

POTENCIAL DE GALICIA NO SECTOR DAS
TECNOLOXÍAS LINGÜÍSTICAS

DOCUMENTO 40 / 2024

COORDINADORES:
SENÉN BARRO AMENEIRO (USC)
LUIS OTERO GONZÁLEZ (USC)

POTENCIAL DE GALICIA NO SECTOR DAS TECNOLOXÍAS LINGÜÍSTICAS

COORDINADORES :

SENÉN BARRO AMENEIRO (USC)

LUIS OTERO GONZÁLEZ (USC)

RESTO DE AUTORES:

CARMEN MAGARIÑOS (INVESTIGADORA DO PROXECTO NÓS)

DIEGO VÁZQUEZ (IMAXIN SOFTWARE)

LUIS IGNACIO RODRÍGUEZ GIL (USC)

MARÍA BAQUEIRO (XESTORA DO PROXECTO NÓS)

MARÍA BASTIDA (USC)

MIGUEL VÁZQUEZ TAÍN (USC)

PABLO GAMALLO (INVESTIGADOR DO CITIUS)

ÍNDICE

- 1 . A INDUSTRIA DAS TECNOLOXÍAS DA LINGUAXE: UNHA REVOLUCIÓN SILENCIOSA.
2. O IMPACTO ECONÓMICO AO REDOR DAS TECNOLOXÍAS LINGÜÍSTICAS
3. O IMPACTO ECONÓMICO DO GALEGO E A SÚA POTENCIALIDADE NO MARCO DAS TECNOLOXÍAS LINGÜÍSTICAS
4. AS TECNOLOXÍAS LINGÜÍSTICAS E O GALEGO: O PROXECTO NÓS
5. EPÍLOGO

RESUMO EXECUTIVO

As tecnoloxías lingüísticas (TL) combinan intelixencia artificial e linguaxe. Máis concretamente abarcan un amplo campo que inclúe ferramentas, modelos, algoritmos e sistemas que facilitan a comprensión, xeración e manipulación da linguaxe humana por parte de máquinas. Isto explica que o sector das TL sexa multidisciplinar, e nos equipos de I+D é frecuente a participación de lingüistas, informáticos, especialistas en intelixencia artificial (IA) e en psicoloxía cognitiva, entre outros. En relación co anterior, é un sector difícil de delimitar pois podemos atopar formando parte do mesmo entidades moi diversas, tanto de carácter público como privado, e de sectores económicos moi distintos como empresas tecnolóxicas, de servizos lingüísticos e de investigación.

O sector das TL está a experimentar un gran crecemento. Entre os factores que, sen dúbida, contribúen a iso están, por unha parte, os avances tecnolóxicos producidos no campo da aprendizaxe automática coa adopción na actualidade de enfoques baseados en redes neuronais, en particular a aprendizaxe profunda e algúns novos modelos, como os “transformers”. Por outra, a dispoñibilidade de datos pois a consultora IDC prevé que no 2025 crearanse uns 175ZB de datos, dos cales o 79% estarán en formato texto. Por último, os múltiples usos e utilidades, tanto a nivel xeral coma empresarial, que teñen os produtos e servizos xerados grazas as TL.

Entre a ampla gama de solucións e casos de uso da aplicación das TL

destacan, entre outros, a tradución automática (TA) e tradución asistida por ordenador (TAO), sistemas de diálogo ou asistentes conversacionais que permiten a interacción entre humanos e máquinas a través da linguaxe natural posible grazas o procesamento da linguaxe natural (PLN), corrección e avaliación lingüística automática, xeración de textos, extracción de información e a análise de sentimentos a partir da linguaxe.

En xeral tódolos axentes sociais teñen nas TL unha aliada que lles facilitan distintas actividades e tarefas, pero sen dúbida, os sectores máis beneficiados son aqueles requiren dunha interacción con outros usuarios e nos que hai unha gran cantidade de datos non estruturados que interesa tratar. Entre estes sectores destacan:

- Educación e a formación, permitindo ó alumnado acceder a contidos en diferentes idiomas, recibir comentarios personalizados e proporcionando ferramentas ó profesorado como modelos xenerativos de linguaxe e de simplificación de texto;
- Saúde e medicina, permitindo a análise e interpretación de datos médicos e informes clínicos, mellorar a precisión dos diagnósticos que permita unha atención médica personalizada e eficiente;
- Financeiro e de seguros, xa que permite a análise de riscos, servizos de atención ao cliente e detección de fraudes.
- Consultaría, publicidade e promoción de vendas, mediante a análise de sentimentos para axudar ás empresas a

expandirse en mercados internacionais e mellorar a experiencia do cliente.

- Administración pública e servizos gobernamentais, pois axudan a automatizar procesos burocráticos, ofrecer asistencia en liña, responder preguntas frecuentes e ofrecer información en distintos idiomas.

Como xa se se indicou o sector das TL está a experimentar un gran crecemento e así o indican distintas fontes a nivel internacional. Se ben a nivel económico poderíase considerar este sector como estratéxico, existe pouca información sobre o mesmo tanto a nivel español, como internacional. Consultoras como Mordor Intelligence (2020) estiman que o mercado mundial crecerá dende os 10.720 millóns USD en 2020 ata os 48.460 millóns USD para 2026, MarketsandMarkets (2022) sinala que dos 15.700 millóns de dólares en 2022, alcanzaranse os 49.400 millóns en 2027 e pola súa parte, Fortune Business Insights (2022) estima que dos 20.900 millóns de dólares en 2021 se acadarán os 160.000 millóns en 2029. Tamén, a gran maioría das propias empresas que están adoptando as TL manifestan que teñen intención de incrementar o seu orzamento no futuro inmediato. O mercado das TL está moi concentrado, así as compañías Amazon, Google, Hewlett Packard Enterprise, IBM Corporation e Microsoft teñen entre o 55% e o 60% da cota de mercado mundial, estando o resto en mans de outros moitos operadores rexionais e locais.

Os mercados cun maior volume de operacións son América do Norte cun 37.5% do total das operacións e Europa

cun 26.2%. De acordo coas previsións existentes a rexión que máis medrará nos vindeiros anos será a de Pacífico. Esta rexión chegará a ter unha cota en 2029 do 43.4% do mercado mundial, e América do Norte e Europa perderán algo de peso ata acadar o 22.1% e o 19.7% respectivamente.

En Europa, nestes momentos, os mercados máis importantes atópanse no Reino Unido (22%), Alemaña (20%), Francia (16%), Italia (11%) e España (9%). Prevese que nos vindeiros anos a distribución do mercado sexa moi semellante. O sector das TL en Europa e España está composto principalmente por pequenas e medianas empresas (pemes).

Non é doado identificar as organizacións que se dedican as TL, polo que neste estudo se utilizaron distintas fontes, en concreto, o “Estudio de caracterización del sector de tecnologías del lenguaje en España” (2018); os datos do Rexistro de Empresas e Centros de Investigación que figura no catálogo de tecnoloxías lingüísticas do Ministerio de Asuntos Económicos e Transformación Dixital; o Mapa de capacidades de tecnoloxías de IA; e o Mapa de capacidades dixitais de Galicia. Despois de depurar os datos o número de axentes atopados foi un total de 214 en España, dos cales, 115 son empresas, 11 son centros de investigación independentes, 47 son centros de investigación universitarios, 17 son institucións privadas sen fins de lucro (IPSFL) e 24 son administracións públicas. A distribución destes axentes non é uniforme no territorio nacional, en concreto a comunidade galega está re-

lativamente ben posicionada no quinto posto, o 7% destes axentes están na nosa comunidade. Nas primeiras posicións atópanse a Comunidade de Madrid (39%), Cataluña (12%), o País Vasco (8%) e Comunidade Valenciana (8%).

En 2021 a facturación total do sector das TL empresas representou o 1,10% do PIB, xerando un alto valor engadido o cal representa un 67% respecto os ingresos totais, sendo a media dos sectores do 25%, e o gasto medio anual por traballador situouse en 84.000€ fronte os 35.000€ de media.

Un dos riscos que se percibe é que as TL discriminen as linguas minoritarias e minorizadas como podería ser o galego. Se iso chegase a ocorrer afectaría, entre outros aspectos, a nosa economía, a nosa cultura e incluso a propia supervivencia do galego como lingua. Entón vemos moi necesario que neste campo sigan a participar as universidades, centros de investigación, administracións públicas e as empresas. A situación do galego no contexto dixital non é equiparable á doutras linguas como o inglés ou o castelán, polo que precisa dun impulso que a sitúe como lingua de uso convencional nos recursos tecnolóxicos.

Cada lingua é un portal cara á tradición cultural, unha entrada á historia e unha ferramenta para a identidade individual e colectiva. Ademais o idioma ten valor económico porque permite proxectar industrias que utilizan a lingua como insumo básico no proceso de xeración de valor, creando un tecido produtivo que doutro xeito non existiría, ou tería unha dimensión menor. No documento "O valor económico do galego" faise

unha cuantificación do valor económico e o impacto da lingua galega. O valor total da produción dos sectores vinculados ao galego en 2019 foi de 4.118 millóns de euros. Este valor componse dun 44,7% de impacto directo (1.841,7 millóns de euros), un 9,7% de impacto indirecto (397,5 millóns de euros), e un 45,6% de efecto inducido (1.878,6 millóns de euros).

Pensamos que está xustificado apostar pola inclusión do galego no mundo tecnolóxico, para deste xeito, non só preservar a diversidade lingüística e cultural, senón para desenvolver capacidades tecnolóxicas que permitan dispoñer das competencias e a cualificación necesarias para impulsar o sector das tecnoloxías lingüísticas. De feito estase a facer xa que Galicia conta co proxecto NOS o cal pretende potenciar a presenza dixital do galego, ó igual que ocorre co catalán (AINA), co euskera (GAITU) ou valenciano (VIVES) que forman parte do PERTE Nova Economía da Lingua. Desta forma, póñense a disposición das empresas recursos que, doutro xeito, serían dificilmente asumibles por empresas de pequeno e mediano tamaño.

A finalidade que persegue o proxecto Nós é acadar unha ampla variedade de recursos lingüísticos en galego de alta calidade e que serán publicados en formato dixital, en xeral baixo licenzas libres, e a través de ferramentas con códigos abertos para ser empregadas en calquera tipo de soporte dixital polas diferentes institucións públicas e, mesmo, polos axentes que conforman o tecido económico empresarial e, asemade, pola sociedade en xeral. O proxecto

contempla oito liñas de traballo, unhas están centradas na voz (lingua falada) e outras no texto (lingua escrita) como son síntese de voz, recoñecemento da fala, sistemas de diálogo, corrección e avaliación lingüística automática, tradución automática, xeración de textos, extracción de información e verificación de información e minaría de opinións.

Este proxecto xa está a obter resultados, e ano 2022 obtivéronse seis corpus, tres demostradores e un prototipo que están dispoñibles libremente e accesibles a través da páxina web do proxecto e o repositorio de GitHub. A elaboración de estes corpus e de gran importancia xa que dispor dunha gran cantidade e variedade de recursos lingüísticos (escritos e orais) de alta calidade é fundamental para o desenvolvemento das tecnoloxías lingüísticas para o galego. En concreto, obtivéronse corpus ASR, TTS paralelos e conversacionais. En canto os demostradores, desenvóléronse un sistema de conversión texto-voz (TTS) de alta

calidade e naturalidade, un demostrador de recoñecemento automático da fala, e un tradutor neuronal español-galego e inglés galego, nas dúas direccións, e no referido ó prototipo conseguiuuse un xerador automático de descrições textuais de datos en tempo real, aplicado ao ámbito específico da predición meteorolóxica.

É requisito indispensable para o proxecto Nós dispor dunha gran cantidade e variedade de recursos lingüísticos de alta calidade para o cal está a realizar unha estreita cooperación coas entidades que dispoñen destes recursos. Entre estas entidades atópanse Xunta de Galicia, Parlamento de Galicia, Consello da Cultura Galega, Real Academia Galega, Centro Ramón Piñeiro, CRTVG, Deputación da Coruña, Deputación de Lugo, Editorial Galaxia, Editorial Hugin e Munin, Editorial Laiovento, Praza Pública, Sermos Galiza, Agrocomunicación, Pasa na costa, Atlántica de información, IGA-DI e Asociación Irimia.

PRÓLOGO

Las Tecnologías Lingüísticas (TL) abarcan un amplio campo que incluye herramientas, modelos, algoritmos y sistemas que facilitan la comprensión, generación y manipulación del lenguaje humano por parte de máquinas.

Estas tecnologías se basan en la idea del Procesamiento del Lenguaje Natural (PLN), que permite a los ordenadores analizar, interpretar y generar lenguaje humano con sentido y utilidad práctica. Las TL abarcan una gran variedad de subcampos, como el reconocimiento y síntesis de voz, la traducción automática o el análisis de sentimientos. El ámbito de las TL es multidisciplinar, y en los equipos de I+D es frecuente la participación de lingüistas, informáticos, especialistas en inteligencia artificial (IA) y en psicología cognitiva, entre otros.

En los últimos años, las LT se han convertido en parte integrante de nuestra vida cotidiana y nuestros entornos profesionales debido a los enormes progresos alcanzados en el desarrollo y aplicación de estas tecnologías. Solo a modo de ejemplo, recogemos algunos de los más presentes en nuestras vidas:

1. Asistentes personales: Dispositivos como Siri, Alexa y Google Assistant, entienden y responden a comandos vocales, ayudando en diversas tareas como establecer alarmas, buscar en la web y gestionar citas.
2. Traducción automática: Herramientas como Google Translate proporcionan traducción en tiempo real entre un gran número de idiomas, eliminando o al menos reduciendo las barreras lingüísticas, y permitiendo una comunicación fluida entre personas que de otro modo no podrían comunicarse.

3. Análisis de sentimientos en las empresas: Permite analizar los comentarios de los clientes, calibrar el sentimiento del público y personalizar las estrategias de marketing.

4. Herramientas de accesibilidad: Los servicios de texto a voz y de voz a texto permiten una mayor accesibilidad a las personas con dificultades o limitaciones para la lectura y/o escritura.

Seguro que estos ejemplos le resultarán familiares, pero no cabe duda de que hoy el paradigma de las TL es ChatGPT. Al margen de la valoración que cada usuario pueda hacer de esta herramienta en función de su experiencia personal, hay que reconocerle a este chatbot el haber universalizado la idea de que la IA ya está plenamente implantada en nuestras vidas.

Llegar aquí ha supuesto, en todo caso, un largo y tortuoso camino. Los primeros sistemas de procesamiento de lenguaje natural se basaban en reglas gramaticales cuidadosamente elaboradas. Una labor extraordinariamente compleja y que dio resultados más bien modestos. A finales del siglo XX, los métodos estadísticos aportaron mejoras significativas. Aquí la disponibilidad de datos, en particular en forma de texto, comenzó a ser explotada. Sin embargo, el verdadero salto se produjo con la adopción de enfoques basados en redes neuronales, en particular el aprendizaje profundo y, tras él, algunos nuevos modelos, como los “transformes”, en los que se basan GPT-3 y GPT-4, por ejemplo. Sin duda, estas tecnologías han supuesto una revolución en el campo, aportando resultados que nos han sorprendido a todos, incluso a quienes investigamos en el campo. Pero se sigue avanzando, y aceleradamente, por lo que

cabe esperar, incluso en el corto y medio plazo, muchas más proezas de las TL. Estaremos muy atentos.

Aunque en sus orígenes, y por décadas, el liderazgo en la investigación en el ámbito de las TL estuvo sobre todo centrado en las universidades y algunos centros de investigación, hace más de una década que dicho liderazgo está claramente asentado, y cada vez más, en las empresas. Es sobre todo una cuestión de inversión, que responde a unas expectativas aún mayores de negocio. Gigantes tecnológicos como Google, Microsoft e IBM, han dedicado y dedican cantidades muy importantes de recursos a la investigación en TL. De entrada, esto es positivo, obviamente, ya que este liderazgo ha impulsado la innovación y la democratización de estas tecnologías. Pero el precio a pagar es grande, en todo caso. Los intereses de las compañías no son los del bien común. Por eso es tan importante que los gobiernos no se desentiendan de la inversión en la inteligencia artificial en general, y en particular en las tecnologías lingüísticas. También es vital invertir en investigación y recursos humanos en el campo de las TL para mantener el progreso y evitar la dependencia tecnológica de otros países. Esto implica cultivar el talento, fomentar la colaboración entre el mundo académico y la industria y promover el desarrollo de una industria tecnológica propia.

