

Plan de Impulso de las Tecnologías del Lenguaje

Octubre 2015



ÍNDICE

1	MOTIVACIÓN.....	5
1.1.	EL PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE.....	8
2	CONTEXTO	10
2.1	INTRODUCCIÓN A LAS TECNOLOGÍAS DEL LENGUAJE	10
2.2	ANÁLISIS DE LA SITUACIÓN ACTUAL.....	10
2.3	ELABORACIÓN DEL PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	12
3	OBJETIVOS ESPECÍFICOS DEL PLAN	18
4	ARTICULACIÓN DEL PLAN	21
5	MEDIDAS	24
6	RELACIÓN CON LOS OBJETIVOS DE LA ADPE	38
7	CALENDARIO.....	41
8	PRESUPUESTO	42
9	INDICADORES.....	43
	ANEXOS.....	44
	ANEXO I: RELACIÓN CON LA ESTRATEGIA EUROPEA.....	44

1 Motivación

Tanto la Agenda Digital para España como la Estrategia Española de Ciencia y Tecnología y de Innovación, establecen el desarrollo de la economía y la sociedad digital como uno de los retos globales que requieren un mayor esfuerzo en materia de Investigación Científica, Desarrollo e Innovación Tecnológica (I+D+i); y **señalan como una fortaleza del país la posición de liderazgo científico, tecnológico y empresarial del sector de las Tecnologías de la Información y la Comunicación (TIC)**. Se le identifica como un ámbito estratégico de innovación con gran potencial para aumentar la competitividad del tejido productivo e impulsar el crecimiento y la generación de empleo. Asimismo, las Comunidades Autónomas coinciden en señalar en sus Estrategias de Especialización Inteligente en Investigación e Innovación (RIS3), el potencial tractor de las TIC en la economía.

En el mismo sentido, el informe de conclusiones y recomendaciones del “ERAC Peer Review of Spanish Research and Innovation System”, presentado el 24 de julio de 2014 y elaborado por un grupo de expertos europeo en I+D+i anima a desarrollar iniciativas, lideradas por las empresas y las Administraciones Públicas, en los denominados **Ámbitos Estratégicos de Innovación**. En ellos **se establecen una serie de prioridades que reflejen las oportunidades de mercado, las fortalezas existentes y el potencial futuro en un entorno competitivo internacional**. El informe señala tres principios de actuación fundamentales:

- 1) La **identificación de proyectos de innovación a gran escala** desde un enfoque de especialización y con un fuerte componente de I+D+i y una amplia contribución de las tecnologías de la información.
- 2) La necesidad de **dotarse de enfoques e instrumentos caracterizados por un horizonte temporal de medio plazo (5-10 años)**
- 3) La **concentración de recursos para incrementar la eficiencia y la eficacia de las ayudas** para que el impacto sea mayor y se produzcan cambios efectivos en el desarrollo económico y social de los sectores o regiones en los que se esté actuando.

La Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información (SETSI) tiene entre otras **competencias**:

- El estudio, propuesta y ejecución de la **política general sobre telecomunicaciones y Sociedad de la Información**.
- La **promoción y desarrollo de las infraestructuras y servicios avanzados** de telecomunicaciones y de la Sociedad de la Información.
- El impulso y la coordinación de los **planes, proyectos tecnológicos y programas** de actuaciones para el fomento de la Sociedad de la Información.

- La elaboración, gestión y seguimiento de **programas orientados a la promoción de la oferta de nuevas tecnologías, servicios y aplicaciones y contenidos** en el ámbito de las telecomunicaciones y la Sociedad de la Información, así como la definición y gestión coordinada de esta política con los correspondientes programas e iniciativas de la Unión Europea y con otros programas internacionales en esta materia.

En el marco de sus competencias, además, la SETSI es responsable de la **ejecución y coordinación de la Agenda Digital para España**, y en este contexto es el órgano gestor de las actividades de la Acción Estratégica de Economía y Sociedad Digital 2013-2016.

El actual desarrollo de Internet y, en general de las TIC, **pone al alcance volúmenes enormes y crecientes de información textual**. Sin embargo, los sistemas informáticos, que procesan fácilmente datos, no pueden procesar directamente el lenguaje humano.

El lenguaje es una de las herramientas distintivas del ser humano. Las matemáticas por su parte, son el soporte de nuestra ciencia y base de las tecnologías de la información y las comunicaciones. **Las tecnologías de procesamiento del lenguaje natural aúnan ambos instrumentos aplicando métodos científicos, por medio de las tecnologías de la información, a la comprensión del lenguaje humano** en la amplia diversidad de lenguas, dialectos y formas de comunicación. Herramientas tales como buscadores, asistentes personales, clasificadores de texto o traductores automáticos se han convertido en indispensables para el desarrollo de nuestra labor cotidiana, sea cual sea nuestro campo de actividad. Pero hay otras muchas aplicaciones donde el Procesamiento del Lenguaje Natural (PLN) y la Traducción Automática (TA) pueden resultar críticas para dotar al ciudadano de nuevos servicios avanzados y de optimizar procesos y recursos productivos tanto en el ámbito empresarial como en las Administraciones Públicas. Cualquier paso en la mejora de la comprensión, síntesis, clasificación o traducción automática de información textual no estructurada **genera valor para la sociedad y resulta de aplicación transversal a todos los sectores productivos**.

El mercado de las tecnologías del lenguaje está creciendo rápidamente y los informes de varias consultoras¹ estiman un **gran crecimiento del mercado mundial en los próximos años** basándose en la explosión de aplicaciones observada en los últimos dos años y en el crecimiento exponencial de los datos textuales digitales.

Junto a este avance tecnológico, existen otros elementos importantes que justifican la oportunidad del plan que aquí se expone.

En el ámbito de la **Unión Europea** hay que destacar que la prioridad dada por la Comisión Europea al logro de un **Mercado Digital Único** (Digital Single Market, DSM) implica, entre sus retos, superar las barreras que imponen la existencia de un gran número de lenguas en el seno de la Unión Europea. Esta diversidad lingüística es, al mismo tiempo, una de sus mayores riquezas culturales. Por ello, las tecnologías del lenguaje tienen un papel muy relevante en la Unión Europea.

¹ *Gartner 2014, LT-Innovate, AltaPlana, etc.*

El español es la lengua más hablada del mundo después del chino, si tenemos en cuenta los hablantes nativos, y la **tercera por número de hablantes después del inglés**. Se prevé que en 2030 sea la segunda lengua de intercambio económico en el mundo, principalmente por el crecimiento del mercado latinoamericano. El español posee una enorme capacidad de internacionalización dado que nueve de cada diez usuarios se encuentra fuera de nuestras fronteras². Que se trate de una lengua compartida con otras naciones representa, sobre todo, una oportunidad de **estrechar los lazos de cooperación con la Comunidad Iberoamericana**.

Por otra parte, España cuenta con **organizaciones con prestigio internacional especializadas en la lengua española**, entre las que destacan la Real Academia Española, el resto de reales academias y el Instituto Cervantes.

Adicionalmente, en España existen **numerosos grupos investigadores de reconocido prestigio internacional** trabajando en procesamiento del lenguaje español y lenguas cooficiales, así como grupos especializados en semántica. Los investigadores y una buena parte de la Industria del sector se encuentra agrupada alrededor de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN), que ha cumplido en 2015 su XXXI aniversario, y de varias agrupaciones empresariales.

Sin embargo, el desarrollo de aplicaciones para una lengua y, en muchos casos, para un ámbito de conocimiento determinado, depende de la disponibilidad de tecnología y recursos para esa lengua y en un ámbito de conocimiento determinado. En el caso de España, la disponibilidad de esos recursos para el idioma del castellano, aunque en menor medida y con algunos vacíos significativos, tiene un nivel parecido al que existe para el alemán o el francés a pesar de tener un número muy superior de hablantes. Para las lenguas cooficiales el nivel es menor. Estos recursos tienen un coste elevado que no pueden asumir pequeñas y medianas empresas. Para garantizar la disponibilidad de aplicaciones en español y lenguas cooficiales es necesario **aumentar el número, la calidad, la variedad y la disponibilidad de los recursos y herramientas** que les dan soporte.

Además, en previsión del crecimiento del sector de procesamiento de lenguaje natural y traducción automática, se considera conveniente ampliar la **formación** en esta tecnología a los profesionales de las TIC en empresas y Administración Pública.

Las **Administraciones Públicas** deben incorporar las tecnologías de procesamiento de lenguaje natural y de la traducción automática para **mejorar la calidad y capacidad del servicio público**, actuando, además, como **tractores** de la demanda. Existen áreas como la Sanidad o la Justicia en las que la Administración tiene un papel determinante en el desarrollo de nuevos servicios basados en la comprensión más profunda o la traducción automática de los contenidos que gestiona.

Por otro lado, el gran valor potencial que tiene buena parte de la información que genera el sector público como recurso lingüístico representa una extraordinaria oportunidad para el

² Información obtenida del Informe 2015 del Instituto Cervantes, "El español: una lengua viva". Capítulo 1. El español en cifras. http://eldiae.es/wp-content/uploads/2015/06/espanol_lengua-viva_20151.pdf

desarrollo de la industria del procesamiento del lenguaje natural³. La política de Reutilización de la Información del Sector Público (RISP) representa un cauce para el desarrollo de estos **datos abiertos de interés lingüístico**, ya que tiene como objetivo poner a disposición de la sociedad la información que el propio sector público genera en el ejercicio de sus funciones y ofrecerla de manera abierta como un recurso que puede ser explotado económicamente.

1.1. El Plan de Impulso de las Tecnologías del Lenguaje

La realidad antes descrita lleva a la Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información a **impulsar el sector de procesamiento de lenguaje natural y traducción automática por medio del presente plan específico**, con un alcance de 5 años y un ámbito geográfico e institucional que se amplía a las distintas Comunidades Autónomas y a las lenguas cooficiales. Este Plan amplía el contenido y alcance de forma parte de la Agenda Digital para España, como se ha venido haciendo desde su aprobación.

Es un Plan **interdepartamental dado el carácter multidisciplinar** de las tecnologías del lenguaje, y tiene como premisa coordinar todas las actuaciones de la Administración General del Estado, en coordinación con las Comunidades Autónomas, para el impulso de las tecnologías del lenguaje.

