



PLAN DE ACCIÓN DE LAS TECNOLOGÍAS DE LA LENGUA 2021-2024



EUSKO JAURLARITZA
GOBIERNO VASCO

KULTURA ETA HIZKUNTZA
POLITIKA SAILA
DEPARTAMENTO DE CULTURA
Y POLÍTICA LINGÜÍSTICA



ÍNDICE

Prólogo	3
1. Introducción. Motivación	5
1.1 Las Tecnologías de la Lengua y la Administración pública	5
1.2 Integración del euskera en las Tecnologías de la Lengua	6
1.3 ¿Por qué un Plan de Acción de Tecnologías de la Lengua?	7
2. Visión	10
3. Objetivos generales	11
4. Plan de Acción	13
4.1 Ejes del Plan de Acción	13
4.1.1 Gestión y principios	13
4.1.2. Recursos lingüísticos: recogida, tratamiento y etiquetado de materiales textuales y orales	14
4.1.3. Herramientas para el procesamiento del lenguaje	15
4.1.4. Adquisición y reutilización de recursos lingüísticos y herramientas	15
4.1.5. Traducción automática	16
4.1.6. Tratamiento del habla: interfaces de voz	16
4.2 Ámbitos estratégicos del Plan de Acción	17
4.2.1 Administración electrónica	18
4.2.2 Educación	19
4.2.3 Euskaldunización de las personas adultas	20
4.2.4 Sanidad	21
4.2.5 Justicia	22
4.2.6 Turismo	23
4.2.7 Cultura, medios de comunicación, humanidades digitales, bibliotecas digitales ...	24
4.2.8 Seguridad	25
4.3 Plan de Acción de las Tecnologías de la Lengua en euskera 2021-2024: proyectos faro	26
5. Previsión presupuestaria inicial del Plan de Acción	52
6. Seguimiento del Plan de Acción de las Tecnologías de la Lengua	53
7. Referencias	57



PRÓLOGO

En los últimos años hemos visto lo importante que es para el euskera estar integrado en el desarrollo de la Inteligencia Artificial. El ejemplo más significativo es el traductor neuronal ITZULI, que ha demostrado claramente que se pueden ofrecer servicios de calidad en euskera tanto a las administraciones públicas como al conjunto de la sociedad.

Precisamente, para avanzar en las Tecnologías de la Lengua y situar el euskera entre las lenguas europeas de alto desarrollo tecnológico, hemos elaborado **GAITU, el Plan de Acción de las Tecnologías de la Lengua 2021-2024**. Es un plan de acción porque se compone de proyectos concretos de la Administración. Y es que, en nuestra opinión, es imprescindible, por un lado, recoger y dar a conocer los proyectos que ya están en marcha y los que se van a poner en marcha a lo largo de esta legislatura en el Gobierno Vasco y sus entidades autónomas, y, por otro, con el fin integrar eficazmente el euskera en las Tecnologías de la Lengua, desarrollar al mismo tiempo recursos lingüísticos (datos y modelos lingüísticos), instrumentos para el tratamiento de la lengua (analizadores, etiquetadores, anonimizadores) y herramientas, además de fomentar su uso. De esta manera, además de responder a las necesidades reales de la Administración, iremos construyendo la infraestructura lingüística necesaria para seguir integrando el euskera en las Tecnologías de la Lengua.

Este plan de acción es el primer paso, el punto de partida. Como ya hemos mencionado, de momento, queremos mostrar la fotografía de los proyectos de la Administración basados en Tecnologías de la Lengua. No se ha recogido, por tanto, ningún proyecto externo a la Administración, ya sea de investigación o de otro tipo. Ese será uno de los objetivos para la siguiente fase. No obstante, en este primer paso, junto con la Administración, ya están implicadas en los proyectos en marcha diferentes entidades o empresas del sector, que seguirán estando presentes en los próximos años: entidades creadoras de contenidos, grupos de investigación, empresas desarrolladoras y empresas usuarias. De hecho, desde la iniciativa pública se pretende hacer una apuesta segura por las Tecnologías de la Lengua en euskera y por la recogida masiva de datos lingüísticos necesarios para su desarrollo. Por ello, uno de los proyectos incluidos en el Plan de Acción es la creación de la 'Oficina del Dato', cuyo fin es recoger los datos y modelos lingüísticos ya disponibles y los que se desarrollen en el futuro, para su puesta a disposición de las empresas, personas investigadoras y demás personas interesadas del sector.

El Plan de Acción es una iniciativa tractora de la Administración que, a través de Tecnologías de la Lengua de calidad en euskera, pretende desarrollar y ofrecer servicios lingüísticos básicos y transversales (traducción automática neuronal, modelos de habla, modelos lingüísticos y datos) para utilizarlos primero en todas las administraciones públicas y, en definitiva, para ofrecer mejores servicios públicos a la ciudadanía.

El Plan es abierto y también incluye el seguimiento de los proyectos recogidos en el mismo. Además, a lo largo de la legislatura se irán incorporando nuevos proyectos y al final de la misma se realizará una evaluación previa a la siguiente fase. De esta manera, obtendremos información detallada sobre la tecnología utilizada, los datos, instrumentos y herramientas lingüísticas desarrolladas y su reutilización, el presupuesto utilizado y el que sería necesario para seguir desarrollándolas.

Cabe señalar que este Plan de Acción de Tecnologías de la Lengua confluye con la “Estrategia para la Transformación Digital de Euskadi 2025” aprobada por el Departamento de Desarrollo Económico, Sostenibilidad y Medio Ambiente en 2021, y que, además, desarrolla dicha estrategia, ya que la Inteligencia Artificial es una de las palancas tecnológicas mencionadas en la misma; así mismo, los proyectos incluidos en el Plan de Acción responden a las necesidades de los ámbitos de aplicación recogidos en dicha Estrategia.

Asimismo, este plan de acción está en línea también con “ARDATZ, Plan Estratégico de Gobernanza, Innovación Pública y Gobierno Digital 2030”, publicado por el Departamento de Gobernanza Pública y Autogobierno en 2021, puesto que ambos planes comparten iniciativas. Y no podemos dejar sin mencionar, asimismo, el “Plan de Ciencia, Tecnología e Innovación Euskadi 2030” de la Presidencia del Gobierno Vasco, plan complementario de nuestro Plan de Acción y con objetivos comunes.

Por último, hay que señalar que también hemos relacionado los proyectos que figuran en el Plan de Acción con los ejes del “Plan de Impulso de las Tecnologías del Lenguaje” publicado por el Gobierno del Estado en 2015.

¡Mi agradecimiento a todas las personas que nos han enviado información para la elaboración del Plan de Acción!

En Vitoria-Gasteiz, a 25 de febrero de 2022

Bingen Zupiria Gorostidi
Consejero de Cultura y Política Lingüística

1. INTRODUCCIÓN. MOTIVACIÓN

1.1 LAS TECNOLOGÍAS DE LA LENGUA Y LA ADMINISTRACIÓN PÚBLICA

El mundo actual está interconectado gracias a las tecnologías y, en ese proceso, han cobrado una gran importancia las Tecnologías de la Lengua o Tecnologías del Lenguaje. La aparición de nuevas necesidades sociales y económicas ha hecho que la tecnología, asociada a las lenguas y, en los últimos años a la Inteligencia Artificial, se haya convertido en un factor clave en la vida de las personas.

El objetivo de las Tecnologías de la Lengua es el procesamiento de la lengua, tanto escrita como oral. Las Tecnologías de la Lengua son el Procesamiento del Lenguaje (PL) y la Traducción Automática (TA), que forman parte de la Inteligencia Artificial. Explotan datos masivos de textos, páginas web y redes sociales. A través de los procesadores lingüísticos se ofrecen actualmente aplicaciones como: la detección de entidades designadas (nombres propios, fechas, marcas, productos, etc.), búsqueda inteligente de documentos, clasificación y agrupación de documentos, realización de resúmenes automáticos, extracción de datos de documentos (análisis de sentimientos y minería de opinión), seguimiento y monitorización de la reputación en redes sociales, generación de alertas, corrección de textos, sistemas de aprendizaje de idiomas, y traducción automática de textos.

El mercado de esas tecnologías está creciendo rápidamente y se prevé un mayor crecimiento para los próximos años.

Las Administraciones Públicas deben incorporar las tecnologías del procesamiento del lenguaje natural, así como de la traducción automática, a su actividad diaria, con el fin de mejorar la calidad y la capacidad de servicio público. Además, debido a los contenidos que gestionan en áreas como la salud o la justicia, las Tecnologías de la Lengua son decisivas, por lo que es necesario desarrollar nuevos servicios para esos ámbitos basados en una comprensión más profunda y en la traducción automática.

Por otro lado, las Tecnologías de la Lengua pueden aumentar la demanda ante la administración y actuar como tracción, acelerando así el desarrollo del sector. La implicación de la Administración proporciona visibilidad, prestigio y garantía para sacar provecho de las aplicaciones y productos desarrollados. Asimismo, la participación de las administraciones permite que los recursos y procesadores del euskera sean abiertos y compartidos, además de fomentar la innovación basada en la reutilización de datos públicos.

La interacción entre diferentes administraciones también se intensifica gracias a las tecnologías. En nuestro contexto, contamos con tres administraciones (general, foral y local) en las que las Tecnologías de la Lengua son de gran ayuda.

En definitiva, las administraciones públicas deben ser impulsoras de las Tecnologías del Lenguaje y, además, tienen que crear plataformas conjuntas desde las que ofrecer y compartir recursos y productos con diferentes sectores interesados.

1.2 INTEGRACIÓN DEL EUSKERA EN LAS TECNOLOGÍAS DE LA LENGUA

La Inteligencia Artificial y las Tecnologías de la Lengua han llegado con fuerza y, por eso, es necesario aprovechar la oportunidad que nos brindan con la aplicación de políticas bien diseñadas y eficaces. Los estados poderosos y, sobre todo, los gigantes tecnológicos, como Google, Microsoft, Facebook, Amazon e IBM, han acaparado espacio en lo concerniente a esas tecnologías para que esa ola no se les escape. La transformación tecnológica se está produciendo en un escenario global, pero no de forma homogénea.

La heterogeneidad es un criterio que hay que tener en cuenta en las políticas de activación de las Tecnologías de la Lengua. Por un lado, el rendimiento y disponibilidad de las herramientas y aplicaciones del área no es en absoluto uniforme, aunque todas ellas forman parte de las tecnologías del lenguaje. Por otro lado, existe una gran diferencia en el desarrollo de unas y otras lenguas, o más concretamente entre las lenguas poderosas y las minoritarias.

Parece que esta realidad múltiple y compleja no sólo va a afectar a la parte tecnológica, sino también a la económica y a la social. Por lo tanto, es necesaria una intervención ordenada y eficaz de las instituciones, especialmente de las instituciones públicas.

En el ámbito del euskera, existe un gran riesgo de que las lenguas hegemónicas acaparen plenamente la atención y el apoyo, y de que las lenguas minoritarias, incluido el euskera, queden relegadas.

Dentro de las Tecnologías de la Lengua, las grandes mejoras que se han producido en los últimos años (con *el Big data* y las redes neuronales, entre otras) pueden dar lugar a un segundo empuje en el proceso de recuperación del euskera. Ese segundo impulso puede ser mucho mayor en nuestra sociedad desde el punto de vista cualitativo. Las mejoras sustanciales que se han producido en el tratamiento del habla, en la traducción automática o en el análisis de los textos pueden contribuir de forma significativa a facilitar y ampliar el uso del euskera.

La integración del euskera en las Tecnologías de la Lengua y en el entorno digital ha sido siempre un objetivo estratégico para el Gobierno Vasco. Por esa razón, esas tecnologías han estado y están presentes en los programas de gobierno, en las agendas estratégicas y en todos los planes elaborados y aprobados. Las Tecnologías de la Lengua han sido impulsadas desde el Gobierno en los últimos años de diversas maneras: subvenciones, programas de investigación y puesta a disposición de las personas usuarias de determinados productos, como el traductor automático, primero traductor gramatical y, hace dos años, gracias a la Inteligencia Artificial, el traductor neuronal ITZULI (www.euskadi.eus/itzuli), que se ha convertido en una herramienta indispensable. O la barra de herramientas ITZULI+ (www.euskadi.eus/itzulplus) que reúne diferentes recursos y herramientas lingüísticas de gran utilidad tanto para el ámbito de la docencia como para cualquier otro.

Pero, sin perder tiempo, hay que seguir desarrollando nuevas herramientas para garantizar que la comunidad vasca pueda consumir en euskera los productos que ya están y los que estarán en el mercado: síntesis y reconocimiento de la voz, contenidos audiovisuales subtítulos automáticamente, asistentes de voz, plataformas de servicios digitales, videojuegos, servicios médicos, traductores neuronales especializados, traducción automática de voz a voz, sistemas conversacionales pregunta-respuesta ...).

1.3 ¿POR QUÉ UN PLAN DE ACCIÓN DE TECNOLOGÍAS DE LA LENGUA?

El euskera no es la única lengua que tiene que hacer frente a la necesidad de desarrollo tecnológico. META-NET,¹ en su estudio de 2013, preveía que al menos 21 lenguas europeas estaban en peligro de extinción digital. A pesar de que META-NET dijo que el euskera era una de estas lenguas, hay que señalar que hay lenguas más grandes, con un número mucho mayor de hablantes, que también tienen que hacer frente a retos similares. De hecho, en un mundo en el que el inglés es dominante, el euskera necesita el soporte indispensable del entorno digital.

La Comisión Europea ha visto en las Tecnologías de la Lengua la mejor opción para superar las barreras derivadas de la existencia de gran cantidad de lenguas en la Unión Europea. Por ello, en los últimos años las cosas se están moviendo mucho en Europa, a pesar de no tener en cuenta de manera directa las lenguas no oficiales. No obstante, las estrategias y documentos surgidos de esas iniciativas de la Unión Europea pueden ser una referencia y una oportunidad importante también para el euskera: *Digital Single Market*² (Mercado Único Digital), *Language equality in the digital age*³ (Igualdad lingüística en la era digital) o *Artificial Intelligence: A European Perspective*⁴ (Inteligencia artificial: visión europea). En 2018 el Parlamento Europeo dijo:

“The European Parliament calls on the Commission and the Member States to develop strategies and policy action to facilitate multilingualism in the digital market; requests, in this context, that the Commission and the Member States define the minimum language resources that all European languages should possess, such as data sets, lexicons, speech records, translation memories, annotated corpora and encyclopaedic content, in order to prevent digital extinction”.

(El Parlamento Europeo insta a la Comisión Europea y a los Estados miembros a desarrollar estrategias y acciones políticas para facilitar la diversidad lingüística en el mercado digital y, en ese contexto, insta a la Comisión y a los Estados miembros a determinar los recursos lingüísticos mínimos que deben tener todas las lenguas europeas para evitar su extinción digital, como colecciones de datos, léxicos, grabaciones orales, memorias de traducción, corpus etiquetados y contenidos enciclopédicos).

Por su parte, el Gobierno español aprobó en 2015 *el Plan de Impulso de las Tecnologías del Lenguaje*.⁵ El objetivo del Plan se concreta así:

1 <https://link.springer.com/content/pdf/bfm%3A978-3-642-30796-6%2F1.pdf>

2 <https://ec.europa.eu/digital-single-market/en/shaping-digital-single-market>

3 http://www.europarl.europa.eu/thinktank/es/document.html?reference=EPRS_STU%282017%29598621

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

5 <https://plantl.mineco.gob.es/tecnologias-lenguaje/PTL/Bibliotecaimpulsotecnologiaslenguaje/Detalle%20del%20Plan/Plan-Impulso->

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

El Plan también establece lo siguiente:

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.

Pues bien, este plan de acción de las Tecnologías de la Lengua debe desarrollarse entre todos los departamentos de la Administración General de Euskadi y organismos autónomos y, además, sería conveniente coordinarlo con el plan del Gobierno de España para aunar esfuerzos y sinergias y evitar duplicidades.

