollo regional e incluso gobiernos), pudiendo tener también una relación muy fuerte con proveedores de servicios fuera de la región que apoyen a empresas con acceso a sus servicios.

En este estudio se ha realizado un análisis de los Digital Innovation Hubs (DIH) existentes a nivel europeo y sus principales características, con el fin de identificar su potencial como herramienta para el impulso de las Tecnologías del Lenguaje (TL).

Por tanto, en una primera sección se **describen los DIHs y sus principales características**: participantes, modelo de gestión y sostenibilidad, detalle de servicios que ofertan a las diferentes entidades que acuden a ellos y qué indicadores serán necesarios para evaluar el impacto de los DIH. Por último, en esta sección también se enumerarán una serie de proyectos europeos y tópicos de programas que abordan esta casuística de los Digital Innovation Hubs.

La sección concluye con el **análisis de los DIHs europeos** que pudieran estar relacionados con las tecnologías del lenguaje. De esta manera se puede observar cómo existen a nivel europeo y a nivel nacional una serie de DIH que tienen como competencias técnicas las tecnologías interactivas (interacción humano-máquina, reconocimiento de movimiento y TL) dónde se podría posicionar el uso de las Tecnologías del Lenguaje. Si bien, cabe decir que, pese a que todos estos DIH indican que tienen estas competencias técnicas en concreto, **ninguno de ellos es propiamente y exclusivamente de TL**. (En el documento de soporte, anexo 3, se incluye un listado de los DIHs relacionados con las TL en Europa y en España).

En la segunda y última sección, se relacionan los 4 ejes del Plan de Impulso de las Tecnologías del Lenguaje (Plan TL) con la aportación de los Digital Innovation Hubs a dichos ejes. El objetivo es analizar cómo **los DIH pueden ser una herramienta útil, una palanca, para implementar los ejes del Plan TL**.

Asimismo, se muestran **los programas nacionales y europeos**, que pueden apoyar a la generación de los DIH (red.es centros de competencia a Interreg) y los programas europeos que apoyan la puesta en red de dichos DIH (Horizon 2020).

8.1 DIH

8.1.1 Concepto

Digital Innovation Hubs son **“one-stop-shops”** que ayudan a las empresas a ser más competitivas en sus procesos de negocio/producción, productos o servicios utilizando tecnologías digitales. Se crean sobre infraestructuras tecnológicas (Centros de Competencia) y proveen acceso a los conocimientos, experiencia y tecnología más avanzados para apoyar a sus clientes a través de pilotos, pruebas y experimentación con innovaciones digitales. DIHs también ofrecen soporte de negocio y financiación para llevar a cabo dichas innovaciones, si es necesario, a través de la cadena de valor.



As **an innovation ecosystem** that provides access to the services, facilities and expertise of a wide range of partners, Digital Innovation Hubs ensure that individual customers get the services they need; that the target market segments receive innovative, scalable solutions; and that DIHs co-operate effectively with each other.

FIGURA 29. MODELO DE DIGITAL INNOVATION HUB

¿QUÉ DIFERENCIA A LOS DIH?

Iniciativas con algunas de estas características son ciertamente evidentes en diferentes lugares de Europa y en varios sectores:

- Universidades y centros tecnológicos de investigación (RTOS), (normalmente conocidos como centros de competencia (CCs)) que proporcionan su experiencia y facilitan el acceso a las instalaciones avanzadas a la industria.
- Empresas privadas (grandes y pequeñas) tienen productos útiles y servicios para la digitalización de procesos, productos y servicios.
- Incubadoras y aceleradoras que ayudan a empresas de nueva creación a crecer y a desarrollarse.

- Clusters, asociaciones industriales que representan a empresas individuales que también juegan un papel importante con respecto a la innovación de un sector. Los inversores ya proporcionan acceso a la financiación.
- Autoridades locales son conscientes de la importancia de la innovación y están desarrollando sus planes inteligentes de especialización.

Lo que diferencia a los DIHs es que ellos agrupan todos estos actores en una región y desarrollan un grupo de servicios coherente y coordinado que es necesario para animar a estas empresas (especialmente pequeñas o empresas de sectores de baja tecnología) que tienen dificultades con la digitalización, a través un único ente. La visión holística de la **digitalización como un proceso de transformación empresarial** permite a las empresas no solamente identificar las soluciones técnicas, sino financiar y alimentar la innovación hasta un nivel que puede ser implementado dentro de la empresa y contribuir a mejorar la competitividad. La figura 30 muestra como los centros de competencia pueden colaborar con otros actores de la innovación en los DIH de cara a proporcionar una serie de servicios de transformación digital a la industria

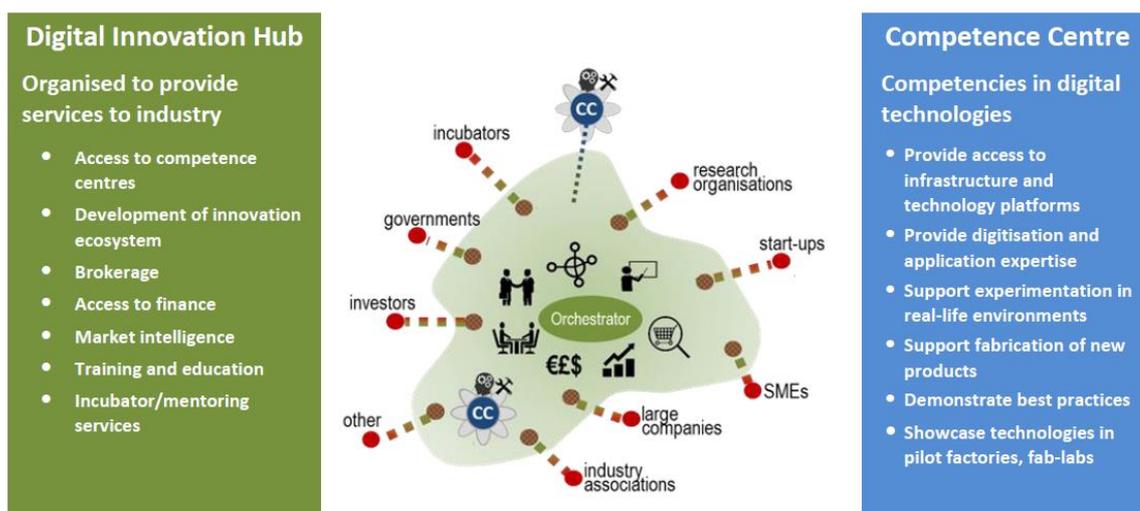


FIGURA 30 CENTROS DE COMPETENCIA COMO UN ELEMENTO DE LOS DIH

8.1.2 Miembros

Un DIH es una **cooperación regional entre diferentes organizaciones** (entidades de investigación universidades, centros tecnológicos), asociaciones industriales, cámaras de comercio, incubadoras/aceleradoras, agencias de desarrollo regional e incluso gobiernos). Pueden, también, tener una relación muy fuerte con proveedores de servicios fuera de la región que apoyen a empresas con acceso a sus servicios.