Todo lo dicho toma un valor muy especial cuando hablamos de las TL en el contexto de una lengua minoritaria y minorizada. En ese caso, el situar dicha lengua en la sociedad y la economía digitales no se ya solo una oportunidad estratégica sino una necesidad para la supervivencia de la lengua. Sin duda este es el caso del gallego. Precisamente, el proyecto Nós, al que dedica-

mos el primer capítulo de este documento, busca dotar al gallego de recursos relacionados con las TL, sacándolo del vagón de cola de las lenguas europeas. Según datos del informe ELE (European Language Equality Project), el gallego estaba en 2022 a la cola del conjunto de lenguas con un soporte tecnológico fragmentado, solo por delante de aquellas lenguas que carecen de soporte o en las que este es débil (gaélico, islandés, serbio o maltés, entre las ellas).

Todo lo dicho nos llevó a dedicar una sesión del Foro Tecnológico del Foro Económico de Galicia a las tecnologías lingüísticas. Vimos la necesidad y la oportunidad de hacerlo, y para hacerlo bien hemos contado con expertos de alto nivel, tanto de la academia, como de la industria y de la administración pública. Además de una jornada dedicada a tratar esta temática desde puntos de vista multidisciplinares y complementarios, hemos querido también elaborar este importante documento a disposición pública. En él hemos abordado, como ya se ha dicho, el proyecto Nós, pero también el papel y el perfil de la industria de las TL y el impacto económico de este sector, que es ya muy importante y de valor creciente, tanto a nivel autonómico, como nacional e internacional. Además, y de modo muy especial, nos interesó conocer el valor de las TL en torno a nuestra lengua propia, el gallego.

Solo hablando desde el rigor, el conocimiento de causa, la cuantificación de lo cuantificable y el valor educativo, cultural, social y sentimental de lo que no lo es, seremos capaces entre todos de situar al gallego en la sociedad y la economía digital y, con ello, no solo ponerlo en valor económico sino garantizar su supervivencia y la de una parte importante de lo que somos.

LIMIAR

As Tecnoloxías Lingüísticas (TL) abarcan un amplo campo que inclúe ferramentas, modelos, algoritmos e sistemas que facilitan a comprensión, xeración e manipulación da linguaxe humana por parte de máquinas.

Estas tecnoloxías baséanse na idea do Procesamento da Linguaxe Natural (PLN), que permite aos computadores analizar, interpretar e xerar linguaxe humana con sentido e utilidade práctica. As TL abarcan unha gran variedade de subcampos, como o recoñecemento e síntese de voz, a tradución automática ou a análise de sentimentos. O ámbito das TL é multidisciplinar, e nos equipos de I+D é frecuente a participación de lingüistas, informáticos, especialistas en intelixencia artificial (IA) e en psicoloxía cognitiva, entre outros.

Nos últimos anos, as LT convertéronse en parte integrante da nosa vida cotiá e as nosas contornas profesionais debido aos enormes progresos alcanzados no desenvolvemento e aplicación destas tecnoloxías. Só a modo de exemplo, recolleamos algúns dos máis presentes nas nosas vidas:

1. Asistentes persoais: Dispositivos como Siri, Alexa e Google Assistant, entenden e responden a comandos vocais, axudando en diversas tarefas como establecer alarmas, buscar na web e xestionar citas.

2. Tradución automática: Ferramentas como Google Translate proporcionan tradución en tempo real entre un gran número de idiomas, eliminando, ou polo menos reducindo, as barreiras lingüísticas, e permitindo unha comunicación fluída entre persoas que doutro xeito non poderían co-

municarse.

3. Análise de sentimentos nas empresas: Permite analizar os comentarios dos clientes, calibrar o sentimento do público e personalizar as estratexias de mercadotecnia.

4. Ferramentas de accesibilidade: Os servizos de texto a voz e de voz a texto permiten unha maior accesibilidade ás persoas con dificultades ou limitacións para a lectura e/ou escritura.

Seguro que estes exemplos resultaranlle familiares, pero non cabe dúbida de que hoxe o paradigma das TL é ChatGPT. Á marxe da valoración que cada usuario poida facer desta ferramenta en función da súa experiencia persoal, hai que recoñecerlle a este chatbot o ter universalizado a idea de que a IA xa está plenamente implantada nas nosas vidas.

Chegar aquí supuxo, en todo caso, un longo e tortuoso camiño. Os primeiros sistemas de procesamento de linguaxe natural baseábanse en regras gramaticais coidadosamente elaboradas. Un labor extraordinariamente complexo e que deu resultados máis ben modestos. A finais do século XX, os métodos estatísticos achegaron mellores significativas. Aquí a dispoñibilidade de datos, en particular en forma de texto, comezou a ser explotada. Con todo, o verdadeiro salto produciuse coa adopción de enfoques baseados en redes neuronais, en particular a aprendizaxe profunda e, tras el, algúns novos modelos, como os “transformers”, nos que se basean GPT-3 e GPT-4, por exemplo. Sen dúbida, estas tecnoloxías supuxeron unha revolución no campo, achegando resultados que nos sorprende-

ron a todos, mesmo a quen investigamos no campo. Pero séguese avanzando, e aceleradamente, polo que cabe esperar, mesmo no curto e medio prazo, moitas máis proezas das TL. Estaremos moi atentos.

Aínda que nas súas orixes, e por décadas, o liderado na investigación no ámbito das TL estivo sobre todo centrado nas universidades e algúns centros de investigación, fai máis dunha década que dito liderado está claramente asentado, e cada vez máis, nas empresas. É sobre todo unha cuestión de investimento, que responde a unhas expectativas aínda maiores de negocio. Xigantes tecnolóxicos como Google, Microsoft e IBM, dedicaron e dedican cantidades moi importantes de recursos á investigación en TL. De entrada, isto é positivo, obviamente, xa que este liderado impulsou a innovación e a democratización destas tecnoloxías. Pero o prezo para pagar é grande, en todo caso. Os intereses das compañías non son os do ben común. Por iso é tan importante que os gobernos non se desentendan do investimento na intelixencia artificial en xeral, e en particular nas tecnoloxías lingüísticas. Tamén é vital investir en investigación e recursos humanos no campo das TL para manter o progreso e evitar a dependencia tecnolóxica doutros países. Isto implica cultivar o talento, fomentar a colaboración entre o mundo académico e a industria e promover o desenvolvemento dunha industria tecnolóxica propia.

Todo o dito toma un valor moi especial cando falamos das TL no contexto dunha lingua minoritaria e minorizada. Nese caso, situar dita lingua na sociedade e a economía dixitais non é só unha oportunidade estratéxica senón tamén unha necesidade para a supervivencia da lingua. Sen dúbida este é o caso do galego. Precisamente, o

proxecto Nós, ao que dedicamos o primeiro capítulo deste documento, busca dotar ao galego de recursos relacionados coas TL, sacándoo do vagón de cola das linguas europeas. Segundo datos do informe ELE (European Language Equality Project), o galego estaba en 2022 á cola do conxunto de linguas cun soporte tecnolóxico fragmentado, só por diante daquelas linguas que carecen de soporte ou nas que este é débil (gaélico, islandés, serbio ou maltés, entre elas).

Todo o dito levounos a dedicar unha sesión do Foro Tecnolóxico do Foro Económico de Galicia ás tecnoloxías lingüísticas. Vimos a necesidade e a oportunidade de facelo, e para facelo ben contamos con expertos de alto nivel, tanto da academia, como da industria e da administración pública. Ademais dunha xornada dedicada a tratar esta temática desde puntos de vista multidisciplinares e complementarios, quixemos tamén elaborar este importante documento que poñemos a disposición pública. Nel abordamos, como xa se dixo, o proxecto Nós, pero tamén o papel e o perfil da industria das TL e o impacto económico deste sector, que é xa moi importante e de valor crecente, tanto a nivel autonómico, como nacional e internacional. Ademais, e de modo moi especial, interesounos coñecer o valor das TL ao redor da nosa lingua propia, o galego.

Só falando desde o rigor, o coñecemento de causa, a cuantificación do cuantificable e o valor educativo, cultural, social e sentimental do que non o é, seremos capaces entre todos de situar ao galego na sociedade e a economía dixital e, con iso, non só poñelo en valor económico senón garantir a súa supervivencia e a dunha parte importante do que somos.

1. A INDUSTRIA DAS TECNOLOXÍAS DA LINGUAXE: UNHA REVOLUCIÓN SILENCIOSA.

DIEGO VÁZQUEZ (IMAXIN SOFTWARE)

1.1. INTRODUCCIÓN

As tecnoloxías da linguaxe (TL) xogan un papel crítico nunha sociedade interconectada e multicultural, permitindo ás computadoras comprender, interpretar e xerar linguaxe humana de maneira moi eficiente. Desde a tradución automática ata distintas áreas do procesamento da linguaxe natural (PLN), as TL conforman un campo interdisciplinario da intelixencia artificial e a lingüística. Engloban diversos campos de coñecemento, como o recoñecemento e síntese de voz, clasificación e análise de texto e a corrección gramatical, entre outras.

O sector das TL tamén está a experimentar unha revolución grazas ao desenvolvemento de novos enfoques de aprendizaxe automática que están a permitir que as computadoras aprendan a realizar tarefas que antes eran imposibles, como conducir automóviles, xogar xogos e diagnosticar enfermidades. Estes desenvolvementos están a abrir unha variedade de novas posibilidades para aplicacións no sector das tecnoloxías lingüísticas. Por exemplo, os novos modelos neuronais estanse a usar para mellorar a precisión dos sistemas de tradución automática, ou desenvolver chatbots cunha capacidade de interlocución en linguaxe natural abraiante.

As actividades económicas encontran nas TL unha gran aliada. De maneira especial, destacan nos sectores que requiren unha interacción co usuario final e no tratamento de datos non estruturados. Ade-

mais, os clientes se volven máis conscientes das vantaxes e beneficios que poden obter. Así en educación, os modelos de linguaxe xenerativos pódense usar para crear materiais de aprendizaxe personalizados para cada estudante. Tamén se poden usar para xerar ensaios, informes e outros tipos de traballos escritos. En entretemento, estes mesmos modelos de linguaxe poden empregarse para crear chatbots que poden manter conversacións cos usuarios e crear historias. En atención médica, os modelos de linguaxe xenerativos pódense usar para xerar resumos de artigos médicos, traducir idiomas e proporcionar apoio emocional aos pacientes. Estas son só algunhas das moitas aplicacións potenciais dos modelos de linguaxe xenerativos. A medida que estes modelos continúen desenvolvéndose, é probable que teñan un impacto aínda maior na industria e vida da xente.

1.2 BENEFICIOS E APLICACIÓNS DAS TL

As TL proporcionan beneficios significativos en diversos sectores da sociedade. A súa capacidade para interpretar e xerar linguaxe natural está a abrir novas oportunidades e a transformar a maneira de traballar, aprender, comunicarse e vivir. Entre os principais ámbitos nos que as TL están a ter un impacto significativo figuran as seguintes:

Educación e formación: As plataformas de aprendizaxe en liña permiten ao estudante acceder a contidos educativos en diferentes idiomas, recibir comentarios personalizados e facilitar a aprendizaxe de novas linguas. Ademais, proporcionan ao profesorado ferramentas para adaptar os materiais ao alumnado, como os modelos de simplificación de texto ou os modelos xenerativos de linguaxe.

Saúde e medicina: No ámbito da saúde, as TL están a desempeñar un papel importante na análise e interpretación de datos médicos, rexistros de pacientes e informes clínicos. A través da análise de grandes volumes de información, as TL axudan a mellorar a precisión dos diagnósticos, permitindo unha atención médica máis personalizada e eficiente.

Comunicación e atención ao cliente: Os chatbots e as interfaces de voz permiten unha interacción máis rápida e eficiente, fornecendo información sobre produtos e servizos, e ofrecendo soporte técnico. As tecnoloxías de análise de sentimentos permiten ás empresas monitorizar as redes sociais e obter información valiosa sobre opinións e necesidades de clientes.

Industria e comercio electrónico: A tradución automática facilita a internacionalización de empresas, permitindo a venda en liña en diferentes idiomas e a comunicación. Ademais, o PLN dálle ás empresas a oportunidade de obter información relevante dos datos non estruturados.

Administración pública e servizos gobernamentais: As TL axudan a automatizar procesos burocráticos, ofrecer asistencia

en liña e responder preguntas frecuentes. A tradución automática tamén é crucial nas comunicacións internacionais e nas relacións diplomáticas entre países.

Medios de comunicación e redes sociais: A análise de sentimentos e a clasificación de textos son clave para monitorizar a opinión pública e obter información sobre as tendencias e preferencias dos usuarios. As TL tamén están a ser utilizadas para a creación de sistemas de xeración de contidos, que poden axudar na redacción de noticias, artigos e contidos personalizados para grupos de usuarios.

Por todo o mencionado, as oportunidades que as TL ofrecen son inmensas en todas as actividades económicas e administrativas. No mundo empresarial actual, as TL desempeñan un papel cada vez máis relevante, xa que se utilizan para optimizar a comunicación e a interacción entre os clientes e as empresas a través dos produtos dixitais. Por exemplo, un dos sistemas máis notables son os asistentes conversacionais, que son capaces de comprender e procesar a linguaxe natural, permitindo unha comunicación máis fluída e efectiva. Isto implica un cambio significativo na forma en que se interactúa cos dispositivos e coas empresas, xa que agora pódense realizar accións cotiás como facer preguntas, obter información e incluso comprar falando directamente co dispositivo.

Por outro lado, a xa mencionada tradución automática facilita a comunicación entre empresas de diferentes países ou culturas, axudando a superar así as barreiras lingüísticas. Hoxe, as empresas poden expandirse a novos mercados internacionais sen as limitacións lingüísticas

tradicionais. Isto fomenta a colaboración a nivel global e abre novas oportunidades de negocio.

Ademais, a clasificación de texto desempeña un papel crucial na análise e minería de datos. Coa crecente cantidade de información dispoñible en liña e nas redes sociais, é esencial ser capaz de extraer información relevante. Grazas a estas técnicas pódense analizar grandes volumes de datos non estruturados e identificar patróns, tendencias e opinións. Por exemplo, mediante a tarefa de análise do sentimento, as empresas poden coñecer, de maneira estatística, as opinións positivas ou negativas dos seus clientes, identificar áreas de mellora nos produtos e servizos, e tomar decisións de negocio baseadas nestes resultados.

1.3 A INDUSTRIA DAS TECNOLOXÍAS DA LINGUAXE

As empresas e organizacións que lideran este cambio conforman un amplo universo multidisciplinar e pódense encontrar formando parte de múltiples sectores. No desenvolvemento das ferramentas e aplicación vinculadas as TL atopamos:

Empresas de software: Estas compañías crean software e ferramentas de tradución automática, correctores gramaticais e ortográficos, sistemas de diálogo, análise e clasificación de texto ou asistentes conversacionais.

Empresas de servizos lingüísticos: Ofrecen servizos de tradución, interpretación e localización en diferentes idiomas, utilizando tecnoloxías como a tradución automática ou tradución asistida para fa-

cilitar a comunicación entre persoas de distintas linguas e culturas.

Institucións académicas e de investigación: Traballan na busca de innovación neste campo e na formación de novos expertos que impulsen o progreso desta rama da intelixencia artificial.

Organizacións gobernamentais e sen fins de lucro: Promoven e financian proxectos de investigación e desenvolvemento de tecnoloxías da lingua. Traballan para garantir o acceso a estas tecnoloxías a comunidades de linguas minoritarias e/ou minorizadas, garantindo a inclusión e igualdade de oportunidades.