El Gobierno Vasco, por su parte, siempre ha tenido muy presente que para el euskera es imprescindible la integración en el entorno digital y en las Tecnologías de la Lengua. La Comisión Especial para Fomentar el Euskera en el Entorno Digital, comisión del Consejo Asesor del Euskera que gestiona la Viceconsejería de Política Lingüística, publicó “El Entorno Digital del Euskera, Recomendaciones 2021-2024”. En el 4.1 se recoge la acción: **“Definir y redactar el Plan de Impulso de las Tecnologías del Lenguaje”**.

Esta comisión especial ya había elaborado en 2016 el trabajo “Recomendaciones para los poderes públicos”, publicado por el Departamento de Cultura y Política Lingüística en el que se recomendaba: “Fomentar el desarrollo de las tecnologías del lenguaje y realizar el plan correspondiente”. Precisamente, con el fin de dar respuesta a esa recomendación a la mayor brevedad posible y con el fin de elaborar un plan de las Tecnologías de la Lengua, el Departamento de Cultura y Política Lingüística encargó en 2018 a Hitz Zentroa de la Universidad del País Vasco un informe. Así, en febrero de 2020 Hitz Zentroa presentó “Hizkuntza-teknologiak sustatzeko planaren txosten teknikoa”(Informe técnico para la elaboración del Plan de Impulso a las Tecnologías del Lenguaje). Dicho informe ha sido tenido muy en cuenta a la hora de elaborar este plan de acción.

El informe señala que el desarrollo de Tecnologías de la Lengua de alta calidad es imprescindible y urgente en el caso del euskera. Imprescindible, si se quiere garantizar la supervivencia de la propia lengua; y urgente, porque el riesgo real de quedar atrás ante la evolución y la velocidad de innovación de esas tecnologías es real.

[Tecnologias-Lenguaje.pdf](#)

Por otro lado, el Plan de Acción de las Tecnologías de la Lengua está conectado con dos importantes planes aprobados y publicados por el Gobierno Vasco en 2021:

- El Departamento de Desarrollo Económico, Sostenibilidad y Medio Ambiente del Gobierno Vasco ha publicado en 2021 la “Estrategia de Transformación Digital de Euskadi 2025”. El Plan de Acción de las Tecnologías de la Lengua debe enmarcarse dentro de esa estrategia, en la que una de las palancas tecnológicas es la Inteligencia Artificial y en cuyos ámbitos de aplicación se encuentran los ámbitos definidos como estratégicos en este plan de acción, entre otros: e-Administración, e-Salud y Bienestar, e-Justicia, Seguridad, Educación, Turismo. Por tanto, el Plan de Acción de las Tecnologías de la Lengua debe entenderse como un desarrollo de esa estrategia más general.
- “ARDATZ, Plan Estratégico de Gobernanza, Innovación Pública y Gobierno Digital 2030”, del Departamento de Gobernanza Pública y Autogobierno. En el apartado 4.3 de dicho plan se han recogido las estrategias y ejes del mismo: I. Gobierno íntegro, igualitario y bilingüe (el euskera, lengua de trabajo y de servicio); II. Gobierno innovador y en aprendizaje permanente; III. Gobierno digital centrado en la ciudadanía; IV. Gobierno basado en datos y que rinde cuentas (Gobernanza de los datos y de la información); V. Gobierno abierto. Como ya se ha mencionado, el Plan de Acción de las Tecnologías de la Lengua es un plan interdepartamental que va a contribuir a la consecución de algunos de los objetivos recogidos en el Plan Ardatz, ya que algunos de sus proyectos son del Departamento de Gobernanza Pública y Autogobierno.

Por último, el Plan de Acción es un paso inicial para la integración transversal de las Tecnologías de la Lengua en euskera en la actividad ordinaria de la Administración General Vasca para el periodo 2021-2024. Se trata de un plan de acción porque recoge los proyectos que se están desarrollando y que se desarrollarán en el Gobierno Vasco a lo largo de la presente legislatura, a través del uso de las Tecnologías de la Lengua. Las tecnologías avanzan rápidamente y es importante que, a la mayor brevedad posible, se ofrezcan a las personas usuarias productos concretos, a la vez que los corpus y recursos lingüísticos desarrollados en los proyectos se guarden y se coloquen en una plataforma para su reutilización.

Detrás de este plan de acción vendrán otros. El punto de partida actual es recoger y hacer seguimiento de los proyectos que ya están en marcha en el Gobierno Vasco y que se van a llevar a cabo a lo largo de la legislatura 2021-2024.

Las Tecnologías de la Lengua son el eje vertebrador del Plan de Acción, por lo que no se incluyen en el mismo las iniciativas del entorno digital, que quedan a cargo de la Comisión Especial para Fomentar el Euskera en el Entorno Digital del Consejo Asesor del Euskera. Tampoco se incluyen los proyectos de investigación que se incorporarán más adelante.

En cuanto a la investigación, cabe señalar que el “Plan de Ciencia, Tecnología e Innovación de Euskadi 2030”, publicado por la Presidencia del Gobierno Vasco en 2021, enlaza directamente con la apuesta estratégica del País Vasco por la investigación y la innovación.

2. VISIÓN

Mejorar la calidad y la capacitación del servicio público en el seno de la Administración General de la Comunidad Autónoma del País Vasco mediante la incorporación de las Tecnologías de la Lengua (mediante la Inteligencia Artificial, las Tecnologías de Procesamiento del Lenguaje Natural y la Traducción Automática), actuando además como impulsoras o tractoras de la demanda. Con este plan de acción se pretende impulsar la recopilación, estandarización, difusión y reutilización de los recursos lingüísticos generados en el contexto de la actividad de la gestión pública bilingüe de la Administración.

De cara al futuro, las tecnologías de la Lengua serán muy importantes para el euskera. En el marco de nuestra pequeña dimensión, la Administración pública debe propiciar la plena integración e impulso de esas tecnologías para garantizar en gran medida la interacción con la ciudadanía y, en consecuencia, la utilización del euskera en el mayor número de situaciones posibles.

Con este plan de acción se pretende priorizar las Tecnologías de la Lengua como única forma de garantizar la supervivencia tecnológica de la lengua. El desarrollo de esas tecnologías es imparable y si no utilizamos la oportunidad que nos brindan, serán otras lenguas (inglés, castellano) las que ocuparán ese espacio.

Las Tecnologías de la Lengua son un sector transversal emergente vinculado a la innovación, con capacidad para fomentar el crecimiento, la competitividad y el empleo de calidad.

Es necesario impulsar y coordinar acciones dentro de la Administración Pública, buscar sinergias y evitar duplicidades de esfuerzos. Las Tecnologías de la Lengua (Inteligencia Artificial) requieren una infraestructura informática potente (supercomputación) en la que, en principio, la Administración Pública tiene más posibilidades de realizar inversiones.

Este plan de acción es el inicio de un plan más largo y continuado en el tiempo. La Administración Pública debe preparar el camino para garantizar una infraestructura sostenible en el tiempo para las generaciones futuras que sitúe al euskera en el centro de la innovación gracias al desarrollo tecnológico.

En definitiva, el objetivo de la Administración debe ser mejorar la vida de las personas y para ello, hoy en día, es imprescindible disponer de todo tipo de tecnologías.

3. OBJETIVOS GENERALES

Los objetivos del Plan de Acción de las Tecnologías de la Lengua son:

- Sentar las bases para el desarrollo e incorporación, de forma transversal, de las Tecnologías de la Lengua en las áreas estratégicas de la Administración General de Euskadi (Administración electrónica, Educación, Euskaldunización de Personas adultas, Sanidad, Justicia, Turismo, Cultura y Medios de Comunicación, y Seguridad), estableciendo así las condiciones para la creación de un ecosistema de Inteligencia Artificial.
- Impulsar la colaboración y coordinación dentro de la Administración General en el desarrollo y aplicación de las Tecnologías de la Lengua para aprovechar sinergias, sincronizar acciones, evitar redundancias y, en definitiva, prestar un mejor servicio público a la ciudadanía.
- La canalización y ejecución de proyectos faro o transversales, tanto en cuanto las administraciones públicas deben ejercer como efecto tractor.
- Facilitar la creación y puesta a disposición de todas las entidades interesadas de los recursos lingüísticos necesarios para el desarrollo de Tecnologías de la Lengua de calidad: corpus textuales y orales (sin tratar, etiquetados, generales y especializados, monolingües, bilingües y plurilingües). Para ello se creará una plataforma de datos y un protocolo de uso de los mismos.
- Reducir la distancia existente entre el castellano y el euskera en lo concerniente al desarrollo tecnolingüístico, al ser, la nuestra, una administración pública bilingüe.
- Recoger y dar a conocer, utilizando las Tecnologías de la Lengua, los proyectos en marcha en las áreas estratégicas de la Administración General de la CAPV y los que se van a desarrollar en el periodo 2021-2024, así como las propuestas que deberían desarrollarse en el futuro en dichas áreas estratégicas u otras.
- Visibilizar las Tecnologías del Lenguaje ante las administraciones y la ciudadanía en general para que vean el impacto positivo que esas tecnologías tienen en su bienestar.
- Fortalecer el sector de las industrias de las Tecnologías de la Lengua de la Comunidad Autónoma del País Vasco a través de la colaboración con la Administración y, por tanto, fomentando y potenciando la colaboración público-privada.



- Promover un desarrollo sostenible, integrador, seguro y equitativo de la Inteligencia Artificial en lo referente al euskera. Además, es necesario aprovechar el boom actual de las Tecnologías de la Lengua para que el euskera esté presente en ellas, de manera que las herramientas desarrolladas en otras lenguas, como pueden ser los asistentes conversacionales y los chatbots, también estén disponibles para las personas euskaldunes que deseen utilizarlas.
- Establecer mecanismos de coordinación y seguimiento del Plan de Acción. El Plan será dinámico (abierto), adaptable y deberá ser actualizado de cara al futuro según la evolución de la situación, aprovechando las experiencias adquiridas. Asimismo, contribuirá a la detección de necesidades en los diferentes ámbitos.



4. PLAN DE ACCIÓN

4.1 EJES DEL PLAN DE ACCIÓN

4.1.1 Gestión y principios

Dado que uno de los objetivos principales del Plan de Acción es sentar las bases para avanzar en el desarrollo, incorporación y desarrollo transversal de las Tecnologías del Lenguaje en las áreas estratégicas de la Administración General de Euskadi, el Plan será gestionado por el Gobierno Vasco. Al ser un plan interdepartamental, se coordinará desde la Viceconsejería de Política Lingüística en colaboración con los responsables de las áreas definidas como estratégicas. Se fomentará el trabajo colaborativo, buscando estrategias comunes, intercambiando ideas, compartiendo recursos y herramientas, analizando conjuntamente las necesidades y proponiendo líneas de trabajo.

Principios:

1. **Participación:** además de las áreas estratégicas de la Administración implicadas en este plan de acción, se impulsará la participación del resto de departamentos. Asimismo, se buscará la de las demás administraciones públicas.
2. **Transparencia:** se visualizarán los proyectos que a lo largo de la legislatura 2021-2024 se están desarrollando y se van a desarrollar en la Administración, y la información sobre su seguimiento.
3. **Apertura y difusión:** los datos, recursos y herramientas derivados de los proyectos promovidos por la Administración se pondrán a disposición de las personas usuarias y de las personas interesadas. Su reutilización será una de las metas del Plan de Acción.
4. **Durabilidad:** este plan de acción es el punto de partida. Recoge los proyectos que se están desarrollando y se van a desarrollar durante la legislatura 2021-2024 y que integran las Tecnologías de la Lengua del euskera. Por primera vez se han recogido en un documento todos esos proyectos, así como los que se incorporarán a lo largo de ese periodo y, además, y además se hará el seguimiento de todos. Por lo tanto, habrá que seguir desarrollando el Plan de Acción legislatura a legislatura.
5. **Dinamicidad y adaptabilidad:** el Plan de Acción no es un documento cerrado, debido a tecnologías tan cambiantes, puede tener modificaciones (para introducir cambios en los proyectos o para integrar nuevos proyectos).

6. Transversalidad: si bien inicialmente el Plan de Acción ha implicado a algunos departamentos del Gobierno Vasco (los considerados estratégicos), se extenderá al resto de departamentos. Asimismo, se buscará la coordinación, por un lado, con la “Estrategia para la Transformación Digital de Euskadi 2025” del Departamento de Desarrollo Económico, Sostenibilidad y Medio Ambiente, y, por otro, con “ARDATZ, Plan Estratégico de Gobernanza, Innovación Pública y Gobierno Digital”. Además, también se alineará con el “Plan de Impulso de las Tecnologías del Lenguaje” del Gobierno del Estado.
7. Colaboración entre las instituciones públicas: para impulsar su digitalización, aprovechar sinergias, compartir recursos y herramientas y, en definitiva, ofrecer a la ciudadanía servicios públicos de calidad en euskera en todos los niveles de la Administración. Porque los servicios públicos no serán buenos si no se ofrecen también en euskera.
8. Colaboración público-privada: el desarrollo conjunto de proyectos potenciará el sector tanto desde el punto de vista tecnológico como económico.
9. Obtener un beneficio global para el conjunto de la sociedad a través de las Tecnologías de la Lengua.

4.1.2. Recursos lingüísticos: recogida, tratamiento y etiquetado de materiales textuales y orales

Los recursos lingüísticos de calidad, indispensables para el desarrollo de las Tecnologías de la Lengua, deben ser recogidos, tratados, etiquetados y puestos en una plataforma pública, en los formatos adecuados y bajo licencia libre. Además, teniendo en cuenta los diferentes usos de dichos recursos, es necesario definir y concretar un protocolo. Se trata, en definitiva, de que las empresas, universidades, agentes tecnológicos y demás personas interesadas que trabajan en el desarrollo tecnolingüístico del euskera puedan utilizar esos recursos para desarrollar, mejorar y entrenar modelos lingüísticos y crear herramientas y aplicaciones.

Los recursos lingüísticos que hay que recoger y etiquetar son:

- Corpus escritos: corpus generales monolingües (etiquetados: morfología, sintaxis, semántica, correferencias, coherencia, roles semánticos, entidades nominativas, expresiones temporales, eventos, etc.), corpus plurilingües paralelos o comparables (memorias de traducción, textos bilingües o multilingües), corpus especializados de diferentes ámbitos (educación, justicia, sanidad, turismo, administración, etc.), corpus históricos o no estándar.
- En cuanto a los corpus no estándar, sería enriquecedor reunir y tratar las formas dialectales y/o los numerosos corpus (de las redes sociales) escritos en los últimos años en un lenguaje no formal. Sin embargo, no se abordarán en esta primera fase del Plan de Acción por sus características, necesidades de tratamiento y criterios diferentes.
- Corpus orales: en la actualidad se generan muchos datos en formato multimedia en diferentes plataformas y sitios (medios de comunicación, Internet ...) y la obtención y etiquetado de esos datos es fundamental para poder utilizar y explotar, en las tecnologías de la voz, de manera adecuada y cómoda, la información que contienen: convertir la voz en texto (reconocimiento de voz, *STT - Speech-to-text*) y convertir el texto en voz (síntesis de voz, *TTS - Text-to-Speech*).

- Entre los corpus orales, cabe destacar el proyecto “Common Voice⁶” de la Fundación Mozilla, que se está desarrollando en diferentes países y lenguas. La Asociación Librezale está recogiendo, a través de Internet, grabaciones de voz en euskera con metadatos. Posteriormente, es necesario validar dichas grabaciones.
- Léxicos, diccionarios, enciclopedias, bancos terminológicos, bases de conocimiento léxico-semánticas (ontologías): proporcionan información léxica, como por ejemplo la categoría gramatical, y también las relaciones semánticas entre las palabras (sinonimia, antonimia, hiperonimia/hiponimia).