Existe una web con un **catálogo** de DIH en Europa: <http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs-tool>



Por cada DIH identificado, hay una ficha con una descripción muy interesante, de entidad coordinadora, socios, lugar y servicios prestados, así como de clientes que disponen. En España hay identificados 48.

Tamaño: Hay tamaños muy variados desde 2 entidades a 20 entidades y bastantes de 5 – 8 entidades.

Cobertura geográfica: regional.

Sector intersectorial: todos ellos están centrados en las tecnologías digitales. Pero, posteriormente se especializan en áreas, en principio áreas que sean acordes a las necesidades de la industria, sectores de la región y de acuerdo a las competencias que ya tenga la región: conocimiento, centros tecnológicos, universidades, etc.

8.1.3 Modelos de Negocio

Modelos de negocio basados en la financiación mixta y que evolucionan con el tiempo.

DIH son servicios enriquecedores de diferentes tipos, de naturaleza mixta, pública y privada. Requieren un modelo de negocio híbrido que combine fuentes de financiación mixtas, públicas y privadas. Dependiendo de la situación, el dinero puede ser necesario para construir y mantener la infraestructura, comprar equipamiento y contratar personal cualificado. La financiación pública puede ser necesaria para implicar a empresas usuarias y proveedoras en la experimentación específica del proyecto. En el lado de los ingresos, tasas de membresía, formación, contratación de I+D, ensayos, y servicio de intermediación son vías de ingresos potenciales.

Bienes públicos y servicios son información y conocimiento que puede ser compartido y que revela externalidades positivas: por consiguiente, pueden ser subvencionadas. Los bienes privados y servicios son adecuados para clientes que deberían de pagar el precio de mercado. Parte de la financiación de infraestructuras podría ser entonces proporcionada por medio de contratos con empresas del sector privado (que también cubrirían el uso de infraestructuras) y por ayudas a la investigación. Esto mismo es válido para la formación que se lleva a cabo en las infraestructuras. La combinación de diferentes servicios y funciones para utilizar completamente la capacidad de la infraestructura puede ser trasladada a un modelo de negocio y plan financiero que podría facilitar el atraer a bancos u otros agentes financieros con una perspectiva de inversión a largo plazo en la implantación de nuevas infraestructuras o la ampliación de las ya existentes. Otro aspecto a considerar es la imagen y reputación, modelos de gestión y gobernanza, propiedad de activos. Durante la primera reunión del grupo de trabajo, 24/10/2016, se presentó un ejemplo de dicho modelo de negocio.

UTILIZACIÓN DE FONDOS PÚBLICOS COMO MEDIO DE INVERSIONES PRIVADAS DE RIESGO

El modelo de negocio solamente puede tener en cuenta los retornos financieros directos, pero el retorno indirecto sobre la inversión proporciona una garantía para los riesgos financieros. El efecto multiplicador en el seguimiento de inversiones privadas será mayor cuanto más satisfactorios sean los servicios de innovación, en lo que respecta a la reducción de los riesgos de la inversión privada en la modernización de la industria, ya que todavía en la actualidad se es reacio a ello.

La utilización de infraestructuras compartidas reduce costes e incrementa los retornos para la experimentación. Cuando más satisfactorias sean las instalaciones, mayor será la capacidad a utilizar y los servicios serán totalmente pagados con fondos privados.

FINANCIACIÓN EUROPEA PARA COLABORACIÓN TRANSFRONTERIZA Y SERVICIOS

La comunicación de la Comisión prevé que la financiación de servicios transfronterizos (tales como el deseo de utilizar los servicios de un DIH ubicado fuera de su región o país) podría ser financiado a nivel europeo a través del H2020. Bienes y servicios públicos locales podrían ser financiados por fondos públicos, por ejemplo, con fondos EIE, especialmente FEDER.

8.1.4 Servicios

Los DIH deben trasladar su propuesta de valor a una única oferta que se dirija a las necesidades específicas de las compañías a las que sirve y no debe replicar el actual suministro de servicios.

En términos generales, los servicios disponibles gracias a los DIH pueden ser categorizados en estas tres categorías, con los servicios divididos como se muestra:

- **Actividades de innovación**, relativas a la identificación de oportunidades de digitalización, desarrollo y validación de soluciones innovadoras basadas en tecnología vanguardista.
- **Desarrollo de negocio**, que ayudará a las empresas a aplicar sus soluciones, evaluará las implicaciones de negocio y gestionará los cambios resultantes.
- **Creación de habilidades**, construcción de capacidad innovadora a través del enriquecimiento del capital humano.

Innovation Activities	Business Development	Skills Creation
<ul style="list-style-type: none"> • Awareness creation • Digital needs/maturity assessment • Innovation scouting • Access to specialist expertise • Access to platforms & infrastructure • Collaborative research 	<ul style="list-style-type: none"> • Envisioning & strategy development • Matchmaking & brokering • Business coaching & mentoring • Start-up support • Access to finance 	<ul style="list-style-type: none"> • Technical training & skills development • Business & finance training & skills development • Management training & skills development

TABLA 6. SERVICIOS DE DIH



8.1.5 Indicadores

De cara a medir el impacto de los DIH, serán necesarios indicadores. Tales medidas deberían abarcar desde las relaciones individuales cliente-hub, al rendimiento global del hub y el impacto en el ecosistema de DIH. Se debería incidir particularmente en la calidad y el impacto de relaciones colaborativas (indicadas más adelante), es principalmente la fortaleza de estos vínculos, más que las actividades no conectadas los que definirán los DIHs. Los indicadores de actuación clave a nivel de hub podrían incluir:

- Número de usuarios de DIH.
- Número de referencias de compañías TIC y otros proveedores de servicios.
- Porcentaje de usuarios con actividades de digitalización satisfactorias.
- Porcentaje de usuarios que vuelven y/o requieren otros servicios.
- Número de eventos y participación en eventos.
- Número de colaboraciones y vínculos transfronterizos.
- Formación proporcionada e incremento de habilidades digitales.
- Cantidad de financiación externa asegurada para empresas clientes.

Indicadores selectivos de medidas podrían ser utilizados para establecer referencias y estándares de servicios, así como intercambio de buenas prácticas.