La iniciativa para elaborar el presenta Plan parte del siguiente **análisis DAFO**:

FORTALEZAS	DEBILIDADES
<ul style="list-style-type: none"> • Alto nivel investigador en procesamiento de lenguaje natural coordinado por la Sociedad Española para el Procesamiento del Lenguaje Natural. • Buena gobernanza del idioma español (RAE, Asociación de Academias de la Lengua Española en Iberoamérica). • Gran experiencia con el multilingüismo por la existencia de lenguas cooficiales. 	<ul style="list-style-type: none"> • El sector está compuesto por pequeñas y medianas empresas que no alcanzan la capacidad industrial para competir en el mercado internacional o completar la cadena de valor en España. • Dificultad de transferencia de conocimiento del sector investigador a la industria debido principalmente al carácter transversal y multidisciplinar del procesamiento de lenguaje natural. •
OPORTUNIDADES	AMENAZAS
<ul style="list-style-type: none"> • Alto potencial de internacionalización del español y de cooperación con Iberoamérica. • Nuevos servicios públicos para ciudadanos y empresas en sectores estratégicos (sanidad, turismo, educación, etc.). • Mercado en fuerte crecimiento asociado 	<ul style="list-style-type: none"> • Pérdida de competitividad económica e industrial de España e Iberoamérica. • Subdesarrollo digital del español. • Extinción digital de las lenguas cooficiales. • Fuga de investigadores y profesionales y deterioro del sector investigador español.

³ Lo ilustra, a modo de ejemplo, que el elemento más descargado del portal de datos abiertos de la UE es el corpus paralelo de traducciones del EURPARL (<https://open-data.europa.eu/es/data/>).

FORTALEZAS	DEBILIDADES
<p>a la innovación y al desarrollo.</p> <ul style="list-style-type: none"> • Potencialidad de la reutilización de la información del Sector Público (RISP) como recursos lingüísticos muy valiosos para la industria y la investigación. 	

Este diagnóstico se podría resumir en las siguientes **ideas clave**:

- El sector de las tecnologías del lenguaje representa un **sector emergente transversal** vinculado a la innovación con capacidad para promover el crecimiento, la competitividad y el empleo de calidad.
- Su **desarrollo es imparable**, pero si no aprovechamos la oportunidad, serán otros los que ocupen este espacio.
- España tiene los medios, pero es necesario **impulsar y coordinar las actuaciones** desde la Administración General del Estado, en coordinación con las Comunidades Autónomas y en colaboración con Iberoamérica y la Unión Europea, para aprovechar la oportunidad.

En consecuencia, el objetivo general del Plan de Impulso de las Tecnologías del Lenguaje es **fomentar el desarrollo del procesamiento del lenguaje natural y la traducción automática** en lengua española y lenguas cooficiales por medio de los siguientes objetivos específicos:

1. **Aumentar el número, calidad y disponibilidad** de las **infraestructuras lingüísticas** en español y lenguas cooficiales.
2. Impulsar la Industria del lenguaje fomentando la **transferencia de conocimiento entre el sector investigador y la industria**. Ayudar a la **internacionalización** de las empresas e instituciones que componen el sector. Mejorar la **difusión** de los proyectos actuales.
3. **Mejorar 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**. Apoyar la **generación, estandarización y difusión** de recursos lingüísticos creados en el contexto de la actividad de gestión pública propia de la Administración.

El Plan pretende que el impulso a las tecnologías del lenguaje se realice de forma coordinada, **buscando sinergias y evitando duplicidad** de esfuerzos, conforme a las recomendaciones de la Comisión para la Reforma de las Administraciones Públicas (**CORA**).

2 Contexto

2.1 Introducción a las tecnologías del lenguaje

El procesamiento del lenguaje natural es el camino hacia una **comprensión automática cada vez más profunda de la mayor producción humana: el lenguaje**. El lenguaje es la forma más común y versátil de comunicación. Los sistemas informáticos procesan fácilmente datos, es decir, información que tiene una estructura y un significado único y explícito (información estructurada). Pueden manejar fácilmente tablas con millones de datos numéricos, por ejemplo. Pero el lenguaje humano es mucho más complejo, sus significados pueden ser variables según el contexto y hacer referencia a información no explícita. Sin embargo, el **volumen de información digital textual no estructurada es abrumador** y crece vertiginosamente, resultando imperioso contar con ayudas para su explotación automática.

A pesar de la dificultad, **el grado de desarrollo de las tecnologías de procesamiento del lenguaje natural permite hoy multitud de aplicaciones de utilidad**, y su rápido desarrollo en la última década augura resultados cada vez más sorprendentes.

Las tecnologías de Procesamiento del Lenguaje Natural (PLN) y Traducción Automática (TA) **son las tecnologías que hacen posible analizar textos y facilitar su explotación en aplicaciones informáticas** de uso muy común en sectores tan dispares como la Sanidad, la Educación o el Turismo. Por ejemplo, la detección de entidades nombradas (nombres propios de personas o empresas, marcas de productos o topónimos), filtrado y clasificación de documentos, creación de resúmenes automáticos, extracción de información, análisis de sentimientos, minería de opinión, seguimiento y monitorización de la reputación en los medios sociales, corrección ortográfica y gramatical, búsqueda inteligente y optimizada, sistemas de respuesta automática a preguntas y asistentes personales, la traducción automática de textos, etc. Todas estas aplicaciones se pueden resumir como la **explotación de información no estructurada que mejora la comprensión de textos en corpora documentales**.

Aplicadas de forma adecuada, estas herramientas de análisis textual pueden ayudar a las empresas y organizaciones a **optimizar muchos de sus procesos productivos y obtener conocimiento muy valioso** de su propia información y de aquella disponible en un mundo cada vez más digitalizado y global. Las tecnologías del lenguaje son una industria habilitadora que participa de forma horizontal en multitud de aplicaciones y dispositivos.

2.2 Análisis de la situación actual

El sector de las tecnologías del lenguaje ha sido retratado en profundidad en diversos informes recientes⁴. Se estima que el mercado mundial de las tecnologías del lenguaje tenía un volumen aproximado de unos 19.300M € en 2011, y **se prevé que en 2015 alcance los 30.000M €⁵**. El mismo informe describe el estado actual del mercado en Europa constatando la existencia de más de 500 empresas relacionadas con este sector. Se trata de PyMEs concentradas en el norte

⁴ Elaborados por organizaciones como LT-Innovate, Meta-Net, Gartner, CRACKER and LT_Observatory.

⁵ "LT2013: Status and Potential of the European Language Technology Markets", LT-Innovate.

de Europa que **no pueden abarcar la totalidad del mercado europeo o internacional**. El informe recomienda el apoyo a estas empresas con infraestructuras específicas para acceder a recursos y tecnología.

Hasta el momento, en traducción automática domina el mercado europeo. Se trata de un **sector estratégico para la Unión Europea** dado que contamos con 24 lenguas oficiales, algunas de ellas en peligro de extinción digital. En la actualidad, el sector industrial se compone de unas 2.500 empresas y 800 centros de investigación en procesamiento de lenguaje natural⁶. Según un informe europeo⁷, el sector de la traducción automática en Europa alcanzó en 2011 la cifra de 8.600M € y **llegará al valor de 14.900 M€ en 2015**. Los ratios de crecimiento de otros mercados mundiales menos maduros serán mucho mayores.

En las convocatorias del **programa marco H2020** y en la definición del programa **Connecting Europe Facility (CEF)** se persigue la puesta en marcha de servicios digitales a nivel paneuropeo, incluyendo servicios lingüísticos para garantizar un mecanismo de coordinación de información y acceso a recursos lingüísticos para las lenguas de Europa.

El español es la lengua más hablada del mundo después del chino, si tenemos en cuenta los hablantes nativos, y también la segunda lengua en un cómputo global de hablantes. **Representa el 6,7% de la población mundial** contando con casi 470 millones de hablantes. Se prevé que en 2030 los hispanohablantes serán el 7,5% de la población mundial⁸.

El uso del mismo idioma aumenta por tres la cuota de exportaciones de España a los países de habla hispana. Nueve de cada diez usuarios de la tecnología que se potencia en el presente Plan se encuentran fuera de nuestras fronteras. Además, existen otras comunidades hispanohablantes en Norteamérica con un enorme potencial de crecimiento. Nuestra proximidad geográfica, histórica y lingüística nos acerca además a la otra gran lengua de la Comunidad Iberoamericana, el portugués.

Por otra parte, España cuenta con **prestigiosas organizaciones internacionales especializadas en la lengua española**, entre las que destacan la Real Academia Española o el Instituto Cervantes, con una enorme capacidad de difusión en Iberoamérica. También se cuenta con prestigiosas academias en áreas de conocimiento especializadas como Medicina, Ingeniería, etc.

Adicionalmente, los investigadores y una buena parte de la Industria del sector, se encuentran agrupados alrededor de la **Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)**, que ha cumplido su XXXI aniversario en 2015. En España existen **numerosos grupos investigadores de reconocido prestigio internacional** trabajando en procesamiento del lenguaje español y lenguas cooficiales así como grupos especializados en semántica. Asimismo, se han configurado algunas agrupaciones científico-empresariales en Cataluña, País Vasco y Madrid alrededor de la industria del lenguaje.

A pesar de que existe el potencial para disponer de una industria de éxito, el estudio realizado por la Multilingual Technology Alliance (META) y la red de excelencia asociada META-NET

⁶ "Strategic Agenda for the Multilingual Digital Single Market", CRACKER and LT_Observatory.

⁷ "LT2013: Status and Potential of the European Language Technology Markets", LT-Innovate.

⁸ "El español: una lengua viva" Informe 2015, Instituto Cervantes.

“Europe's Languages in the Digital Age”⁹, señala que **la carencia de recursos para el procesamiento del lenguaje natural en una lengua particular son la barrera más importante** para el desarrollo en Europa de aplicaciones de la industria del lenguaje, tales como:

- **Optimización de procesos industriales** de gestión lingüística de documentación: traducción de documentos y herramientas de autor (correctores, generación de documentos, etc.)
- **Comunicación y asistencia personal** (asistentes virtuales, comunicación hombre-máquina para coches, atención al cliente e interacción con robots; buscadores inteligentes y respuesta automática de preguntas).
- **Procesamiento inteligente de información y conocimiento** (extracción y minería de información de textos y contenidos, clasificación de documentos, resumen automático, etc.).
- **Asistencia en el aprendizaje de lenguas.**

En los programas europeos, no obstante, **el suministro efectivo de recursos que constituyen la infraestructura de apoyo para cada lengua no queda cubierta**. Es necesario garantizar el suministro de recursos para el español y las lenguas cooficiales porque, a pesar de que existe una industria de las tecnologías del lenguaje, las características de la misma (las empresas son PyMEs en su gran mayoría) hacen imposible pensar que la industria se autoabastezca con la rapidez necesaria.