1.4 ACTIVIDADES DAS EMPRESAS DO SECTOR DAS TECNOLOXÍAS DA LINGUA

Os tipos de empresas e organizacións anteriormente citadas ofrecen unha ampla gama de solucións e casos de uso da aplicación das TL, e comparten un conxunto de técnicas, solucións e ferramentas que articulan as súas actividades ou servizos, entre as que destacan:

Tradución automática (TA) e tradución asistida por ordenador (TAO): A TA e a TAO son actividades esenciais nas TL. As empresas especializadas nestas áreas desenvolven software e ferramentas para a tradución automática de textos entre diferentes idiomas, así como para asistir aos tradutores humanos no seu traballo diario. Estas ferramentas permiten unha tradución máis rápida e precisa, axudando a superar as barreiras lingüísticas en diversos contextos.

Sistemas de diálogo ou asistentes conversacionais: Esta actividade está relacionada co desenvolvemento de sistemas que permiten a interacción conversacional entre humanos e máquinas. As empresas nesta área traballan na creación de asistentes virtuais avanzados capaces de entender e responder preguntas complexas, facer recomendacións personalizadas e realizar tarefas específicas a través da linguaxe natural.

Recoñecemento (ASR) e síntese de voz (TTS): As compañías que se dedican ao desenvolvemento de tecnoloxías de ASR e TTS traballan no recoñecemento da fala humana e na xeración de voz sintetizada. Estas tecnoloxías permiten a conversión de voz en texto e viceversa, o que é fundamental para aplicacións como a transcripción automática, a accesibilidade para persoas con discapacidade auditiva e a xeración de voz en aplicativos.

Tarefas do procesamento da linguaxe natural (PLN): As empresas que se dedican a esta actividade traballan no desenvolvemento de algoritmos e ferramentas que facilitan a análise de texto oral e escrito en diferentes ámbitos. Algúns exemplos son a análise morfolóxica, análise sintáctica, análise semántica e as tarefas pragmáticas e discursivas.

Creación de recursos para as tecnoloxías da linguaxe: Un elemento fundamental no desenvolvemento das TL son os recursos lingüísticos. Estes inclúen corpus de texto, léxicos, ontoloxías e outras fontes de información que permiten o estudo e a análise da linguaxe. As empresas deste campo traballan na creación, mantemento e enriquecemento destes recursos, que

son esenciais para o desenvolvemento de novas ferramentas e aplicacións.

A pesar da importancia e do potencial das TL, é necesario abordar os desafíos e as preocupacións asociadas á súa evolución. Algúns dos retos que a industria debe afrontar inclúen:

Precisión e fiabilidade: Aínda que as TL acadaron grandes avances, a súa precisión e fiabilidade deben mellorar. A comprensión e a interpretación da linguaxe humana son tarefas complexas, normalmente ben desenvolvidas en linguas e culturas dominantes, pero é necesario un desenvolvemento maior en linguas minorizadas, culturas pouco representadas ou, incluso, variantes dialectais dunha lingua. Polo tanto, resulta clave seguir investigando e desenvolvendo novos modelos e ferramentas para incrementar a precisión e reducir os erros nas aplicacións lingüísticas. É neste punto onde o traballo das pequenas empresas é crucial, xa que as grandes multinacionais non chegan a desenvolver este tipo de adaptacións.

Privacidade e seguridade: A protección da privacidade e a seguridade dos usuarios converteuse nunha preocupación importante. É esencial establecer medidas de protección de datos adecuadas e garantir a confidencialidade das comunicacións. As empresas e administracións deben adoptar políticas e prácticas que aseguren unha xestión responsable e ética dos datos lingüísticos. A este respecto, cobra especial importancia para Galicia a creación da Agencia Española de Supervisión de la Inteligencia Artificial (AESIA) na Coruña.

Inclusión e equidade: Para que as TL sexan verdadeiramente transformadoras, é necesario que sexan accesibles para todas as persoas e que non reforcen as desigualdades existentes para parte das comunidades minoritarias, persoas con discapacidade ou áreas rurais. A inclusión e a equidade deben ser consideracións centrais no desenvolvemento e na implementación das tecnoloxías da lingua.

Formación e capacitación: O desenvolvemento e a adopción das TL requiren profesionais capacitados e preparados, polo que hai que investir na formación de expertos en lingüística computacional e PLN. As institucións académicas e as empresas deben traballar en colaboración para ofrecer programas de formación e investigación que preparen aos futuros profesionais para os retos e as oportunidades deste campo en constante evolución. Estes retos son complexos e requiren un enfoque colaborativo e multidisciplinario para seren superados. É necesario fomentar a cooperación entre empresas, institucións académicas, organismos gobernamentais e a sociedade en xeral para avanzar na investigación, o desenvolvemento e a aplicación das tecnoloxías da lingua.

A través da colaboración entre os diversos axentes do ecosistema é posible avanzar cara a solución de desafíos e á creación de novas oportunidades de negocio. Este crecemento, impulsado polo uso da IA, presenta unha competencia intensa, con numerosas empresas que buscan destacar na creación de solucións lingüísticas innovadoras. Neste contexto, as multinacionais desempeñan un papel importante, xa que grazas á súa potencia comercial e capacidade para ofrecer solucións integrais, contribúen á divulgación e

adopción das TL en diferentes ámbitos e sectores.

Por un lado, xogan un papel positivo ao crear necesidades e concienciar ás empresas sobre as vantaxes das tecnoloxías da linguaxe. Pero, por outro lado, a súa posición dominante tamén pode supoñer un desafío para as empresas máis pequenas, que deben esforzarse por competir neste mercado. Mais, como se mencionou anteriormente, a capacidade de desenvolver tecnoloxías que sexan eficaces, precisas e adaptadas aos distintos idiomas e contextos culturais será un diferencial importante para o éxito das pemes no mercado.

1.5 CARACTERIZACIÓN DO SECTOR EN ESPAÑA

A falta de datos actualizados é un desafío ao analizar este sector, tanto a nivel español e europeo como global. Algúns dos escasos estudos ao respecto son o Estudio de caracterización del sector de tecnologías del lenguaje en España (ACAP e OESIA (2018) e o Final study report on CEF Automated Translation value proposition in the context of the European LT market/ecosystem (Vasiljevs, A., et al., 2019).

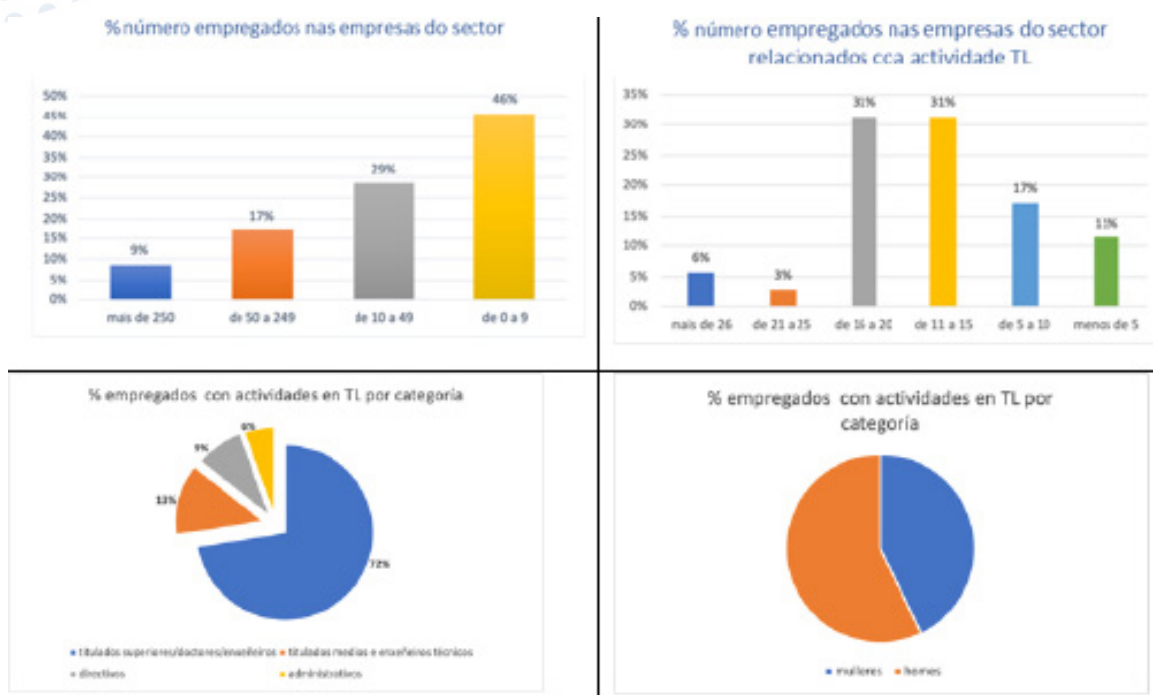
Estes estudos ofrecen algunhas pautas sobre o sector, pero a obtención de datos precisos é complexa debido a varios motivos. En primeiro lugar, non hai un código CNAE específico para este tipo de actividades. Ademais, a definición dos axentes considerados no sector pode ser variada, xa que moitas empresas desenvolven solucións de TL pero non son consideradas parte do sector. Por último, o sector está fragmentado e atomizado, o que dificulta a consolidación de datos uniformes e representativos

Porén, analizados os datos aportados polos informes anteriormente citados, pódense destacar varios aspectos relevantes:

Tamaño das empresas: O sector das TL en España está composto principalmente por pequenas e medianas empresas (pemes), aínda que tamén hai algunhas grandes compañías especializadas. A maioría das empresas ten menos de 49 empregados, o que reflicte a fragmentación e atomización do sector. As pemes

son áxiles e flexibles, o que lles permite adaptarse ás necesidades cambiantes do mercado e ofrecer solucións personalizadas aos seus clientes. Por outro lado, as grandes corporacións teñen recursos significativos e capacidade para desenvolver proxectos complexos e a gran escala. No referido ao tamaño das empresas do sector en España en comparación ás empresas homólogas en Europa, podemos observar semellanzas en termos de tamaño.

Gráfico 1. Datos resumen do sector das TL en España.



Fonte: Estudio de caracterización del sector de tecnologías del lenguaje en España (2018).

No mercado europeo das TL, un número relativamente pequeno de empresas multinacionais xogan un papel dominante. Estas empresas, con sede en diferentes países europeos, teñen unha ampla presenza e influencia no sector, dominando o mercado europeo das TL grazas á súa ampla experiencia, capacidade de investimento en investigación e desenvolvemento, e a súa capacidade para ofrecer solucións nos principais idiomas de Europa.

É importante ter en conta que, a pesar da presenza dominante destas empresas multinacionais, o sector das TL tamén conta con empresas de menor tamaño e startups innovadoras que contribúen á diversidade e competencia do mercado. A súa participación é esencial para promover a innovación e a evolución constante das TL en Europa e prestar servizo ás linguas minorizadas.

Ámbito de actuación: As empresas do sector en España operan principalmente nun ámbito local ou rexional, aínda que ofrecen solucións multilingües. Son moi poucas as compañías que lograron expandirse e ofrecer os seus servizos e produtos en mercados internacionais.

Tipoloxía do persoal: O sector das TL require profesionais altamente cualificados e multidisciplinares. Os equipos de traballo adoitan estar compostos por lingüistas, enxeñeiros informáticos, especialistas en intelixencia artificial e outros expertos da área. Ademais, debido á natureza interdisciplinaria deste campo, é común atopar perfís que combinan coñecementos lingüísticos e técnicos, o que incrementa aínda máis a dificultade para captar empregados cualificados, sendo

esta unha das principais preocupacións das empresas do sector. Por outra banda, segundo o estudo de caracterización do sector, o 42,9% dos traballadores do sector das TL en España son mulleres, mentres que o 57,1% son homes, o que supón unha menor brecha de xénero que noutros sectores tecnolóxicos.

Mercados verticais: As TL teñen aplicacións en diversos sectores económicos, como a administración pública, a banca, as telecomunicacións, os servizos profesionais, etc. Cada sector ten necesidades específicas en termos de procesamento da lingua e comunicación multilingüe, o que abre oportunidades para as empresas de TL.

Actualización e adaptación: O sector está en constante evolución e actualización. As empresas deben estar ao tanto dos avances tecnolóxicos e as novas tendencias no procesamento da linguaxe natural, a intelixencia artificial e outras áreas relacionadas. A capacidade de adaptarse rapidamente aos cambios e ofrecer solucións innovadoras é fundamental para ser competitivo neste sector.

Multilingüismo: Ademais da diversidade de actividades e perfís profesionais presentes no sector en España, é importante resaltar a súa natureza multilingüe. Segundo datos do Estudo de caracterización do sector, constatouse que o 86,7% das empresas traballan co castelán e as súas variantes dialectais, o 76,7% traballan con linguas internacionais e o 56,7% con linguas cooficiais. Esta realidade reflicte a necesidade de contar con solucións tecnolóxicas capaces de abordar

diferentes idiomas e contextos lingüísticos, o que implica un enfoque plurilingüe e adaptado á diversidade lingüística existente no ámbito nacional e internacional. O dominio de múltiples linguas e a capacidade de ofrecer solucións eficaces en distintos idiomas son factores clave para o éxito das empresas do sector e para a súa adaptación ás necesidades do mercado global. A pesar de contar cos produtos e coñecementos necesarios, a materia pendente deste sector é precisamente a súa internacionalización.

Actividades principais: En canto ás actividades principais, as empresas do sector das TL en España abarcan un amplo espectro. Algúns exemplos destas actividades son:

- Desenvolvemento de software e ferramentas de procesamento da linguaxe natural.
- Tradución automática
- Tradución e localización de contido en diferentes idiomas.
- Desenvolvemento de sistemas de recoñecemento e síntese de voz.
- Análise de texto e minería de datos.
- Desenvolvemento de chatbots e sistemas de diálogo automatizado.
- Servizos de consultoría e formación especializada en tecnoloxías da linguaxe.

Sectores de aplicación: As empresas

do sector atenden a diversos sectores, entre os que destacan:

- **Sector da saúde:** Fornecen solucións tecnolóxicas para a mellora da asistencia sanitaria, como sistemas de recoñecemento de voz para documentación clínica e análise de texto para diagnóstico e investigación médica.
- **Sector financeiro:** Desenvolven ferramentas de procesamento da linguaxe natural para análise de riscos, servizos de atención ao cliente e detección de fraudes.
- **Sector do comercio electrónico:** Ofrecen solucións de tradución automática e análise de sentimentos para axudar ás empresas a expandirse en mercados internacionais e mellorar a experiencia do cliente.
- **Sector público:** Prestan servizos de tradución e interpretación en administracións públicas e institucións gobernamentais, así como solucións de procesamento da linguaxe natural para a mellora dos servizos públicos.

1.6 O FUTURO DAS TL

A demanda de solucións de tecnoloxías da linguaxe está aumentando significativamente a medida que a dixitalización avanza e a necesidade de comunicación en distintos idiomas aumenta, a demanda continuará a crecer. A competencia neste mercado é intensa e a evolución tecnolóxica continua, o que obriga ás empresas a unha constante innovación e

adaptación ás novidades e necesidades do mercado.

O sector das tecnoloxías lingüísticas é un campo dinámico e en rápido crecemento que está a cambiar a forma en que interactuamos coas computadoras. Os novos modelos de linguaxe xenerativos e os enfoques de aprendizaxe automática están a revolucionar o sector abrindo unha variedade de novas posibilidades para aplicacións en distintos campos. O seu potencial é tan amplo que aínda estamos a comprender plenamente as súas implicacións en diversos ámbitos, desde a asistencia ao cliente ata a investigación e desenvolvemento de novas solucións. Pero esta evolución trae aparelhada unha serie de riscos:

Competencia para as pequenas e medianas empresas do sector de TL: As TL estanse volvendo cada vez máis accesibles, o que fai moito máis complicado competir para as pequenas empresas do sector. As grandes empresas, como Google e Microsoft, teñen unha gran cantidade de recursos que lles permiten desenvolver e comercializar novas tecnoloxías da linguaxe a un ritmo máis rápido que as pequenas empresas. Isto dá ás grandes empresas unha vantaxe competitiva significativa, o que dificulta que as pequenas empresas se manteñan no mercado.