Se podría medir un mayor impacto utilizando **medidas econométricas**, tales como la concienciación creciente, competitividad, mejora y evaluación de la maduración digital. Los ejemplos incluyen: incremento de la cuota de mercado de la empresa; creación de valor vía nuevos mercados y modelos de negocio; establecimiento de nuevas cadenas de valor; aumento del ratio de facturación entre servicios y productos; cuantificación de la reducción de costes de servicios y optimización de recursos gracias a la digitalización; número de patentes y otro tipo de protección intelectual (p.ej. diseños registrados); número de proyectos innovadores (p.ej. hackatones); número de personas formado en habilidades digitales.

La monitorización sistemática no debería ser complicada ni cara. Mientras que algunos parámetros de rendimiento son ciertamente necesarios, **la valoración del usuario y redes sociales** deberían ser también observados. Los usuarios deberían ser capaces de compartir sus experiencias en el hub a través de un motor tipo “TripAdvisor”, para así crear un entorno de usuarios relativo a servicios de digitalización. Las redes sociales también deberían ser utilizadas como medio de evaluación del rendimiento del hub.



8.2 Listado de proyectos europeos y líneas del H2020 en los que aspectos relacionados con DIH han sido financiados o se pueden financiar

Financiados por el H2020,

- Aceleradoras FIWARE: aunque no es exactamente el concepto de DIH: pero se partía de una infraestructura tecnológica que se había desarrollado en otros proyectos europeos que dio lugar a FIWARE y con estas aceleradoras lo que se quería eran desarrollar aplicaciones informáticas que utilizaran esta tecnología. La Comisión financió 16. El listado de las 16 están en: <https://www.fiware.org/fiware-accelerator-programme/>. Una media de 4M€ a repartir en cada aceleradora a SMEs, Start Ups y emprendedores. En estas aceleradoras, además de financiación para llevar a cabo sus aplicaciones sobre FIWARE, se les prestaban servicios de apoyo al desarrollo de negocio, mentorización, formación, apoyo a la búsqueda de financiación privada y pública. El número de socios medio por aceleradora eran unos 5.
- En los ICT Workprogrammes 2014 – 2015 y 2016 – 2017 como Digital Innovation Hubs, no hay tópicos. Pero en el Anexo 3.1 proyectos con la palabra “Hub” dentro de ICT y FoF que están publicados en Cordis. Están incluidos proyectos hasta las convocatorias de 2016.
- En el ICT Workprogramme 2018 – 2020. Hay los siguientes tópicos para DIH
 - *DT-ICT-01-2019: Smart Anything Everywhere* – Presupuesto hasta 8.000.000. No pone el % dedicado a Cascade Funding, pero para cada proyecto financiado por cascade funding estiman entre 20.000€ y 100.000€
 - *DT-ICT-02-2018: Robotics - Digital Innovation Hubs (DIH)*. Presupuesto hasta 16.000.000. El 50% para Cascade funding (a cada Proyecto hasta 300.000)
 - *DT-ICT-03-2020: I4MS (phase 4) - uptake of digital game changers and digital manufacturing platforms*
 - *DT-ICT-04-2020: Photonics Innovation Hubs*
 - *DT-ICT-05-2020: Big Data Innovation Hubs*
 - *DT-ICT-06-2018: Coordination and Support Activities for Digital Innovation Hub network*

8.3 DIH en tecnologías del lenguaje

Si se realiza una búsqueda de DIH en el catálogo europeo de DIH, y se selecciona la “competencia técnica”: *Interaction technologies (e.g. human-machine Interaction, motion recognition and language technologies)*, se encuentra la distribución de DIHs que se muestran en el gráfico

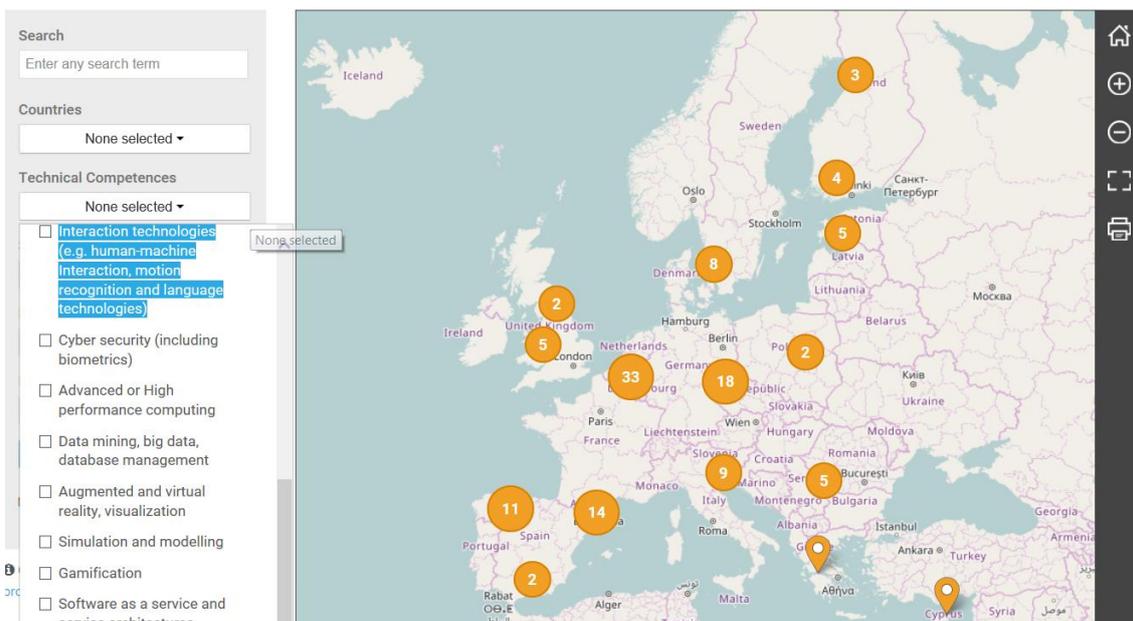


FIGURA 31. MAPA DE DIH EN EUROPA RELACIONADOS CON TECNOLOGÍAS DEL LENGUAJE

En el Anexo 3.2 y 3.3 del documento de soporte, se adjunta un fichero con la lista de todos los DIH que han indicado “*Technical Competences*” con: *interacción technologies (e.g. human machine interaccion, motion recognition and language Technologies)* en Europa y en España.