2.3 Elaboración del Plan de Impulso de las Tecnologías del Lenguaje

La iniciativa para elaborar un Plan de Impulso de las Tecnologías del Lenguaje parte de la Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información. Para ello, se constituyó un **Comité Directivo para la elaboración del Plan** con la participación de los siguientes órganos:

- Secretaría General de Industria y de la Pequeña y Mediana Empresa.
- Secretaría de Estado de Investigación, Desarrollo e Innovación.
- Secretaría de Estado de Cultura.
- Subsecretaría de Educación, Cultura y Deporte.
- Secretaría de Estado de Cooperación Internacional y para Iberoamérica.
- Subsecretaría Ministerio de Presidencia.
- Dirección de Tecnologías de la Información y de las Comunicaciones de la Administración General del Estado.
- Secretaría de Estado de Turismo.

⁹ <http://www.meta-net.eu/>

- Subdirección General de Tecnologías de la Información y de las Comunicaciones de la Administración General del Estado del Ministerio de Sanidad, Servicios Sociales e Igualdad.

Se espera además la próxima incorporación de otros organismos que han mostrado su interés en formar parte de éste Comité.

El Comité Directivo creó, a su vez, un **Comité de Expertos para la elaboración de un Informe** previo que sirviera de apoyo para la creación del Plan de Impulso de las Tecnologías del Lenguaje¹⁰. Dicho Comité de Expertos está formado por:

- Real Academia Española de la Lengua.
- Instituto Cervantes.
- Instituto de España.
- Biblioteca Nacional de España.
- Real Academia de Ingeniería.
- Grupos investigadores especializados en Procesamiento del Lenguaje Natural y Traducción Automática (SEPLN).
- Grandes empresas del sector TIC español.
- Agrupaciones empresariales especializadas en Procesamiento del Lenguaje Natural y Traducción Automática.
- Representantes de la política de Reutilización de la Información del Sector Público (RISP).
- Dirección de Tecnologías de la Información y la Comunicación de la Administración (DTIC).
- Comisión para la Reforma de las Administraciones Públicas (CORA).
- Cooperación Internacional y para Iberoamérica.

El análisis del Comité de Expertos se resume en el siguiente **Análisis DAFO**:

FORTALEZAS

- Desarrollo de líneas de **investigación** en procesamiento de lenguaje natural en España que **abarcen casi todos los ámbitos en los que se trabaja actualmente a nivel internacional**. Existen **recursos y herramientas propias, consolidadas y robustas** para hacer el procesamiento básico de procesamiento de lenguaje natural y traducción automática para el castellano, catalán, vasco y gallego, además del inglés y se dispone de amplia **información del sector público**

¹⁰ Informe sobre el estado de las tecnologías del lenguaje en España dentro de la Agenda Digital para España.

FORTALEZAS

susceptible de convertirse en recursos lingüísticos.

- **Disponibilidad de investigadores españoles** que participan en proyectos, asociaciones y grupos de estandarización europeos e internacionales. Existen **más de 30 grupos de investigación** consolidados y organizados en asociaciones y redes, que han sido el origen de 9 spin-off que abarcan casi todos los ámbitos en los que se trabaja actualmente a nivel internacional.
- Posibilidad de establecer **modelos de colaboración con centros investigadores** de forma rápida gracias a los programas nacionales de transferencia (CENIT, CIEN, Doctorados industriales).
- **Gran experiencia en gestión del multilingüismo** de las empresas españolas y las Administraciones Públicas, lo que puede ser un modelo exportable.
- Posición de liderazgo de España **en el mercado potencial del español** con cerca de 470 millones de hablantes, 54 en EEUU, por su situación económica. Se han abierto **oficinas en EEUU e Iberoamérica** por parte de varias empresas españolas, demostrando que la internacionalización de la tecnología desarrollada en España es viable y que existe un sector TIC español potente que está demostrando su capacidad para competir globalmente.
- España pertenece a las redes Iberoamericanas, con prestigio en la sociedad, **de las instituciones relacionadas con las lenguas españolas** (RAE, IEC, etc.) que colaboran internacionalmente en la labor de regulación del lenguaje.
- Bajo **coste de reutilizar y visibilizar** todos los materiales existentes para los lenguajes de especialidad, dada la estructura y organización ya existente y demostrada eficacia de la **experiencia en acciones estratégicas** en los planes nacionales.
- Existencia de **bases de datos terminológicas**, sinónimos, thesaurus y topónimos, exportables globalmente. Existencia de un **tejido industrial en traducción automática** con experiencia que tiene como clientes prioritarios las Administraciones Públicas.
- Existencia de iniciativas de traducción automática ya en marcha, como la plataforma Plata¹¹.
- Adhesión de España a la directiva europea de reutilización de datos de las Administraciones Públicas (RISP).
- Experiencia en la **participación en iniciativas europeas** como Meta-Share, Clarin-Eric, ELRA, etc. y en proyectos europeos como OPENER, NEWSREADER, QT-Leap, etc.

DEBILIDADES

- **Insuficiente colaboración entre empresas y entre empresas y grupos de investigación** impidiendo la reutilización de datos y herramientas y multiplicando la inversión que realizan las empresas, restándole efectividad en otros ámbitos como la promoción comercial. **Falta de conocimiento e inversión coordinada y escasa compartición** de herramientas de amplia

¹¹ <http://administracionelectronica.gob.es/ctt/plata>

DEBILIDADES

cobertura entre las empresas y las academias, lo que dificulta la implantación de métodos que garanticen la re-utilización, generando duplicidades y dispersión de esfuerzos en la construcción de corpus, herramientas, etc.

- **Déficit de interdisciplinariedad** debido a la baja interacción entre lingüistas e informáticos en la creación y compartición de recursos y aplicaciones, motivado por la rigidez estructural universitaria y la falta de centros de investigación sobre procesamiento de lenguaje natural y traducción automática.
- **Discontinuidad en la financiación de la investigación** en procesamiento de lenguaje natural y traducción automática, lo que dificulta el progreso en la investigación y el mantenimiento de equipos de trabajo especializados.
- Insuficiente **inversión en formación** de especialistas altamente cualificados. Falta de oportunidades educativas vinculadas a postgrados, masters y becas específicas lo que se traduce en un **escaso número de spin-off** en la universidad española.
- Insuficiencia de **líneas de financiación que ayuden a la internacionalización** de las empresas y a acometer los proyectos de I+D+I. No obstante, la **necesidad de inversión es alta** en tiempo y coste para el desarrollo de nuevas herramientas tecnológicas.
- Insuficiente **capacidad de inversión** en investigación e innovación y **poca capacidad de internacionalización** de sus productos por falta de capitalización, ya que la mayoría de empresas son PyMEs y microempresas.
- **Insuficiente investigación básica y desarrollo tecnológico** en torno al procesamiento de lenguaje natural o a la traducción automática por parte de las Agencias Públicas de Investigación y **disminución del número de grupos de investigación en procesamiento de lenguaje natural y traducción automática** en grandes empresas en España (IBM, TIC, etc.)
- Reducida disponibilidad de recursos y herramientas para el **español hispanoamericano**.
- Existencia de una debilidad en cuanto al tratamiento de textos especializados ya que el castellano **no es un idioma mayoritario en literatura científica y patentes**.
- **Escaso conocimiento de los estándares, licencias y modelos** de negocio ya consensuados en Europa por las empresas en España.
- **Falta de reconocimiento de esta disciplina en la academia nacional** pese a ser un ámbito de investigación interdisciplinar. La formación está incluida en diferentes especialidades, lo que dificulta su visibilidad.
- **Falta de un mayor conocimiento de las características de esta tecnología** con poco coste de licencia de programas, pero alto coste de inversión en localización y adaptación (lengua y dominios de actividades) para lograr procesos mucho más eficientes en múltiples áreas de aplicación en la Administración.
- Carencia de una **norma específica de interoperabilidad**.
- **Reducida definición de la estrategia de comercialización** generando grandes dificultades para acceder al mercado.

DEBILIDADES

- **Falta de mayor cultura RISP** en los diferentes colectivos.

OPORTUNIDADES

- **Gran momento en Europa para el desarrollo del procesamiento de lenguaje natural y traducción automática** con la última incorporación de nuevos países. Infraestructura, asociaciones, fundaciones y distribuidores han trabajado ya de forma colaborativa los aspectos formales: estándares, licencias y modelos de negocio. En la actualidad, existen **modelos europeos e internacionales** de oferta de datos lingüísticos en portales de datos abiertos a los que sumarse.
- **Demanda urgente de aplicaciones** relacionadas con social media, big data y los datos abiertos, creando objetivos a corto plazo que ya pueden ser satisfechos en consorcios mixtos de desarrollo, lo que proporcionará gran visibilidad al área.
- **Disponibilidad de múltiples áreas de aplicación:** Turismo, Sanidad, Justicia, Educación, etc. donde optimizar y sistematizar procesos horizontales que pueden servir de efecto demostrador y aprendizaje para proyectos futuros, por su posibilidad de generar **recursos reutilizables**.
- Es posible generar valor mediante la definición de conjuntos de **datos abiertos de interés lingüístico** dentro de la estrategia RISP.
- Existencia de más de **300 investigadores en Iberoamérica**, la mayoría en Méjico, con los que poder colaborar para el desarrollo de infraestructuras en la región.
- **Auge de las redes sociales junto al procesamiento del Big Data** que sitúan las industrias de la lengua en una excelente posición para, por un lado, explorar nuevos ámbitos de actuación y, por otro, obtener recursos básicos para avanzar en la mejora de los sistemas.
- Gran **número de agentes** implicados en el sector. De manera específica, se identifican **necesidades horizontales y globales** de esta industria para todas las Administraciones Públicas.
- Existencia de **programas de doctorado en España** con temáticas de procesamiento de lenguaje natural.
- Existencia de **programas de I+D+i de la UE** que pueden ayudar a financiar nuevos proyectos en este campo.
- Posibilidad de **mejorar la accesibilidad** de colectivos con limitaciones funcionales.
- Existencia del **Centro de Interoperabilidad Semántica** y disponibilidad en el Centro de Transferencia Tecnológica de material para definir **soluciones comunes en la Administración**.

AMENAZAS

- **Pérdida de competitividad** frente a terceros países, como EEUU, en el desarrollo de recursos y herramientas para el procesamiento de lenguaje natural y traducción automática para el

AMENAZAS

castellano y español de América.

- Falta de acuerdo de estándares y modelos de licencias por parte de diferentes **asociaciones y organizaciones europeas** que pueden minorar a la industria española si no participa más activamente.
- Posibilidad de desaparición del español y lenguas cooficiales como lenguas de **dominios especializados** si no se fomentan las publicaciones en español y su disponibilidad y uso en el mundo digital.
- Requisito de **inversión y planificación** para ofrecer **datos lingüísticos de calidad** en portales de datos abiertos, convirtiéndose en una actividad poco sostenible sin financiación específica.
- **Competencia por parte de grandes empresas** en el ámbito de la investigación y desarrollo con los grupos de investigación de procesamiento de lenguaje natural y traducción automática, tanto en español como en lenguas cooficiales.