Linguas minorizadas: As tecnoloxías da linguaxe tamén poden ter un impacto negativo nas linguas minorizadas. Isto débese a que están deseñadas na súa maioría para as linguas máis faladas do mundo. Como resultado, as linguas minorizadas a miúdo non teñen presenza nas aplicacións ou ferramentas lingüísti-

cas. Isto pode facer que sexa máis difícil para os seus falantes utilízalas e acceder a información e servizos na súa lingua. Como xa se mencionou anteriormente, este é o potencial papel das pequenas empresas que traballan en local, pero para iso é imprescindible o apoio das administracións para que fomenten, promociónen e axuden estes desenvolvementos.

Perda de emprego en sectores cualificados: As tecnoloxías lingüísticas tamén poden conducir á perda de emprego en sectores cualificados. Isto débese a que poden automatizar moitas tarefas que actualmente realizan os humanos. Por exemplo, os sistemas de tradución automática poden traducir textos dun idioma a outro cunha alta calidade, o que pode conducir á perda de empregos no sector da tradución.

Estes son só algúns dos riscos potenciais das tecnoloxías lingüísticas. Hai que tomar conciencia deles para tomar decisións informadas sobre como utilizar estas tecnoloxías. Igualmente, existen outros riscos potenciais, como a posibilidade de uso indebido dos datos ou a violación da privacidade, os cales é importante coñecer para podelos mitigar. En xeral, as tecnoloxías lingüísticas teñen o potencial de ser unha ferramenta poderosa e beneficiosa para a sociedade, sempre que estes aspectos se teñan en conta.

Mais o futuro do sector das TL é prometededor. A aparición de novos modelos de linguaxe xenerativos, como *ChatGPT* de *OpenAI*, *Bard* de Google, *LLaMA* (*Large Language Model Meta AI*) ou *Vicuna*, creado por investigadores de UC Berke-

ley, CMU, Stanford e UC San Diego, está a revolucionar a forma en que interactuamos coas computadoras. Estes modelos son capaces de xerar texto que é indistinguible do texto escrito por humanos, o que abre unha variedade de novas posibilidades para aplicacións en educación, entretemento, atención médica e moitos outros campos.

Un exemplo paradigmático de iniciativa coa capacidade de propiciar un avance significativo, tanto para as empresas como para a propia lingua galega é o proxecto Nós, que se describe con certo detalle en outro capítulo deste documento. Este proxecto aspira a potenciar a presenza dixital do idioma galego por medio da creación de ferramentas e, o máis destacable, recursos de excelente calidade que se liberarán para a súa implementación en calquera produto tecnolóxico, aplicación ou servizo que incorpore a nosa lingua. A expectativa é que este proxecto xere un impacto relevante na proxección futura do galego.

De forma paralela, existen proxectos con características semellantes para outras linguas cooficiais, como o catalán (AINA), o euskera (GAITU) e o valenciano (VIVES). Desta forma, estanse a disponibilizar para as empresas do sector un conxunto de recursos que, doutro xeito, sería difícilmente asumible por empresas de pequeno e mediano tamaño. Estes recursos permitirán ás empresas crear, optimizar e evolucionar os produtos e servizos que ofrecen nas linguas cooficiais, contribuíndo deste xeito ao seu fortalecemento e visibilidade no panorama di-

xital actual.

Non podemos ignorar que este é o momento idóneo para que as administracións públicas miren cara o futuro e se comprometan con estes proxectos. Apoiar iniciativas como o proxecto Nós non só significa impulsar a industria tecnolóxica local, senón tamén contribuír decisivamente á supervivencia e promoción das linguas minorizadas no ecosistema dixital. As administracións teñen o deber de salvagardar e fomentar a diversidade lingüística, unha riqueza que se pode ver ameazada pola expansión de linguas maioritarias no ámbito dixital. O respaldo a este tipo de proxectos supón un investimento a longo prazo que asegura unha presenza sólida das nosas linguas no mundo dixital, favorece o desenvolvemento da economía local e, o que é máis importante, garante o dereito dos cidadáns a usar a súa lingua en todos os ámbitos da vida, tamén no dixital.

Para facer fronte a este desafío, é necesario que o sector das TL traballe na súa profesionalización e especialización. É preciso promover a formación de expertos neste campo, así como a investigación e o desenvolvemento de novas TL que aporten valor engadido e se distingan no mercado. Ademais, é fundamental fomentar a colaboración e a cooperación entre as empresas e organizacións do sector, creando redes de traballo e impulsando proxectos conxuntos que impulsen o avance das TL. Non sabemos o que nos deparará o futuro, pero seguro que será aínda máis increíble.

2. O IMPACTO ECONÓMICO AO REDOR DAS TECNOLOXÍAS LINGÜÍSTICAS

LUIS OTERO GONZÁLEZ (USC)
LUIS IGNACIO RODRÍGUEZ GIL (USC)

2.1. INTRODUCCIÓN

As tecnoloxías lingüísticas (TL) están a adquirir cada día un maior peso económico a nivel internacional. Por unha banda, hai empresas e institucións públicas e privadas, que ofertan produtos e servizos baseados nas TL (recoñecemento de voz, de imaxe, chatbots, etc.). Ademais os avances tecnolóxicos e a dispoñibilidade de datos, unido a un potencial de mercado moi amplo, está a contribuír ó crecemento do número de competidores e o nivel de facturación, cun importante aumento na oferta deste tipo de produtos e servizos. Por outra banda, o uso das TL por parte das empresas está tamén aumentando e todo apunta a que vai a seguir nesta tendencia, xa que estanse a incorporar nas operacións fabrís, administrativas e comerciais. O principal beneficio das solucións das TL para as empresas é utilizar a automatización para reducir custos e mellorar as operacións comerciais para maximizar a produtividade e a rendibilidade. Neste sentido, as empresas que levan varios anos empregando estas solucións argumentan que estas tecnoloxías melloran a eficacia, xera vantaxes competitivas, reduce o tempo de produción, reduce riscos, custos e permite acceder a novas oportunidades de negocio (Expert.ai, 2022)¹.

O potencial económico das linguas en España ponse de manifesto co impulso do “PERTE: La nueva economía de la lengua²”, que é un proxecto de carácter estratéxico con gran capacidade de arrastre para o crecemento económico, o emprego e a competitividade, neste caso para a economía española. A nova revolución industrial baseada na intelixencia artificial sitúa a lingua como un dos factores clave para o desenvolvemento tecnolóxico e empresarial xa que o tratamento da lingua é unha parte fundamental da dixitalización e das novas tecnoloxías e, por conseguinte, da propia IA. O presuposto inicial estimouse en 1.100 millóns de euros para o período 2021 – 2026, anunciándose tamén que cando menos 30 millóns íanse a destinar en exclusiva a proxectos relacionados con linguas cooficiais do Estado, algo que de momento non foi así. A esta inversión sumárase a inversión privada que no seu caso puidese xurdir.

A proceso de dixitalización das empresas está moi vinculada cas TL. Segundo publica o Observatorio ADEI³ no estudo “Dixitalización e produtividade” o efecto de incrementar en 10 puntos porcentuais o grao de dixitalización do tecido empresarial (medido polo uso de ferramentas dixitais) traduciríase nun incremento próximo ao 0,4% da produtividade do traballo. Na mesma liña, o estudo “Dixitalización

1- <https://www.expert.ai/wp-content/uploads/2022/12/The-2023-Expert-NLP-Survey-Report-Trends-driving-NLP-Investment-and-Innovation.pdf>

2- <https://planderecuperacion.gob.es/como-acceder-a-los-fondos/pertes/perte-nueva-economia-de-la-lengua>

3- <http://www.observatorioadei.es/publicaciones/NotaTecnica-Digitalizacion-y-productividad.pdf>

e desempeño empresarial” do mesmo Observatorio⁴, indica que, en España, a evolución do tamaño empresarial, a capacidade de xeración de ingresos de explotación, contía de ingresos de explotación por cada empregado e de EBITDA e a produtividade laboral, entre outras variables, foi superior nas empresas dixitalizadas que nas que non o están.

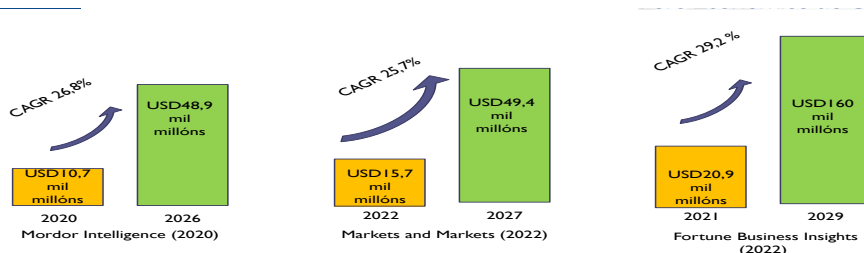
A pesar do potencial económico, non existen apenas estudos que avalíen esta cuestión. Os únicos informes que abordan a cuestión, son feitos por consultoras que adoitan un enfoque internacional. No ámbito nacional, só contamos cun estudo do ano 2018⁵ elaborado para o Observatorio nacional de tecnoloxía e sociedade (ONT-SI), de aí a necesidade, neste senso, resulta interesante avaliar o impacto económico que está a xerar esta tecnoloxía. Para coñecer este impacto total faise necesario por unha parte, determinar o impacto que xera nas empresas que prestan estes produtos ou servizos, tanto directos como indirectos, e por outra determinar o impacto nas empresas que implantan e usan tecnoloxías vinculadas coas TL.

2.2 PREVISIÓNS SOBRE A EVOLUCIÓN DAS TECNOLOXÍAS LINGÜÍSTICAS (TL)

A nivel mundial distintas fontes cuantifican e prognostican un crecemento considerable do mercado . Máis en detalle e tal e como figura no Gráfico 2, pódese observar como Mordor Intelligence (2020) estima que o mercado global crecerá dende os 10.720 millóns USD en 2020 ata os 48.460 millóns USD para 2026, rexistrando unha CAGR (taxa de crecemento anual composta) do 26,84 % durante o período de prognóstico (2021-2026). Na mesma liña, o informe de MarketsandMarkets (2022) indica que a taxa de crecemento anual sostida do sector ata 2027 sería dun 25,7%. Dos 15.700 millóns de dólares en 2022, espérase que medre ata alcanzar os 49.400 millóns en 2027. Por a súa parte, Fortune Business Insights (2022) estima que dos 20.900 millóns de dólares en 2021 se acadarán os 160.000 millóns en 2029

Para o desenvolvemento destas tecnoloxías necesítanse grandes cantidades de datos para poder adestrar as máquinas.

Gráfico 2. Previsións do mercado das TL a nivel internacional



Fonte: Elaboración propia a partir das fontes indicadas.

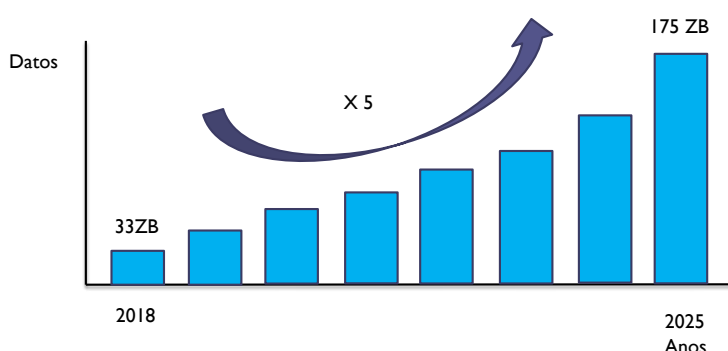
4- http://observatorioadei.es/publicaciones/digitalizacion_y_desempeno_empresarial.pdf

5- Estudio de Caracterización del Sector de las Tecnologías del Lenguaje en España 2018. Madrid : ACAP e OESIA, 2018.

Actualmente isto é máis doado porque se están a crear grandes bases e centros de datos, e os procesos de dixitalización favorecen a consecución da materia prima. Este é un dos motivos que están a contribuír ó desenvolvemento tan espectacular das TL.

Tal e como se recolle no informe da empresa Seagate e a consultora IDC⁶, prevese que no horizonte do 2025 crearanse uns 175ZB⁷ de datos, dos cales o 79% estarán en formato texto. Pódese ver a evolución no Gráfico 3

Gráfico 3. Xeración de datos a nivel internacional



Fonte: IDC (2018)

Tamén as redes sociais xogan un papel importante, pois a gran maioría dos datos que se comparten son linguaxe natural. Como mostra, indicamos na Táboa 1 algúns dos datos que proporciona o informe Data Never Sleeps 8.0 (2020)⁸ sobre esta cuestión: en 2020 cada minuto publícanse 347.000 stories en Instagram; 147.000 fotos en Facebook; 500 horas de vídeo en Youtube; 69.500 usuarios en LinkedIn publicaban a unha oferta de traballo; e enviáanse case 42 millóns de mensaxes por WhatsApp.

Por parte das empresas que están adop-

tando estas tecnoloxías, os datos dos que dispoñemos indican que están a dedicar cada vez unha maior cantidade de recursos. Así, a enquisa levada a cabo por Gradient Flow (2021) a 655 líderes tecnolóxicos de 55 países revela que o 60 % manifestou que os seus orzamentos en PLN, comparando 2020 e 2021, aumentaron alo menos nun 10 %. Tamén o 33 % indicou un aumento dun 30 % nos orzamentos e o 15 % informou que duplicaron os recursos. As previsións a futuro tamén confirman unha evolución favorable no investimento en PLN. Dita información apa-

6- <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

7- ZB, un zettabyte equivale 1021 bytes

8- <https://www.domo.com/learn/infographic/data-never-sleeps-8>

Táboa 1. Datos xerados nun minuto en redes sociais a nivel internacional

Instagram	347.000 stories
Facebook	147.000 fotos, 150.000 mensaxes
LinkedIn	69.500 usuarios publican unha oferta de traballo
WhatsApp	42 millóns de mensaxes
YouTube	500 horas de videos cargados

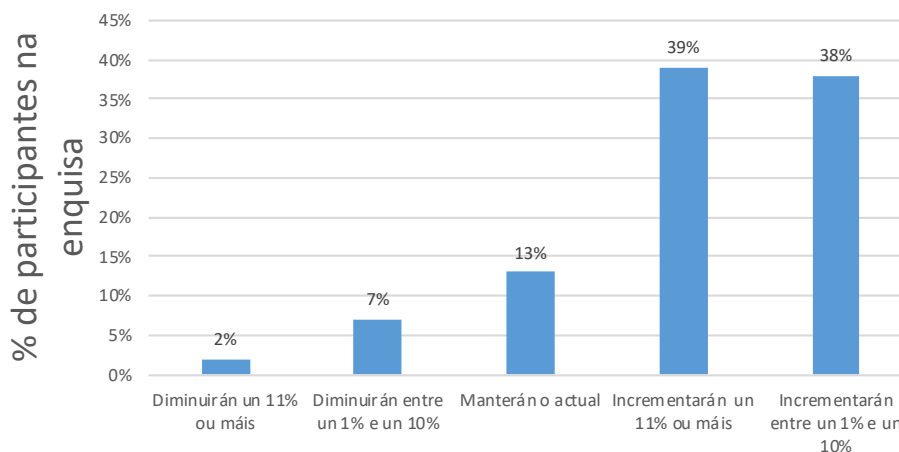
Fonte: Domo (2020)

rece recollida no informe “The 2023 Expert NLP Survey Report”⁹, onde o 77% dos enquisados¹⁰ manifestan que incrementarán o seu presuposto en PLN entre os próximos 12 e 18 meses, tal e como se pode apreciar no Gráfico 4. Esta inversión en tecnoloxías lingüísticas contribuirá a que o mercado siga medrando, pois a demanda deste tipo de solucións cada vez é maior.

Polo que acabamos de ver as empresas do sector das TL teñen unha gran capa-

cidade para seguir crescendo, impulsadas tanto polos avances tecnolóxicos que se están producindo como pola dispoñibilidade de datos. En consecuencia, a oferta deste tipo de produtos / servicios está garantida e pode seguir incrementándose. Polo que respecta ás empresas usuarias, as previsións son tamén moi positivas, e a demanda aumentará debido as vantaxes que con leva o uso das tecnoloxías de procesamento da linguaxe.