4.1.3. Herramientas para el procesamiento del lenguaje

Para integrar el euskera en las Tecnologías de la Lengua y, por consiguiente, para desarrollar herramientas y aplicaciones para procesar recursos lingüísticos y prepararlos para que sean útiles, es decir, para crear modelos, se requieren:

Herramientas de análisis y etiquetado. Estos son, a menudo, los más complejos:

- Herramientas relacionadas con textos: analizadores morfológicos, sintácticos, semánticos (para identificar personas, organizaciones, lugares ...) y analizadores del discurso. Son muy necesarios para realizar búsquedas. Las técnicas para su desarrollo han ido cambiando y actualmente se está usando cada vez más la Inteligencia Artificial.
- Herramientas relacionados con el habla o la voz que permiten etiquetar los sistemas de reconocimiento del habla y los materiales multimedia mediante transcripción enriquecida. Son imprescindibles para obtener instrumentos de transcripción automática. Facilitan la gestión y búsqueda de contenidos multimedia. Se pueden crear modelos genéricos y posteriormente adaptarlos y utilizarlos para **ámbitos** específicos.

Herramientas de anonimización de datos personales y entidades. Son indispensables para limpiar corpus y memorias de traducción.

4.1.4. Adquisición y reutilización de recursos lingüísticos y herramientas

Actualmente existe una gran masa de información accesible (textos, productos multimedia ...) que está aumentando constantemente tanto en la administración pública como en los medios de comunicación y en Internet en general. Sería muy beneficioso que se dieran las condiciones adecuadas para su reutilización.

Así, en este primer paso, dentro del Plan de Acción, se pretende impulsar una plataforma que recoja los recursos lingüísticos ya existentes en la Administración Pública, los que se creen en los próximos años (tanto textuales como orales), así como las herramientas que se desarrollen. A ese fin, se creará en los servidores de la Sociedad Informática EJIE la plataforma “Oficina del

⁶ <https://voice.mozilla.org/en/datasets>; <https://commonvoice.mozilla.org/es>

Dato”. En ella se depositarán los corpus, recursos, modelos lingüísticos y herramientas ya desarrollados y los que se desarrollarán para los diferentes proyectos, con el fin de que las empresas del sector y los centros de innovación desarrollen aplicaciones y herramientas con menor coste y tecnologías competitivas. Tal y como se ha mencionado anteriormente, se deberá establecer un protocolo de distribución y acceso a los recursos respetando las licencias de cada producto.

Además, dado que la explotación de esos recursos lingüísticos y el aprovechamiento para las Tecnologías de la Lengua requiere de máquinas con un alto poder computacional, la ventaja para el sector es que podrán obtener los datos ya procesados (etiquetados) y listos para su uso.

4.1.5. Traducción automática

Es evidente que en la actualidad los traductores automáticos se han convertido en herramientas imprescindibles para las actividades cotidianas (administración pública, escuela, empresa, negocios, Banca, Sanidad, ámbito socio-económico), así como para el ocio (deporte, cultura ...). Desde la integración en redes neuronales (Inteligencia Artificial) se ha producido un salto importante en la calidad de los traductores automáticos. Estas herramientas facilitan en este momento que el bilingüismo y el plurilingüismo se conviertan en un reto cotidiano.

En los últimos años se han desarrollado algunos traductores neuronales que ofrecen el euskera. El Gobierno Vasco ha puesto a disposición de toda la ciudadanía el traductor neuronal genérico ITZULI (<https://www.euskadi.eus/itzuli>) y el traductor neuronal jurídico-administrativo; además, en el caso del traductor jurídico-administrativo, a través de la red Euskalsarea, las administraciones públicas pueden traducir también documentos. Estas herramientas están siendo cada vez más utilizadas en el ámbito profesional y, poco a poco, se están convirtiendo en herramientas de ayuda imprescindibles.

A partir de ahora, además de seguir mejorando la calidad de la traducción, será necesario desarrollar traductores automáticos para actividades específicas (Sanidad, Justicia ...), entrenados con corpus y terminología de esas áreas. Además, en el ámbito de la traducción automática, el reto de futuro cercano será la traducción voz-voz.

4.1.6. Tratamiento del habla: interfaces de voz

La Inteligencia Artificial también está siendo utilizada en tecnologías de voz (síntesis y reconocimiento), lo que ha permitido un aumento considerable de la calidad. En los últimos años, cada vez están más presentes en los dispositivos que utilizamos a diario, teléfonos móviles, televisión y altavoces inteligentes, entre otros. Ofrecen facilidad de uso y todas las personas nos estamos acostumbrando a esa comodidad.

Por ese motivo, cada vez se están usando más los asistentes virtuales y los sistemas conversacionales automáticos (chatbots, incluso textuales). No existen todavía en euskera, por tanto, este será uno de los objetivos a conseguir.

El reto es diseñar interfaces de voz adecuadas y obtener voces naturales, espontáneas, que expresen personalidad y emociones; para ello, será necesario trabajar voces diversas y expresivas, teniendo en cuenta diferentes tonos y dialectos.



4.2 ÁMBITOS ESTRATÉGICOS DEL PLAN DE ACCIÓN

Los ámbitos considerados estratégicos para este Plan de Acción son:

1. Administración electrónica
2. Educación
3. Euskaldunización de personas adultas
4. Sanidad
5. Justicia
6. Turismo
7. Cultura, medios de comunicación, humanidades digitales, bibliotecas digitales ...
8. Seguridad

A continuación se recogen las razones para la integración de las Tecnologías de la Lengua en esos ámbitos, los objetivos que se podrían alcanzar y las tareas que se podrían desarrollar. Además, en cada caso se enumeran las tecnologías y herramientas que se deberían utilizar.



4.2.1 Administración electrónica

MOTIVACIÓN	OBJETIVOS	TAREAS, TECNOLOGÍAS Y HERRAMIENTAS
<p>- “ARDATZ, Plan Estratégico de Gobernanza, Innovación Pública y Gobierno Digital 2030” del Departamento de Gobernanza Pública y Autogobierno, tiene entre sus objetivos: impulsar el uso del euskera como lengua de trabajo y de servicio; contar con sistemas de gestión pública avanzada e innovación; disponer de herramientas adaptadas a los retos que se planteen en cada momento; ofrecer de forma proactiva a la ciudadanía un servicio público y personalizado; la seguridad de los datos de la gobernanza y de la información.</p> <p>- Requisitos de las Administraciones:</p> <ul style="list-style-type: none"> · Transparencia: Ley 19/2013 · Comunicación y participación: de la ciudadanía y de los agentes. · Protección de datos: Reglamento 2016/679 del Parlamento Europeo y la Ley Orgánica 3/2018. · Inlusividad y universalidad: consideración de la diversidad funcional en la prestación de los servicios públicos. 	<p><u>Configuración de la administración electrónica</u>: el manejo en soporte electrónico de la información (en formato de texto o de voz) que gestionan las administraciones, de forma que se facilite una gestión ordenada y eficaz.</p>	<ol style="list-style-type: none"> a. Tramitación semiautomática de solicitudes en ventanilla única. b. Procesamiento automático de expedientes. c. Transparencia. d. Actas automáticas: transcripción, resumen y estructuración. e. Anonimización de datos. f. Digitalización y consulta del archivo y del patrimonio. g. Inlusividad y universalidad: trabajar la voz y el texto con perspectiva universal. h. Seguimiento de la información y vigilancia tecnológica. <p>TECNOLOGÍAS Y HERRAMIENTAS</p> <ul style="list-style-type: none"> • Reconocimiento de voz: transcripción • Traductor automático • Sistemas conversacionales automáticos (chatbots) • Búsqueda, extracción, etiquetado y clasificación automática de documentos • Corrector ortográfico, gramatical y de lenguaje no sexista • Estructuraciones automáticas por plantillas • Resúmenes automáticos • Anonimización automática de datos personales • Alimentación automática de bases de datos • Sistema de ayuda automático para facilitar la lectura de textos

4.2.2 Educación

MOTIVACIÓN	OBJETIVOS	TAREAS, TECNOLOGÍAS Y HERRAMIENTAS
<ul style="list-style-type: none"> - En 2018 la Comisión Europea comenzó a desplegar el Plan de Acción de la Educación Digital (<i>Digital Education Action Plan</i>): la tecnología digital (Inteligencia Artificial, Robótica y Computación en la nube) enriquece de diferentes maneras los procesos de aprendizaje y ofrece oportunidades de aprendizaje que deben estar al alcance de todas las personas. - El Plan Heziberri 2020 en el marco pedagógico (Decreto 236/2015 BOPV): el alumnado que ha finalizado la Educación Básica debe contar con una competencia digital y mediática, en línea con el marco europeo de la competencia digital, que garantice la plena alfabetización y capacitación funcional que demanda la ciudadanía actual. - Educación inclusiva. - ISEI-Instituto Vasco de Evaluación e Investigación Educativa (IVEI) 2018: “se está ralentizando la euskaldunización del alumnado”. - Decreto 236/2015, de 22 de diciembre, por el que se establece el currículo de Educación Básica y se implanta en la Comunidad Autónoma del País Vasco. 	<ul style="list-style-type: none"> - Aplicación e integración de técnicas avanzadas de procesamiento de textos y tratamiento de voz: generación y análisis automático de materiales de diferentes áreas de conocimiento, aprendizaje personalizado, fomento de la motivación del alumnado, reducción de desigualdades sociales. - Trabajar el uso de las redes sociales desde el punto de vista educativo: búsqueda de información y recursos, desarrollo del pensamiento crítico, el empoderamiento crítico como parte de los procesos educativos. 	<p>[En todos los niveles educativos]</p> <ol style="list-style-type: none"> a. Desarrollar herramientas para el alumnado y el profesorado. b. Crear entornos virtuales de aprendizaje. c. Crear la plataforma de la comunidad educativa (para compartir recursos didácticos ...) d. Elaborar instrumentos de colaboración entre los diferentes agentes educativos. e. Desarrollar herramientas para trabajar el nivel lingüístico alcanzado en las diferentes áreas de conocimiento. f. Garantizar la anonimización de los datos. g. Detección automática de noticias falsas que aparecen en redes sociales y portales de noticias. h. Desarrollar herramientas para la detección del cyberbulling. <p>TECNOLOGÍAS Y HERRAMIENTAS</p> <ul style="list-style-type: none"> • Resúmenes automáticos • Preguntas y respuestas • Herramientas de ayuda a la escritura, corrección de escritos, etiquetado de materiales audiovisuales, transcripción y gestión • Recopilar y procesar corpus educativos • Herramientas especializadas de aprendizaje y formación • Análisis automático de sentimientos en entrevistas en redes sociales

4.2.3 Euskaldunización de las personas adultas

MOTIVACIÓN	OBJETIVOS	TAREAS, TECNOLOGÍAS Y HERRAMIENTAS
<ul style="list-style-type: none"> - Enseñanza en contexto digital: el desarrollo tecnológico y las posibilidades de desarrollo de competencias digitales no se explotan suficientemente. - Es necesario mejorar los sistemas o recursos que actualmente se utilizan en euskera (como Ikasbil, BOGA o ARIAN) e integrar los recursos y Tecnologías de la Lengua en entornos digitales como PRIMERAN o ARIAN. - Diseñar herramientas de autoaprendizaje centradas en el alumnado y que posibiliten la autoevaluación. 	<p>[HABE + euskaltegis + grupos de investigación]</p> <ul style="list-style-type: none"> - Formación del profesorado de los euskaltegis en Tecnologías de la Lengua. - Reutilización de los datos y materiales que se generan en los euskaltegis y creación de recursos lingüísticos (corpus, guías ...). - Elaborar un plan de disponibilidad de la tecnología lingüística en los euskaltegis, adaptando y facilitando la utilización de las herramientas desarrolladas, así como detectando y diseñando las herramientas estratégicas necesarias para fomentar la enseñanza de la lengua a partir de los datos y elaborar y simplificar los sistemas de evaluación de acuerdo con los estándares. 	<ul style="list-style-type: none"> a. Elaboración y etiquetado de un corpus de textos escritos y orales (tareas y variables de comunicación). b. Crear un sitio (banco) donde se recojan materiales didácticos (documentación, vídeos, audios). c. Desarrollar un entorno amplio y accesible para el aprendizaje de la lengua online. d. Crear recursos que ayuden a desarrollar herramientas de autoaprendizaje y materiales didácticos. e. Análisis automático de textos. f. Crear herramientas robustas de búsqueda y visualización de datos en corpus multimodales. g. Subtitulado automático de materiales audiovisuales mediante el reconocimiento de la voz. h. Clasificación automática de textos (A1-2/B1-2/C1-2) <p>TECNOLOGÍAS Y HERRAMIENTAS</p> <ul style="list-style-type: none"> • Corrección automática • Reconocimiento de voz: transcripción • Traductor automático • Sistemas conversacionales automáticos (chatbots) • Búsqueda, extracción, etiquetado y clasificación automática de documentos • Corrector ortográfico, gramatical y del lenguaje no sexista

4.2.4 Sanidad

MOTIVACIÓN	OBJETIVOS	TAREAS, TECNOLOGÍAS Y HERRAMIENTAS
<p>El euskera y el castellano son lenguas cooficiales en la CAPV. Aunque a menudo la comunicación con el personal sanitario se realiza en euskera, los informes sanitarios se redactan en castellano. Osakidetza cuenta con un sistema centralizado y la comprensión de los informes sanitarios por parte de todo el personal es muy importante para la seguridad de los y las pacientes.</p> <p>Entre los retos y proyectos estratégicos de Osakidetza 2017-2020 se establecieron dos acciones concretas para incrementar la presencia del euskera en la documentación e historia clínica:</p> <ul style="list-style-type: none"> - Incrementar la presencia del euskera en la documentación clínica y en la historia clínica. - Avanzar hacia una historia clínica única y bilingüe. <p>El personal sanitario profesional demanda cada vez más el uso del euskera en la historia clínica, pero una gran parte de los profesionales no conocen las dos lenguas.</p> <p>Las tecnologías del lenguaje pueden aportar luz porque desarrollan técnicas de extracción de información para diferentes tipos de lengua y textos.</p>	<p>Utilizar en Osakidetza sin problemas las dos lenguas cooficiales, haciendo frente a los problemas de comprensión a través de las tecnologías.</p> <p>Existen herramientas adecuadas para el procesamiento de textos sanitarios en castellano, pero se necesita una mayor adaptación y desarrollo para el euskera. Las herramientas deben ser bilingües para garantizar los derechos lingüísticos de los y las pacientes.</p> <p>La traducción automática puede ser una herramienta muy útil. Pero hay que desarrollar un traductor automático específico, creando un corpus paralelo de informes sanitarios amplio y completo. Los informes de salud monolingües también son muy útiles para mejorar los resultados, pero son escasos.</p> <p>El reconocimiento y procesamiento de la voz también es indispensable para llegar a más profesionales sanitarios y pacientes.</p> <p>La Sanidad es un ámbito de gran impacto, ya que, en definitiva, los servicios sanitarios son utilizados por toda la sociedad.</p>	<ul style="list-style-type: none"> a. Creación de corpus paralelos (informes sanitarios, memorias de traducción) y explotación de corpus monolingües. b. Procesar la documentación clínica. c. Desarrollar una cadena abierta y multilingüe para el procesamiento de la documentación clínica (pipeline). d. Creación de herramientas de ayuda para la redacción de informes sanitarios (dictado). e. Elaborar un sistema automático de traducción de informes sanitarios. f. Crear recursos terminológicos y léxicos. <p>TECNOLOGÍAS Y HERRAMIENTAS</p> <ul style="list-style-type: none"> • Búsqueda, extracción, etiquetado y clasificación automática de documentos • Anonimización automática de datos personales • Herramientas de procesamiento de textos • Tecnologías de la voz: síntesis y reconocimiento • Traducción automática y postedición automática • Correctores: ortográfico, gramatical • Estandarización y creación de plantillas tipo • Diccionarios bilingües y plurilingües • Extracción automática de terminología, expansión y desambiguación de abreviaturas