3 Objetivos específicos del plan

El objetivo general del Plan de Impulso de las Tecnologías del Lenguaje es **desarrollar la industria del procesamiento del lenguaje natural y la traducción automática** en España, especialmente en lengua española y lenguas cooficiales.

Este objetivo general puede descomponerse en los siguientes **objetivos específicos**:

1. Desarrollo de infraestructuras lingüísticas:

- Impulsar la industria del procesamiento del lenguaje natural en español y lenguas cooficiales poniendo a su disposición **infraestructuras lingüísticas** (recursos y procesadores) de propósito general.
- **Reducir la distancia** que separa las infraestructuras lingüísticas en español y las lenguas cooficiales **respecto del inglés** en lo que se refiere a la cantidad, calidad y disponibilidad de infraestructuras lingüísticas de propósito general para procesamiento del lenguaje natural y traducción automática.
- Asegurar la **disponibilidad pública gratuita o a bajo coste** (al menos para PyMEs innovadoras, sector investigador y Administraciones Públicas) de infraestructuras lingüísticas de calidad en español y lenguas cooficiales.
- Asegurar la **coordinación en el desarrollo de infraestructuras lingüísticas**, evitando duplicidades y buscando sinergias. Disponer de herramientas comunes para la generación así como campañas de evaluación de infraestructuras lingüísticas.
- Adoptar **normas técnicas de interoperabilidad**, una política de licencias adecuada, y mecanismos de protección de datos personales en la generación de recursos lingüísticos.
- Potenciar **métodos de generación automática** de recursos lingüísticos.

2. Impulso de la industrias de las **tecnologías del lenguaje**:

2.1. Mejorar la **visibilidad y la transferencia** de conocimiento del sector:

- Mejorar la **visibilidad** del sector del procesamiento de lenguaje natural y de la traducción automática.
- Trasladar la **excelencia investigadora española** a la industria. Recuperar doctores especialistas y formar nuevos investigadores y desarrolladores en estas áreas.

2.2. Apoyo a la **internacionalización y la comercialización** del sector:

- Mejorar la **internacionalización** de las empresas del sector, especialmente en el mercado iberoamericano y norteamericano.

- Hacer uso de las **herramientas proporcionadas por el Plan de Internacionalización** de la Agenda Digital para España, que contiene un programa de financiación directa a las empresas para actividades de internacionalización (ICEX, AECID, ...).
- Apoyar el **papel tractor de las grandes empresas** españolas asentadas en el mercado iberoamericano y otros mercados emergentes.
- Estrechar la **cooperación con la comunidad iberoamericana** para liderar la implantación de las tecnologías de procesamiento de lenguaje natural y traducción automática en español. Apoyo y participación institucional de los distintos países iberoamericanos en el proyecto.

3. La Administración Pública como impulsora de la industria del lenguaje:

3.1. Creación de las **Plataformas comunes** de procesamiento de lenguaje natural y de traducción automática de las Administraciones Públicas:

- **Mejorar la calidad y capacidad** del servicio público incorporando las tecnologías de procesamiento de lenguaje natural y traducción automática.
- Dotar de **herramientas comunes** para el procesamiento del lenguaje natural y la traducción automática a las Administraciones Públicas.
- En concordancia con las recomendaciones **CORA**, simplificar, lograr sinergias y aplicar la economía de escala en la puesta en marcha de nuevos servicios basados en tecnologías de procesamiento de lenguaje natural y traducción automática. Se busca el ahorro de costes mediante la compartición de recursos entre las Administraciones Públicas, y la eliminación de duplicidades y redundancias; además de mediante la mejora de la eficiencia de los procedimientos y la mejora del conocimiento. La plataforma se basará en el empleo de componentes reutilizables e interoperables, y, preferentemente, con licencias no restrictivas de código abierto.
- Disponer de herramientas comunes para la **generación, evaluación y explotación** de los recursos lingüísticos.
- Compartir infraestructuras y sistemas entre los **proyectos innovadores** que sean apoyados desde las instituciones públicas. También podrán ser usados tras la finalización de los proyectos como centros demostradores online.
- Emplear infraestructuras y sistemas comunes para su aplicación en el **sector investigador** facilitando el desarrollo de nuevos componentes y el análisis de grandes corpus documentales.
- Compartir servicios con otras **instituciones del sector**: Real Academia Española, Instituto de España, organismos iberoamericanos, etc.

3.2. Recursos lingüísticos de las Administraciones Públicas y Política de **Reutilización de la Información del Sector Público**:

- Impulsar la industria del procesamiento del lenguaje natural en español y lenguas cooficiales poniendo a su disposición **recursos lingüísticos generados a partir de la información del sector público**.
- Asegurar la **disponibilidad pública gratuita o a bajo coste** de los recursos lingüísticos generados a partir de la información del sector público.
- Asegurar la **coordinación en el desarrollo de recursos lingüísticos** generados a partir de la información del sector público, evitando duplicidades y buscando sinergias.
- Desarrollo y empleo de **herramientas comunes** para la generación y evaluación de recursos lingüísticos generados a partir de la información del sector público.
- Adoptar **normas técnicas de interoperabilidad**, una política de licencias adecuada y mecanismos de protección de datos personales.

El Plan pretende que el impulso a la industria del lenguaje se realice de forma coordinada, buscando **sinergias y evitando duplicidades y redundancias**, conforme a las recomendaciones de la Comisión para la Reforma de las Administraciones Públicas (CORA).

4 Articulación del Plan

Para la elaboración del Plan de Impulso de las Tecnologías del Lenguaje se ha constituido un **Comité Directivo** interdepartamental que reúne a los órganos de la Administración General del Estado con protagonismo en esta materia. El Comité Directivo, a su vez, requirió un informe previo de un **Comité de Expertos** que reunía a representantes de los sectores implicados (investigador, industria, Academias, Administraciones Públicas) sobre el estado de las tecnologías del lenguaje en España que incluía, además, recomendaciones para el desarrollo de este sector en España.

Las medidas propuestas en este Plan pretenden contribuir al desarrollo de los objetivos estratégicos de la **Agenda Digital para España**, así como a aquellos objetivos concretos establecidos en el presente Plan. Estas medidas se organizan en los siguientes ejes:

Eje 0: Gobernanza

Este eje inicial pretende **definir un mecanismo de coordinación, colaboración e intercambio de ideas, documentación e información entre los diferentes agentes implicados en esta industria** con el objetivo de evaluar la situación actual del sector, evaluar el avance del Plan periódicamente y planificar las sucesivas actuaciones al respecto. Esta herramienta de coordinación y control estará compuesta por el **Comité Directivo** y por un **Comité de Expertos**.

Eje I: Apoyo al desarrollo de infraestructuras lingüísticas

El procesamiento de lenguaje natural y la traducción automática de textos se basa en el uso de **procesadores lingüísticos** que permiten, por ejemplo, realizar automáticamente la extracción de lemas, reconocimiento de entidades (nombres de personas, organizaciones, lugares,...), desambiguación de términos, cálculo de la similitud semántica entre documentos, clasificación automática de los mismos, simplificación del contenido textual, generación de resúmenes, etc.

Para realizar dichas tareas son necesarios **recursos lingüísticos** específicos que dependen de la lengua procesada o traducida, el dominio de información tratada y otros factores específicos. Estos recursos lingüísticos son, por ejemplo, corpus paralelos (colecciones de documentos en varias lenguas), listas de nombre propios (de personas, organizaciones, marcas, topónimos, etc.), terminologías, diccionarios, etc.; todos ellos en los formatos adecuados para ser explotados por los procesadores lingüísticos.

El objetivo de este eje es el **desarrollo de procesadores y recursos lingüísticos** de propósito general.

Eje II: Impulso de la Industria del Lenguaje

Se pretende apoyar la **transferencia de conocimiento entre el sector investigador y la industria**. Debe apoyarse firmemente la captación y generación de talento en torno al sector de las tecnologías del lenguaje con el fin de preservar el conocimiento en nuestras universidades y su transferencia a empresas del sector o de nueva creación en el marco del apoyo a la innovación.

Ello comprende también la necesidad de formación especializada a la industria tanto del sector de las Tecnologías de la Información como de los distintos sectores verticales que son potenciales usuarios de esta tecnología. Ha de lograrse dar una mayor difusión a los proyectos actuales y a las capacidades actuales de la técnica.

Por otra parte, se plantea el **apoyo a la internacionalización de las empresas e instituciones** que componen el sector. Se ha de fortalecer el mantenimiento y creación de recursos lingüísticos con alta capacidad de internacionalización y su adaptación a los países objeto de exportación. Esta tarea puede verse amplificada por el alto prestigio y capacidad difusora de instituciones como la Real Academia Española. La **mejora de la visibilidad internacional de empresas e investigadores españoles** debe realizarse de forma coordinada con el trabajo del ICEX, por medio de misiones comerciales, presencia en congresos internacionales, proyectos demostradores, etc. Se prestará una atención especial a la cooperación con la Comunidad Iberoamericana.

Eje III: La Administración como impulsor de la Industria del Lenguaje

El servicio público que prestan las administraciones es un potencial beneficiario de las tecnologías del lenguaje. Por su carácter transversal, sus aplicaciones son muy numerosas. Ejemplos que no agotan estos potenciales beneficios son:

- Aumentar la capacidad, calidad y automatismo de los **servicios de atención al ciudadano** por los canales telefónico e Internet; y cada vez más, en movilidad.
- Mejorar la capacidad y la calidad, al tiempo que se reduce su coste, de la **traducción entre el castellano y las lenguas cooficiales**.
- Mejorar la capacidad, calidad y rapidez de **la tramitación administrativa** incorporando etapas de procesamiento automático de información en lenguaje natural.
- Mejorar la **accesibilidad** a personas discapacitadas.
- Mejorar el **conocimiento profundo y actualizado** de los distintos sectores para la dirección de políticas públicas.

Siguiendo, además, el espíritu propuesto por la Comisión para la Reforma de las Administraciones Públicas (CORA), la Administración española se plantea la necesidad de la eliminación de duplicidades administrativas y de proporcionar servicios públicos comunes de calidad. En consecuencia, en este eje se propone la creación de sendas **plataformas comunes de procesamiento del lenguaje y de traducción automática para las Administraciones Públicas**.

Por otro lado, la política de Reutilización de la Información del Sector Público (RISP) representa una extraordinaria oportunidad para el desarrollo de la industria del procesamiento del lenguaje natural por el gran valor potencial como recurso lingüístico que tiene buena parte de la información que genera el sector público. Dichos recursos comprenden, por ejemplo, textos legales y contractuales, traducciones juradas, diccionarios terminológicos especializados en dominios específicos, nombres de entidades (personas, topónimos, organizaciones, empresas,

publicaciones, ...), terminologías, corpus paralelos de traducción, etc. Se propone el apoyo para la creación y explotación de **datos abiertos de interés lingüístico**.