Gráfico 4. Previsión sobre a variación no orzamento dedicado a PLN



Fonte: The 2023 Expert NLP Survey Report (2022)

9- <https://www.expert.ai/wp-content/uploads/2022/12/The-2023-Expert-NLP-Survey-Report-Trends-driving-NLP-Investment-and-Innovation.pdf>

10- Enquisa realizada en EEUU e Europa a 150 profesionais con proxectos de PLN activos ou planificados.

2.3 O IMPACTO ECONÓMICO DAS EMPRESAS DO SECTOR DAS TECNOLOXÍAS DA LINGUAXE (TL)

Neste apartado intentaremos coñecer cal é o impacto económico daquelas empresas que desenvolven e ofrecen produtos ou solucións vinculados ás TL. Unha das variables de interese é o volume de negocio que xeran están actividades coa comercialización de produtos e servizos.

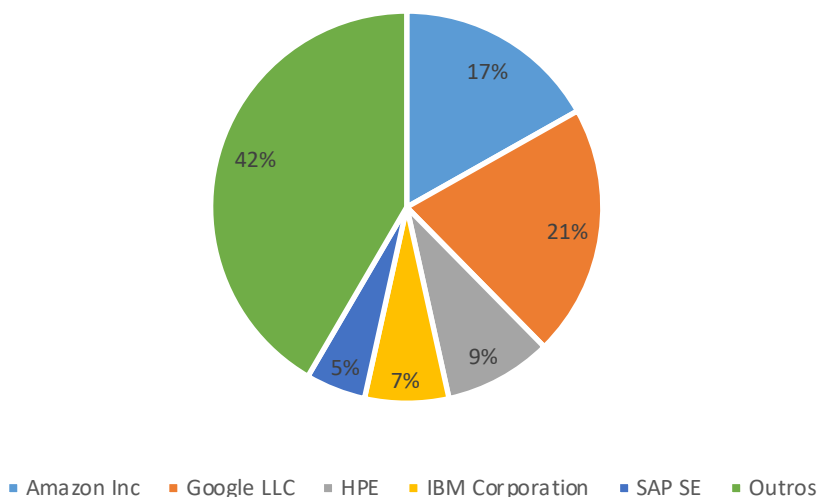
2.3.1 O MERCADO A NIVEL MUNDIAL DAS TL

Unha cuestión relevante é coñecer como se configura e cales son os principais actores deste mercado. En base os datos proporcionados por Fortune Business Insights (2022)¹¹ e que figuran no Gráfico 5 o mercado está moi concentrado, unhas poucas empresa repártense entre o 55% e o 60% da cota de mercado mundial, que-

dando a restante cota (entre o 40 % e o 45%) en mans de outros moitos operadores rexionais e locais.

Como se pode apreciar no gráfico anterior, os grandes líderes a nivel mundial son grandes empresas como é o caso de Amazon, Google, Hewlett Packard Enterprise e IBM Corporation. Actualmente tamén habería que ter en conta a compañía Microsoft que si aparece noutros informes como no de Acumen Research and Consulting (2023). Esta compañía nos últimos tempos está a manter unha estreita colaboración coa compañía OpenIA que alcanzou gran popularidade co lanzamento do ChatGPT. Estas empresas seguiron distintas estratexias para chegar á posición que ocupan hoxe. No Gráfico 6 podemos observar as distintas estratexias así como a súa relevancia. Vemos que as dúas estratexias máis seguidas son as adquisicións e o lanzamento de novos produtos. Este tipo de estratexias están xustificadas, sobre todo,

Gráfico 5. Cota de mercado dos principais axentes das TL.



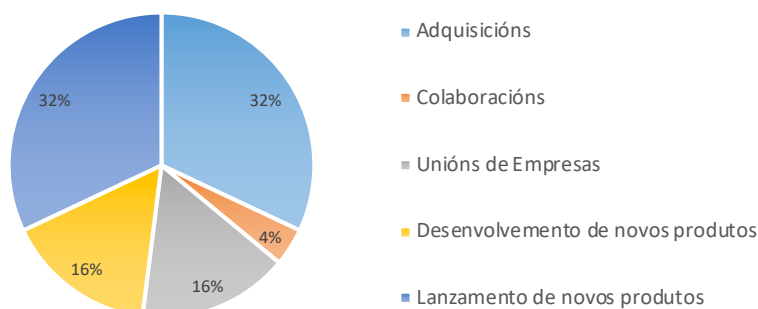
Fonte: Fortune Business Insights (2022)

11- Natural Language Processing (NLP) (2022)

nas empresas que teñan un alto compoñente tecnolóxico. Por unha parte, a tecnoloxía avanza de forma moi rápida entón esixe que os produtos-solucións se teñan que estar adaptando constantemente. Por

outra parte, sobre todo as empresas grandes cando identifican que outra empresa ten produtos /solucións cun gran potencial, é habitual que rematen comprando dita empresa.

Gráfico 6. Estratexias dos principais axentes das TL

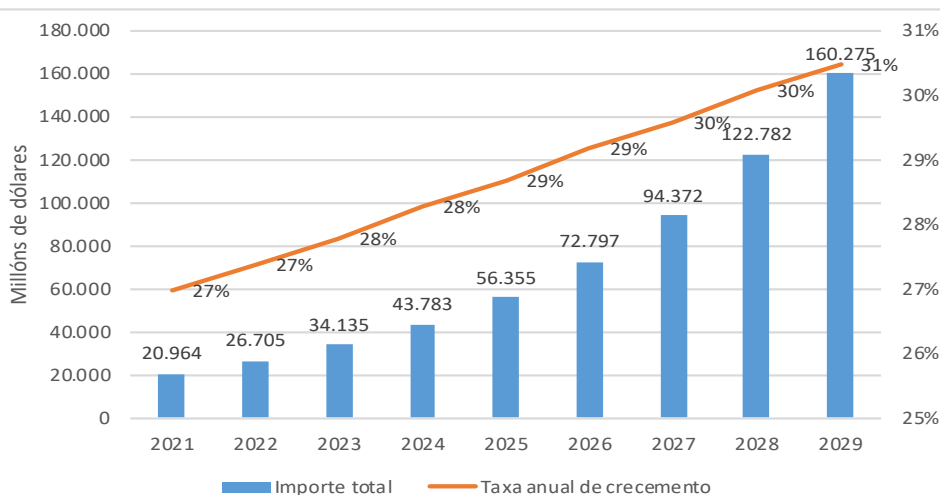


Fonte: Allied Market Research (2022)

Como xa indicamos anteriormente, todas as fontes consultadas prognostican un incremento mundial do mercado, e no Gráfico 7 pódese observar un crecemento

continuo a taxas moi elevadas, e unha estimación de máis de 160.000 millóns para o ano 2029. Dito potencial de mercado pon de manifesto a importancia de dispor

Gráfico 7. Evolución do mercado mundial das TL



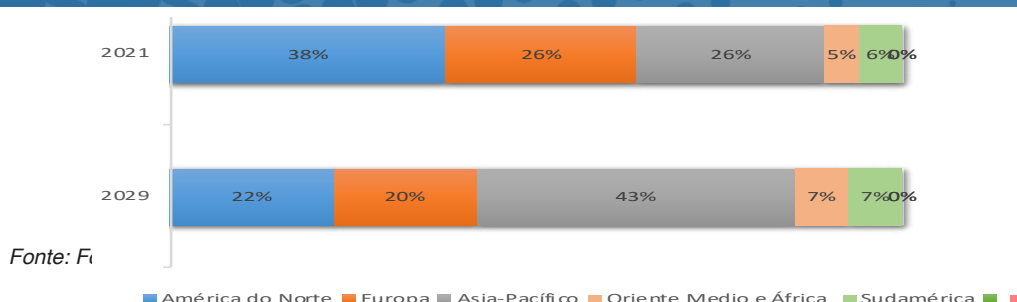
Fonte: Fortune Business Insights (2022)

dunha estratexia que permita dispor a nivel local das capacidades necesarias para desenvolver tecnoloxías propias.

Con respecto á distribución do mercado, actualmente as rexións cun maior mercado son América do Norte e Europa. Esta situación parece que cambiará, xa que no

ano 2029 prevese que a rexión máis importante sexa Asia – Pacífico, principalmente debido á puxanza de China e India. En todo caso, unha cota de mercado próxima ó 20% supón un mercado potencial igualmente moi relevante tal e como se aprecia no seguinte gráfico.

Gráfico 8. Cota de mercado mundial das TL(%) por rexión.

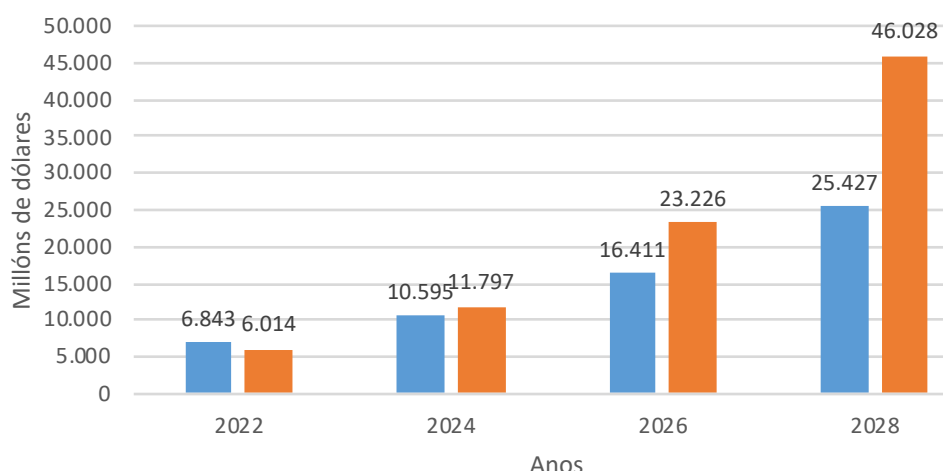


2.3.2 O MERCADO EN EUROPA DAS TL

Cando abordamos a situación no mercado europeo, ó igual que o que se observa para o mercado mundial, a tendencia

esperada é de crecemento sostido durante o período 2020 -2028. Como se pode observar no Gráfico 9, as distintas fontes consultadas prognostican un aumento no mercado, se ben difiren no volume de negocio esperado.

Gráfico 9. Evolución do mercado europeo das TL



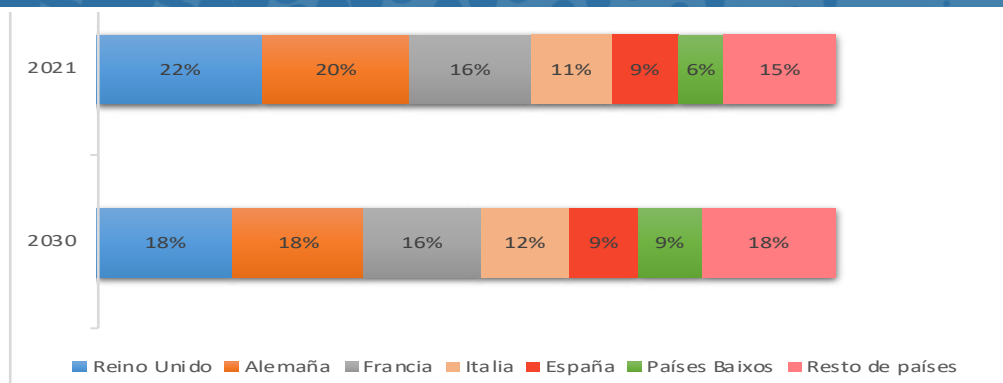
Fonte: Fortune Business Insights (2022) e Allied Market Reserarch (2022)

Dentro de Europa o mercado repártese de forma desigual entre os distintos países. Como podemos apreciar, os mercados máis importantes atópanse no Reino Unido, Alemaña e Francia. A previsión é que estes países sigan tendo o maior peso en 2030, aínda que en Reino Unido e Alemaña descenda lixeiramente. España encóntrase no posto 5º cun 9% da cota de mercado, e considérase que a situación manterase para 2030.

Outras fontes consultadas, en particular o estudo a nivel europeo de 2019¹² que analiza o mercado das tecnoloxías lingüísticas, chegan a conclusións semellantes.

Este estudo, en concreto, prognostica un aumento do mercado das tecnoloxías en Europa e un predominio de certos países sobre o resto. O 73% dos vendedores enquisados neste estudo estaban en Reino Unido, Francia, Alemaña, Italia e España. En concreto, analizouse todo o referido ás tecnoloxías de tradución, tecnoloxías da fala, comprensión da linguaxe natural, analítica e sistemas de procura. Tendo en conta estes 5 segmentos, os ingresos proveñen principalmente das tecnoloxías de tradución e das tecnoloxías da fala. No documento detállase máis polo miúdo as orixes dos ingresos destas compañías.

Gráfico 10. Cota de mercado en Europa das TL (%) por país.

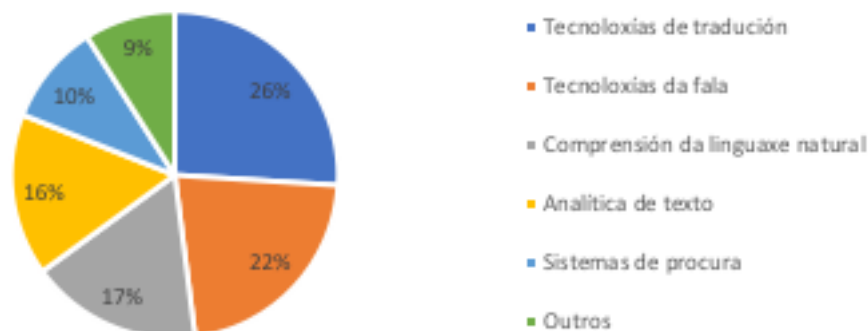


Fonte: Allied Market Research (2022)

Con respecto ás tecnoloxías lingüísticas emerxentes e cunha maior proxección, os enquisados deste último estudo manifestaron que entre as tecnoloxías considera-

das, o procesamento da linguaxe natural é de gran interese para o 88%, a análise de texto para o 63%, e o recoñecemento da fala para o 61%.

12- Final study report on CEF Automated Translation value proposition in the context of the European LT market/ecosystem (2019), dispoñible en <https://op.europa.eu/en/publication-detail/-/publication/8494e56d-ef0b-11e9-a32c-01aa75ed71a1/language-en/format-PDF/source-106906783>

Gráfico 11. Principais orixes dos ingresos das compañías de TL (%)


Fonte: Final study report on CEF Automated Translation value proposition in the context of the European LT market/ecosystem (2019).

2.4 AS TECNOLOXÍAS LINGÜÍSTICAS EN ESPAÑA

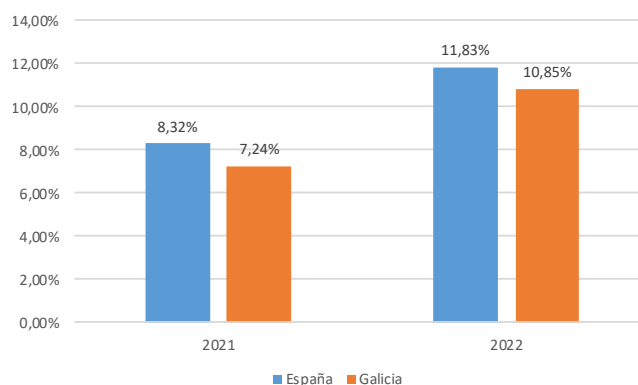
2.4.1 USO AS TL NO CONTEXTO DA IA

As TL forman parte da intelixencia artificial, polo que a situación na que se atope a IA vai a determinar en boa medida a propia das TL. Os datos referidos a 2022 reflicten que o uso por parte das empresas galegas de 10 ou máis persoas empregadas de tecnoloxías de IA, estase aproximando á media nacional (Gráfico 12). Tal e como se pode apreciar, Galicia experimentou unha evolución algo máis favorable, cun aumento de case un 50% fronte ó 40% de España.