Si se detalla la búsqueda sobre España, el resultado es el siguiente:

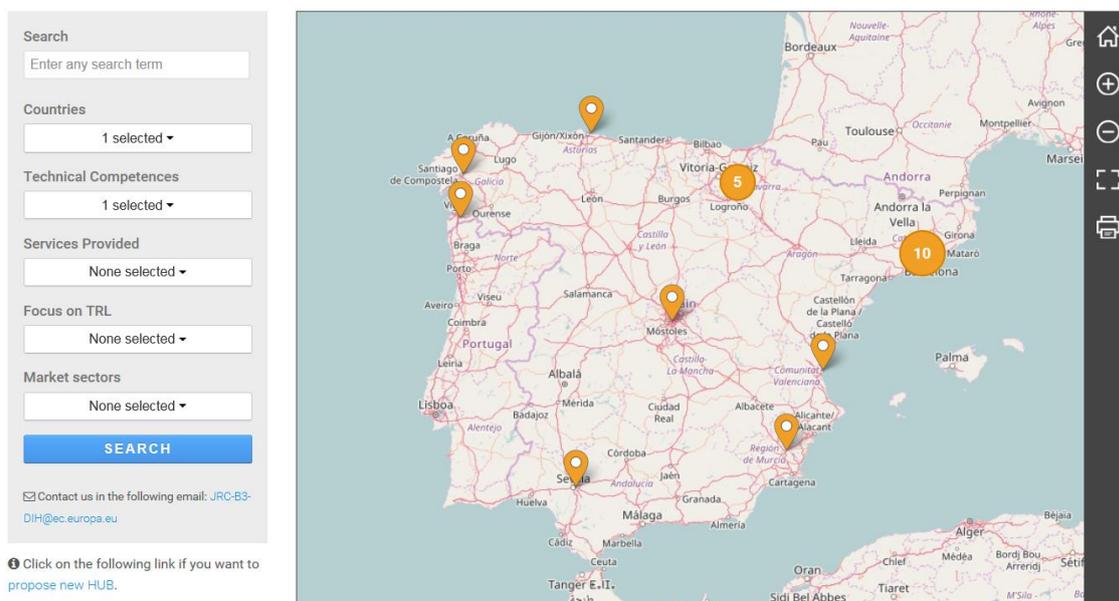


FIGURA 32. MAPA DE DIH EN ESPAÑA RELACIONADOS CON TECNOLOGÍAS DEL LENGUAJE

A continuación, se muestra la lista de las entidades identificadas en el mapa:

Hub Name
Barça Innovation Hub (BIHUB)
Basque Digital Innovation Hub (BDIH)
Catalan Robotics Hub
Catalunya Industry 4.0
Centre d'Innovació i Tecnologia, UPC
Digital Innovation Hub , D-LIGHT
Digital Urban Development (DUD)
DIH on Healthcare – Basque Country (Data Science Alkmaar)
DIH on Traceability and Big data (TBD)
EFFIRENP (Efficient Ren Power HUB)
eHealth Catalonia
Experience-based industries Hub (e xpérience)
Galician Advance Manufacturing Innovation Consortia
Galician Automotive Hub
HPC-Cloud and Cognitive Systems for Smart Manufacturing processes, Robotics and Logistics.
Hub 4.0 of Manufacturing Sectors in Valencian Region
iAsturias 4.0
Innovation for Manufacturing in the South (I4MSOUTH)
RIOHUB
RoboCity2030
SmartCityTech
Sustainable Construction Hub, SCH

TABLA 7. LISTADO DE DIH ESPAÑOLES RELACIONADOS CON TECNOLOGÍAS DEL LENGUAJE

En cualquier caso, cabe destacar que **ninguno de estos DIH es propiamente y exclusivamente de Tecnologías del Lenguaje.**

8.4 Los DIH como herramienta para el PLAN TL

La siguiente tabla, establece las relaciones entre los objetivos del Plan de Impulso y los DIH indicando qué pueden aportar los DIH a los ejes del Plan:

EJES DEL PLAN DE IMPULSO TL	APORTACIÓN DE LOS DIH A LOS EJES
<p><i>Eje 1: Apoyo al desarrollo de infraestructuras lingüísticas</i></p> <p>Este eje pretende facilitar el desarrollo de procesadores (reconocedores de entidades, desambiguadores, calculadores de proximidad semántica, etc.) y recursos lingüísticos (corpus paralelos, diccionarios, taxonomías, etc.) que sirvan de combustible para el desarrollo de la industria española de procesamiento de lenguaje natural y de traducción automática.</p>	<p>Como se ha identificado por varios estudios, como la RIA o en entregables del proyecto LT_Observatory, el desarrollo de infraestructuras lingüísticas es necesario, pero también es necesario que dichas infraestructuras lingüísticas estén disponibles para todos los actores (investigadores e industria) del sector de las tecnologías del lenguaje. Por ello, disponer de Centros de Competencia, de Tecnologías del Lenguaje,</p>

EJES DEL PLAN DE IMPULSO TL	APORTACIÓN DE LOS DIH A LOS EJES
	<p>que permitieran disponer de infraestructuras de tecnologías del lenguaje, vivas y sostenibles, para las empresas del sector, podría ser de gran ayuda.</p>
<p><i>Eje II: Impulso de la Industria de las Tecnologías del Lenguaje</i></p> <p>El segundo de los ejes responde al objetivo de apoyar la transferencia de conocimiento entre el sector investigador y la industria, así como la internacionalización de las empresas e instituciones que componen el sector.</p>	<p>Los DIH son “one-stop-shops” que ayudan a las empresas a ser más competitivas en sus procesos de negocio/producción, productos o servicios utilizando tecnologías digitales. Se crean sobre infraestructuras tecnológicas (Centros de Competencia) y proveen acceso a los conocimientos, experiencia y tecnología más avanzados para apoyar a sus clientes a través de pilotos, pruebas y experimentación con innovaciones digitales.</p> <p>Es decir, el objetivo de los DIH es que las empresas mejoren su competitividad, ofreciéndoles una serie de servicios, que les ayuden a aprovechar las ventajas que las nuevas tecnologías digitales en un ámbito les ofrece. Para ello les ofrecen servicios como:</p> <ul style="list-style-type: none"> • Actividades de innovación, relativas a la identificación de oportunidades de digitalización, desarrollo y validación de soluciones innovadoras basadas en tecnología vanguardista. • Desarrollo de negocio, que ayudará a las empresas a aplicar sus soluciones, evaluará las implicaciones de negocio y gestionará los cambios resultantes.