Eje IV: Proyectos Faro

Los proyectos faro son proyectos emprendidos por las Administraciones Públicas o por el sector privado de **aplicación de las tecnologías del lenguaje en sectores estratégicos** que pretenden servir de demostración de sus capacidades y beneficios, generar industria y generar recursos reutilizables en otros proyectos. También servirán de aprendizaje para futuros proyectos en otros sectores.

Los proyectos faro buscan una sinergia con las demás medidas del Plan de Impulso de las Tecnologías del Lenguaje, que son de naturaleza horizontal, transversal (infraestructuras lingüísticas generales, reutilización de información del sector público de interés lingüístico, plataformas de procesamiento de lenguaje natural y traducción automática de las Administraciones Públicas), con actuaciones en sectores concretos que **abarquen toda la cadena de valor y den lugar a productos y servicios acabados**.

Los requisitos empleados para seleccionar los primeros proyectos faro del sector público han sido:

- Sinergia con sectores estratégicos de la economía o los servicios públicos españoles.
- Elevado impacto económico y social.
- Desarrollo de la industria del lenguaje nacional.
- Concreción y compromiso de los órganos competentes.
- Generación de recursos reutilizables.
- Sinergias con las demás medidas del Plan de Impulso de las Tecnologías del Lenguaje, y en particular, con la de generación de recursos lingüísticos.
- Adquisición de experiencia para futuros proyectos. Carácter demostrador de las capacidades y beneficios de las tecnologías del lenguaje.

Entre los sectores verticales en los que se empezará a trabajar en los próximos meses se encuentran **Sanidad, Turismo y Educación**. En el espíritu del Plan está el conseguir la máxima difusión y reutilización de los desarrollos realizados, por lo que se prevé la próxima incorporación de otros proyectos verticales en áreas estratégicas.

5 Medidas

A continuación, se exponen todas las **medidas que componen el Plan organizadas en función de ejes y líneas de actuación específicas**. Dichas medidas serán planificadas por medio de planes operativos y se establecerán indicadores para conocer su progreso. Se estudiará de forma periódica el impacto y el estado del sector así como la adopción de las tecnologías del lenguaje en las Administraciones Públicas.

Eje 0: GOBERNANZA DEL PLAN

Motivo:

La realización del Plan exige el **concurso de numerosos actores** y comprende medidas de diversa índole. Por ello es necesario establecer un **mecanismo de coordinación, control y seguimiento** del Plan que asegure la actuación concertada de los actores y la sinergia entre las actuaciones del Plan para el logro de sus objetivos.

Objetivos:

Los objetivos de este Eje son los siguientes:

- **Identificar las Administraciones y Órganos competentes**, y unidades operativas que van a participar en la realización del Plan.
- Identificar las **cadena de mando** y definir los **procedimientos de toma de decisiones**. Asegurar el compromiso de los actores y dar cauce a su participación.
- Crear un **órgano de coordinación de los Órganos competentes** y dotarlo de la infraestructura necesaria. Asegurar la actuación concertada de los actores y la sinergia entre las actuaciones del Plan, evitando redundancias, sincronizando actuaciones y compartiendo recursos.
- Realizar la **planificación operativa del Plan** y de los planes operativos de cada Eje del Plan. Realizar la **evaluación** de la situación inicial y de los resultados de las sucesivas fases del Plan.
- Dotar de **flexibilidad a la implementación** del Plan que permita una buena adaptación a la evolución de las circunstancias y el aprovechamiento de la experiencia adquirida. Asegurar la transparencia de las actuaciones.
- Crear un **órgano de interlocución** con los agentes implicados en las tecnologías del procesamiento de lenguaje natural y la traducción automática.

Medidas:

El Ministerio de Industria, Turismo y Comercio, a través de la **Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información**, será el órgano de la Administración General del Estado que ejercerá la **coordinación de la realización del Plan** de Impulso de las

Tecnologías del Lenguaje.

– **Medida 1: Creación del Comité de Dirección:**

Se creará el *Comité de Dirección del Plan de Impulso de las Tecnologías del Lenguaje* con las siguientes funciones:

- Establecer un **mecanismo de coordinación, colaboración, intercambio de experiencias y ayuda** mutua entre los órganos competentes de la Administración General del Estado y de otras Administraciones, que evite el solapamiento de actuaciones y busquen su complementariedad, en aras de la optimización conjunta del beneficio esperado, derivado de todas las actuaciones del Plan de Impulso de las Tecnologías del Lenguaje.
- **Evaluar** de forma periódica **la situación en los distintos ejes de actuación del Plan** para planificar en detalle (planificación operativa) y evaluar el avance del Plan.
- **Nombrar y cesar** a los miembros del Comité de Expertos que se define más adelante. Establecer un proceso ágil de toma de decisiones.
- El Comité de Dirección podrá crear **grupos de trabajo** para tareas específicas.

El Comité de Dirección del Plan de Impulso de las Tecnologías del Lenguaje estará **constituido** inicialmente por:

- Secretaría de Estado de Turismo.
- Secretaría de Estado de Investigación, Desarrollo e Innovación.
- Secretaría de Estado de Cooperación Internacional y para Iberoamérica.
- Secretaría de Estado de Cultura.
- Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información.
- Subsecretaría de Educación, Cultura y Deporte.
- Subsecretaría de la Presidencia.
- Subsecretaría de Sanidad, Servicios Sociales e Igualdad.
- Dirección de Tecnologías de la Información de la Administración General del Estado.
- Secretaría General de Industria y de la Pequeña y Mediana Empresa.

El desarrollo de las diversas fases del Plan de Impulso de las Tecnologías del Lenguaje, y su propia evolución especialmente en relación con los proyectos verticales, determinarán **qué órganos de la Administración General del Estado son competentes en cada momento.**

– **Medida 2: Creación del Comité de Expertos:**

Se creará el Comité de Expertos del Plan de Impulso de las Tecnologías del Lenguaje con las **siguientes funciones**:

- **Asesoramiento técnico** al Comité de Dirección para el desarrollo del Plan.
- Servir de **mecanismo de interlocución** entre los actores implicados en la consecución de los objetivos del Plan de Impulso de las Tecnologías del Lenguaje.
- Facilitar la **colaboración, y el intercambio de experiencias y buenas prácticas** en el proceso de realización del Plan. Colaborar en la difusión del Plan en sus respectivos sectores.

En la composición del Comité de Expertos del Plan de Impulso de las Tecnologías del Lenguaje se conjugará **representatividad y eficacia**. Su composición podrá variar a lo largo de la realización del Plan. El Comité de Expertos podrá crear grupos de trabajo para tareas específicas. En el Comité de Expertos deberán estar **representados al menos los siguientes sectores**:

- Sector investigador en procesamiento de lenguaje natural.
- Sector industrial del procesamiento de lenguaje natural.
- Sector académico e institucional relacionado con el lenguaje.
- Sector público implicado en la realización del Plan.

– **Medida 3:** Planificación Operativa

Entre los cometidos del Comité de Dirección se encuentra la **planificación operativa** del Plan a lo largo de su existencia. Es responsable de cada uno de los ciclos de planificación y de las fases de cada ciclo (evaluación de la situación inicial, objetivos operativos, diseño y evaluación de alternativas, propuesta de actuación, coordinación de actuaciones, evaluación de sus resultados). La aprobación del plan operativo corresponderá a los órganos competentes. Los planes que se citan en los diversos ejes y líneas forman parte de esta planificación operativa que es responsabilidad del Comité de Dirección.

– **Medida 4:** Evaluación de la situación y de la realización del plan

Un elemento principal de los ciclos de planificación operativa es una evaluación periódica objetiva y suficientemente precisa tanto de la situación, como de los resultados de cada fase de la implementación del Plan. Estas tareas de evaluación se coordinarán, en la medida de lo posible, con otros planes de la Agenda Digital para España.

– **Medida 5:** Coordinación con otras administraciones

Es imprescindible asegurar una estrecha coordinación y cooperación con otras administraciones en varios ejes: Administraciones Autonómicas, Unión Europea e Iberoamérica.

A continuación, se expone una **tabla** en la que se indican aquellas líneas de actuación que requieren una especial coordinación entre ellas.

LÍNEAS DE ACTUACIÓN DEL PLAN							
LÍNEAS DE ACTUACIÓN DEL PLAN		Línea 1.1	Línea 2.1	Línea 2.2	Línea 3.1	Línea 3.2	Línea 4.1
	Línea 1.1				X	X	X
	Línea 2.1					X	X
	Línea 2.2						X
	Línea 3.1	X				X	X
	Línea 3.2	X	X		X		X
	Línea 4.1	X	X	X	X	X	

En cada uno de los Ejes descritos a continuación se reiteran las acciones de gobernanza, enunciadas en el eje 0, que les son específicas.

EJE I: DESARROLLO DE INFRAESTRUCTURAS LINGÜÍSTICAS

Línea 1: Desarrollo de infraestructuras lingüísticas

Motivo:

Se entiende por infraestructura lingüística a la terna constituida por **procesadores, recursos y campañas de evaluación de carácter lingüístico**.

El procesamiento de lenguaje natural y la traducción automática requieren **procesadores** (reconocedores de entidades, desambiguadores, calculadores de proximidad semántica, etc.) y **recursos lingüísticos** (corpus paralelos, diccionarios, taxonomías, etc.) Estos pueden ser de propósito general (ej. función en la oración de la palabra) o depender del dominio que se analice (ej. diccionario términos médicos). Asimismo, es necesario **evaluar la calidad** de estos procesadores y recursos por medio de campañas de evaluación sobre tareas concretas. Se priorizará su desarrollo en aquellas áreas de interés en proyectos verticales.

El motivo de esta línea de actuación del Plan es que el idioma español, a pesar de ser la tercera lengua del mundo en número de hablantes, y la segunda en hablantes nativos, **está muy lejos de**

disponer de las infraestructuras lingüísticas que tiene el inglés. La situación de las lenguas cooficiales es aún peor. Por ello, es necesario desarrollar infraestructuras lingüísticas en español y lenguas cooficiales para que sirvan de **combustible para el desarrollo de la industria española** de procesamiento de lenguaje natural y de traducción automática.

Estas infraestructuras lingüísticas deberán ponerse a disposición de los **usuarios con licencias abiertas** para lograr los ahorros económicos derivados de la reutilización de componentes y la compartición de costes, así como los beneficios cualitativos del mantenimiento colaborativo de los recursos abiertos.

Las **Administraciones Públicas** son el actor idóneo para desarrollar infraestructuras lingüísticas por su tamaño, por las sinergias con la política de Reutilización de la Información del Sector Público (RISP) y por garantizar el acceso abierto a las mismas.