En canto ós usos da IA, tal e como se aprecia no Táboa 2, estes son moi variados. En 2021 os usos máis frecuentes en España teñen que ver coa identificación de obxectos ou persoas en función de imaxe, mentres que en Galicia están relacionados coa auto-

matización de fluxos de traballo ou a axuda na toma de decisións. En 2022 tamén o uso máis frecuente en España ten que ver coa automatización de fluxos de traballo ou axuda na toma de decisións, mentres que en Galicia mudou á conversión da linguaxe falada en formato lexible por unha máquina ou á análise da linguaxe escrita. Deste xeito, ponse de manifesto a importancia que ten este segmento xa na actualidade dentro das aplicacións da intelixencia artificial. En termos xerais, as empresas galegas ocupan una boa posición. En concreto, dos tres usos da IA vinculados coa linguaxe, en dous deles a posición en Galicia supera á nacional, concretamente no emprego da IA para a análise da linguaxe escrita e no emprego da IA para converter a linguaxe falada en formato lexible por unha máquina.

Gráfico 12. Porcentaxe de empresas que empregan tecnoloxías de IA.



Fonte: elaboración propia a partir da Enquisa do Uso de TIC e comercio electrónico na empresas. INE (VVAA)

Táboa 2. Usos da IA por parte das empresas.

	2021		2022	
	Nacional	Galicia	Nacional	Galicia
% empresas con IA de análise da linguaxe escrita (1)	29,75	26,70	33,66	27,48
% empresas con IA que converte a linguaxe falada en formato lexible por unha máquina (1)	31,70	29,14	35,07	46,04
% empresas con IA que xera linguaxe escrita ou falada (1)	19,14	29,86	24,04	31,95
% empresas con IA de identificación de obxectos ou persoas en función de imaxes (1)	40,56	44,59	41,98	38,68
% empresas con IA de análises de datos (Aprendizaxe automática) (1)	30,42	35,23	34,07	24,86
% empresas con IA de automatización de fluxos de traballo ou axuda na toma de decisións (1)	38,57	55,40	45,97	37,27
% empresas con IA que permite o movemento físico de máquinas (1)	12,64	24,23	16,97	16,07

(1) Porcentaxe sobre o total de empresas que empregan IA

Fonte: elaboración propia a partir da Enquisa do Uso de TIC e comercio electrónico na empresas. INE (VVAA)

En Galicia apreciamos un aumento considerable (38%) no uso da IA que converte a linguaxe falada en formato lexible por unha máquina, pasando a ser o máis importante na nosa comunidade.

Os usos principais da IA presentan unha alta heteroxeneidade segundo a actividade económica (Táboa 3). Se nos fixamos nos usos vinculados lin-

guaxe', tanto a nivel nacional como a nivel galego, as empresas do sector servizos e da construción fan un uso máis intensivo.

Táboa 3. Porcentaxe de empresas por sectores que empregan IA segundo o seu uso. España e Galicia (2022)

	Nacional			Galicia		
	Industria	Construción	Servizos	Industria	Construción	Servizos
% empresas que empregan tecnoloxías de Intelixencia Artificial (IA)	9,89	6,26	14,17	9,61	7,05	12,69
% empresas con tecnoloxía IA de análise da linguaxe escrita (1)	17,31	38,65	34,93	10,11	0,00	35,28
% empresas con tecnoloxía IA que converte a linguaxe falada en formato lexible por unha máquina (1)	17,15	52,86	35,08	7,71	90,99	45,69
% empresas con tecnoloxía IA que xera linguaxe escrita ou falada (1)	12,99	16,85	27,71	7,33	41,98	34,75
% empresas con tecnoloxía IA de identificación de obxectos ou persoas en función de imaxes (1)	46,15	25,61	39,21	45,95	0,00	37,14
% empresas con tecnoloxía IA de análise de datos (Aprendizaxe automática) (1)	24,74	20,29	37,33	18,35	4,12	32,76
% empresas con tecnoloxía IA de automatización de fluxos de traballo ou axuda na toma de decisións (1)	51,39	29,47	46,41	46,17	9,01	43,63
% empresas con tecnoloxía IA que permite o movemento físico de máquinas (1)	33,31	9,25	13,34	39,09	0,00	22,63

(1) Porcentaxe sobre o total de empresas que empregan IA

Fonte: elaboración propia a partir da Enquisa do Uso de TIC e comercio electrónico na empresas. INE (VVAA)

2.4.2 O MERCADO DAS TL EN ESPAÑA

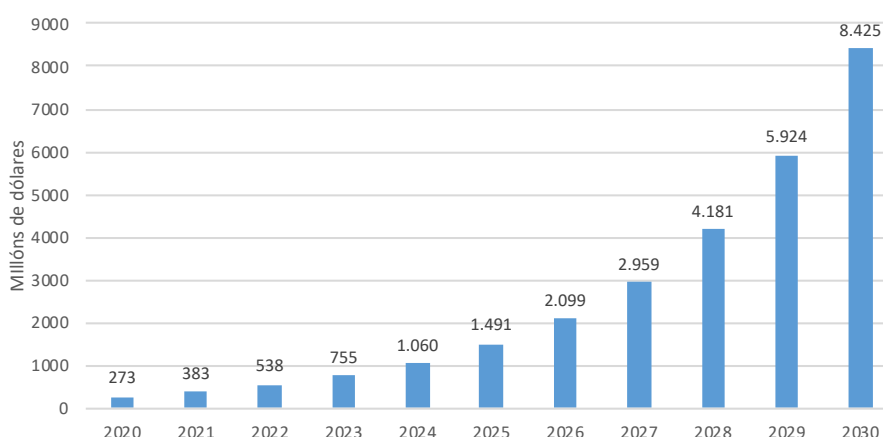
Cando nos referirmos a España, é necesario facer mención ó único estudo de mercado feito: Estudo de caracterización do sector de tecnoloxías

lingüísticas en España (2018)¹³, elaborado por ACAP e OESIA, onde se indicaba un volume de negocio para 2016 de 205 millón de euros. As estimacións feitas recentemente pola consultora Allied Market Research para España

13- <https://www.ontsi.es/es/publicaciones/Estudio-de-caracterizacion-del-sector-de-tecnologias-del-lenguaje-en-Espana>

tamén indican que o mercado medrará de forma intensa, chegando a alcanzar 8.425 millóns de dólares en 2030 (Gráfico 13).

Gráfico 13. Evolución do mercado en España das TL



Fonte: Allied Market Reserarch (2022)

Na actualidade á hora de identificar os distintos axentes implicados no PLN en España tivéronse en conta distintas fontes de información e, máis concretamente: o “Estudio de caracterización del sector de tecnologías del lenguaje en España” (2018); os datos do Rexistro de Empresas e Centros de Investigación¹⁴ que figura no catálogo de tecnoloxías lingüísticas do Ministerio de Asuntos Económicos e Transformación Dixital; o Mapa de capacidades de tecnoloxías de IA; e o Mapa de capacidades dixitais de Galicia¹⁵. Des-

pois de depurar os datos das fontes anteriores, o número de axentes atopados foi un total de 214, dos cales, 115 son empresas. Na Táboa 4 pode verse un maior detalle.

Como se pode apreciar no Gráfico 14, o peso dos axentes do PLN non é o mesmo en todas as comunidades autónomas. A maioría dos axentes do sector concéntranse en Madrid, Cataluña, País Vasco e Valencia, mentres que Galicia está relativamente ben posicionada, ó situarse na quinta posición.

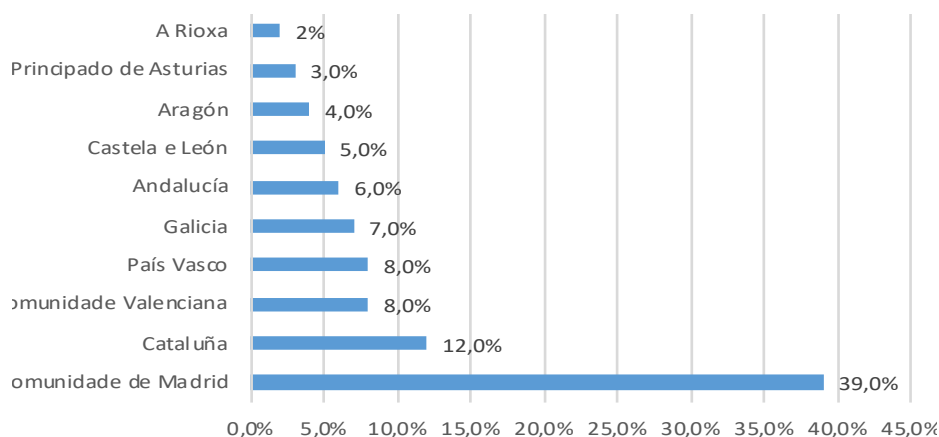
14- <https://plantl.mineco.gob.es/tecnologias-lenguaje/catalogo-TL/Paginas/catalogo-TL.aspx>

15- <https://gaiastech.xunta.gal/es/mapa-de-capacidades-digitales-de-galicia>

Táboa 4. Número de axentes das TL en España (2023)

Empresas	115
Centro de investigación independente	11
Centro de investigación universitario	47
Institución Privada Sen Fins de Lucro (IPSFL)	17
Administración pública	24
TOTAL	214

Fonte: elaboración propia a partir do Rexistro de empresas e centros de investigación, Mapa de capacidades de tecnoloxías de IA e o Mapa de capacidades dixitais de Galicia.

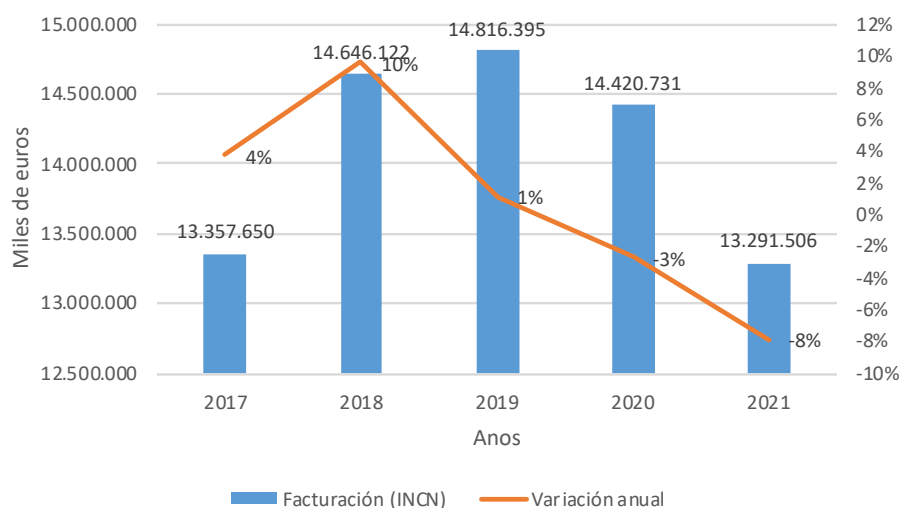
Gráfico 14. Axentes das TL (%) por comunidade autónoma

Fonte: elaboración propia a partir do Rexistro de empresas e centros de investigación, Mapa de capacidades de tecnoloxías de IA e o Mapa de capacidades dixitais de Galicia

Para valorar a importancia económica do procesamento da linguaxe natural é preciso dispoñer do volume de vendas por produtos e servizos directamente relacionados cas TL e do persoal que traballa tamén de forma específica neste ámbito, así como determinar os efectos directos e indirectos. A día de hoxe non se dispón desta información. Os datos mostrados fan referencia ó total das empresas, e das 115 identificadas só se obtiveron datos para 103. Por outra banda,

non é posible saber a partir da información pública ó volume de negocio específico das TL, xa que as empresas non detallan a porcentaxe que representa do total de ingresos. O mesmo sucede para os datos referidos ó persoal. Con respecto á facturación, entre os anos 2017 e 2021, as 103 empresas consideradas tiveron unha cifra de negocio preto dos 15.000 millóns de euros en 2019, contía que diminuíu ata os 13.300 millóns de euros, por mor da pandemia.

Gráfico 15. Facturación das empresas do sector das TL

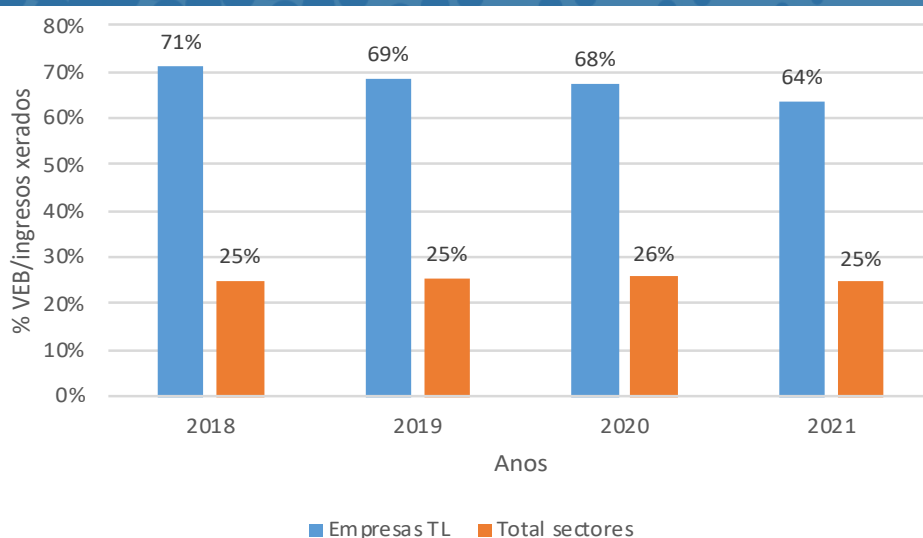


Fonte: elaboración propia a partir de SABI.

O ano de maior facturación foi 2019, mentres que no ano 2021 a facturación destas empresas representou o 1,10% do PIB de España. Como se pode apreciar, dito sector ten un peso relevante

na economía española. Ademais, tal e como podemos apreciar no Gráfico 16, son empresas que xeran un alto valor engadido se o comparamos co total de sectores da economía.

Gráfico 16. Relación entre valor engadido bruto e ingresos xerados.

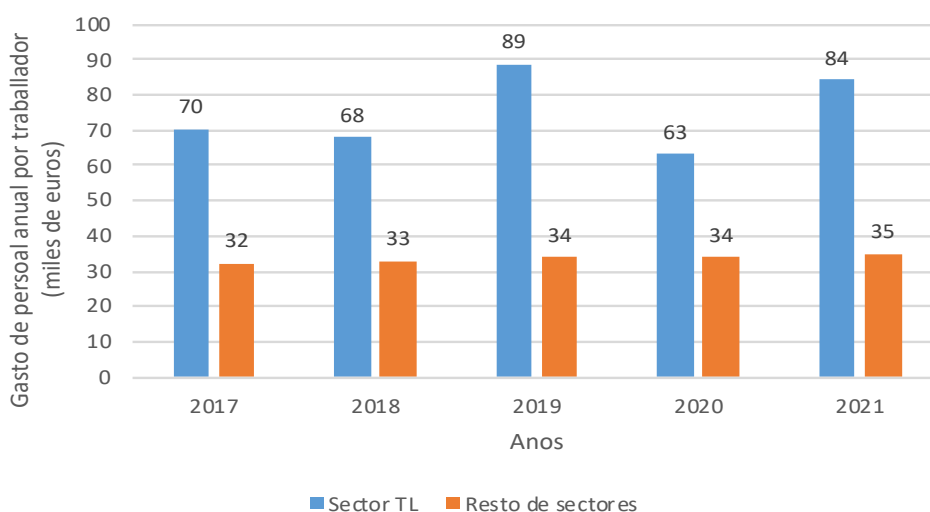


Fonte: elaboración propia a partir de SABI e Central de Balances do Banco de España.

Outra das características destas empresas é que, en xeral, o emprego que xeran é un emprego cualificado, xa que se temos en conta o gasto de persoal anual por traballador (soldos e salarios

+ cargas sociais + outros conceptos salariais), vemos que é maior neste sector en relación co total de sectores da economía, tal e como podemos apreciar no seguinte gráfico.

Gráfico 17. Gasto de persoal anual por traballador.



Fonte: elaboración propia a partir de SABI e Central de Balances do Banco de España.

2.5 O IMPACTO ECONÓMICO DAS EMPRESAS QUE USAN E APLICAN TL

Como xa anticipamos, os beneficios que as TL poden aportar ás empresas que o utilizan son moi variados, e poden conlevar claras vantaxes competitivas. Entre os principais sinalamos: a redución do tempo de produción, de costes e de riscos; o incremento dos ingresos, da produtividade e da eficiencia; e a creación de marca.