4.2.5 Justicia

MOTIVACIÓN	OBJETIVOS	TAREAS, TECNOLOGÍAS Y HERRAMIENTAS
<p>El Centro de Documentación Judicial lleva años trabajando en profundidad la selección, organización, tratamiento, difusión y publicación de información legislativa y jurisprudencial.</p> <p>Entre los recursos desarrollados por el centro destaca el corpus jurisprudencial compuesto por seis millones de documentos. Las personas miembros de la Carrera Judicial disponen íntegramente del corpus a través de medios telemáticos. Sin embargo, al público general se le ofrece una parte del corpus tras un proceso de anonimización previo. Además, se ha desarrollado un tesoro jurídico que se ha utilizado en el proyecto LEGIVOC para la construcción de una interfaz semántica entre las diferentes terminologías nacionales del derecho.</p> <p>Por tanto, en la actualidad, la gestión de grandes depósitos con documentación jurídica y jurisprudencia ya incorpora técnicas avanzadas para el tratamiento automático del lenguaje, aprendizaje automático y análisis semántico. Mediante ellos, la extracción de información, catalogación y clasificación de los documentos digitalizados se realiza de forma homogénea y automática. Sin embargo, debido al elevado coste de estos procesos, por el momento, solo se procesan las sentencias de los órganos colegiados. Sería conveniente, por tanto, procesar todas las sentencias (incluidas las de los órganos unipersonales) y los millones de documentos jurídicos que se generan y gestionan en la administración.</p>	<p>Contribuir a la mejora de la eficiencia (eliminación de duplicidades) y eficacia (mejora de los procesos productivos en marcha e incorporación de nuevas funcionalidades y capacidades) en el ámbito jurídico. Se pretende aplicar técnicas avanzadas de procesamiento de textos, tratamiento de voz y traducción automática en los depósitos de documentación jurídica existentes y en las lenguas cooficiales. El procesamiento de la documentación judicial y legal, mediante técnicas de procesamiento de textos y traducción automática, puede tener un gran impacto en el ámbito judicial y, también, en el conjunto de la sociedad por la marca transversal de justicia.</p> <p>Además, para acceder a las diferentes fases del trabajo profesional de la justicia es imprescindible utilizar las tecnologías de la voz (reconocimiento y procesamiento).</p>	<ol style="list-style-type: none"> Alinear corpus paralelos (memorias de traducción) y procesar corpus monolingües. Preparar el procesamiento judicial y documental. Crear una cadena abierta y multilingüe para el procesamiento de textos (pipeline) del ámbito de la Justicia. Desarrollar un sistema de traducción automática específico para dicho ámbito. Crear recursos terminológicos y léxicos. Impulsar la digitalización de las vistas judiciales. Desarrollar la verificación automática de personas y documentos. <p>TECNOLOGÍAS Y HERRAMIENTAS</p> <ul style="list-style-type: none"> Búsqueda, extracción, etiquetado y clasificación automática de documentos Anonimización automática de datos personales Resumen automático Analizadores morfosintácticos y semánticos Traducción automática Diccionarios especializados Extracción automática de terminología Tecnologías de voz: transcripción, indexación, interfaces de voz, identificación de personas, certificación y verificación (técnicas de detección de defraudadores e intrusos– <i>spoofing</i> y <i>antispoofing</i> –)

4.2.6 Turismo

MOTIVACIÓN	OBJETIVOS	TAREAS, TECNOLOGÍAS Y HERRAMIENTAS
<p>El turismo se caracteriza por la necesidad de comunicación y, por tanto, de las lenguas. El sector turístico tiene además una enorme capacidad económica.</p> <p>Hoy en día, la recogida de la opinión de turistas y personas usuarias se considera imprescindible en Internet, sobre todo en las redes sociales, y para ello se necesitan tecnologías.</p> <p>La información turística en muchos idiomas requiere sistemas gestores de contenidos. La traducción automática es una herramienta importante ya que la lengua local y otras se pueden procesar de forma sencilla. Los conversores de texto a voz y los de voz a texto son también importantes en este ámbito, ya que aumentan la accesibilidad y la comunicación.</p> <p>Por otra parte, el informe “Plan Estratégico del Turismo Vasco 2020” publicado por el Gobierno Vasco, en la definición de las líneas de acción e iniciativas estratégicas, realizaba una apuesta clara por el uso de la tecnología lingüística en el ámbito turístico. Recogía en tres apartados la forma de abordar la estrategia de innovación: innovación tecnológica, marca y comunicación y fomento de la formación lingüística.</p>	<p>En el ámbito del turismo se pueden prever dos grupos de aplicaciones concretas: la de analizar el valor añadido sobre los museos y el patrimonio, y la extracción de la opinión de los y las turistas en las redes sociales. Ambas en el escenario de un entorno multilingüe. Las páginas web de las empresas turísticas también pueden ser un objetivo importante.</p> <p>Ofrecer esas herramientas en forma de servicio en la nube a disposición de agentes turísticos (administración, consorcios, empresas tecnológicas) puede ser la solución más adecuada. Habría que dar especial importancia al euskera para poder aportar un valor añadido al patrimonio en euskera. La persona visitante debe ser consciente de que ha venido al País Vasco y de que está en el país del euskera.</p> <p>También se podrían desarrollar aplicaciones de ayuda relacionadas con el idioma para el trabajo de guía. Una innovación interesante sería un chatbot multilingüe para el servicio de información turística del País Vasco.</p> <p>En el paisaje lingüístico también es importante la señalización. Disponer de aplicaciones adecuadas en aeropuertos, estaciones o en los mensajes de autopistas supondría un avance importante, ya que en un entorno plurilingüe es de gran ayuda que cada persona reciba los mensajes en su propio idioma, de forma accesible y comprensible.</p>	<ol style="list-style-type: none"> Crear un entorno para la gestión multilingüe de museos y demás patrimonio. Crear un servicio en la nube para conocer la opinión de las personas turistas (y mejorar los servicios). Elaborar un chatbot multilingüe sobre información turística del País Vasco. Análisis y puesta en marcha de servicios de traducción automática para guías de turismo. Desarrollar una app multilingüe orientada a la señalización. <p>TECNOLOGÍAS Y HERRAMIENTAS</p> <ul style="list-style-type: none"> Herramientas para la gestión de contenidos plurilingües Traducción automática Tecnologías de voz: síntesis y reconocimiento, interfaces de voz Sistemas conversacionales automáticos (chatbots) Procesar información sobre opiniones (twitter) Una api de integración en sitios web (idiomas, geolocalización, festivales, eventos, fiestas ... con ubicación en mapas) Resumen automático

4.2.7 Cultura, medios de comunicación, humanidades digitales, bibliotecas digitales ...

MOTIVACIÓN	OBJETIVOS	TAREAS, TECNOLOGÍAS Y HERRAMIENTAS
<p>En relación con la Cultura, en Europa está adquiriendo gran importancia el área de las Humanidades Digitales (o eHumanidades), así como la digitalización del patrimonio y el valor añadido al mismo. Gracias a la tecnología se está produciendo un cambio de paradigma en las Humanidades (datos).</p> <p>Con la digitalización cada vez hay más textos disponibles (actuales e históricos), y, además, la voz es cada vez más utilizada en el contexto digital. Aparte de los textos y registros del habla, hay otros elementos que también hay que tener en cuenta: música, versos, fotografías, audiovisuales...</p> <p>La radio y la televisión son también fuentes de información de vital importancia. Los medios de comunicación tienen un gran reto con relación al uso del euskera.</p> <p>Además, la forma de consumir está cambiando mucho con la digitalización, sobre todo entre las personas jóvenes.</p> <p>La digitalización viene acompañada de nuevas fuentes de información relacionadas con la Cultura: blogs, podcasts y redes sociales. Se están desarrollando herramientas para la detección de hábitos, opiniones, intereses, curiosidades, etc.</p>	<ul style="list-style-type: none"> - Impulsar la digitalización, corregida por OCR: algunas colecciones no están completas, otras no han pasado por OCR o no se han corregido los errores de dicho proceso OCR. - Facilitar la disponibilidad de metadatos basados en estándares y preparados para la interacción internacional. - Realización del procesado lingüístico básico: llevar a cabo los pasos básicos en los textos y ofrecer los resultados. Esto es fundamental para poder realizar búsquedas efectivas (lematizadas, en función de la época ...) - Añadir valor añadido: utilizar nuevos métodos de investigación, medir los riesgos y el impacto de la tecnología, aplicar herramientas lingüísticas avanzadas como el enriquecimiento con metadatos, así como gráficos y mapas, entre otros. <p>Los medios de comunicación (explotación avanzada de la hemeroteca/mediateca) y las redes sociales (aplicaciones relacionadas con la sociolingüística) son ámbitos interesantes de aplicación del procesamiento. Traducir las noticias y los contenidos de cultura en euskera al castellano a través de la traducción automática (los resultados están siendo muy buenos en la dirección eu → es) y así facilitar su difusión por el mundo.</p>	<p>[Archivos, bibliotecas, fondos patrimoniales, medios de comunicación]</p> <ol style="list-style-type: none"> a. En el ámbito de las bibliotecas digitales, ofrecer una cadena de procesamiento lingüístico de documentos y otros elementos para colaborar en la digitalización. b. Promover las Humanidades Digitales. c. Aportar valor añadido a los textos históricos a través de la digitalización: además del facsímil, realizar la transcripción y normalización del texto y enlazar sus elementos con Wikipedia o con otras obras (o traducciones) de la biblioteca. d. Desarrollar herramientas de valor añadido para los medios de comunicación. e. Análisis sociolingüístico de las redes sociales: el comportamiento lingüístico de las personas hablantes, la evolución de la lengua o el fenómeno de la alternancia de códigos. <p>TECNOLOGÍAS Y HERRAMIENTAS</p> <ul style="list-style-type: none"> • OCR (optical character recognition); herramientas para completar metadatos y realizar el procesado lingüístico básico • Datos y corpus, y su procesamiento multimodal automático • Agrupaciones y clasificaciones automáticas de documentos • Tecnología de voz: conversores de texto a voz y transcripción • Traductor automático • Creación e inserción de subtítulos • Enlace automático a noticias asociadas

4.2.8 Seguridad

MOTIVACIÓN	OBJETIVOS	TAREAS, TECNOLOGÍAS Y HERRAMIENTAS
<p>El II Plan de normalización del uso del euskera en la Ertzaintza 2018-2022:</p> <p><i>“Continuar con el incremento de la capacitación en euskera de las y los ertzainas; continuar con la promoción del uso del euskera con sus compañeros y compañeras de trabajo y con la ciudadanía; integrar la gestión del euskera en el sistema de gestión habitual de la Ertzaintza; y desarrollar estrategias eficaces de información y comunicación.”</i></p> <p>El III Plan de la Ertzaintza tiene vigencia hasta el año 2027 y se centra en cuatro ejes principales:</p> <ol style="list-style-type: none"> 1. Capacitación lingüística en euskera de las y los ertzainas. 2. Uso del euskera con la ciudadanía y con los y las compañeras de trabajo. 3. Gestión del euskera. 4. Información y comunicación en torno al plan de uso. <p>Los recursos y herramientas que ofrecen las tecnologías del lenguaje son imprescindibles para mejorar la competencia lingüística de los y las ertzainas, y sobre todo, son de gran ayuda para el uso del euskera en la comunicación escrita y oral (uso del corrector ortográfico, diccionarios especializados y traducción automática).</p>	<p>En las relaciones de la Ertzaintza con la ciudadanía se promoverá la garantía de los derechos lingüísticos y la designación del euskera como lengua habitual de trabajo en las unidades y ámbitos de trabajo prioritarios. Esto puede tener un gran impacto en el uso del euskera, tanto dentro de la Ertzaintza como en el conjunto de la sociedad.</p> <p>Con el fin de cumplir el objetivo principal del III Plan, se pueden aplicar técnicas de procesamiento de textos, tratamiento de voz y traducción automática. La puesta en marcha de este tipo de proyectos contribuiría a mejorar la eficacia. Con el fin de llegar a las diferentes fases del trabajo profesional, además del procesamiento de textos, es imprescindible el reconocimiento y procesamiento del habla.</p> <p>Los recursos y herramientas de ámbito general deberán adaptarse a este dominio concreto, teniendo en cuenta sus características y necesidades.</p>	<ol style="list-style-type: none"> a. Tratamiento de la documentación. b. Desarrollar una cadena abierta y multilingüe para el procesamiento de la documentación (pipeline). c. Crear herramientas de ayuda para la redacción de informes. d. Traducción automática de informes. e. Elaborar recursos terminológicos y léxicos. f. Digitalización de las relaciones con la ciudadanía. <p>TECNOLOGÍAS Y HERRAMIENTAS</p> <ul style="list-style-type: none"> • Proceso de anonimización estándar automático (para eliminar información sensible de los textos) • Extracción automática de la información y terminología de los informes • Procesamiento lingüístico automático: etiquetado ... • Creación y estandarización de plantillas tipo • Corrector ortográfico y gramatical • Diccionarios multilingües especializados • Traductor automático y postedición automática • Corpus paralelos (memorias de traducción) • Tecnologías de voz: conversión de voz a texto, indexación de transcripciones y búsquedas avanzadas, implementación de interfaces de voz

4.3 PLAN DE ACCIÓN DE LAS TECNOLOGÍAS DE LA LENGUA EN EUSKERA 2021-2024: PROYECTOS FARO

Los proyectos faro suelen definirse como proyectos emprendidos por las administraciones públicas que aplican Tecnologías de la Lengua en sectores estratégicos, muestran el potencial y beneficios de esas tecnologías, crean e impulsan la industria y generan recursos reutilizables en otros proyectos. Además, son una vía de aprendizaje para, en el futuro, poner en marcha proyectos en otros sectores.

Los proyectos faro buscan la sinergia con el resto de medidas de impulso a las tecnologías del lenguaje que ya están en marcha. Estas medidas son horizontales y transversales (infraestructuras lingüísticas generales, reutilización de la información del sector público de interés desde el punto de vista lingüístico, plataformas para el procesamiento del lenguaje natural y traducción automática de las administraciones públicas) y, en determinados sectores, se trata de actuaciones concretas que abarcan toda la cadena de valor y proporcionan productos y servicios terminados.

En las áreas definidas como estratégicas en este plan de acción, se han recogido los proyectos que ya están en marcha o los previstos para la legislatura 2021-2024 y que integran las Tecnologías de la Lengua.

Como se ha dicho anteriormente, este plan de acción es el punto de partida, ya que tendrá continuidad. Se realizará el seguimiento de los proyectos faro aquí recogidos con el fin de:

- Conocer la evolución de los proyectos,
- Conocer los resultados de los mismos,
- Analizar las posibilidades de reutilización de los recursos lingüísticos, modelos y herramientas creadas y utilizadas para el desarrollo de los proyectos,
- Integrar en el plan los nuevos proyectos que surjan a lo largo de la legislatura en esos ámbitos estratégicos,
- Extender el plan de acción y las tecnologías de la lengua a más ámbitos **estratégicos**, y
- Difundir las tecnologías de la lengua, los proyectos y las herramientas disponibles en la administración pública, en el ámbito educativo, en los medios de comunicación, en el mundo laboral y en el conjunto de la sociedad.

En algunos casos no se ha podido completar toda la información de la ficha por no estar todavía definida. Dicha información se completará y actualizará junto con el seguimiento de los proyectos.

Un apartado importante es el conocimiento del presupuesto, lo que facilitará la previsión del presupuesto necesario para toda la legislatura y, por tanto, la prospección de cara al futuro.

En total se han recogido en fichas 25 proyectos faro.