Objetivos:

Los objetivos principales que persigue este eje son los siguientes:

- **Impulsar la industria del procesamiento del lenguaje natural** en español y lenguas cooficiales poniendo a su disposición infraestructuras lingüísticas de propósito general.
- **Reducir la distancia** que separa las infraestructuras lingüísticas en español y las lenguas cooficiales **respecto del inglés** en lo que se refiere a la cantidad, calidad y disponibilidad de infraestructuras lingüísticas de propósito general para procesamiento de lenguaje natural y traducción automática.
- Asegurar la **disponibilidad pública gratuita o a bajo coste** (al menos para PyMEs innovadoras, sector investigador y Administraciones Públicas) de infraestructuras lingüísticas de calidad en español y lenguas cooficiales.
- Llevar la industria a la **frontera de la innovación** facilitando el uso abierto de infraestructuras lingüísticas básicas.
- Asegurar la **coordinación en el desarrollo de infraestructuras lingüísticas**, evitando duplicidades y buscando sinergias. Disponer de **herramientas comunes** para la generación y evaluación de infraestructuras lingüísticas.
- Adoptar **normas técnicas de interoperabilidad**, una política de licencias adecuada, y mecanismos de protección de datos personales en la generación de recursos lingüísticos.
- Potenciar **métodos de generación automática** de recursos lingüísticos.

Medidas:

- **Medida 1:** Seleccionar normas técnicas de **interoperabilidad**, políticas de **licencias** y mecanismos de **protección de datos personales** adecuados para la generación de recursos lingüísticos.
- **Medida 2:** Adquirir o desarrollar **herramientas comunes** para la generación y evaluación de infraestructuras lingüísticas.

- **Medida 3:** Elaborar y ejecutar un **plan** de desarrollo de infraestructuras lingüísticas. Realizar un **inventario** de infraestructuras lingüísticas actualmente disponibles. Evaluar la **evolución** de la cantidad, calidad y disponibilidad de infraestructuras lingüísticas.
- **Medida 4:** Facilitar el acceso público a las infraestructuras lingüísticas existentes.

Eje II: IMPULSO DE LA INDUSTRIA DE LAS TECNOLOGÍAS DEL LENGUAJE

Línea 1: Mejora de la visibilidad y de la transferencia del sector

Motivo:

El sector de la traducción automática y más aún el sector del procesamiento de lenguaje natural es **poco conocido por las Administraciones Públicas** y por la mayoría de los **sectores productivos** de nuestro país. Por otro lado, se hace necesario **incrementar el número de expertos** en procesamiento de lenguaje natural y traducción automática para garantizar el crecimiento del sector. En el ámbito académico la situación es similar, no hay una oferta formativa adecuada.

Objetivos:

Los objetivos perseguidos por esta línea de actuación son los siguientes:

- **Mejorar la visibilidad del sector** del procesamiento de lenguaje natural y de la traducción automática.
- Trasladar la **excelencia investigadora española** a la industria. Garantizar la disponibilidad de doctores especialistas y formar nuevos investigadores y desarrolladores en estas áreas.

Medidas:

- **Medida 1:** Elaborar un **plan** de mejora de la visibilidad y de la transferencia.
- **Medida 2:** Planificar y coordinar acciones encaminadas a la mejora de la **visibilidad** y la formación en el sector con otras Administraciones y especialmente con Iberoamérica.
- **Medida 3:** Elevar la **formación**: Inclusión de asignaturas específicas sobre tecnologías del lenguaje en el currículum académico, impulso de la creación de contenidos formativos en línea [MooCs], hackathones, apoyo a programas de doctorado industrial, másters y becas especializadas.
- **Medida 4:** Elevar la **visibilidad**: Jornadas de formación básica para PyMES y profesionales; Congresos, foros y participación en ferias nacionales e internacionales. Impulso a centros demostradores en línea (SaaS). Coordinación con el portal propuesto en el Eje 1.

Línea 2: Apoyo a la internacionalización y comercialización del sector

Motivo:

El español es una lengua transnacional con cerca de 470 millones de hablantes, lo que representa un **gran mercado**. Además, nueve de cada diez usuarios de la tecnología potenciada por este Plan se encuentran fuera de nuestras fronteras. Esto representa una gran **oportunidad de ampliar el área de internacionalización**, así como una oportunidad de ampliar la cooperación institucional y económica con los países de Iberoamérica.

Objetivos:

Los objetivos principales de esta línea de actuación son los siguientes:

- **Mejorar la internacionalización** de las empresas del sector, especialmente en el mercado iberoamericano y norteamericano.
- **Hacer uso de las herramientas proporcionadas por el Plan de Internacionalización** de la Agenda Digital para España, que contiene un programa de financiación directa a las empresas para actividades de internacionalización (ICEX, AECID, etc.).
- Apoyar el **papel tractor de las grandes empresas españolas** asentadas en el mercado iberoamericano y otros mercados emergentes.
- **Mejorar la cooperación** con la comunidad iberoamericana para liderar la implantación de las tecnologías de procesamiento de lenguaje natural y traducción automática en español. Apoyo y participación institucional de los distintos países iberoamericanos en el proyecto. Trabajar conjuntamente en la creación de recursos y datos abiertos de interés lingüístico.

Medidas:

- **Medida 1:** Elaborar un plan de internacionalización.
- **Medida 2:** Cooperación con Iberoamérica (Cumbre Iberoamericana, IODC 2016 Madrid, ...). Colaboración con la Secretaría General Iberoamericana (SEGIB) y otras instituciones iberoamericanas.
- **Medida 3:** Coordinación con los **programas de la AECID** y utilización de la Red de Consejeros Científicos en el Exterior y de las asociaciones ya creadas de científicos españoles en el exterior.
- **Medida 4:** Posibilidad de integrar el procesamiento de lenguaje natural y traducción automática dentro de los ámbitos que se financian actualmente en el marco de la **Acción Estratégica de Economía y Sociedad Digital**.
- **Medida 5:** El sector TIC es uno de los prioritarios en **Invest in Spain** para ampliar la inversión extranjera en España. Se debe incluir el procesamiento de lenguaje natural y la

traducción automática entre los subsectores en los que España ofrece oportunidades de inversión más interesantes.

- **Medida 6:** Posibilidad de incluir el procesamiento de lenguaje natural y traducción automática en los acuerdos (**MoU**) que se firmen con países Iberoamericanos (o de otras regiones) en el futuro.
- **Medida 7:** Favorecer el desarrollo de infraestructuras lingüísticas y apertura de datos públicos en **variantes del español**.
- **Medida 8:** Identificar ferias, congresos o eventos en los que, con la colaboración del ICEX, pueda darse **difusión** a los productos y proyectos realizados por las empresas españolas del sector.
- **Medida 9:** Estudio sobre el **estado del arte** de la internacionalización de las **empresas españolas** del sector en los países con mayor mercado de internacionalización.
- **Medida 10:** Estudiar la posibilidad de ayudas a incubadoras o aceleradoras, o de plantear proyectos de hermanamiento entre empresas pequeñas y empresas grandes.

Eje III: LA ADMINISTRACIÓN PÚBLICA COMO IMPULSOR DE LA INDUSTRIA DEL LENGUAJE

Línea 1: Plataformas de procesamiento de lenguaje natural y traducción automática en las Administraciones Públicas

Motivo:

El servicio público que prestan las Administraciones Públicas es un **claro beneficiario de las tecnologías del lenguaje**. Por su carácter transversal, sus aplicaciones en la mejora de los servicios al ciudadano son numerosas. Ejemplos que no agotan estos beneficios son:

- Aumentar la capacidad, calidad y automatismo de los servicios de **atención al ciudadano** por los canales telefónico e Internet; y cada vez más, en movilidad. Prestación de servicios avanzados basados en el procesamiento del lenguaje natural y en la traducción automática.
- Mejorar la capacidad y reducir el coste de la **traducción entre el castellano, las lenguas cooficiales y las lenguas comunitarias**.
- Mejorar la **capacidad, calidad y rapidez de la tramitación** administrativa incorporando etapas de procesamiento automático de información en lenguaje natural.
- **Mejorar la accesibilidad** a personas con discapacidad.
- **Mejorar el conocimiento profundo y actualizado de la información** gestionada por los distintos sectores para la dirección de políticas públicas. Este conocimiento se puede

extraer mediante análisis automático de los grandes corpora documentales (historiales clínicos, solicitudes de ayudas, contratación pública, patentes, etc.).

Siguiendo, además, las recomendaciones de la Comisión para la Reforma de las Administraciones Públicas (CORA), la Administración española se plantea la **necesidad de evitar duplicidades, simplificar, aplicar economías de escala y proporcionar servicios públicos comunes de calidad**. En consecuencia, en esta línea se propone la creación de sendas plataformas comunes de procesamiento de lenguaje natural y de traducción automática, basadas en componentes reutilizables de procesamiento del lenguaje natural y traducción automática para las Administraciones públicas.

Además, estas plataformas pueden fortalecer el papel **tractor de la demanda** de las Administraciones Públicas y proporcionar servicios demostradores que muestren aplicaciones de las tecnologías de procesamiento de lenguaje natural y de la traducción automática. Así mismo, estas plataformas pueden ser de utilidad para los sectores innovadores e investigadores.

Objetivos:

Los objetivos de esta línea son los siguientes:

- **Mejorar la calidad y capacidad del servicio público** incorporando las tecnologías de procesamiento de lenguaje natural y de la traducción automática.
- Dotar de **herramientas comunes** para facilitar el procesamiento del lenguaje natural y la traducción automática a las Administraciones públicas.
- En concordancia con las recomendaciones **CORA**, simplificar, lograr sinergias y aplicar la economía de escala en la puesta en marcha de nuevos servicios basados en tecnologías de procesamiento de lenguaje natural y traducción automática. Se busca el ahorro de costes mediante la **compartición de recursos entre las Administraciones Públicas**, y la eliminación de duplicidades y redundancias; además de mediante la mejora de la eficiencia de los procedimientos y la mejora del conocimiento. La plataforma se basará en el empleo de componentes reutilizables e interoperables, y, preferentemente, con licencias no restrictivas de código abierto.
- Disponer **herramientas comunes para la generación**, evaluación y explotación de los recursos lingüísticos.
- Los **flujos de procesamiento se deberán ejecutar** tanto en la plataforma física común como mediante su reproducción sencilla en plataformas locales, para atender a necesidades de altos niveles de protección de la confidencialidad o casos de volúmenes muy grandes de información difíciles de desplazar.
- Compartir infraestructuras y sistemas entre los **proyectos innovadores** que sean apoyados desde las instituciones públicas. También podrán ser usados tras la finalización de los proyectos como centros demostradores online.