Por unha parte, en Europa as TL son empregadas por unha gran variedade de

empresas, se ben hai uns sectores máis intensivos que outros, tal e como se indica no traballo “Final study report on CEF automated translation value proposition in the context of the European LT market/ecosystem” (2019). A Administración, a banca, as telecomunicacións, os servizos profesionais e a sanidade son os principais mercados industriais para estes provedores. Con todo, os ingresos repártense entre 18 mercados ou sectores de actividade. Por outra parte, as TL empréganse nunha gran cantidade de actividades dentro da empresa. En concreto,

nun dos informes de Expert.ai de 2022¹⁶ indícase que entre as aplicacións máis frecuentes atópase a atención ó cliente, operacións de seguridade, desenvolvemento empresarial, vendas, márketing, recursos humanos e finanzas.

Se nos referimos á intelixencia artificial en xeral, o estudo “Notes from the AI frontier: Modeling the impact of AI on the world economy” (2018), de McKinsey Global Institute, analiza o posible efecto que terá a adopción da IA e a non adopción, tanto na economía en xeral como nas empresas. Con respecto á economía, dito estudo indica que a adopción da IA podería xerar unha actividade económica mundial adicional duns 13 billóns de dólares en 2030, o que supón un 16% máis de PIB acumulado que na actualidade. Isto equivale a un crecemento adicional do PIB de ao redor do 1,2% anual. Se o análise se realiza a nivel empresa, as súas previsións son que as empresas pioneiras poderían chegar a duplicar o seu fluxo de caixa en 2030 (tendo en conta os beneficios económicos obtidos e os costes de inversión e transición asociados). Pola súa parte, as empresas que non adopten a IA poderían experimentar un descenso do seu fluxo de caixa de ao redor do 20% con respecto aos niveis actuais. Existen toda una serie de estudos empíricos que conclúen que o despregue de distintas solucións de IA dan lugar a melloras no rendemento das empresas (Wamba-Taguimdje et al., 2020; Mishra e Pani, 2021 e Mikalef e Gupta, 2021). No traballo de Noy e Zhang (2023) que analiza o impacto do uso do ChatGPT na produtividade nas empresas, os resultados mostran un incremento substancial. Máis

concretamente, diminúe o tempo que se dedica a certas tarefas e aumenta a calidade da produción. Ademais, aumenta a satisfacción laboral e a autoeficacia, aumentando tanto a preocupación como o entusiasmo polas tecnoloxías da automatización.

A nivel internacional existe un caso de estudo que levou a cabo a consultora Forrester Consulting para IBM¹⁷ en 2021. Neste estudo analizáronse catro clientes que empregaban solucións de IBM no ámbito do PLN. As empresas eran de Europa, Asia e Estados Unidos, pertencentes ós sectores industrial, servizos legais, provedor de servizos na nube e servizos profesionais vinculados á industria financeira. Analizouse o efecto durante 3 anos. No 1º ano a aplicación foi usada por 200 persoas das empresas e o 2º e 3º anos xa por 400. Para a análise tivéronse en conta, por unha parte, os custes xerados polo uso e implantación da solucións referidas (tales como os das licencias, de implantación e de formación do persoal). Pola outra, consideráronse os beneficios que supuxo o seu uso (como foron os derivados daqueles traballadores que antes dedicaban o 20 % do seu tempo a tarefas de análise / procura de texto, e que reduciron ese tempo nun 50 %. Por outra parte, a eficiencia e a precisión das ferramentas de PLN xeraron un 5% adicional de crecemento comercial por ano e viu-se mellorada a satisfacción do cliente e a calidade do traballo. Tendo en conta todos estes factores, os beneficios obtidos foron de 6,1 millóns de dólares no período analizado (3 anos) fronte a un coste de 1,27 millóns, dando como resultado neto 4,86 millóns de dólares.

16- The 2023 Expert NLP Survey Report: Trends driving NLP Investment and Innovation.

17- <https://www.ibm.com/downloads/cas/XMRMP7XK>

2.6 CONCLUSIONES

A adopción das TL por parte das actividades fabrís, administrativas e comerciais, a todos os niveis, cada vez é maior. Tanto a demanda coma a oferta deste tipo de solucións non para de medrar. Por unha parte, as empresas están en pleno proceso de dixitalización e demandan solucións vinculadas as TL xa que contribúen en gran medida a esta dixitalización o que lles permite obter importantes melloras a nivel de custos, eficiencia, vantaxes competitivas e rendibilidade entre outras. Por outra parte, as empresas do sector das TL ofrecen cada vez máis e mellores solucións, motivado tanto por as melloras tecnolóxicas coma pola dispoñibilidade de datos.

Todo isto fai presaxiar que o impacto económico destas tecnoloxías pode ser importante. Neste senso os estudos son moi escasos e na maioría dos casos son levados a cabo por grandes consultoras a nivel internacional. Resulta interesante polo tanto avaliar o impacto económico que está a xerar esta tecnoloxía. Para coñecer este impacto total faise necesario por unha parte, determinar o impacto que xera nas empresas que prestan estes produtos ou servicios, tanto directos como indirectos, e por outra determinar o impacto nas empresas que implantan e usan TL.

Aínda que non coinciden na taxa de crecemento, tódolos estudos existentes prognostican un incremento do mercado das TL, e os maiores crecementos aínda se producirán nos vindeiros anos, podendo alcanzar en 2029 os 160.000 millóns de dólares. Este incremento producirase en tódalas áreas xeográficas se ben,

destaca o crecemento que se producirá en Asia, motivado por China e India, pasando de ter unha cota no mercado mundial actualmente de 25,6% a unha do 43% en 2029. O mercado internacional caracterízase por estar dominado por grandes compañías como Amazon, Google, Hewlett Packard Enterprise, e IBM Corporation e Microsoft que se reparten preto do 60% do mercado.

A identificación das empresas que se dedican as TL non é doada. Ademais estas empresas tamén fan outras actividades polo que aínda unha vez identificadas é difícil coñecer e obter información sobre as TL, polo que identificamos un campo interesante para investigar máis en profundidade. En España contamos con algún rexistro que é voluntario, polo tanto podemos coñecer algunha pero non todas. Empregando distintas fontes en España identificáronse 214 axentes que están distribuídos por todo o territorio se ben, a Comunidade de Madrid concentra o 39% destes axentes seguida por Cataluña cun 12%. Galicia ocupa o quinto posto cunha porcentaxe do 7% destes axentes polo que se pode considerar en unha boa posición. O impacto económico destas empresas é salientable xa que a súa facturación representa o 1% do PIB nacional. A facturación de 2021 alcanzou os 13.200 millón de euros. Ademais son empresas que xeran gran valor engadido, xa que representa o 64% dos ingresos fronte o 25% se consideramos todos os sectores da economía. Outra das características destas empresas é que, en xeral, o emprego que xeran é un emprego cualificado, xa que se consideramos o gasto salarial anual por traballador do sector das TL en 2021

é de 84.000 euros fronte os 35.000 se consideramos todos os sectores da economía. Polo tanto, poderíamos dicir que é un sector que aporta moitos beneficios a economía dun país.

Ademais o resto de sectores económicos e a sociedade en xeral necesitan de produtos e solucións vinculadas as TL para manterse actualizadas e seguir

sendo competitivos. Os datos existentes revelan que cada vez máis as empresas adoitan TL e aquelas que o fan obteñen vantaxes nos seus procesos, laborais e económicos.

En definitiva, o sector das TL é un sector estratéxico da economía polo que é imprescindible contar cun sector propio forte.

3. O IMPACTO ECONÓMICO DO GALEGO E A SÚA POTENCIALIDADE NO MARCO DAS TECNOLOXÍAS LINGÜÍSTICAS

MARÍA BASTIDA (USC)
MIGUEL VÁZQUEZ TAÍN (USC)

3.1. INTRODUCCIÓN

Na complexa interacción entre a sociedade e a economía, os idiomas desenvolven un papel fundamental que vai moito máis alá da comunicación verbal. Cada lingua é un portal cara á tradición cultural, unha entrada á historia e unha ferramenta para a identidade individual e colectiva. Con todo, o valor dos idiomas transcende a súa dimensión sociocultural e esténdese ao ámbito económico.

O estudo centra a análise no caso da lingua galega, seguindo os azos doutros realizados previamente sobre o impacto económico de linguas coma a vasca, o valenciano, ou incluso o castelán. Na liña destes estudos, conclúese que o galego é o foco principal dun sector de actividade que xira ao redor da lingua, da educación, da cultura e as comunicacións e, polo tanto, ten repercusións significativas na actividade económica de Galicia, en forma de xeración de valor engadido, produción e emprego. Poren, non debemos esquecer que o mundo dixital está a transformar moitas destas actividades económicas, de xeito que as tecnoloxías lingüísticas están chamadas a ser un dos principais sectores de actividade económica arredor da lingua, pero que agora apenas entran dentro da ecuación.

De aí que abordemos a necesidade de preservar e impulsar a lingua galega no ámbito dixital. Todo elo precisa impulsar

o sector das tecnoloxías lingüísticas en galego como elemento fundamental para axudar a preservar a cultura, comunicación e o desenvolvemento económico. A inclusión do galego no ámbito tecnolóxico non só é necesaria para garantir a súa supervivencia, senón para desenvolver un sector cunha enorme capacidade de crecemento e de creación de valor.

3.2 ASPECTOS SOCIO-CULTURAIS DOS IDIOMAS

A diversidade lingüística é unha expresión directa da riqueza cultural. Cada idioma ten unha riqueza e unha bagaxe única que reflicten as formas únicas en que diferentes comunidades interpretan a súa contorna. Os idiomas son transmisores de identidade cultural, polo que a relación entre os idiomas e a cultura é intrínseca e mutuamente enriquecedora.

Á hora de estimar o impacto económico dunha lingua, debe terse presente que a súa cuantificación é complexa, moitas veces pola propia dificultade de obtención e cuantificación de datos. Esta dificultade é maior no caso do cálculo de intanxibles, como acontece co valor dun idioma.

En termos xerais, pode afirmarse que a posesión dunha lingua propia importa, e moito, na vida económica dun país. Pese á súa consideración coma activo intanxible, o idioma ten valor económico porque

permite proxectar industrias que utilizan a lingua como insumo básico no proceso de xeración de valor (industrias culturais, de ensino do idioma ou das comunicacións, por exemplo), creando un tecido produtivo que doutro xeito non existiría, ou tería unha dimensión menor.

Esta visión económica acada maior relevancia cando a lingua ten proxección internacional. Ademais de potenciar os efectos previamente comentados, permite tamén reducir os custos de transacción das operacións realizadas entre países pertencentes á mesma comunidade idiomática, potenciando a internacionalización da economía e a proxección dos axentes económicos. Este efecto é de particular interese na análise das implicacións potenciais da incorporación da IA ás empresas galegas vencelladas ao idioma.

Polo tanto, a lingua é fonte de valor económico, cuxa achega é susceptible de ser cuantificada, aínda que de forma imperfecta. Neste contexto, unha lingua é un instrumento indispensable no desenvolvemento das actividades humanas, polo que se lle poden atribuír características económicas como valor, utilidade, custos e beneficios.

Tradicionalmente, a introdución dos enfoques de tipo económico nos estudos da lingua atribúese a Marschak (1965), quen sentou as bases dun campo específico de especialización académica denominado “Economía e Linguaxe”. Esta disciplina implica un esforzo sistemático por empregar unha variedade de conceptos teóricos da economía para establecer vínculos entre o uso dun idioma e a actividade económica, incluíndo a dimensión macroeconómica como unha análise individual de comportamento.

Para iso, identifícanse unha serie de características propias da linguaxe e se relaciona con determinados conceptos económicos. Cómpre destacar que, facendo uso da mesma analoxía, poderíamos preguntarnos se o idioma presenta as características definitorias propias dos activos estratéxicos: únicos, valiosos, escasos, non apropiables nin substituíbles, e sustentables. Neste caso, estaríamos perante as bases de conformación dunha vantaxe competitiva.

3.3 ACTIVIDADES E IMPACTO ECONÓMICO DO GALEGO

Para cuantificar a importancia do galego coma motor de impulso económico, é fundamental identificar as actividades económicas que gardan relación co uso da lingua. Neste grupo, coma é obvio, inclúense as estruturas e servizos das administracións públicas para o impulso e desenvolvemento do galego, pero tamén o emprego e o valor económico xerado polas actividades relacionadas co galego no sector privado.

A partir do estudo pioneiro de Martín Municio (2003) sobre o valor económico do español, establécese un procedemento de identificación das actividades que xiran en torno a unha lingua. O dito procedemento consiste en atribuír a certos produtos e servizos unha vinculación esencial co idioma e, a partir de aí, cuantificar como eses produtos, encadrados en diversas actividades ou ramas da economía, entran a formar parte doutros, incorporando a eles un «valor» que se considera tamén «devido á lingua».

A identificación de actividades é froito

de alternativas de selección das actividades soportadas pola lingua, baixo a hipótese de que «o idioma sexa parte esencial do produto/servizo principal», e como se pondera nelas este compoñente (é dicir, o coeficiente de lingua). Tamén é froito dunha decisión a selección das fontes estatísticas empregadas. Coma é obvio, estas decisións implican limitacións nos cálculos, ademais de engadir dificultade na obtención de datos. Todo isto parte dunha concepción da lingua como ben privado que, ocasionalmente, é obxecto de transacción mercantil (como no caso do ensino da lingua) ou soporte de comunicación esencial dos bens e servizos en sectores económicos diversos (industrias da lingua).

Dentro da industria da lingua pódese, á súa vez, establecerse unha clasificación útil para os efectos da delimitación económica. Así, identifícanse as actividades nas que a lingua é un input presente, directa ou indirectamente, en calquera actividade (e, polo tanto, susceptible de valoración): input tecnolóxico (como tecnoloxía social de comunicación), e input laboral, considerada coma unha destreza máis que atesouran os traballadores. Seguindo esta delimitación, hai un núcleo central de actividades ocupado polos servizos lingüísticos (ensino de lingua, edición para o ensino), un sector estratéxico (industria da lingua) e outros tres de difusión (os sectores editorial, audiovisual e musical –englobados na cultura– e os medios de comunicación).

Seguindo estudos precedentes, identifícanse catro grandes grupos de actividade vinculados á lingua: dous no eido privado (industria da lingua, cultura e medios de

comunicación); un no público (Administración Pública) e outro mixto (educación):

- Actividades relacionadas coa industria da lingua. Neste apartado teñen cabida institucións ligadas ao ensino do galego (do ámbito público, como as escolas oficiais de idiomas, e privado, como as academias) e os organismos que regulan o seu uso normativo, así como empresas que ofrecen servizos de tradución e interpretación ou asesoría lingüística.
- Industrias da cultura e medios de comunicación cuxo uso vehicular é o galego: industria editorial, musical, audiovisual, artes escénicas etc., e outras actividades relacionadas co mundo da cultura. Igualmente, medios de Comunicación.
- Actividades relacionadas co ámbito educativo, nas que o idioma é un elemento crave para desenvolvese e sobre as cales recae o labor de promover o uso e o coñecemento do galego dentro dunha contorna de aprendizaxe plurilingüe. Inclúe todos os ámbitos educativos de formación regrada.
- Servizos para a difusión e coñecemento do galego desde as administracións públicas. Inclúense as actividades que emanan directamente do sector público ligadas á promoción lingüística de maneira directa e indirecta (ámbito cultural).

A partir dos datos incluídos no documento “O valor económico do galego”¹⁸

18- Consello da Cultura Galega (2023). *O valor Económico do galego*, DOI.10.17075/veg.2023. Dispoñible en: *O valor económico do galego unha publicación do Consello da Cultura Galega*

Táboa 5. Actividades vinculadas a lingua.



Fonte: Elaboración propia.

do Consello da Cultura Galega, pódese facer unha aproximación á cuantificación do valor económico e o impacto da lingua galega, abarcando diversos aspectos e utilizando metodoloxías de medición económica. Para dita estimación, identifícanse varios sectores vinculados ao idioma xa mencionados aquí, como as industrias da cultura e os medios de comunicación, o ensino de idiomas, a tradución e interpretación, e as actividades de promoción lingüística e cultural das administracións públicas.