EJES DEL PLAN DE IMPULSO TL	APORTACIÓN DE LOS DIH A LOS EJES
	<ul style="list-style-type: none"> • Creación de habilidades, construcción de capacidad innovadora a través del enriquecimiento del capital humano.
<p><i>Eje III: La Administración Pública como impulsor de la Industria del Lenguaje</i></p> <p>Este eje se propone la creación de sendas plataformas comunes de procesamiento del lenguaje y de traducción automática para las Administraciones Públicas. Además, la política de reutilización de la información del sector público (RISP) representa un cauce para el desarrollo de estos recursos lingüísticos.</p>	<p>En este caso los DIH, consideran a la Administración Pública como impulsores de la industria, ya que la Administración Pública apoya inicialmente, en los primeros años al desarrollo y lanzamiento de los DIH, hasta su consolidación como una estructura sostenible.</p> <p>DIH son servicios enriquecedores de diferentes tipos, de naturaleza mixta, pública y privada. Bienes públicos y servicios son información y conocimiento que puede ser compartido y que revela externalidades positivas: por consiguiente, pueden ser subvencionadas. Los bienes privados y servicios son adecuados para clientes que deberían de pagar el precio de mercado.</p>
<p><i>Eje IV: Proyectos faro de tecnologías de procesamiento de lenguaje natural</i></p> <p>El cuarto de los ejes se orienta al impulso de los proyectos faro, que son proyectos de aplicación de las tecnologías del lenguaje natural, emprendidos inicialmente por las Administraciones Públicas en sectores estratégicos (Sanidad, Turismo y Educación), y abiertos a la futura iniciativa público-privada, que pretenden servir de demostración de sus capacidades y beneficios, generar industria y crear recursos reutilizables en otros proyectos. También servirán de aprendizaje para futuros desarrollos</p>	<p>Los DIH, pueden contribuir a la tracción en el sector. Impulsando la transferencia de la tecnología, nuevos proyectos innovadores en el sector, y un mayor conocimiento en la industria y en la sociedad de las ventajas que pueden aportar las tecnologías del lenguaje.</p>

TABLA 8. CÓMO APOYAN LOS DIH A LOS EJES DEL PLAN IMPULSO TECNOLOGÍAS DEL LENGUAJE



Como conclusión, se podría decir que los DIH, pueden ser una herramienta útil, una palanca, para implementar los ejes del Plan TL.

8.5 Herramientas para apoyar la generación de DIH

En el ámbito europeo, las ayudas existentes relacionadas con la generación de DIH en H2020 y INTERREG son:

H2020:

- **ICT Workprogramme 2018 – 2020.**
 - DT-ICT-01-2019: Smart Anything Everywhere – Presupuesto hasta 8.000.000€. No pone el % dedicado a Cascade Funding, pero para cada proyecto financiado por cascade funding estiman entre 20.000€ y 100.000€
 - DT-ICT-02-2018: Robotics - Digital Innovation Hubs (DIH). Presupuesto hasta 16.000.000€. El 50% para Cascade funding (a cada Proyecto hasta 300.000€)
 - DT-ICT-03-2020: I4MS (phase 4) - uptake of digital game changers and digital manufacturing platforms
 - DT-ICT-04-2020: Photonics Innovation Hubs
 - DT-ICT-05-2020: Big Data Innovation Hubs
 - DT-ICT-06-2018: Coordination and Support Activities for Digital Innovation Hub network

De estos tópicos, el que más relación tiene con las tecnologías del lenguaje sería el [DT-ICT-02-2018: Robotics - Digital Innovation Hubs \(DIH\)](#). Propiamente no es un tópico de tecnologías del lenguaje, pero el uso de las mismas podría tener sentido. La Comisión no apoya la generación de nuevos DIH, sino la puesta en red y la prestación de servicios en red.

INTERREG – Programa de cooperación regional

Para las entidades españolas, los programas INTERREG interesantes son:

- [SUDOE](#)
- [ATLANTIC](#)
- [POCTEP](#)
- [POCTEFA](#)



En cada uno de ellos hay ejes/prioridades principales en los que se deben enmarcar los proyectos. Dentro de estos ejes, normalmente uno tiene relación con la promoción de las PYMES, y es en ese marco que la generación de un DIH podría ser posible. Para los Interreg indicados anteriormente los ejes serían:

- SUDOE: [Eje 2 Competitividad de las PYMES](#)
- ATLANTIC: [Prioridad 1, estimulando la innovación y la competitividad](#)
- POCTEP: Eje 2: [Mejorar la competitividad de las pequeñas y medianas empresas](#)
- POCTEFA: Eje 1: [Dinamizar la innovación y la competitividad](#)

Los proyectos suelen ser de 2-3 años y con una financiación entre el 70-80% en subvención a fondo perdido y con presupuestos entre 1-3 millones de euros.

En el **ámbito nacional**: existen los convenios entre Red.es con Ayuntamientos o Entidades Locales para la generación de “Centros de Competencia” como es el ejemplo de “Hub de contenidos digitales de Málaga”.

Si se consideran los DIH, una herramienta útil para la potenciación del sector de tecnologías del lenguaje, **desde el Plan TL se podría potenciar la creación de dichos DIH, a través:**

- Estableciendo convenios entre red.es y los Ayuntamientos
- Fomentando a nivel regional, con las entidades regionales públicas, la participación en proyectos Interreg.
- Financiando directamente la generación de DIH



9 Conclusiones

En base a la información mostrada en las secciones anteriores, a continuación, se describen las principales conclusiones

El peso de las TL dentro de los programas de Tecnologías de la Información y las Comunicaciones (TIC) europeos. Áreas financiadas y a financiar de las TL.

El impulso, fomento y apoyo a las tecnologías del lenguaje, a su desarrollo y su adopción ha sido y es una realidad a nivel europeo. De hecho, desde 2007 la EC ha venido apoyando a través de sus instrumentos de financiación proyectos del ámbito de las tecnologías del lenguaje. Las líneas objeto de los proyectos han sido:

- hasta 2007 en el marco del 7PM (7º Programa Marco): *semantic technologies, knowledge management technologies, integration of language-enabled content technologies (combination of content authoring, management, translation and publishing tasks and tools within typical production processes and translation/localisation workflows, in real-life multilingual settings), advancing machine translations, natural spoken interaction, Information access and mining, Cross-media content analytics.*
- Entre 2009-2010 con el programa CIP ICT PSP (2009-2010) (*Competitive and Innovation Framework Programme*) más que el apoyo a la investigación en sí, se dio prioridad al desarrollo de nuevos productos que potenciaran el multilingüismo y los servicios en internet. Es por ello, que los usuarios finales muy vinculados a la validación, jugaban un papel fundamental.
- A partir de 2014, en el marco del H2020 (Horizon 2020), el único tópico del área estaba orientado a la Traducción Automática. A partir de entonces, hasta el año 2017, las tecnologías del lenguaje pierden fuerza como área propia de investigación, y se incluyen dentro del área de *Big Data*. Las líneas apoyadas fueron: *language understanding technologies for carrying out data analytics, data quality assessment and improvement, prediction and visualization tasks at extremely large scale and with diverse structured and unstructured data, advanced machine translation, re-use, sharing of language resources.*
- CEF TELECOM (2014-2020) (*Connecting European Facilities*): desde su inicio ha mantenido sus líneas de financiación, sin introducir cambios relevantes. Cada año, en las convocatorias que financia, sigue considerando una convocatoria para “*Automated Translation*” orientada a aumentar la recopilación de recursos lingüísticos en diferentes dominios, y a mejorar y fomentar el servicio de la Comisión Europea MT@EC.