- Compartir infraestructuras y sistemas comunes para su aplicación en el **sector investigador** facilitando el desarrollo de nuevos componentes y el análisis de grandes corpus documentales.
- **Compartir servicios con otras instituciones del sector:** Real Academia Española, Instituto de España, organismos iberoamericanos.

Medidas:

- **Medida 1:** Elaborar un **plan** de desarrollo de las plataformas de procesamiento de lenguaje natural y traducción automática de las Administraciones Públicas.
- **Medida 2:** Clarificar el esquema competencial y de financiación que asegure su mantenimiento más allá de la vida del Plan.
- **Medida 3:** Creación de una **plataforma común de procesamiento del lenguaje natural y de traducción automática para las Administraciones Públicas** con los siguientes requisitos esenciales:
 - Facilitar la puesta en marcha de **servicios avanzados** basados en procesamiento de lenguaje natural y traducción automática en la Administración General del Estado y las Comunidades Autónomas.
 - Desarrollar una infraestructura **escalable** basada en componentes para el procesamiento paralelo de grandes corpus documentales.
 - Mantener las garantías de **confidencialidad** propias de los servicios públicos.
 - Agregar distintos **componentes** y recursos lingüísticos al flujo de procesamiento lingüístico con diferentes modelos de licenciamiento y modos de ejecución del procesamiento.
 - **Herramientas comunes** de anonimización, editores, post-edición de traducciones automáticas, etc.
 - Esta plataforma pondrá a disposición **recursos lingüísticos** de propósito **general** (Eje 1) y específicos de **dominio** (principalmente aquellos recursos especializados necesarios para el desarrollo de los proyectos verticales potenciados en el Eje 4).
 - Será un lugar para la explotación y estandarización de los recursos lingüísticos generados al amparo de la política de Reutilización de la Información del Sector Público (**RISP**).
 - Permitirá distintos modelos de ejecución y distribución (embebido, cluster local, remoto y ejecución en centros de supercomputación).

Línea 2: Recursos lingüísticos de las Administraciones Públicas y política de Reutilización de la

Motivo:

El gran valor potencial que tiene buena parte de la información que genera el sector público como recurso lingüístico representa una **extraordinaria oportunidad para el desarrollo de la industria del procesamiento del lenguaje natural**. Lo ilustra, a modo de ejemplo, que el elemento más descargado del portal de datos abiertos de la UE es el corpus paralelo de traducciones del EURPARL (<https://open-data.europa.eu/es/data/>).

Entre otros recursos, la Administración dispone de: topónimos, nombres de personas, marcas, nombres de organizaciones, nombres de empresas, taxonomías, glosarios, textos traducidos con alta calidad (v.g. traducciones juradas), corpus de dominio (legales, médicos, etc.), textos clasificados, etc. Es necesario **adaptar estos valiosos recursos a formatos reutilizables por procesadores lingüísticos**. Además, la política de Reutilización de la Información del Sector Público (RISP) representa **un cauce para el desarrollo de estos recursos lingüísticos**, ya que tiene como objetivo poner a disposición de la sociedad, la información que el propio sector público genera en el ejercicio de sus funciones y ofrecerla de manera abierta como un recurso que puede ser explotado económicamente.

Objetivos:

Los objetivos de esta línea de actuación son los siguientes:

- **Impulsar la industria del procesamiento del lenguaje natural** en español y lenguas cooficiales poniendo a su disposición recursos lingüísticos generados a partir de la información del sector público.
- Asegurar la **disponibilidad pública gratuita o a bajo coste** de los recursos lingüísticos generados a partir de la información del sector público.
- Asegurar la **coordinación en el desarrollo de recursos lingüísticos** generados a partir de la información del sector público, evitando duplicidades y buscando sinergias.
- Desarrollar y emplear de **herramientas comunes** para la generación y evaluación de recursos lingüísticos generados a partir de la información del sector público.
- **Adoptar normas técnicas de interoperabilidad**, una política de licencias adecuada, y mecanismos de protección de datos personales.

Medidas:

- **Medida 1:** Encuadrar estas actuaciones en el marco de la **política de Reutilización de la Información del Sector Público (RISP)**.
 - Introducir en RISP el concepto de **datos abiertos de interés lingüístico**.
 - Difundir este concepto en la Administración.

- Introducir el concepto de datos abiertos de interés lingüístico en la IODC 2016, en colaboración con **Iberoamérica**, para dar visibilidad a la política de datos abiertos lingüísticos.
- **Medida 2:** Seleccionar normas técnicas de **interoperabilidad**, políticas de licencias abiertas y mecanismos de protección de datos personales adecuados.
- **Medida 3:** Disponer de las **herramientas comunes** necesarias para generar y explotar estos recursos lingüísticos (anonimizadores, alineadores de texto, flujos de procesamiento...) en la plataforma de procesamiento del lenguaje natural de las Administraciones Públicas prevista en este Plan.
- **Medida 4: Identificar** aquellos corpus de información del sector público susceptibles de convertirse en recursos lingüísticos.
- **Medida 5: Catalogar** estos recursos lingüísticos abiertos dentro del portal de datos abiertos introduciendo una experiencia de usuario avanzada.
- **Medida 6:** Facilitar la **disponibilidad** a estos recursos en las plataformas de procesamiento de lenguaje natural y traducción automáticas de las Administraciones Públicas previstas en este Plan.
- **Medida 7:** Elaborar un **plan** de generación de recursos lingüísticos a partir de la información del sector público.

Eje IV: PROYECTOS FARO DE TECNOLOGÍAS DE PROCESAMIENTO DEL LENGUAJE NATURAL

Línea 1: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas

Motivo:

Poner en marcha las demás medidas del Plan, que son de naturaleza horizontal, transversal, con 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.

Los proyectos de la Administración en el sector del procesamiento de lenguaje natural y la traducción automática requieren en general de **capacidades de escalada horizontal y vertical para procesamiento de grandes volúmenes de información y servicio simultáneo a numerosos usuarios, orientación a componentes y empleo de estándares que garanticen la máxima reutilización e interoperabilidad**. Por otra parte, es necesario facilitar también la posibilidad de desplazar el procesamiento al dato por diversas razones, como pueden ser el volumen excesivo de los datos o las restricciones de confidencialidad.

Objetivos:

Los objetivos perseguidos por esta línea de actuación son los siguientes:

- Ofrecer **nuevos servicios públicos** o **mejorar la capacidad y calidad de servicios públicos existentes** mediante la aplicación de las tecnologías del procesamiento del lenguaje natural.
- Servir de **demostración de las capacidades y beneficios** de las tecnologías del procesamiento del lenguaje natural.
- **Generar recursos reutilizables** para otros proyectos del mismo u otro dominio.
- Servir de **aprendizaje para futuros proyectos**.
- **Servir de aplicación inmediata** de medidas horizontales del Plan; uso de infraestructuras y plataformas comunes.

Medidas:

- **Medida 1:** Realizar un conjunto limitado de proyectos de aplicación de las tecnologías del lenguaje natural en servicios públicos estratégicos de gran impacto social. Se seleccionarán los proyectos con los **siguientes requisitos**:
 - **Compromiso de los órganos competentes.** Asegurar el liderazgo de los que conocen bien el problema y tienen competencias para resolverlo.
 - **Concreción.** Responder a problemas ya identificados que justifiquen la idoneidad y el momento de la puesta en marcha del proyecto.
 - Elevado **impacto económico y social**.
 - Desarrollo de la **cadena de valor completa**.
 - Generación de **recursos reutilizables**. Evitar la generación de recursos cautivos de las implementaciones y las tecnologías propietarias que no permiten la portabilidad de recursos.
 - **Sinergias con las demás medidas** del Plan de Impulso de las Tecnologías del Lenguaje, y en particular, con la generación de recursos lingüísticos y con la plataforma de procesamiento del lenguaje natural y de traducción automática de las Administraciones Públicas.
 - Carácter **demostrador de las capacidades y beneficios** de las tecnologías del lenguaje.
 - Especial atención a la **adquisición de experiencia** para futuros proyectos.

Los sectores en los que se empezará a trabajar son Sanidad, Turismo y Educación (primeros proyectos 2016-17). Pero se prevé la posibilidad de incorporar otros proyectos faro más adelante aplicados a otras áreas (Justicia, Atención al Ciudadano, y Vigilancia

Sectorial son claros candidatos). Se prevé, asimismo, la incorporación de las Comunidades Autónomas.

6 Relación con los objetivos de la ADpE

A continuación se presenta una tabla que expone la relación entre las líneas de actuación del Plan de Impulso de las Tecnologías del Lenguaje y los objetivos de la Agenda Digital para España.

Objetivo	Subobjetivo	Líneas de actuación	Identificador (ID)	Líneas PIL
2. Desarrollar la economía digital	2.1. Incentivar el uso transformador de las TIC en nuestras empresas	2.1.3. Fomentar el desarrollo de soluciones TIC específicas adaptadas a las necesidades de sectores productivos insuficientemente atendidos por la oferta TIC actual	ADpE- 2.1.3.	Línea I.1.
	2.3. Impulsar la producción y distribución a través de Internet de contenidos digitales.	2.3.3. Simplificar las condiciones para la reutilización de la información del Sector Público.	ADpE- 2.3.3.	Línea III.2.
	2.7. Fortalecer la industria TIC mediante el desarrollo de proyectos tecnológicos en servicios públicos.	2.7.6. Desarrollo de servicios para profesionales y ciudadanos, basados en la disponibilidad de la Historia Clínica Digital en el Sistema Nacional de Salud.	ADpE- 2.7.6.	Línea IV.1.
	2.7. Fortalecer la industria TIC mediante el desarrollo de proyectos tecnológicos en servicios públicos.	2.7.10. Impulso de los estándares para favorecer la interoperabilidad de las TIC sanitarias, la teleasistencia y la telemedicina, mediante mecanismos de colaboración con la industria.	ADpE- 2.7.10.	Línea IV.1
	2.7. Fortalecer la industria TIC mediante el desarrollo de proyectos tecnológicos en servicios públicos.	2.7.15 Utilización de entornos virtuales de aprendizaje para la aplicación de planes educativos específicos y para la extensión del concepto de aula en el tiempo y en el espacio.	ADpE- 2.7.15.	Línea IV.1
	2.7. Fortalecer la industria TIC mediante el desarrollo de proyectos tecnológicos en servicios públicos.	2.7.16. Establecimiento de los formatos que deberán ser soportados por las herramientas y sistemas de soporte al aprendizaje en el ámbito de los contenidos educativos digitales públicos.	ADpE- 2.7.16.	Línea IV.1
	2.7. Fortalecer la industria TIC mediante el desarrollo de proyectos tecnológicos en servicios públicos.	2.7.17. Utilización de plataformas digitales y tecnológicas y de recursos didácticos de calidad compartidos por toda la comunidad educativa.	ADpE- 2.7.17.	Línea IV.1