Proyecto nº 1

NOMBRE	Mejorar la atención a la ciudadanía a través de la Inteligencia Artificial
ÁMBITO DE APLICACIÓN	ADMINISTRACIÓN ELECTRÓNICA
DESCRIPCIÓN	<p>La idea principal del proyecto es desarrollar un robot de atención a la ciudadanía que ofrezca respuestas automáticas (ventanilla única). El núcleo debería ser alimentado con respuestas anteriores de agentes del servicio <i>Zuzenean</i> y el sistema debería ser capaz de dar una respuesta correcta entre las diferentes opciones a través del <i>Maching Learning</i>.</p> <p>El sistema sería bilingüe. Recogería la solicitud mediante texto y devolvería la respuesta mediante texto. En una capa superior la voz podría convertirse en texto; de esa manera, también se podría utilizar mediante el teléfono. El sistema sería útil no sólo para la ciudadanía, sino también para el personal del servicio <i>Zuzenean</i>.</p>
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: corpus escritos - Herramientas para el procesamiento del lenguaje: analizadores semánticos, analizadores del discurso - Tratamiento del habla: interfaces de voz, transcripción automática
PALABRAS CLAVE	Ventanilla única, servicio automático de atención a la ciudadanía
DEPARTAMENTO RESPONSABLE	GOBERNANZA PÚBLICA Y AUTOGOBIERNO Dirección de Atención Ciudadana y Servicios Digitales
ENTIDAD(ES) DESARROLLADORA(S)	Zuzenean, Área Web (Dirección de Atención a la Ciudadanía y Servicios Digitales), EJIE
AÑO(S) DE DESARROLLO	<input type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	sin determinar

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	3.2 INTELIGENCIA ARTIFICIAL 5.7 E-ADMINISTRACIÓN 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas



Proyecto nº 2

NOMBRE	Aplicación para la anonimización de datos personales
ÁMBITO DE APLICACIÓN	TODAS (ADMINISTRACIÓN ELECTRÓNICA, SANIDAD, JUSTICIA, EDUCACIÓN, SEGURIDAD ...)
DESCRIPCIÓN	El IVAP está trabajando en el desarrollo de una herramienta automática para la anonimización y sustitución de datos personales de las memorias de traducción. Esta herramienta está en fase de desarrollo y, una vez anonimizadas, todas las memorias de traducción se publicarán en el catálogo de Open Data para que cualquier persona pueda reutilizarlas. Esos datos permitirían a cualquier empresa o usuario externo desarrollar o mejorar traductores automáticos y fomentar el sector de las Tecnologías del Lenguaje.
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: corpus escritos - Herramientas para el procesamiento del lenguaje: analizadores morfosintácticos y semánticos - Traducción automática
PALABRAS CLAVE	Garantizar la protección de datos, reutilización de los recursos lingüísticos del sector público
DEPARTAMENTO RESPONSABLE	GOBERNANZA PÚBLICA Y AUTOGOBIERNO Dirección de Atención Ciudadana y Servicios Digitales, IVAP, EJE
ENTIDAD(ES) DESARROLLADORA(S)	EJIE, IVAP
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input type="checkbox"/> 2023 <input type="checkbox"/> 2024
PRESUPUESTO	75.000 € (15.000 + 60.000)

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	3.2 INTELIGENCIA ARTIFICIAL 5.5 SEGURIDAD 5.7 E-ADMINISTRACIÓN 5.8 EDUCACIÓN 5.11 E-SALUD 5.12 E-JUSTICIA 5.13 INDUSTRIAS DE LA LENGUA 5.14 TURISMO Y COMERCIO
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	Eje III: La Administración Pública como impulsora de la Industria del Lenguaje - Línea 2: Recursos lingüísticos de las Administraciones Públicas y política de reutilización de la Información del Sector Público



Proyecto nº 3

NOMBRE	Fomento de la Inteligencia Artificial
ÁMBITO DE APLICACIÓN	TODAS (ADMINISTRACIÓN ELECTRÓNICA, SANIDAD, JUSTICIA, EDUCACIÓN, SEGURIDAD ...)
DESCRIPCIÓN	<p>Hay mucho trabajo por realizar en el campo de la Inteligencia Artificial y hay que plantearlo de manera amplia. Es necesario impulsar el tejido empresarial y contar con nuevas ideas. Explorar las posibilidades existentes en los distintos departamentos del Gobierno Vasco y, con ello, que las empresas tecnológicas avancen, es decir, que la Administración y las empresas caminen juntas. Las líneas principales de trabajo del proyecto son:</p> <ul style="list-style-type: none"> - Subvenciones para el fomento de ideas aplicables en la Administración - Probar nuevas posibilidades técnicas (Question & Answering - <i>Conversational Artificial Intelligence</i>): pruebas de concepto <p>Por ello, será importante:</p> <ul style="list-style-type: none"> - Realizar vigilancia tecnológica - Observar el sector (mercado) (detección de nuevas necesidades)
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos - Herramientas para el procesamiento del lenguaje - Adquisición y reutilización de recursos y herramientas lingüísticas - Traducción automática - Tratamiento del habla: interfaces de voz
PALABRAS CLAVE	Transparencia, garantía de protección de datos, ventanilla única, procesamiento de documentos, recepción automática de actas, digitalización y consulta del archivo, inclusividad y universalidad, custodia tecnológica de la información
DEPARTAMENTO RESPONSABLE	GOBERNANZA PÚBLICA Y AUTOGOBIERNO Dirección de Tecnologías de la Información y la Comunicación (DTIC)
ENTIDAD(ES) DESARROLLADORA(S)	Convocatoria de ayudas Deusto Seidor, Ibermática, LKS, Teknei, Legítimo, Serikat Pruebas de tecnología (Q & A): <i>Hitz Zentroa</i> de la UPV
AÑO(S) DE DESARROLLO	<input type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input type="checkbox"/> 2023 <input type="checkbox"/> 2024
PRESUPUESTO	200.000 € - Convocatoria de ayudas 90 k+90 k - Pruebas de tecnología (Q & A): 20k

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	<p>3.2 INTELIGENCIA ARTIFICIAL 5.5 SEGURIDAD 5.7 E-ADMINISTRACIÓN 5.8 EDUCACIÓN 5.11 E-SALUD 5.12 E-JUSTICIA 5.13 INDUSTRIAS DE LA LENGUA 5.14 TURISMO Y COMERCIO</p>
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje II: Impulso de la Industria del Lenguaje</p> <ul style="list-style-type: none"> - Línea 1: Mejora de la visibilidad y la transferencia del sector - Línea II: Apoyo a la internacionalización y comercialización del sector <p>Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas.</p>



Proyecto nº 4

NOMBRE	Servicios generales transversales de Tecnologías de la Lengua
ÁMBITO DE APLICACIÓN	TODAS (ADMINISTRACIÓN ELECTRÓNICA, SANIDAD, JUSTICIA, EDUCACIÓN, SEGURIDAD ...)
DESCRIPCIÓN	<p>En 2019 se puso en marcha ITZULI, traductor neuronal y actualmente está arraigado en la sociedad. Además de la traducción, los siguientes pasos son la transcripción y la síntesis. Las líneas principales de trabajo del proyecto son:</p> <ul style="list-style-type: none"> - Adquisición de tecnología para su uso en toda la Comunidad y puesta en marcha de servicios de traducción, transcripción y síntesis. - Organizar la gestión técnica de los servicios en función tanto del uso de los departamentos como de la evolución de la actividad.
EJES	<ul style="list-style-type: none"> - Herramientas para el procesamiento del lenguaje: analizadores del discurso - Adquisición y reutilización de recursos y herramientas lingüísticas - Traducción automática - Tratamiento del habla: interfaces de voz, transcripción automática
PALABRAS CLAVE	Servicios transversales, procesamiento de documentos, inclusividad y universalidad, custodia tecnológica de la información, reutilización de los recursos lingüísticos del sector público
DEPARTAMENTO RESPONSABLE	GOBERNANZA PÚBLICA Y AUTOGOBIERNO Dirección de Tecnologías de la Información y la Comunicación (DTIC)
ENTIDAD(ES) DESARROLLADORA(S)	EJIE, VICOMTECH
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	<p>800.000 € (4 años)</p> <ul style="list-style-type: none"> - Tecnología 200k+200k+200k+200k <p>Nota: 100.000 €, ITZULI mejora y actualización de modelos del traductor neuronal (ver ficha 17)</p>

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	<p>3.2 INTELIGENCIA ARTIFICIAL 5.5 SEGURIDAD 5.7 E-ADMINISTRACIÓN 5.8 EDUCACIÓN 5.11 E-SALUD 5.12 E-JUSTICIA 5.13 INDUSTRIAS DE LA LENGUA 5.14 TURISMO Y COMERCIO</p>
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje I: Apoyo al desarrollo de infraestructuras lingüísticas</p> <p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje - Línea 1: Plataformas de aprendizaje del lenguaje natural y traducción automática en las administraciones públicas</p> <p>Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas.</p>



Proyecto nº 5

NOMBRE	INGURA: plataforma de aprendizaje integrada para la euskaldunización de personas adultas
ÁMBITO DE APLICACIÓN	EUSKALDUNIZACIÓN DE PERSONAS ADULTAS
DESCRIPCIÓN	Creación de una plataforma de aprendizaje integrada para la euskaldunización de personas adultas que permita alcanzar los seis niveles de competencia en euskera siguiendo las directrices del HEOC (<i>Helduen Euskalduntzearen Oinarritzko Curriculum</i>). Incluirá tanto cursos supervisados como no supervisados para su desarrollo por niveles de manera autónoma y comunitaria, herramientas de comunicación para el desarrollo de una enseñanza de aprendizaje participativa y compartida y ámbitos de gestión en función de los diferentes tipos de personas usuarias (alumnado, profesorado, administrador/a ...).
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de textos y materiales orales - Herramientas de procesamiento del lenguaje: analizadores, etiquetadores, herramientas de anonimización - Adquisición y reutilización de recursos y herramientas lingüísticas - Traducción automática - Tratamiento del habla: interfaces de voz, síntesis, transcripción automática
PALABRAS CLAVE	Materiales didácticos, formación del profesorado, entorno de aprendizaje del idioma, recursos para la creación de materiales, análisis automático de textos, herramientas de búsqueda en corpus, subtitulación de vídeos a través del reconocimiento de voz, chatbots
DEPARTAMENTO RESPONSABLE	CULTURA Y POLÍTICA LINGÜÍSTICA HABE, Viceconsejería de Política Lingüística
ENTIDAD(ES) DESARROLLADORA(S)	EJIE
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	2.000.000 € (2020-2026)

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	3.2 INTELIGENCIA ARTIFICIAL 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje I: Apoyo al desarrollo de infraestructuras lingüísticas</p> <p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje</p> <ul style="list-style-type: none"> - Línea 1: Plataformas de aprendizaje del 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. <p>Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas.</p>



Proyecto nº 6

NOMBRE	APP INGURA: aplicación para temarios de aprendizaje de euskera
ÁMBITO DE APLICACIÓN	EUSKALDUNIZACIÓN DE PERSONAS ADULTAS
DESCRIPCIÓN	Crear una aplicación para teléfonos móviles que ofrezca temarios para el aprendizaje de euskera, con el fin de que cualquier persona pueda acercarse al euskera y alcanzar la competencia comunicativa básica. El objetivo es la creación de una app similar a las existentes actualmente en el mercado, teniendo en cuenta las directrices del currículo básico de HABE (HEOC). La aplicación puede ser independiente o formar parte de la plataforma INGURA.
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de textos y materiales orales - Herramientas de procesamiento del lenguaje: analizadores, etiquetadores, herramientas de anonimización - Adquisición y reutilización de recursos y herramientas lingüísticas - Traducción automática - Tratamiento del habla: síntesis, transcripción automática, interfaces de voz
PALABRAS CLAVE	Materiales didácticos, formación del profesorado, entorno de aprendizaje del idioma, recursos para la creación de materiales, análisis automático de textos, herramientas de búsqueda en corpus, subtítulos de vídeos a través del reconocimiento de voz, chatbots
DEPARTAMENTO RESPONSABLE	CULTURA Y POLÍTICA LINGÜÍSTICA HABE, Viceconsejería de Política Lingüística
ENTIDAD(ES) DESARROLLADORA(S)	EJIE
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input type="checkbox"/> 2023 <input type="checkbox"/> 2024
PRESUPUESTO	Integrado en el proyecto de la plataforma INGURA.

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	3.2 INTELIGENCIA ARTIFICIAL 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje I: Apoyo al desarrollo de infraestructuras lingüísticas</p> <p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje</p> <ul style="list-style-type: none"> - Línea 1: Plataformas de aprendizaje del 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. <p>Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas.</p>



Proyecto nº 7

NOMBRE	Desarrollo de una herramienta de tratamiento de textos por niveles
ÁMBITO DE APLICACIÓN	EUSKALDUNIZACIÓN DE PERSONAS ADULTAS
DESCRIPCIÓN	Este proyecto está siendo desarrollado en colaboración entre HARE y el centro de investigación HITZ Taldea, en el marco del convenio de colaboración VPL y UPV de 15 de septiembre de 2020 (línea de trabajo de una beca de investigación adjudicada por HARE en 2020 y finalizada en noviembre de 2021). El objetivo del proyecto es desarrollar una herramienta para la clasificación y tratamiento automático de textos existentes en la plataforma Ikasbil. A partir de esos textos, se tratarán automáticamente, bien visualizando en la web determinadas características de complejidad por niveles (por ejemplo, el plugin VIEW desarrollado por la universidad de Tübingen), bien creando ejercicios en función de las necesidades del alumnado (por ejemplo, la herramienta FLAIR de la universidad de Tübingen).
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de textos y materiales orales; léxicos, terminologías, ontologías. - Herramientas para el procesamiento del lenguaje: analizadores, etiquetadores - Adquisición y reutilización de recursos y herramientas lingüísticas - Traducción automática - Tratamiento del habla: síntesis, transcripción automática, interfaces de voz
PALABRAS CLAVE	Programas de investigación, entorno de aprendizaje de la lengua, recursos para la creación de materiales, análisis automático de textos, herramientas de búsqueda en corpus, colaboración con organismos internacionales
DEPARTAMENTO RESPONSABLE	CULTURA Y POLÍTICA LINGÜÍSTICA Viceconsejería de Política Lingüística UPV Facultad de Educación de Bilbao
ENTIDAD(ES) DESARROLLADORA(S)	HARE, HITZ Taldea (2021)
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input type="checkbox"/> 2024
PRESUPUESTO	35.400 € (beca anual)

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	3.2 INTELIGENCIA ARTIFICIAL 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas.



Proyecto nº 8

NOMBRE	Elaboración del corpus de pruebas de verificación
ÁMBITO DE APLICACIÓN	EUSKALDUNIZACIÓN DE PERSONAS ADULTAS
DESCRIPCIÓN	Este proyecto está siendo desarrollado en colaboración entre HARE y el centro de investigación HITZ Taldea, en el marco del convenio de colaboración VPL y UPV de 15 de septiembre de 2020 (línea de trabajo de una beca de investigación adjudicada por HARE en 2020 y finalizada en noviembre de 2021). El objetivo del proyecto es completar un corpus con las pruebas del proceso de verificación de HARE. El corpus estará formado por dos secciones o bases de datos: la base de datos de pruebas de expresión escrita y la base de datos de pruebas de expresión oral. Se analizará cuál podría ser la estructura más adecuada para el corpus y se realizará una selección de los metadatos más relevantes. También se analizará cómo se pueden obtener las transcripciones de las pruebas.
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de textos y materiales orales - Herramientas para el procesamiento del lenguaje: analizadores, etiquetadores - Adquisición y reutilización de recursos y herramientas lingüísticas - Tratamiento del habla: transcripción automática, interfaces de voz
PALABRAS CLAVE	Programas de investigación, etiquetado de la colección de textos, análisis automático de textos
DEPARTAMENTO RESPONSABLE	CULTURA Y POLÍTICA LINGÜÍSTICA Viceconsejería de Política Lingüística UPV Facultad de Educación de Bilbao
ENTIDAD(ES) DESARROLLADORA(S)	HARE, HITZ Taldea (2021)
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input type="checkbox"/> 2024
PRESUPUESTO	35.400 € (mediante beca 2021, 2023), 70.000 € (2022)

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	3.2 INTELIGENCIA ARTIFICIAL 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas.