- En el programa de trabajo TIC 2018 – 2020 del H2020 vuelve a aparecer un tópico relacionado con las tecnologías del lenguaje. Ya no está encuadrado dentro de Big Data, sino en el área de “Next Generation Internet”. En este caso no está tan dirigido a la traducción automática, sino más a las tecnologías del lenguaje en un ámbito más amplio y dando importancia a su aplicación en sectores industriales de diferentes ámbitos, es decir proyectos de investigación e innovación con aplicación en el mercado.

Lo que ocurrirá en el FP9, todavía es una incertidumbre, pero consideramos que el área de “Next Generation Internet” se está posicionando en el FP9 y si las tecnologías del lenguaje están ahí enmarcadas, pueden tener un lugar en el FP9. Eso sí consideramos que el ámbito de las convocatorias será más amplio, y continuará y aumentará los proyectos de investigación e innovación, pero con aplicación real al mercado. Asimismo, las áreas de actuación, entendemos que están centradas en el desarrollo de tecnologías que permitan un Mercado Único Digital multilingüe, con ello, creemos que el ámbito de actuación será más amplio que la traducción automática, pero es cierto que la traducción automática tiene un papel fundamental en la evolución de las tecnologías del lenguaje.

Un punto de inflexión podría ser si la propuesta presentada a la convocatoria [FETFLAG-01-2018](#) del H2020 que serviría para la preparación del gran proyecto *Human Language Project* si fuera financiada. En ese caso se englobaría, en dicho proyecto, gran parte del presupuesto destinado a la investigación y el desarrollo. Por las entidades que han promovido esta propuesta (DFKI), entendemos que las líneas de investigación estarán en la línea de las líneas de investigación de la Agenda Estratégica SRIA: *Cross-lingual Big Data Language Analytics, High-Quality Machine Translation, Meaning, Semantics, Knowledge y Conversational Technologies*.

Importancia de las TL en las estrategias europeas.

En la estrategia europea del Mercado Único Digital, se han identificado 4 iniciativas que afectan en mayor medida al sector de las tecnologías del lenguaje:

- Un marco de derechos de autor (Copyright) moderno y más europeo.
- Abordar las barreras de la economía de datos europea.
- Definir prioridades para los estándares y la interoperabilidad.
- Apoyar una sociedad digital inclusiva.

Por cada una de ellas se han identificado los siguientes puntos a tener en cuenta en el desarrollo del PLAN TL:

- Seguimiento de la Regulación y la directiva de “Copyright” en un Mercado Único Digital y cómo ésta afecta a la utilización de contenidos en Internet así como a resultados de investigación,



para su utilización en investigación o en productos/servicios comercializables, con el fin de poder divulgar dicha información al sector de las Tecnologías del Lenguaje español.

- Dada la importancia que tiene para el desarrollo de las tecnologías del lenguaje, principalmente de la traducción automática, el compartir recursos lingüísticos, resultantes de la investigación, desde la Secretaría de Estado se podría fomentar la participación de entidades de investigación españolas, tanto en las iniciativas EOSC (European Open Science Cloud) como EUDAT (European Data Infrastructure), así como en las convocatorias de e-infraestructuras, para que las necesidades propias del área de tecnologías del lenguaje, estén recogidas dentro de estas grandes iniciativas.
- El MSP ([European Multi Stakeholder Platform on ICT Standardisation](#)) está compuesto por representantes de autoridades nacionales de los estados miembros y países de la “Asociación Europea de libre comercio – EFTA”, autoridades internacionales y por organizaciones de partes interesadas que representan a la industria, las pequeñas y medianas empresas y los consumidores. Está copresidido por las Direcciones Generales: [Internal Market, Industry, Entrepreneurship and SME](#) and [CONNECT](#). En el desarrollo de las tecnologías del lenguaje, la estandarización, al igual que en otras áreas TIC, es esencial para el desarrollo de productos y servicios que lleguen al mercado. Propiamente, las tecnologías del lenguaje no aparecen como tal en las prioridades, pero sí las tecnologías de datos, que tienen relación. Por tanto, sería recomendable conocer los avances por parte de estos organismos y darlos a conocer, al sector de tecnologías del lenguaje.
- En estrategia del Mercado Único Digital se da importancia a la capacitación de la sociedad en habilidades digitales, desde los ciudadanos a los profesionales. En el sector de las tecnologías del lenguaje, también hace falta una especialización de profesionales. Los instrumentos más directos para apoyar a esta especialización, serían conocer y divulgar las iniciativas y seguir las iniciativas identificadas dentro del [Digital Skills and Jobs Coalition](#) para darlas a conocer a los agentes interesados (colegios, universidades, etc.) y en algún caso participar. Otra opción interesante es participar en proyectos europeos dentro del programa Erasmus +, específicamente en los proyectos tipo “Sector Skill Alliances”.

Importancia de resultados cercanos y aplicables al mercado en los proyectos europeos de I+D+i

Es una realidad que, en la Comisión Europea, **se otorga, cada vez más importancia a que los resultados de I+D+i lleguen al mercado**. Esto es debido a que es una realidad, que en Europa, hay una investigación de alta calidad, pero que los resultados de la misma, no se traducen en productos que lleguen al mercado y aumenten o generen nuevos negocios y mayor riqueza. Por ello, en el H2020, el



criterio de evaluación de impacto ha adquirido más valor. En algunos instrumentos como las “Acciones de Innovación” se requiere, que la nota mínima sea un 4 sobre 5. Anteriormente se pedía un 3 sobre 5.

Además, en los **consorcios**, la presencia de **usuarios finales**, que permitan definir sus necesidades y los requerimientos que los resultados de las investigaciones y desarrollos deben tener, está muy bien valorado. Asimismo, los **socios encargados de explotar**, desde un punto de vista comercial, los resultados del proyecto también tienen que formar parte del consorcio, y demostrar que durante el proyecto estudiarán cómo será la futura explotación comercial de sus resultados.

En el H2020, también se han incluido dos **nuevos instrumentos**, el *SME Instrument* y el *Fast Track to Innovation (FTI)*, claramente orientados a facilitar que los resultados de investigación lleguen al mercado, tal y como se ha explicado en el punto de Instrumentos de Financiación. La financiación en cascada, que ayuda a que publiquen convocatorias para que, normalmente, PYMES, start ups, emprendedores, lleven a cabo proyectos de desarrollo de productos/servicios basados en tecnologías desarrolladas durante los proyectos es otra muestra clara de esta tendencia de la Comisión Europea. Asimismo, varias convocatorias dentro del H2020, han financiado la creación de aceleradoras, que a su vez apoyan a que las PYMES y las start ups, lleven sus productos/servicios innovadores al mercado.