Objetivo	Subobjetivo	Líneas de actuación	Identificador (ID)	Líneas PIL
3. Mejorar la e-Administración y adoptar soluciones digitales para una prestación eficiente de los servicios públicos	3.1. Avanzar hacia una Administración integrada en la sociedad con servicios públicos de calidad centrados en ciudadanos y empresas.	3.1.2. Evolucionar los servicios públicos actuales hacia servicios orientados al ciudadano, de forma que sean personalizables, proactivos, accesibles desde diferentes plataformas, adaptados a las necesidades de los usuarios y fáciles de utilizar, que estén orientados a eventos vitales y que tengan garantizada la calidad y seguridad.	ADpE- 3.1.2.	Línea IV.1
	3.1. Avanzar hacia una Administración integrada en la sociedad con servicios públicos de calidad centrados en ciudadanos y empresas.	3.1.5. Propiciar la Reutilización de la Información del Sector Público para permitir el desarrollo de servicios de alto valor que contribuyan al impulso de la actividad económica y la generación de servicios de valor para ciudadanos y empresas.	ADpE- 3.1.5.	Línea III.2.
	3.4. Promover la cooperación y la colaboración con organizaciones, empresas y agentes sociales en materia de Administración Electrónica.	3.4.1. Promover la compartición de experiencias, proyectos, servicios y aplicaciones de Administración Electrónica desarrollados por todas las Administraciones Públicas, las empresas y las organizaciones y establecer foros de intercambio de las mismas.	ADpE- 3.4.1.	Línea IV.1
	3.4. Promover la cooperación y la colaboración con organizaciones, empresas y agentes sociales en materia de Administración Electrónica.	3.4.2. Establecer un nuevo marco de relación con las organizaciones, las empresas y los agentes sociales que contribuya a dinamizar el mercado de las TIC, especialmente mediante el estudio de mecanismos que faciliten la colaboración público-privada.	ADpE- 3.4.2.	Línea IV.1.
5. Impulsar el sistema de I+D+i en Tecnologías de la Información y las Comunicaciones	5.1. Incrementar la eficacia de la inversión pública en I+D+i en TIC.	5.1.1. Coordinar con el conjunto de agentes públicos implicados en el fomento de la I+D+i en TIC las líneas estratégicas que se apoyan.	ADpE- 5.1.1.	Línea III.2.
	5.1. Incrementar la eficacia de la inversión pública en I+D+i en TIC.	5.1.2. Facilitar la colaboración entre empresas y organismos públicos de investigación mediante iniciativas que fortalezcan el conocimiento mutuo de las capacidades y las necesidades, como por ejemplo los Mapas de Conocimiento, las plataformas tecnológicas o la innovación abierta.	ADpE- 5.1.2.	Línea I.1. Línea III.1. Línea IV.1.

Objetivo	Subobjetivo	Líneas de actuación	Identificador (ID)	Líneas PIL
	5.1. Incrementar la eficacia de la inversión pública en I+D+i en TIC.	5.1.3. Elaborar un Plan de adecuación de los sistemas de gestión de la I+D+i en TIC para incrementar la transparencia, fomentar la participación y colaboración de las entidades solicitantes y facilitar el acceso a los recursos públicos.	ADpE- 5.1.3.	Línea III.2.
	5.2. Fomentar la inversión privada en I+D+i en TIC.	5.2.1. Estimular la inversión privada en I+D+i en la industria electrónica y las TIC mediante el uso estratégico de la compra pública y la colaboración público-privada.	ADpE- 5.2.1.	Línea IV.1.
	5.2. Fomentar la inversión privada en I+D+i en TIC.	5.2.2. Potenciar los fondos de co-inversión con el sector privado en I+D+i aplicadas a las TIC.	ADpE- 5.2.2.	Línea IV.1.
	5.4. Ampliar la participación española en I+D+i en TIC en el ámbito internacional.	5.4.1. Desarrollo de esquemas de cofinanciación e impulso a la participación española en programas europeos e internacionales de I+D+i en TIC.	ADpE- 5.4.1.	Línea II.2.
	5.4. Ampliar la participación española en I+D+i en TIC en el ámbito internacional.	5.4.3. Incremento de la representación española en programas e iniciativas internacionales.	ADpE- 5.4.3.	Línea II.2.
6. Promover la inclusión digital y la formación de nuevos profesionales TIC	6.2. Capacitación digital y formación de nuevos profesionales TIC	6.2.2. Maximizar la eficiencia en la gestión y asignación de los Fondos de Formación destinados a la formación continua en materia TIC, tanto de personal del sector privado como del sector público.	ADpE- 6.2.2.	Línea II.1.
	6.2. Capacitación digital y formación de nuevos profesionales TIC	6.2.3. Asignar parte de los recursos disponibles para la formación continua a la capacitación y adquisición de habilidades digitales de profesionales TIC.	ADpE- 6.2.3.	Línea II.1.
	6.2. Capacitación digital y formación de nuevos profesionales TIC	6.2.4. Reorientar la Formación Profesional relacionada con las TIC.	ADpE- 6.2.4.	Línea II.1.
	6.2. Capacitación digital y formación de nuevos profesionales TIC	6.2.5. Potenciar una mejora de la oferta universitaria destinada a la formación de profesionales TIC	ADpE- 6.2.5.	Línea II.1.

7 Calendario

Medida- Fase (Trimestre/Año)	4ºT 2015	1ºT 2016	2ºT 2016	3ºT 2016	4ºT 2016	1ºT 2017	2ºT 2017	3ºT 2017	4ºT 2017	1ºT 2018	2ºT 2018	3ºT 2018	4ºT 2018	1ºT 2019	2ºT 2019	3ºT 2019	4ºT 2019	1ºT 2020	2ºT 2020	3ºT 2020	4ºT 2020	
Eje 0: Gobernanza																						
Línea I: Gobernanza																						
Eje I: Desarrollo de Infraestructuras Lingüísticas																						
Línea 1: Desarrollo de infraestructuras lingüísticas																						
Eje II: Impulso de la Industria de las Tecnologías del Lenguaje																						
Línea 1: Mejora de la visibilidad y de la transferencia del sector																						
Eje III: La Administración Pública como Impulsor de la Industria del Lenguaje																						
Línea 1: Plataformas de procesamiento de lenguaje natural y traducción automática en las Administraciones Públicas																						
Línea 2: Recursos lingüísticos de las Administraciones Públicas y política de Reutilización de la Información del Sector Público																						
Eje IV: Proyectos Faro de Tecnologías de Procesamiento del Lenguaje Natural																						
Línea 1: Proyectos faro																						

Contratación
 Ejecución

8 Presupuesto

Ejes y Medidas	SETSI		OTROS ORGANISMOS		TODOS	
	Total € (miles)	2016 € (miles)	Total € (miles)	2016 € (miles)	Total € (miles)	2016 € (miles)
Eje 0: Gobernanza	1.260	420	0	0	1.260	420
Línea I: Gobernanza	1.260	420	0	0	1.260	420
Eje I: Desarrollo de Infraestructuras Lingüísticas	14.083	3.231	16.349	0	30.432	3.231
Línea 1: Desarrollo de infraestructuras lingüísticas	14.083	3.231	16.349	0	30.432	3.231
Eje II: Impulso de la Industria de las Tecnologías del Lenguaje	1.780	480	400	0	2.180	480
Líneas 1 y 2: Mejora de la visibilidad y de la transferencia del sector	1.780	480	400	0	2.180	480
Eje III: La Administración Pública como Impulsor de la Industria del Lenguaje	6.180	1.540	0	0	6.180	1.540
Línea 1: Plataformas de procesamiento de lenguaje natural y traducción automática en las Administraciones Públicas	4.140	1.100	0	0	4.140	1.100
Línea 2: Recursos lingüísticos de las Administraciones Públicas y política de Reutilización de la Información del Sector Público	2.040	440	0	0	2.040	440
Eje IV: Proyectos Faro de Tecnologías de Procesamiento del Lenguaje Natural	49.090	8.378	0	0	49.090	8.378
Línea 1: Proyectos faro	49.090	8.378	0	0	49.090	8.378
Total	72.393	14.048	16.749	0	89.142	14.048

9 Indicadores

Ejes y Medidas	Indicador	Fecha
Eje I: Desarrollo de Infraestructuras Lingüísticas		
Línea 1: Desarrollo de infraestructuras lingüísticas	Indicadores europeos armonizados de recursos generales (v.g. Meta-Net ¹²).	2020
Eje II: Impulso de la Industria de las Tecnologías del Lenguaje		
Líneas 1 y 2: Mejora de la visibilidad y de la transferencia del sector	Indicadores del Plan de internacionalización. Catálogo del sector de las tecnologías del lenguaje.	2020
Eje III: La Administración Pública como Impulsor de la Industria del Lenguaje		
Línea 1: Plataformas de procesamiento de lenguaje natural y traducción automática en las Administraciones Públicas	Medidas de uso de las plataformas.	2020
Línea 2: Recursos lingüísticos de las Administraciones Públicas y política de Reutilización de la Información del Sector Público	Indicadores europeos armonizados de recursos reutilizables de las AA	2020
Eje IV: Proyectos Faro de Tecnologías de Procesamiento del Lenguaje Natural		
Línea 1: Proyectos faro	Indicadores europeos armonizados de recursos específicos. Medidas de uso.	2020
Objetivo global del Plan	Crecimiento del sector de las tecnologías del lenguaje.	2020

¹² Fuente: DG Translation, European Commission: Translation figures 2014-LT-Innovate, META-FORUM 2015, Meta-Net.
Plan de Impulso de las Tecnologías del Lenguaje

Anexos

Anexo I: Relación con la estrategia europea

Eje de la UE	Línea de la UE	Eje del PIL	Línea del PIL
I. Soluciones tecnológicas innovadoras para el mercado multilingüe digital único	1.2. Soluciones tecnológicas para el sector público	Eje V: Proyectos Faro	Línea 1: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas
II. Servicios de tecnologías lingüísticas, plataformas e Infraestructuras	2.1. Desarrollo de servicios de tecnologías lingüísticas, plataformas e Infraestructuras	Eje I: Desarrollo de Infraestructuras lingüísticas	Línea 1: Desarrollo de infraestructuras lingüísticas
IV. Medidas Horizontales	4.2. Estándares e interoperabilidad	Eje III: La Administración Pública como impulsor de la industria del lenguaje	Línea 2: Recursos lingüísticos de las Administraciones Públicas y política de Reutilización de la Información del Sector Público.
	4.3. Datos abiertos		

Fuente: "Strategic Agenda for the Multilingual Digital Single Market. Technologies for Overcoming Language Barriers towards a truly integrated European Online Market" META-NET