Proyecto nº 9

NOMBRE	HOBE (HABERen Ordenagailu Bidezko Egiaztagintza - Verificación por Ordenador de HABE)
ÁMBITO DE APLICACIÓN	EUSKALDUNIZACIÓN DE PERSONAS ADULTAS
DESCRIPCIÓN	El sistema HOBE es una herramienta informática que se está desarrollando para acreditar el nivel de competencia comunicativa en euskera. Por el momento, permite la realización de pruebas de destreza de exámenes en soporte digital, pero podría ser una buena base para la recopilación de textos (escritos y orales) para el desarrollo de las tecnologías del lenguaje, para el procesamiento de datos mediante la Inteligencia Artificial (análisis automático de la producción de pruebas, por ejemplo) o para la implementación de funcionalidades de comunicaciones digitales (para la realización de pruebas orales mediante videollamadas, por ejemplo).
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de textos y materiales orales; léxicos, terminologías, ontologías - Herramientas para el procesamiento del lenguaje: analizadores, etiquetadores, transcripción automática - Adquisición y reutilización de recursos y herramientas lingüísticas - Tratamiento del habla: reconocimiento y transcripción
PALABRAS CLAVE	Entorno de aprendizaje del idioma, reconocimiento de voz, subtitulación de vídeos, creación de entornos para procesar el euskera
DEPARTAMENTO RESPONSABLE	CULTURA Y POLÍTICA LINGÜÍSTICA Viceconsejería de Política Lingüística HABE
ENTIDAD(ES) DESARROLLADORA(S)	HABE, EJIE
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	30.000 € anuales (desarrollo del sistema)

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	3.2 INTELIGENCIA ARTIFICIAL 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje</p> <ul style="list-style-type: none"> - Línea 2: Recursos lingüísticos de las Administraciones Públicas y política de reutilización de la información del Sector Público. <p>Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas.</p>



Proyecto nº 10

NOMBRE	ADITU-Osakidetza
ÁMBITO DE APLICACIÓN	SANIDAD
DESCRIPCIÓN	Utilización de la herramienta ADITU de reconocimiento del habla por parte de los equipos de trabajo del Servicio Corporativo de Euskera de Osakidetza para la realización de transcripciones, subtítulos y traducciones automáticas. Las transcripciones, subtítulos y traducción automática se utilizarán principalmente para la creación de la versión en euskera de los vídeos producidos en castellano, pero también se prevé su utilización cuando se quiera utilizar el euskera en las reuniones y no se pueda disponer de servicio de interpretación.
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: corpus orales - Herramientas para el procesamiento del lenguaje - Traducción automática - Tratamiento del habla: interfaces de voz, transcripción automática
PALABRAS CLAVE	Servicios de salud digitales, tecnologías de voz, traducción automática
DEPARTAMENTO RESPONSABLE	OSAKIDETZA Subdirección de Desarrollo y Formación de Recursos Humanos Servicio de Euskera Corporativo
ENTIDAD(ES) DESARROLLADORA(S)	Fundación Elhuyar
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input type="checkbox"/> 2023 <input type="checkbox"/> 2024
PRESUPUESTO	4.840 € anuales

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	3.2 INTELIGENCIA ARTIFICIAL 5.11 E-SALUD Y BIENESTAR 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	Eje I: Apoyo al desarrollo de infraestructuras lingüísticas Eje III: La Administración Pública como impulsora de la Industria del Lenguaje



Proyecto nº 11

NOMBRE	IDITE-Osakidetza
ÁMBITO DE APLICACIÓN	SANIDAD
DESCRIPCIÓN	Se pretende personalizar el verificador profesional de léxico IDITE según las recomendaciones del libro de estilo de Osakidetza, inicialmente para el Servicio de Euskera Corporativo y, posteriormente, para su extensión al conjunto de la organización (dos fases). La tecnología que utiliza IDITE está desarrollada, pero hay que adaptarla a las recomendaciones del libro de estilo de Osakidetza. El objetivo de la utilización de la herramienta IDITE en Osakidetza es garantizar el uso correcto del léxico en euskera en los textos producidos o traducidos al euskera en la Dirección General de Osakidetza (1ª fase). Los correctores ortográficos en euskera no siempre están disponibles (por ejemplo para Wordfast PRO), además, puede ocurrir que un término sea ortográficamente correcto, pero que no esté utilizado de manera adecuada en un contexto determinado, y, en eso, IDITE puede ser de gran ayuda.
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: léxicos, terminologías, ontologías - Herramientas para el procesamiento del lenguaje: analizadores morfosintácticos y semánticos
PALABRAS CLAVE	Servicios de salud digitales, recursos terminológicos y léxicos
DEPARTAMENTO RESPONSABLE	OSAKIDETZA Dirección de Recursos Humanos – Servicio de Euskera Corporativo
ENTIDAD(ES) DESARROLLADORA(S)	UZEI, Osakidetza
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	7.260 € anual

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	5.11 E-SALUD Y BIENESTAR 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje - Línea 2: Recursos lingüísticos de las Administraciones Públicas y política de reutilización de la información del Sector Público.</p> <p>Eje IV: Proyectos faro de tecnologías del procesamiento del lenguaje natural de las Administraciones Públicas. Proyectos de aplicación de las tecnologías del lenguaje natural en servicios públicos estratégicos de gran impacto social</p>



Proyecto nº 12

NOMBRE	Itzulbide
ÁMBITO DE APLICACIÓN	SANIDAD
DESCRIPCIÓN	Osakidetza está desarrollando un traductor automático neuronal de textos clínicos que se encuentra en fase de validación. Posteriormente, se implantará en las estaciones clínicas de Osakidetza y se marcarán las líneas para seguir desarrollando la herramienta (corrector médico, terminología, abreviaturas en euskera ...)
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de materiales textuales y orales; terminologías, ontologías - Traducción automática
PALABRAS CLAVE	Procesamiento de la documentación clínica, herramienta de apoyo para la redacción de informes sanitarios, traducción automática de informes sanitarios, recursos terminológicos y léxicos
DEPARTAMENTO RESPONSABLE	OSAKIDETZA Dirección de Recursos Humanos – Servicio de Euskera Corporativo
EMPRESA(S) DESARROLLADORA(S)	HITZ Zentroa de la Facultad de Informática de la UPV
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	90.000 € anuales (previsto)

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	3.2 INTELIGENCIA ARTIFICIAL 5.11 E-SALUD Y BIENESTAR 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje I: Apoyo al desarrollo de infraestructuras lingüísticas</p> <p>Eje II: Impulso de la Industria del Lenguaje</p> <p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje</p> <p>Eje IV: Proyectos faro de tecnologías del procesamiento del lenguaje natural de las Administraciones Públicas. Proyectos de aplicación de las tecnologías del lenguaje natural en servicios públicos estratégicos de gran impacto social</p>



Proyecto nº 13

NOMBRE	Desarrollo e implantación del traductor neuronal judicial
ÁMBITO DE APLICACIÓN	JUSTICIA INNOVADORA
DESCRIPCIÓN	Se aprovechará la experiencia de traductor neuronal jurídico-administrativo y las memorias de traducción realizadas por el equipo de Normalización Lingüística en la Administración de Justicia para realizar el traductor neuronal judicial euskera-castellano-euskera. Esta herramienta se pondrá a disposición del personal de la Administración de Justicia, del personal auxiliar y de los operadores jurídicos externos. Se vinculará a las funcionalidades del nuevo gestor judicial electrónico y sus aplicaciones complementarias.
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de corpus textuales (memorias de traducción); léxicos, terminologías, ontologías - Instrumentos de procesamiento del lenguaje: analizadores morfosintácticos y semánticos, instrumentos de anonimización - Traducción automática
PALABRAS CLAVE	Justicia digital, traducción automática
DEPARTAMENTO RESPONSABLE	DEPARTAMENTO DE IGUALDAD, JUSTICIA Y POLÍTICAS SOCIALES Dirección de Justicia Digital e Infraestructuras
ENTIDAD(ES) DESARROLLADORA(S)	EJIE
AÑO(S) DE DESARROLLO	<input type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input type="checkbox"/> 2024
PRESUPUESTO	Por determinar

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	3.2 INTELIGENCIA ARTIFICIAL 5.12 E-JUSTICIA 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	Eje III: La Administración Pública como impulsora de la Industria del Lenguaje Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas



Proyecto nº 14

NOMBRE	Desarrollo de una herramienta de textualización de videoconferencias e implantación integrada en el gestor procesal electrónico (GPE)
ÁMBITO DE APLICACIÓN	JUSTICIA INNOVADORA
DESCRIPCIÓN	Realizar la textualización automática de la voz recogida en una videograbación (actuaciones judiciales orales y vistas judiciales grabadas) para realizar búsquedas que permitan acudir a momentos concretos de la videograbación, agilizando así la labor de los jueces en la elaboración de sentencias o la de los fiscales en la calificación. La textualización realizada deberá integrarse con el nuevo GPE de Avantius.
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de materiales orales - Herramientas para el procesamiento del lenguaje: transcripción automática; búsquedas automáticas - Tratamiento del habla: interfaces de voz, reconocimiento de voz y transcripción
PALABRAS CLAVE	Justicia digital, digitalización de vistas judiciales
DEPARTAMENTO RESPONSABLE	DEPARTAMENTO DE IGUALDAD, JUSTICIA Y POLÍTICAS SOCIALES Dirección de Justicia Digital e Infraestructuras
ENTIDAD(ES) DESARROLLADORA(S)	EJIE, Servicio de Desarrollo de la Dirección de Justicia Digital e Infraestructuras
AÑO(S) DE DESARROLLO	<input type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input type="checkbox"/> 2024
PRESUPUESTO	Por determinar

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	3.2 INTELIGENCIA ARTIFICIAL 5.12 E-JUSTICIA 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	Eje III: La Administración Pública como impulsora de la Industria del Lenguaje Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas



Proyecto nº 15

NOMBRE	Subtitulado y traducción automática en contenidos audiovisuales
ÁMBITO DE APLICACIÓN	CULTURA, MEDIOS DE COMUNICACIÓN, HUMANIDADES DIGITALES, BIBLIOTECAS DIGITALES
DESCRIPCIÓN	<p>El proyecto está en marcha desde 2019.</p> <p>Tiene tres objetivos principales:</p> <ol style="list-style-type: none"> 1.- Obtención de la subtitulación automática en contenidos audiovisuales (euskera y castellano). 2.- Una vez obtenida y mejorada la subtitulación, realizar traducciones automáticas del euskera al castellano y del castellano al euskera. 3.- Enriquecer la documentación de los contenidos y facilitar la búsqueda de contenidos audiovisuales digitales a través de cualquier buscador, aprovechando la subtitulación realizada a través del reconocimiento de voz.
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de textos y materiales orales - Herramientas para el procesamiento del lenguaje: transcripción automática - Tratamiento del habla: reconocimiento, transcripción - Traducción automática
PALABRAS CLAVE	Digitalización de documentos, reconocimiento de voz, subtitulación y traducción de audiovisuales, búsquedas automáticas, herramientas de valor añadido para los medios de comunicación
DEPARTAMENTO RESPONSABLE	<p>Dirección de Ingeniería y Explotación de EITB</p> <p>VICOMTECH</p> <p>Instituto de investigación aplicada</p>
ENTIDAD(ES) DESARROLLADORA(S)	EITB, Vicomtech, Instituto de Investigación Aplicada
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	400.000 € (2021-2024) (2020: 124.000 €)

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	<p>3.2 INTELIGENCIA ARTIFICIAL</p> <p>5.13 INDUSTRIAS DE LA LENGUA</p>
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje</p> <p>Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas</p>



Proyecto nº 16

NOMBRE	Creación de un corpus de voz para el proyecto Common Voice
ÁMBITO DE APLICACIÓN	TODAS (ADMINISTRACIÓN ELECTRÓNICA, SANIDAD, JUSTICIA, EDUCACIÓN, SEGURIDAD ...)
DESCRIPCIÓN	<p>Para esta iniciativa de la Fundación Mozilla se creará una serie de datos de grabaciones de audio de código abierto. Objetivo: que pueda ser reutilizado por cualquier empresa o entidad que desarrolle tecnologías de voz para crear productos concretos a partir de dicho corpus. Se realizará en tres fases:</p> <p>Fase 1: completar el corpus de texto de 900.000 frases (sin datos personales, en formato adecuado y en licencia libre) y subirlo a la plataforma Common Voice.</p> <p>Fase 2: las grabaciones de estas frases se realizarán mediante la participación de personas voluntarias. Se deberán utilizar todo tipo de voces.</p> <p>Fase 3: las grabaciones se validarán una a una.</p>
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de textos y materiales orales - Herramientas para el procesamiento del lenguaje: analizadores, etiquetadores, herramientas de anonimización - Tratamiento del habla: interfaces de voz
PALABRAS CLAVE	Procesamiento de documentos, garantía de protección de datos, inclusividad y universalidad, desarrollo de herramientas lingüísticas, tratamiento del habla
DEPARTAMENTO RESPONSABLE	<p>CULTURA Y POLÍTICA LINGÜÍSTICA Dirección de Promoción del Euskera</p> <p>Librezale.eus</p>
ENTIDAD(ES) DESARROLLADORA(S)	Vicomtech (1ª fase)
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input type="checkbox"/> 2024
PRESUPUESTO	18.016,90 € (2021: fase 1)

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	<p>3.2 INTELIGENCIA ARTIFICIAL</p> <p>5.5 SEGURIDAD</p> <p>5.7 E-ADMINISTRACIÓN</p> <p>5.8 EDUCACIÓN</p> <p>5.11 E-SALUD</p> <p>5.12 E-JUSTICIA</p> <p>5.13 INDUSTRIAS DE LA LENGUA</p> <p>5.14 TURISMO Y COMERCIO</p>
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje I: Apoyo al desarrollo de infraestructuras lingüísticas</p> <p>Eje II: Impulso de la Industria del Lenguaje</p> <ul style="list-style-type: none"> - Línea 1: Mejora de la visibilidad y la transferencia del sector - Línea II: Apoyo a la internacionalización y comercialización del sector <p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje</p>



Proyecto nº 17

NOMBRE	Continuar mejorando el traductor automático neuronal ITZULI (www.euskadi.eus/itzuli)
ÁMBITO DE APLICACIÓN	TODAS (ADMINISTRACIÓN ELECTRÓNICA, SANIDAD, JUSTICIA, EDUCACIÓN, SEGURIDAD ...)
DESCRIPCIÓN	<p>Se continuará con la mejora del traductor automático neuronal puesto a disposición de todas las personas usuarias en 2019, mediante el entrenamiento con nuevos corpus:</p> <ul style="list-style-type: none"> - francés↔euskera: con el corpus de <i>Euskararen Erakunde Publikoa</i> - demás pares lingüísticos: a medida que se vayan adquiriendo corpus. <p>Se analizará la integración de nuevas funcionalidades: a medida que se escribe un texto, mostrar su traducción ...</p>
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de textos y materiales orales - Herramientas para el procesamiento del lenguaje - Adquisición y reutilización de recursos y herramientas lingüísticas - Traducción automática
PALABRAS CLAVE	Herramientas lingüísticas, herramientas para alumnos-profesores, plataforma educativa, herramientas de aprendizaje, servicios para humanidades digitales, herramientas de valor añadido para los medios de comunicación, materiales didácticos, recursos para la creación de materiales, traducción automática
DEPARTAMENTO RESPONSABLE	<p>GOBERNANZA PÚBLICA Y AUTOGOBIERNO Dirección de Tecnologías de la Información y la Comunicación EJIE</p> <p>KULTURA ETA HIZKUNTZA POLITIKA - Dirección de Promoción del euskera</p>
ENTIDAD(ES) DESARROLLADORA(S)	EJIE
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	Ver ficha 4

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	<p>3.2 INTELIGENCIA ARTIFICIAL</p> <p>5.5 SEGURIDAD</p> <p>5.7 E-ADMINISTRACIÓN</p> <p>5.8 EDUCACIÓN</p> <p>5.11 E-SALUD</p> <p>5.12 E-JUSTICIA</p> <p>5.13 INDUSTRIAS DE LA LENGUA</p> <p>5.14 TURISMO Y COMERCIO</p>
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje I: Apoyo al desarrollo de infraestructuras lingüísticas</p> <p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje</p> <ul style="list-style-type: none"> - Línea 1: Plataformas de aprendizaje del lenguaje natural y traducción automática en las administraciones públicas <p>Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas.</p>