En el marco de la Comisión Europea también se ha introducido el concepto de los **Digital Innovation Hubs**, que tal como se ha explicado antes, son “one-stop-shops” que ayudan a las empresas a ser más competitivas en sus procesos de negocio/producción de productos o servicios, utilizando tecnologías digitales. Se crean sobre infraestructuras tecnológicas (Centros de Competencia) y proveen acceso a los conocimientos, experiencia y tecnología más avanzados para apoyar a sus clientes a través de pilotos, pruebas y experimentación con innovaciones digitales. DIHs también ofrecen soporte de negocio y financiación para llevar a cabo dichas innovaciones, si es necesario, a través de la cadena de valor. Este nuevo instrumento, puede ser muy útil para potenciar el sector. Para su financiación se podrían utilizar los programas Interreg.

Por todo ello, se puede concluir, que en los programas europeos la I+D es importante, pero no se puede olvidar y hay que tener en cuenta el potencial mercado y dependiendo del instrumento que se utilice en las convocatorias, el mercado tiene mayor o menor importancia.

Aplicación sectorial de los proyectos de TL

Tal y como se ha descrito en el punto anterior, el mercado cada vez tiene una mayor importancia en los programas europeos. Por ello, en los proyectos la aplicación de la investigación y los desarrollos tecnológicos a un sector concreto se va a dar, porque es consecuencia de un mayor acercamiento al mercado. Esta aplicación sectorial, es fomentada por la Comisión Europea de dos formas:



- En convocatorias tecnológicas, por ejemplo de TICS, señala que los casos de uso deben dirigirse a unos sectores concretos.
- En convocatorias de retos sociales: salud, transporte, seguridad, indica que el desarrollo de soluciones basadas en la tecnología “X” podrían ser muy interesantes.

Desde el punto de vista de nuestra experiencia, el indicar que se hagan casos de uso e indicar ciertos sectores, obliga a que los aspectos de mercado se tengan más en cuenta y que los resultados puedan estar más cerca del mercado. Consideramos que la I+D básica es necesaria, pero también es necesario que la I+D llegue al mercado, y por ello programas, como el Plan TL, que promuevan que la investigación se aplique, suelen ser de gran interés para las empresas y suelen tener un mayor impacto.

Interés de los Países Europeos en las TL

De acuerdo a los datos, obtenidos de la participación en los programas relacionados con tecnologías del lenguaje en el 7PM y el H2020, parece que la participación de entidades de Alemania, Italia, España, UK, son fuertes. En países como España tiene sentido por la existencia de varias lenguas dentro del propio país. En el caso de Alemania, consideramos que es por la fuerte presencia que tiene DFKI en los programas europeos. En el caso de UK, la presencia se debe principalmente a la Universidad de Edimburgo. El número de proyectos relacionados con las tecnologías del lenguaje, no consideramos que sea tan grande como para indicar que realmente en el país, existen una estrategia o interés por fomentar las tecnologías del lenguaje. Lo que demuestra es que hay centros de investigación o empresas fuertes y sobre todo con interés en participar en programas europeos, pero no implica que sea una prioridad para el país. De hecho, en el estudio sobre las políticas y estrategias en tecnologías del lenguaje que se realizó en el entregable [D3.1](#) del proyecto LT_Observatory, en gran parte de los países de la Unión Europea, se concluyó que a nivel Nacional, los países no tienen una estrategia relacionada con las Tecnologías del Lenguaje, a excepción de España con el Plan TL, y países como Irlanda, donde quieren promover el “Gaélico”, o en países Bálticos, donde quieren dar mayor fuerza a sus propias lenguas, frente a lenguas como el ruso, que fueron impuestas. En estos casos, la estrategia es más relacionada con políticas de educación y promoción de la lengua que de tecnologías del lenguaje.

De acuerdo a nuestra experiencia, que en un país exista una política de apoyo a cierta tecnología, no implica que la participación de ese país sea significativa en programas europeos relacionados con dicha tecnología. La participación en programas europeos se fomenta más por actuaciones de difusión, de los programas, y de las ventajas de financiación que aportan y con actividades concretas de apoyo a dicha participación, mediante, por ejemplo, bonos para apoyar la preparación de propuestas, etc. Pero puede darse el caso de que entidades muy fuertes en ciertas temáticas, sobre todo empresas, si tienen

una financiación interesante en su país, no acudan a programas europeos, que requieren mayor colaboración y mayor complejidad, normalmente.

Situación de España en los Programas Europeos relacionados con las TL

España tiene una presencia relevante en los programas más relacionados con tecnologías del lenguaje que se centrarían en el:

- 7PM en los tópicos: ICT-2007-4.2, ICT-2011-4.1, ICT-2011-4.2, ICT-2013-4.1, ICT-2013-4.3
- CIP-ICT-PSP en los tópicos: CIP-ICT-PSP.2009.5.1, CIP-ICT-PSP.2009.5.2, CIP-ICT-PSP.2009.5.3, CIP-ICT-PSP.2010.6.2, CIP-ICT-PSP.2010.6.1
- H2020 en los tópicos: ICT-15-2014, ICT-17-2014, ICT-16-2015,

A continuación, se muestran unas estadísticas de la participación de las entidades españolas, por tipología de entidad en el 7PM y en el H2020 para ver la evolución. Del CIP-ICT-PSP, no se analiza puesto que este programa no se ha repetido, y por tanto no se puede ver su evolución. Los datos son obtenidos del portal abierto de datos de la Comisión Europea (Ver detalle del listado de entidades en los Anexos 4.1 y 4.2).

7PM	Nº de entidades	Nº de proyectos en los que participan	Nº de proyectos coordinados
Entidades de investigación Españolas	14 (12%)	21 (30%)	10 (18%)
Entidades de investigación Europeas	116	69	55
Empresas Españolas	24 (11%)	22 (32%)	3 (21%)
Empresas Europeas	213	68	14
Otros (entidades públicas, fundaciones,...) Españolas	0 (0%)	0 (0%)	0 (0%)
Otros (entidades públicas, fundaciones,	11	12	1

7PM	Nº de entidades	Nº de proyectos en los que participan	Nº de proyectos coordinados
asociaciones...) Europeas			
Total entidades españolas	34	30	13
Total entidades	340	70	70

TABLA 9. ESTADÍSTICAS DE LA PARTICIPACIÓN DE LAS ENTIDADES ESPAÑOLAS, POR TIPOLOGÍA DE ENTIDAD EN EL 7PM

El número de entidades de investigación y de empresas que participan, sobre todo de empresas, es elevado y el número de proyectos coordinados es bastante alto, comparado con las tasas en otros programas. A continuación, se indican las entidades españolas que participan en más de 2 proyectos financiados del 7PM. En color azul se han resaltado las que coordinan proyectos:

Nombre de entidad	Nº de participaciones en proyectos
INTELLIGENT SOFTWARE COMPONENTS S.A.	5
UNIVERSIDAD DEL PAIS VASCO EHU UPV	4
FUNDACION TECNALIA RESEARCH & INNOVATION	3
INMARK ESTUDIOS Y ESTRATEGIAS S.A.	3
UNIVERSIDAD POLITECNICA DE MADRID	3
FUNDACIO BARCELONA MEDIA UNIVERSITAT POMPEU FABRA	2
FUNDACION CENTRO DE TECNOLOGIAS DE INTERACCION VISUAL Y COMUNICACIONES VICOMTECH	2
TELEFONICA INVESTIGACION Y DESARROLLO SA	2
UNIVERSITAT POLITECNICA DE VALENCIA	2
UNIVERSITAT POMPEU FABRA	2
YAHOO IBERIA SL	2

TABLA 10. ENTIDADES ESPAÑOLAS PARTICIPANTES EN PROYECTOS EUROPEOS RELACIONADOS CON LAS TL Y FINANCIADOS EN EL 7PM

La siguiente tabla muestra los datos correspondientes al H2020:

H2020	Nº de entidades	Nº de proyectos en los que participan	Nº de proyectos coordinadores
Entidades de investigación Españolas	6 (7%)	6 (20%)	1 (5%)
Entidades de investigación Europeas	82	29	21
Empresas Españolas	15 (15%)	12 (41%)	2 (40%)
Empresas Europeas	100	29	5
Otros (entidades públicas, fundaciones,...) Españolas	1 (5%)	1 (7%)	0
Otros (entidades públicas, fundaciones, asociaciones...) Europeas	20	14	0
Total entidades españolas	22 (11%)	15 (51%)	3 (12%)
Total entidades	202	29	26

TABLA 11. ESTADÍSTICAS DE LA PARTICIPACIÓN DE LAS ENTIDADES ESPAÑOLAS, POR TIPOLOGÍA DE ENTIDAD EN EL H2020

La participación en el H2020, no es tan grande, en número, porque el número de convocatorias ha sido menor, sólo se dispone de información de tres años: 2014, 2015 y 2016. Sin embargo, **de los porcentajes se puede deducir que la participación española es adecuada, y la de las empresas españolas es muy fuerte, tanto en el rol de socios, como en el rol de coordinadores de propuestas. Además, se muestra un aumento de la participación de empresas en el H2020, comparado con el 7PM, mientras que la participación de las entidades de investigación ha bajado, principalmente en coordinadores de proyectos.**

A continuación, se indican las entidades españolas que participan en 2 proyectos financiados del H2020.



Nombre de entidad	Nº de participaciones en proyectos
ATOS SPAIN SA	2
TELEFONICA INVESTIGACION Y DESARROLLO SA	2

TABLA 12. ENTIDADES ESPAÑOLAS PARTICIPANTES EN PROYECTOS EUROPEOS RELACIONADOS CON LAS TL Y FINANCIADOS EN EL H2020

Las entidades que actúan como coordinadores son:

Nombre de entidad	Nº de participaciones en proyectos
FUNDACION CENTRO DE TECNOLOGIAS DE INTERACCION VISUAL Y COMUNICACIONES VICOMTECH	1
TREELOGIC TELEMATICA Y LOGICA RACIONAL PARA LA EMPRESA EUROPEA SL	1
ZABALA INNOVATION CONSULTING, S.A.	1

TABLA 13. ENTIDADES ESPAÑOLAS COORDINADORAS EN PROYECTOS EUROPEOS RELACIONADOS CON LAS TL Y FINANCIADOS EN EL H2020

Se destaca, que los actores, no parecen ser los mismos, a excepción de Telefónica Investigación y Desarrollo, S.A., aunque todavía el H2020 no ha llegado a su fin.

En resumen, se podría decir que la participación española en proyectos directamente relacionados con tecnologías del lenguaje es elevada tanto desde el punto de vista de entidades de investigación como por empresas. Esta alta participación, es un indicativo de que a nivel estatal existen entidades de investigación y empresas de alta calidad, con capacidad de generar nuevas tecnologías, productos y servicios relacionados con las TL en un mercado europeo.

El impacto de estos proyectos en el sector industrial y de investigación, es difícil de medir, puesto que gran parte de estos proyectos están todavía en marcha y por tanto sería necesario hacer un análisis de sus resultados y evaluar el impacto en investigación (publicaciones, patentes, etc.) así como en el mercado, con nuevos productos o servicios. Hasta el momento de los proyectos finalizados, no conocemos ningún estudio donde se haya analizado el impacto de estos proyectos.



10 Glosario de siglas y acrónimos

7PM	7º Programa Marco
9PM	9º Programa Marco
ADpE	Agenda Digital para España
CE	Comisión Europea
CEF	Connecting European Facilities
CIP ICT PSP	Competitive Innovation Programme: ICT Policy Support Programme
CORDIS	Community Research and Development Information Service
CPI	Compra Pública de Innovación
CPP	Compra Pública Precomercial
CSA	Coordination and Support Action
DG	Directorate General
DIH	Digital Innovation Hubs
DSM	Digital Single Market
EC	European Commission
FTI	Fast Track to Innovation
EOSC	European Open Science Cloud
EUDAT	European Data Infrastructure
H2020	Horizon 2020
IA	Innovation Action
ICT	Information and Communication Technologies
INEA	Innovation and Networks Executive Agency
IoT	Internet of Things
ONTSI	Observatorio Nacional de las Telecomunicaciones y de la Sociedad de la Información
ML	Machine Learning
MSP	European Multi Stakeholder Platform on ICT Standardisation
PLAN TL	Plan de Impulso de las Tecnologías del Lenguaje
PLN	Procesamiento del Lenguaje Natural
PYME	Pequeñas y Medianas Empresas
RIA	Research and Innovation Action
RISP	Reutilización de la Información del Sector Público
SESIAD	Secretaría de Estado para la Sociedad de la Información y la Agenda Digital
SRIA	Strategic Research Innovation Agenda
TA	Traducción Automática



TICs	Tecnologías de la Información y las Comunicaciones
TL	Tecnologías del Lenguaje
TRL	Technology Readiness Level