Proyecto nº 18

NOMBRE	Continuar con la promoción y difusión del traductor automático neuronal ITZULI (www.euskadi.eus/itzuli)
ÁMBITO DE APLICACIÓN	TODAS (ADMINISTRACIÓN ELECTRÓNICA, SANIDAD, JUSTICIA, EDUCACIÓN, SEGURIDAD ...)
DESCRIPCIÓN	<p>Es necesario seguir ampliando el uso del traductor para que llegue a un mayor número de personas usuarias:</p> <ul style="list-style-type: none"> - Web www.euskadi.eus/itzuli: traductor general y jurídico-administrativo - Plugin para Trados y Wordfast: se ha integrado en los servicios de traducción de las Administraciones Públicas (SOT del IVAP, las tres diputaciones forales, ayuntamientos de las tres capitales, Gobierno de Navarra). - Integración mediante API: <i>Euskararen Erakunde Publikoa</i> (Iparralde) - Apps para Android e iOS: se actualizan continuamente <p>En 2022 se han previsto nuevas integraciones: Ertzaintza, Parlamento de Navarra ...</p>
EJES	<ul style="list-style-type: none"> - Herramientas para el procesamiento del lenguaje - Adquisición y reutilización de recursos y herramientas lingüísticas - Traducción automática
PALABRAS CLAVE	Herramientas lingüísticas, herramientas de ayuda, traducción automática
DEPARTAMENTO RESPONSABLE	<p>GOBERNANZA PÚBLICA Y AUTOGOBIERNO Dirección de Tecnologías de la Información y la Comunicación EJIE</p> <p>KULTURA ETA HIZKUNTZA POLITIKA - Dirección de Promoción del euskera</p>
ENTIDAD(ES) DESARROLLADORA(S)	EJIE
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	<p>API 8.000 € (2021+2022)</p> <p>Apps 20.000 € (2022)</p>

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	<p>3.2 INTELIGENCIA ARTIFICIAL</p> <p>5.5 SEGURIDAD</p> <p>5.7 E-ADMINISTRACIÓN</p> <p>5.8 EDUCACIÓN</p> <p>5.11 E-SALUD</p> <p>5.12 E-JUSTICIA</p> <p>5.13 INDUSTRIAS DE LA LENGUA</p> <p>5.14 TURISMO Y COMERCIO</p>
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje I: Apoyo al desarrollo de infraestructuras lingüísticas</p> <p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje</p> <ul style="list-style-type: none"> - Línea 1: plataformas de aprendizaje del lenguaje natural y traducción automática en las administraciones públicas <p>Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas.</p>



Proyecto nº 19

NOMBRE	Añadir nuevas funcionalidades a la barra de herramientas ITZULI+
ÁMBITO DE APLICACIÓN	TODAS (ADMINISTRACIÓN ELECTRÓNICA, SANIDAD, JUSTICIA, EDUCACIÓN, SEGURIDAD ...)
DESCRIPCIÓN	<p>Herramienta desarrollada en 2020. La barra de herramientas ITZULI+ (www.euskadi.eus/itzulplus) se instala en Chrome o Firefox, a modo de extensión y, al navegar por una página web, ofrece las siguientes opciones: traducir un texto (6 pares lingüísticos de ITZULI), buscar palabras en 3 diccionarios (euskera↔castellano, inglés, alemán), o buscar en la Wikipedia en euskera.</p> <p>Pues bien, a la barra de herramientas se le añadirán nuevas funcionalidades: voz sintética desarrollada mediante Inteligencia Artificial, transcripción, lematización, más diccionarios y opciones para trabajar con textos.</p>
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos y herramientas - Traducción automática - Tratamiento del habla: interfaces de voz, síntesis, reconocimiento
PALABRAS CLAVE	Herramientas lingüísticas, herramientas para alumnado-profesorado, plataforma educativa, herramientas de aprendizaje, inclusividad y universalidad
DEPARTAMENTO RESPONSABLE	CULTURA Y POLÍTICA LINGÜÍSTICA Dirección de Promoción del Euskera
ENTIDAD(ES) DESARROLLADORA(S)	EJIE, Teknei
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input type="checkbox"/> 2023 <input type="checkbox"/> 2024
PRESUPUESTO	32.000 € (2021) + 72.600 € (lematizadores, diccionario de sinónimos) (2021) 25.000 € (2022)

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	<p>3.2 INTELIGENCIA ARTIFICIAL</p> <p>5.5 SEGURIDAD</p> <p>5.7 E-ADMINISTRACIÓN</p> <p>5.8 EDUCACIÓN</p> <p>5.11 E-SALUD</p> <p>5.12 E-JUSTICIA</p> <p>5.13 INDUSTRIAS DE LA LENGUA</p> <p>5.14 TURISMO Y COMERCIO</p>
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje</p> <ul style="list-style-type: none"> - Línea 1: plataformas de aprendizaje del 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



Proyecto nº 20

NOMBRE	Sobre la base de plugins de diccionarios, marcar palabras en los textos y traducirlas
ÁMBITO DE APLICACIÓN	SEGURIDAD
DESCRIPCIÓN	En las aplicaciones del paquete de Ofimática, en un texto, al colocar el cursor en una palabra, obtener la traducción en otro idioma, así como algún ejemplo de uso.
EJES	<ul style="list-style-type: none"> - Herramientas para el procesamiento del lenguaje - Recursos lingüísticos y herramientas
PALABRAS CLAVE	Recursos léxicos, herramientas de ayuda lingüística
DEPARTAMENTO RESPONSABLE	SEGURIDAD Dirección de Gestión de Telecomunicaciones y Sistemas Informáticos (DGTSI) – Sistemas de Ayuda al Desarrollo
ENTIDAD(ES) DESARROLLADORA(S)	Por determinar
AÑO(S) DE DESARROLLO	<input type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input type="checkbox"/> 2023 <input type="checkbox"/> 2024
PRESUPUESTO	Por determinar

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	5.5 SEGURIDAD 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	Eje III: La Administración Pública como impulsora de la Industria del Lenguaje - Línea 1: Plataformas de aprendizaje del lenguaje natural y traducción automática en las Administraciones Públicas



Proyecto nº 21

NOMBRE	Desarrollo de un sistema de transcripción y traducción simultánea de mensajes de voz
ÁMBITO DE APLICACIÓN	SEGURIDAD
DESCRIPCIÓN	Cuando se recibe un mensaje de voz en euskera (por ejemplo, en la recepción de las comisarías), transcribir dicho mensaje, convertirlo en texto y, a la vez, traducirlo al castellano. Sería muy conveniente que el sistema captara más idiomas.
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: corpus oral y escrito - Herramientas lingüísticas: transcripción automática - Traducción automática - Tratamiento del habla: interfaces de voz, síntesis, reconocimiento
PALABRAS CLAVE	Tecnologías de voz, transcripción automática, traducción automática
DEPARTAMENTO RESPONSABLE	SEGURIDAD Dirección de Gestión de Telecomunicaciones y Sistemas informáticos Sección Informática de la Ertzaintza (SIP)
ENTIDAD(ES) DESARROLLADORA(S)	Por determinar
AÑO(S) DE DESARROLLO	<input type="checkbox"/> 2021 <input type="checkbox"/> 2022 <input type="checkbox"/> 2023 <input type="checkbox"/> 2024 <input checked="" type="checkbox"/> por determinar
PRESUPUESTO	Por determinar

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	5.5 SEGURIDAD 5.13 INDUSTRIAS DE LA LENGUA
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	Eje IV: Proyectos faro de tecnologías del procesamiento del lenguaje natural de las Administraciones Públicas. Proyectos de aplicación de las tecnologías del lenguaje natural en servicios públicos estratégicos de gran impacto social



Proyecto nº 22

NOMBRE	Recopilación de los corpus lingüísticos creados por las Administraciones Públicas
ÁMBITO DE APLICACIÓN	TODAS (ADMINISTRACIÓN ELECTRÓNICA, SANIDAD, JUSTICIA, EDUCACIÓN, SEGURIDAD ...)
DESCRIPCIÓN	<p>Los datos masivos son imprescindibles para seguir integrando el euskera en las tecnologías que utilizan Inteligencia Artificial y, en consecuencia, para poner en cada vez más ámbitos, a disposición de las personas usuarias herramientas y aplicaciones. Esos datos se obtienen de corpus (tanto orales como escritos; monolingües, bilingües y plurilingües) y es importante gestionar su calidad.</p> <p>Las Administraciones Públicas crean diariamente ese tipo de corpus. Por ello, se pretende ir recopilándolos, para, en una siguiente fase, tratarlos y prepararlos para su uso en la Inteligencia Artificial y, así poder compartirlos.</p>
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de textos y materiales orales - Herramientas para el procesamiento del lenguaje: analizadores, etiquetadores, herramientas para la anonimización - Adquisición y reutilización de recursos lingüísticos y herramientas
PALABRAS CLAVE	Garantizar la protección de datos, la vigilancia tecnológica de la información, la reutilización de los recursos lingüísticos del sector público
DEPARTAMENTO RESPONSABLE	DEPARTAMENTO DE CULTURA Y POLÍTICA LINGÜÍSTICA Dirección de Promoción del Euskera
ENTIDAD(ES) DESARROLLADORA(S)	
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	En principio, sin gasto

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	<p>3.2 INTELIGENCIA ARTIFICIAL</p> <p>5.5 SEGURIDAD</p> <p>5.7 E-ADMINISTRACIÓN</p> <p>5.8 EDUCACIÓN</p> <p>5.11 E-SALUD</p> <p>5.12 E-JUSTICIA</p> <p>5.14 TURISMO Y COMERCIO</p>
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje III: la Administración Pública como impulsora de la Industria del Lenguaje</p> <ul style="list-style-type: none"> - Línea 2: Recursos lingüísticos de las Administraciones Públicas y política de reutilización de la Información del Sector Público



Proyecto nº 23

NOMBRE	Tratamiento y preparación de datos para su uso en Inteligencia Artificial + adquisición de infraestructuras
ÁMBITO DE APLICACIÓN	TODAS (ADMINISTRACIÓN ELECTRÓNICA, SANIDAD, JUSTICIA, EDUCACIÓN, SEGURIDAD ...)
DESCRIPCIÓN	<p>El desarrollo de herramientas basadas en la inteligencia artificial requiere datos lingüísticos que deben ser tratados y preparados. Por lo tanto, los datos recogidos de los corpus se tratarán y prepararán en función del desarrollo que se vaya a hacer. También se prevé la generación de datos ad hoc cuando sean necesarios.</p> <p>Las líneas principales de trabajo del proyecto son:</p> <ul style="list-style-type: none"> - tipificación de datos (memorias de traducción, ...) - preparación de datos (anonimización, etiquetas, transcripciones ...), y - adquisición de infraestructuras técnicas (tarjetas gráficas-GPU)
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de textos y materiales orales - Herramientas para el procesamiento del lenguaje: analizadores, etiquetadores, herramientas para la anonimización - Adquisición y reutilización de recursos lingüísticos y herramientas - Traducción automática - Tratamiento del habla
PALABRAS CLAVE	Garantizar la protección de datos, hacer vigilancia tecnológica de la información, reutilización de los recursos lingüísticos del sector público
DEPARTAMENTO RESPONSABLE	<p>GOBERNANZA PÚBLICA Y AUTOGOBIERNO Dirección de Tecnologías de la Información y la Comunicación (DTIC)</p> <p>DEPARTAMENTO DE CULTURA Y POLÍTICA LINGÜÍSTICA Dirección de Promoción del Euskera</p>
ENTIDAD(ES) DESARROLLADORA(S)	Dirección de Tecnologías de la Información y la Comunicación (DTIC) EJIE, IVAP
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	200.000 € anuales

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	<p>3.2 INTELIGENCIA ARTIFICIAL</p> <p>5.5 SEGURIDAD</p> <p>5.7 E-ADMINISTRACIÓN</p> <p>5.8 EDUCACIÓN</p> <p>5.11 E-SALUD</p> <p>5.12 E-JUSTICIA</p> <p>5.13 INDUSTRIAS DE LA LENGUA</p> <p>5.14 TURISMO Y COMERCIO</p>
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje I: Apoyo al desarrollo de infraestructuras lingüísticas</p> <p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje</p> <ul style="list-style-type: none"> - Línea 2: recursos lingüísticos de las Administraciones Públicas y política de reutilización de la Información del Sector Público <p>Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas.</p>



Proyecto nº 24

NOMBRE	Oficina del Dato
ÁMBITO DE APLICACIÓN	TODAS (ADMINISTRACIÓN ELECTRÓNICA, SANIDAD, JUSTICIA, EDUCACIÓN, SEGURIDAD ...)
DESCRIPCIÓN	<p>Se creará un espacio para albergar los datos que constituyen la base de la Inteligencia Artificial (en este caso, los de las Administraciones Públicas) y, así, poder ofrecerlos para su reutilización.</p> <p>La Oficina recogerá los datos (recursos lingüísticos, corpus), modelos y herramientas creados para llevar a cabo los proyectos de la Administración y elaborará protocolos de reutilización de los mismos. Los protocolos fijarán los criterios para reutilizar, difundir o compartir dichos datos.</p> <p>El proyecto pretende dar los primeros pasos en la creación del ecosistema <i>linked open data</i> de recursos en euskera. Para ello, tratará de aplicar los criterios de Linguistic Linked Open Data.</p> <p>Por tanto, una vez en marcha, se estudiará la incorporación de recursos de otras entidades.</p>
EJES	<ul style="list-style-type: none"> - Recursos lingüísticos: recopilación y etiquetado de textos y materiales orales - Herramientas para el procesamiento del lenguaje: analizadores, etiquetadores, herramientas para la anonimización - Adquisición y reutilización de recursos lingüísticos y herramientas
PALABRAS CLAVE	Garantizar la protección de datos, hacer vigilancia tecnológica de los recursos lingüísticos, reutilizar los recursos lingüísticos del sector público
DEPARTAMENTO RESPONSABLE	<p>GOBERNANZA PÚBLICA Y AUTOGOBIERNO Dirección de Tecnologías de la Información y la Comunicación (DTIC) Dirección de Atención Ciudadana y Servicios Digitales</p> <p>DEPARTAMENTO DE CULTURA Y POLÍTICA LINGÜÍSTICA Dirección de Promoción del Euskera</p>
ENTIDAD(ES) DESARROLLADORA(S)	Dirección de Tecnologías de la Información y la Comunicación (DTIC) EJIE
AÑO(S) DE DESARROLLO	<input type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	100.000 € (2022)

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	<p>3.2 INTELIGENCIA ARTIFICIAL</p> <p>5.5 SEGURIDAD</p> <p>5.7 E-ADMINISTRACIÓN</p> <p>5.8 EDUCACIÓN</p> <p>5.11 E-SALUD</p> <p>5.12 E-JUSTICIA</p> <p>5.13 INDUSTRIAS DE LA LENGUA</p> <p>5.14 TURISMO Y COMERCIO</p>
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje I: Apoyo al desarrollo de infraestructuras lingüísticas</p> <p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje</p> <p>- Línea 2: recursos lingüísticos de las Administraciones Públicas y política de reutilización de la Información del Sector Público</p> <p>Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas.</p>



Proyecto nº 25

NOMBRE	Herramienta para la medición del uso oral
ÁMBITO DE APLICACIÓN	TODAS (ADMINISTRACIÓN ELECTRÓNICA, SANIDAD, JUSTICIA, EDUCACIÓN, SEGURIDAD ...)
DESCRIPCIÓN	<p>La aplicación mide el uso oral, distinguiendo y clasificando las conversaciones en función de la lengua.</p> <p>La unidad que mide es la conversación y recoge la lengua (euskera, castellano o cualquier otra) en la que se ha realizado.</p> <p>Recoge mediante un micrófono y con una frecuencia previamente programada, las conversaciones que se están produciendo en el lugar determinado para el estudio, las transcribe y, sin guardarlas, las separa y cuantifica según el idioma.</p>
EJES	<ul style="list-style-type: none"> - Herramientas para el procesamiento del lenguaje - Adquisición y reutilización de recursos lingüísticos y herramientas - Tratamiento del habla: interfaces de voz
PALABRAS CLAVE	Medir el uso oral, conocer el lenguaje del habla, garantizar la protección de datos, vigilancia tecnológica de la información
DEPARTAMENTO RESPONSABLE	<p>GOBERNANZA PÚBLICA Y AUTOGOBIERNO Dirección de Tecnologías de la Información y la Comunicación (DTIC)/EJIE</p> <p>DEPARTAMENTO DE CULTURA Y POLÍTICA LINGÜÍSTICA Dirección de Promoción del Euskera</p>
ENTIDAD(ES) DESARROLLADORA(S)	EJIE
AÑO(S) DE DESARROLLO	<input checked="" type="checkbox"/> 2021 <input checked="" type="checkbox"/> 2022 <input checked="" type="checkbox"/> 2023 <input checked="" type="checkbox"/> 2024
PRESUPUESTO	50.000 (2022 y 2023)

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	<p>3.2 INTELIGENCIA ARTIFICIAL</p> <p>5.5 SEGURIDAD</p> <p>5.7 E-ADMINISTRACIÓN</p> <p>5.8 EDUCACIÓN</p> <p>5.11 E-SALUD</p> <p>5.12 E-JUSTICIA</p> <p>5.13 INDUSTRIAS DE LA LENGUA</p> <p>5.14 TURISMO Y COMERCIO</p>
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<p>Eje I: Apoyo al desarrollo de infraestructuras lingüísticas</p> <p>Eje III: La Administración Pública como impulsora de la Industria del Lenguaje</p> <p>- Línea 2: Recursos lingüísticos de las Administraciones Públicas y política de reutilización de la Información del Sector Público</p> <p>Eje IV: Proyectos faro de tecnologías de procesamiento del lenguaje natural de las Administraciones Públicas.</p>

5. PREVISIÓN PRESUPUESTARIA INICIAL DEL PLAN DE ACCIÓN

Este presupuesto no es definitivo, ya que en el caso de algunos proyectos, todavía no se ha precisado, mientras que en otros casos, aunque aparezca una cantidad, será necesario concretar la cantidad exacta ejecutada cada año. Por lo tanto, se actualizará anualmente y además se prevé la integración de nuevos proyectos. El presupuesto inicial previsto es el siguiente:

Nº	NOMBRE DEL PROYECTO	2021	2022	2023	2024	TOTAL
1	Mejorar la atención a la ciudadanía a través de la Inteligencia Artificial					
2	Aplicación para la anonimización de datos personales	15.000	60.000			75.000
3	Fomento de la Inteligencia Artificial		200.000			200.000
4	Servicios generales transversales de Tecnologías de la Lengua	200.000	200.000	200.000	200.000	800.000
5	INGURA: plataforma de aprendizaje integrada para la euskaldunización de personas adultas	333.000	333.000	333.000	333.000	1.332.000
6	APP INGURA: aplicación para temarios de aprendizaje de euskera					
7	Desarrollo de una herramienta de tratamiento de textos por niveles	35.400	35.400	35.400		106.200
8	Elaboración del corpus de pruebas de verificación	35.400	70.000	35.400		140.800
9	HOBE (HABEren Ordenagailu Bidezko Egiazatzegintza - Verificación por Ordenador de HABE)	30.000	30.000	30.000	30.000	120.000
10	ADITU-Osakidetza	4.840	4.840			9.680
11	IDITE-Osakidetza	7.260	7.260	7.260	7.260	29.040
12	Itzulbide	90.000	90.000	90.000	90.000	360.000
13	Desarrollo e implantación del traductor neuronal judicial					
14	Desarrollo de una herramienta de textualización de videoconferencias e implantación integrada en el gestor procesal electrónico (GPE)					
15	Subtitulado y traducción automática en contenidos audiovisuales	100.000	100.000	100.000	100.000	400.000
16	Creación de un corpus de voz para el proyecto Common Voice	18.016,90				18.017
17	Continuar mejorando el traductor automático neuronal ITZULI (www.euskadi.eus/itzuli)					
18	Continuar con la promoción y difusión del traductor automático neuronal ITZULI (www.euskadi.eus/itzuli)	4.000	24.000			28.000
19	Añadir nuevas funcionalidades a la barra de herramientas ITZULI+	104.600	25.000			129.600
20	Sobre la base de plugins de diccionarios, marcar palabras en los textos y traducirlas					
21	Desarrollo de un sistema de transcripción y traducción simultánea de mensajes de voz					
22	Recopilación de los corpus lingüísticos creados por las Administraciones Públicas					
23	Tratamiento y preparación de datos para su uso en Inteligencia Artificial + adquisición de infraestructuras	200.000	200.000	200.000	200.000	800.000
24	Oficina del Dato	100.000				100.000
25	Herramienta para la medición del uso oral		25.000	25.000		50.000
	TOTAL	1.277.517	1.404.500	1.056.060	960.260	4.698.337

6. SEGUIMIENTO DEL PLAN DE ACCIÓN DE LAS TECNOLOGÍAS DE LA LENGUA

Se realizará un seguimiento de los proyectos faro que constituyen el punto de partida de este plan de acción, de forma que se conozca el resultado final de cada proyecto y las posibilidades claras de reutilización de los recursos lingüísticos y de los datos que se han tenido que tratar para llevarlo a cabo, así como de las tecnologías y productos desarrollados.

Además, será importante conocer realmente los proyectos que se están desarrollando y se van a desarrollar en los diferentes departamentos del Gobierno Vasco, incluidos los organismos autónomos, para aprovechar sinergias y aunar esfuerzos. Para ello, también será un dato esclarecedor recoger cada año las inversiones que se van a realizar en los diferentes proyectos y el presupuesto utilizado.

Además del seguimiento de los proyectos faro recogidos en el Plan de Acción, a lo largo de la legislatura 2021-2024 se crearán más proyectos faro, por lo que se tratará de incorporarlos al Plan de Acción. De hecho, este plan de acción no es cerrado, sino abierto, ya que se va a ir completando y actualizando constantemente. En este sentido, se mantendrá un contacto permanente con todos los Departamentos y Organismos Autónomos del Gobierno Vasco.

La información de los proyectos se actualizará una vez al año, en el mes de noviembre. No obstante, está prevista la celebración de dos reuniones anuales con las dos comisiones de coordinación constituidas para la elaboración del Plan de Acción. Las comisiones de coordinación son:

1ª Comisión de Coordinación:

- DEPARTAMENTO DE CULTURA Y POLÍTICA LINGÜÍSTICA: Viceconsejería de Cultura; Viceconsejería de Política Lingüística; HABE; EITB
- DEPARTAMENTO DE GOBERNANZA PÚBLICA Y AUTOGOBIERNO: Dirección de Tecnologías de la Información y la Comunicación-EJIE; Dirección de Atención a la Ciudadanía y Servicios Digitales; Dirección de Gobierno Abierto; IVAP
- DEPARTAMENTO DE DESARROLLO ECONÓMICO, SOSTENIBILIDAD Y MEDIO AMBIENTE: Dirección de Transformación Digital y Emprendimiento
- DEPARTAMENTO DE EDUCACIÓN: Dirección de Aprendizaje e Innovación Educativa; Beritzegune Nagusia

2ª Comisión de Coordinación:

- OSAKIDETZA: Dirección de Recursos Humanos - Servicio de Euskera Corporativo
- DEPARTAMENTO DE IGUALDAD, JUSTICIA Y POLÍTICAS SOCIALES: Dirección de Justicia Digital e Infraestructuras
- DEPARTAMENTO DE TURISMO, COMERCIO Y CONSUMO: Viceconsejería de Turismo y Comercio
- DEPARTAMENTO DE SEGURIDAD: Dirección de Recursos Humanos

Si se incorporaran al Plan de Acción nuevos proyectos de otros Departamentos u Organismos Autónomos, se integrarían en dichas comisiones.

Además, se mantendrá una reunión anual especial entre las personas gestoras del Plan de Acción y la Dirección de Transformación Digital y Emprendimiento del Departamento de Desarrollo Económico, Sostenibilidad y Medio Ambiente con el fin de compaginar y coordinar con la Estrategia para la Transformación Digital de Euskadi 2025.

Asimismo, con carácter anual, se informará a la Comisión Especial para Fomentar el Euskera en el Entorno Digital del Consejo Asesor del Euskera, sobre el seguimiento del Plan de Acción, para que ésta a su vez lo integre en el seguimiento de la iniciativa “Recomendaciones 2021-2024”.

Tal y como se preveía, el Plan de Acción de Tecnologías de la Lengua se ha planteado a largo plazo, por lo que un seguimiento exhaustivo de todos los proyectos y la incorporación de nuevos contribuirán a la integración del euskera en las Tecnologías de la Lengua y, en general, en la Inteligencia Artificial. Eso facilitará conocer las necesidades y avanzar en esos ámbitos. Además, sabemos que el desarrollo de esas tecnologías va a ser rápido.

Así, se añaden dos modelos de fichas:

- 6.1 2021-2024 Plan de Acción de las Tecnologías de la Lengua: seguimiento de los proyectos faro.
- 6.2 2021-2024 Plan de Acción de las Tecnologías de la Lengua: nuevos proyectos.

Para hacer el seguimiento de todos los proyectos y tener acceso a la información relacionada con ellos, la Viceconsejería de Política Lingüística creará en 2022 una página web. Además, los resultados de este Plan de Acción del periodo 2021-2024 se recogerán en un informe técnico, a modo de evaluación y balance, que permitirá abordar la siguiente fase en la próxima legislatura. Dicho informe técnico será publicado y comunicado.

6.1 2021-2024 PLAN DE ACCIÓN DE LA TECNOLOGÍAS DE LA LENGUA: MODELO DE FICHA DE SEGUIMIENTO DE PROYECTOS FARO

Nº Proyecto	<i>(el de este Plan de Acción)</i>
Fecha seguimiento	
NOMBRE	<i>(el de este Plan de Acción)</i>
ÁMBITO DE APLICACIÓN	<i>(la de este Plan de Acción)</i>
DEPARTAMENTO RESPONSABLE	<i>(el de la ficha correspondiente en este Plan de Acción)</i>
ENTIDAD(ES) DESARROLLADORA(S)	<i>(actualizar si es necesario)</i>
AÑO(S) DE DESARROLLO	<input type="checkbox"/> 2021 <input type="checkbox"/> 2022 <input type="checkbox"/> 2023 <input type="checkbox"/> 2024 <i>(actualizar si es necesario)</i>
PRESUPUESTO	Previsto: € <i>(actualizar si es necesario)</i> Real: €
Estado del proyecto	<input type="checkbox"/> fase inicial <input type="checkbox"/> fase intermedia <input type="checkbox"/> fase final <input type="checkbox"/> FINALIZADO
Si está finalizado, resultado final del proyecto	
Dónde y cómo está disponible	
Datos lingüísticos (recursos) y herramientas desarrolladas para la ejecución del proyecto	<input type="checkbox"/> Corpus textuales (unilingües, bilingües, multilingües) <input type="checkbox"/> Corpus orales (bases de datos acústicas, grabaciones, ...) <input type="checkbox"/> Léxicos, terminologías ... <input type="checkbox"/> Analizadores <input type="checkbox"/> Herramientas del habla (voz): etiquetadores para la transcripción <input type="checkbox"/> Modelos lingüísticos (traducción automática, tecnologías de la voz) <input type="checkbox"/> Herramientas de anonimización <input type="checkbox"/> Otros (especificar)
¿Los datos generados para el proyecto son reutilizables? En caso afirmativo, las personas gestoras del Plan de Acción gestionarán las licencias con quien corresponda.	<input type="checkbox"/> SÍ <input type="checkbox"/> NO
Añadir datos de uso del producto	
Añadir información complementaria	

6.2 2021-2024 PLAN DE ACCIÓN DE LAS TECNOLOGÍAS DE LA LENGUA: MODELO DE FICHA PARA NUEVOS PROYECTOS

NOMBRE	
ÁMBITO DE APLICACIÓN	<input type="checkbox"/> Administración electrónica <input type="checkbox"/> Educación <input type="checkbox"/> Euskaldunización de personas adultas <input type="checkbox"/> Sanidad <input type="checkbox"/> Justicia <input type="checkbox"/> Turismo <input type="checkbox"/> Cultura, medios de comunicación, humanidades digitales, bibliotecas digitales... <input type="checkbox"/> Seguridad <input type="checkbox"/> Otros (<i>especificar</i>)
DESCRIPCIÓN	
EJES	<i>(Ver apartados 4.1.2, 4.1.3, 4.1.4, 4.1.5 y 4.1.6 del Plan de Acción)</i>
PALABRAS CLAVE	
DEPARTAMENTO RESPONSABLE	
ENTIDAD(ES) DESARROLLADORA(S)	
AÑO(S) DE DESARROLLO	<input type="checkbox"/> 2021 <input type="checkbox"/> 2022 <input type="checkbox"/> 2023 <input type="checkbox"/> 2024
PRESUPUESTO	€

ESTRATEGIA PARA LA TRANSFORMACIÓN DIGITAL DE EUSKADI 2025	<i>(A cumplimentar por las personas gestoras del Plan de Acción)</i>
PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE	<i>(A cumplimentar por las personas gestorAs del Plan de Acción)</i>

7. REFERENCIAS

GOBIERNO VASCO Departamento de Desarrollo Económico, Sostenibilidad y Medio Ambiente (2021). *Estrategia para la Transformación Digital de Euskadi 2025*.

https://bideoak2.euskadi.eus/2021/03/30/news_67948/EEDE2025_Estrategia_EUS.pdf

https://bideoak2.euskadi.eus/2021/03/30/news_67948/ETDE2025_Estrategia_ES.pdf

GOBIERNO VASCO Departamento de Gobernanza Pública y Autogobierno

(2021). *ARDATZ Plan Estratégico de Gobernanza, Innovación Pública y Gobierno Digital 2030*.

https://www.euskadi.eus/contenidos/informacion/ardatz_2030/eu_def/adjuntos/ardatz-2030-eu.pdf

https://www.euskadi.eus/contenidos/informacion/ardatz_2030/es_def/adjuntos/ardatz-2030-es.pdf

GOBIERNO VASCO Presidencia

(2021). *Plan de Ciencia, Tecnología e Innovación Euskadi 2030*.

<https://www.ikerbasque.net/sites/default/files/files/ZTBP%20EUSKADI%202030%20.pdf>

https://www.innobasque.eus/uploads/attachment_files/pcti_euskadi_2030-60acacb14bf62.pdf

CONSEJO ASESOR DEL EUSKERA. VICECONSEJERÍA DE POLÍTICA LINGÜÍSTICA

(2021). *El entorno digital del euskera. Recomendaciones 2021-2024*". Vitoria-Gasteiz:

Servicio Central de Publicaciones del Gobierno Vasco. ISBN978-84-457-3592-3

<https://www.euskadi.eus/eusko-jauriaritza/hizkuntza-politika/> (Argitalpenak)

<https://www.euskadi.eus/gobierno-vasco/politica-linguistica/> (Publicaciones)

GOBIERNO DE ESPAÑA Ministerio de Asuntos Económicos y Transformación Digital

(2015). *Plan de Impulso de las Tecnologías del Lenguaje*.

<https://plantl.mineco.gob.es/Paginas/index.aspx>

INFORME INTERNO

UPV-Centro HITZ. *Hizkuntza-teknologiak sustatzeko planaren txosten teknikoa* (2020)

(Informe técnico del Plan de Impulso a las tecnologías de la Lengua)



Garena izaten jarraitzeko
hizkuntza-teknologiak