En función destes cálculos, o valor total da produción dos sectores vinculados ao galego en 2019 foi de 4.118 millóns de euros. Este valor componse dun 44,7% de impacto directo (1.841,7 millóns de eu-

ros), un 9,7% de impacto indirecto (397,5 millóns de euros), e un 45,6% de efecto inducido (1.878,6 millóns de euros). Esta produción vinculada ao impacto do galego representa o 3,4% do total da economía galega. O impacto total dos sectores do galego en termos de VEB foi de 2.694 millóns de euros, distribuído entre un 55,4% de impacto directo (1.439,2 millóns de euros), un 8,5% de impacto indirecto (209,9 millóns de euros), e un 36% de efecto inducido (1.044,9 millóns de euros). Este VEB vinculado ao impacto do galego constitúe o 4,7% do VEB total da economía galega. En termos de emprego, o impacto total dos sectores do galego foi de 49.893 postos de traballo equivalentes a tempo completo. Isto inclúe un 55,5% de impac-

to directo (27.702 empregos), un 8,5% de impacto indirecto (4.242 empregos), e un 36,0% de efecto inducido (17.949 empregos). O emprego vinculado ao impacto total do galego representa un 4,7% do total da economía galega.

3.4 O GALEGO NO ÁMBITO DAS TECNOLOXÍAS DA LINGUAXE

A expansión das tecnoloxías da linguaxe realizouse fundamentalmente ao redor das linguas dominantes, deixando en clara desvantaxe ás linguas minoritarias, que apenas aparecen representadas nas plataformas tecnolóxicas e que non dispoñen dos desenvolvementos necesarios para implementalas nos recursos tecnolóxicos. No caso do eúskaro, un 90% da poboación ten acceso a internet, pero menos do 0,1% das webs analizadas utilizan o idioma¹⁹. O acceso a internet no 75% dos casos se fai en 10 linguas, e o 90% dos africanos²⁰ teñen que usar un segundo idioma cando acceden a plataformas de creación de contidos.

Sen a incorporación ao ámbito das tecnoloxías lingüísticas a transmisión da cultura, o coñecemento e as tradicións podería enfrontar serias dificultades. Ademais, a ausencia no mundo dixital e tecnolóxico podería limitar as oportunidades educativas e económicas, provocando un maior declive e, no seu caso, a súa desaparición.

De aí a necesidade de apostar pola inclusión das linguas no mundo tecnolóxico, para deste xeito, non só preservar a

diversidade lingüística e cultural, senón para desenvolver capacidades tecnolóxicas que permitan dispoñer das competencias e a cualificación necesarias para impulsar o sector das tecnoloxías lingüísticas. A singularidade lingüística dunha comunidade presenta desafíos específicos para os desenvolvedores de tecnoloxías da linguaxe. Se ben o desenvolvemento tecnolóxico pode ser unha ameaza para unha lingua como ou galego, tamén é posible aproveitar a oportunidade que xurde arredor de todo o desenvolvemento que se precisa para incorporalo ó mundo das tecnoloxías lingüísticas. Ter unha lingua propia abre a porta á creación de solucións específicas adaptadas ás necesidades lingüísticas locais, permitindo unha mellor precisión do procesamento da linguaxe natural. Ademais, o desenvolvemento de aplicacións específicas favorece a adopción de tecnoloxías e fomenta a diversidade e inclusión na era dixital. De aí que o galego pode ser un activo estratéxico que está servindo para potenciar o desenvolvemento e a innovación no ámbito das tecnoloxías da linguaxe.

Un feito que pon de manifesto a relevancia de ter unha lingua propia é o ranking da actividade económica vencellada ás tecnoloxías da linguaxe desglosada por comunidades autónomas. Despois de Madrid, á industria das tecnoloxías lingüísticas española se sitúa en Cataluña, Valencia, País Vasco e Galicia, todas elas comunidades bilingües. O desenvolvemento de corpus lingüísticos, de tradutores automáticos, de procesadores de

19- González-Dios, I., y Altuna, B. (2022). *Natural language processing and language technologies for the Basque Language*. Cuadernos Europeos de Deusto. DOI: <https://doi.org/10.18543/ced042022>

20- Association for progressive communications (2023). *The language of power: how your trendy chatbot is leaving marginalised communities behind*.

linguaxe natural, e chatbots, entre outros, pode axudar a impulsar un sector tecnolóxico arredor da propia lingua que impulse o crecemento económico, a innovación e a retención de talento. En todos os casos, tamén hai un denominador común, a existencia de proxectos financiados pola administración para axudar a desenvolver ferramentas e recursos que poden ser utilizados no desenvolvemento de produtos asociados ás TL. En Galicia o proxecto NOS pretende potenciar a presenza dixital do galego, ó igual que ocorre co catalán (AINA), co euskera (GAITU) ou valenciano (VIVES). Desta forma, póñense a disposición das empresas recursos que, doutro xeito, serían difícilmente asumibles por empresas de pequeno e mediano tamaño.

A presenza do galego no eido dixital permitirá preservar a riqueza cultural e lingüística de Galicia. Pero manter ou incrementar dito valor dependerá da estratexia da rexión relativa ás tecnoloxías da linguaxe aplicadas ó galego. Nun mundo máis dixitalizado e automatizado, o desenvolvemento das tecnoloxías da linguaxe precisa da integración do galego para asegurar que non quede ó marxe da revolución tecnolóxica. Os primeiros pasos están dados, agora queda seguir consolidando un sector que pode axudar a reter talento, a crear equipos multidisciplinares, a xerar verticais no ámbito das TIC, e, a aplicar o coñecemento adquirido co galego nas TL a outros idiomas. Para elo se precisa que a administración siga impulsando a actividade a través do acceso a datos, recursos financeiros e tecnolóxicos, así como do

talento necesario para o seu desenvolvemento e implementación. No caso do Euskaro, o departamento de cultura e política lingüística puxo en marcha o Plan de Acción de Tecnoloxías Lingüísticas (GAITU) 2021-2024. O PERTE de “nueva economía de la lengua”, é unha clara aposta que vai na dirección deste documento. E tamén, mediante o apoio da administración, o impulso do gaélico²¹ en Irlanda permitiu a creación de startups centradas nas tecnoloxías lingüísticas, aproveitando o potencial cultural e económico da lingua. Todo elo ten contribuído a xerar oportunidades de emprego, investigación e desenvolvemento no campo das tecnoloxías da linguaxe e o mesmo tempo se fortaleceu o uso do gaélico en contextos moi diversos. O caso irlandés pon de manifesto que paga a pena promocionar o uso da lingua minoritaria de forma activa xa que se beneficia tanto a economía rexional como a propia lingua.

3.5 CONCLUSIÓNS

Una parte da cultura galega está vinculada a nosa lingua e non se entendería de non ser pola súa existencia. Seguir mantendo e reforzando esa cultura dependerá do mantemento máis ou menos activo do galego.

Ademais o galego xera riqueza económica e unha parte importante da economía galega depende da nosa lingua. Máis concretamente, o VEB vinculado ao impacto do galego constitúe o 4,7% do VEB total

21- *Digital plan for the Irish language: Speech and language technologies 2023-2027. Ireland Government.*

da economía galega e o emprego vinculado ao impacto total do galego representa un 4,7% do total de Galicia.

Un dos riscos das linguas minoritarias e por conseguinte do galego é que a tecnoloxía a exclúa. Se isto chega a suceder vai afectar o seu uso e por conseguinte a cultura e a economía da nosa rexión. Polo contrario, se o galego logra facerse un oco

nas TL, tanto a cultura como a economía saíran beneficiadas. No caso específico de este tipo de linguas pode que a iniciativa privada non sexa suficiente para promover a súa incorporación a tecnoloxía, polo que vai ser decisiva que a administración promova proxectos para desenvolver ferramentas e recursos que poden ser aplicados polos axentes económicos.



4. AS TECNOLOXÍAS LINGÜÍSTICAS E O GALEGO: O PROXECTO NÓS.

CARMEN MAGARIÑOS (INVESTIGADORA DO PROXECTO NÓS)

MARÍA BAQUEIRO (XESTORA DO PROXECTO NÓS)

PABLO GAMALLO (INVESTIGADOR DO CITIUS)

4.1. INTRODUCCIÓN

Nos últimos anos realizáronse importantes esforzos para dotar o galego de ferramentas tecnolóxicas e poñelas á disposición do conxunto da cidadanía. Porén, a situación do galego no contexto dixital non é equiparable á doutras linguas como o inglés ou o castelán, polo que precisa dun impulso decidido que a sitúe como lingua de uso convencional nos recursos tecnolóxicos de uso común pola sociedade.

Para conseguir que o galego dea un paso decidido cara adiante no eido das tecnoloxías intelixentes, o “proxecto Nós” (de-Dios-Flores et al., 2022; Vladu et al., 2022) naceu coa virtualidade de configurarse como o alicerce do ecosistema dixital que permitirá xerar recursos dixitais e lingüísticos transversais que proporcionarán ao galego a posibilidade de se incorporar a plataformas tecnolóxicas, das cales, até agora, estivera excluído. A finalidade que persegue o proxecto Nós co seu desenvolvemento é acadar unha ampla variedade de recursos lingüísticos en galego de alta calidade e que serán publicados en formato dixital, en xeral baixo licenzas libres, e a través de ferramentas con códigos abertos para ser empregadas en calquera tipo de soporte dixital polas diferentes institucións públicas e, mesmo, polos axentes que conforman o tecido económico empresarial e, asemade, pola

sociedade en xeral. Ademais dos recursos lingüísticos, a elaboración de demostradores será un aspecto de especial relevancia no proxecto, xa que vai permitir visibilizar as posibilidades dos recursos, animando así a empresas e institucións a desenvolver casos de uso ao tempo que contribuirá á creación dun ecosistema galego innovador arredor das tecnoloxías lingüísticas.

Por todo iso, no ano 2021, a Consellería de Cultura, Educación e Universidade, a través da Secretaría Xeral de Política Lingüística (SXPL) e do Centro Ramón Piñeiro para a Investigación en Humanidades (CRPIH) dela dependente, a Axencia para a Modernización Tecnolóxica de Galicia (en adiante AMTEGA) e a Universidade de Santiago de Compostela (en adiante USC), a través do Instituto da Lingua Galega (ILG) e o CiTIUS (Centro Singular de Investigación en Tecnoloxías Intelixentes), asinaron un convenio para definir, deseñar e elaborar a planificación pormenorizada do proxecto Nós para o período 2022-2025.

Posteriormente, en 2022, asinouse un novo convenio de colaboración entre a Consellaría de Cultura, Educación e Universidade, AMTEGA e a USC para desenvolver as actividades previstas para dito ano. Este convenio entrou en vigor o 29 de abril de 2022 e tivo vixencia ata o 20 de decembro do mesmo ano.

En 2023, o proxecto Nós incorporouse ao PERTE Nova Economía da Lingua, for-

mando parte dun proxecto conxunto cos outros proxectos análogos no ámbito do catalán e éuscaro: AINA (Cataluña), GAITU (País Vasco) e Vives (Valencia). Trátase dun proxecto de cooperación para fornecer as tecnoloxías lingüísticas nas outras linguas, ademais do castelán, que conforman as linguas oficiais do Estado español. O conxunto do proxecto está financiado con 7,5 millóns de euros, dos que dous son para o proxecto Nós.

4.2 LIÑAS DE TRABALLO DO PROXECTO

Para colocar a lingua galega na vanguarda da Intelixencia Artificial (IA) e das Tecnoloxías Lingüísticas (TL), o proxecto contempla oito liñas de traballo, das que unhas están centradas na voz (lingua falada) e outras no texto (lingua escrita):

1. Síntese de voz: desenvolveranse sistemas de conversión texto-voz (TTS) de alta calidade e naturalidade, con capacidade para producir voces con distintas identidades, estilos de fala e expresións emocionais.

2. Recoñecemento da fala: crearase un sistema de recoñecemento automático da fala (ASR) de propósito xeral, con capacidade de desempeño en diferentes dominios.

3. Sistemas de diálogo: facilitarán a interacción natural mediante texto ou voz entre unha persoa e un sistema informático.

4. Corrección e avaliación lingüística automática: desenvolveranse correctores avanzados ortográficos, léxicos, gramaticais e de estilo.

5. Tradución automática: serán desenvolvidos tradutores neuronais desde o galego para outras linguas e viceversa, permitindo entrada textual ou por voz.

6. Xeración de textos: xeración automática de textos en galego, tanto con estratexias data-to-text baseadas en regras e lexicóns, como con técnicas end-to-end baseadas en redes neuronais e aprendizaxe profunda.

7. Extracción de información: consistirá na identificación e extracción das principais entidades, relacións semánticas e eventos dun texto ou conxunto de textos. Unha aplicación desta área concretarase na creación de sistemas automáticos de pregunta-resposta, tanto a través de texto como de voz.

8. Verificación de información e minaría de opinións: permitirá clasificar información en relación á veracidade do contido e á polaridade (desde moi negativa até moi positiva) do propio texto.

A día de hoxe os traballos están principalmente enfocados en tres destas liñas: tradución, síntese de voz e recoñecemento da fala, se ben se está a traballar tamén nos ámbitos da xeración de texto e no campo da corrección e avaliación lingüística automática, aínda que en menor medida.

4.3 RESULTADOS DO PRIMEIRO ANO

Os esforzos do equipo investigador no primeiro ano deron como resultado seis corpus, tres demostradores e un prototipo, cumprindo deste xeito cos obxectivos establecidos no convenio 2022. Gran parte do traballo estivo dirixido á elaboración dos corpus, dado que dispor dunha gran cantidade e variedade de recursos lingüísticos (escritos e orais) de alta calidade é fundamental para o desenvolvemento das tecnoloxías lingüísticas para o galego.

A figura 1 mostra de forma gráfica o avance que se ten feito no primeiro ano de traballo no desenvolvemento dos corpus, conseguindo incrementar de forma

significativa o seu tamaño que, como mínimo, se duplicou. A táboa 1 recolle de forma resumida os datos máis relevantes destes corpus. Con relación aos demostradores, desenvóléronse un sistema de conversión texto-voz (TTS) de alta calidade e naturalidade, un demostrador de recoñecemento automático da fala, e un tradutor neuronal español-galego e inglés galego, nas dúas direccións, así como un prototipo de xeración automática de descrições textuais de datos en tempo real, aplicado ao ámbito específico da predición meteorolóxica.

Todos os resultados (corpus, modelos e datos de avaliación) están dispoñibles libremente e accesibles a través da páxina web do proxecto²² e o repositorio de GitHub.²³

Gráfico 18. Tamaño dos corpus antes do proxecto Nós e tras o primeiro ano do proxecto.



Fonte: Elaboración propia.

Táboa 6. Corpus elaborados durante o primeiro ano de desenvolvemento do proxecto Nós (2022).

CORPUS	Tamaño	Utilidade	Orixe dos datos	Linguaxe
Corpus ASR (aliñado automaticamente)	+1.670 h	Recoñecemento da fala	Parlamento (2015-2022)	Lingua estándar (formal)
Corpus ASR (aliñado manualmente)	53 h	Recoñecemento da fala	Entrevistas e discursos	Lingua estándar (formal)
Corpus para ASR (datos brutos)	3.700 h	Recoñecemento da fala	TX 2019-2022 + AGO	Lingua estándar + popular
Corpus TTS	30 h	Síntese de voz	20.000 frases (extraídas)	Lectura
Corpus paralelo galego-español	30 millóns de frases	Tradución automática	Parlamento europeo + subtítulado de cinema	Lingua estándar formal (parlamento) + informal
Corpus paralelo galego-inglés	25 millóns de frases	Tradución automática	Parlamento europeo + subtítulado de cinema	Lingua estándar formal (parlamento) + informal
Corpus conversacional	+585.000 frases	Xeración de textos	Subtitulado	Lingua informal

Fonte: Elaboración propia.

Corpus ASR

Elaboráronse dous corpus de voz e texto aliñados, de utilidade para a tarefa de recoñecemento da fala. Estes corpus poden empregarse para adestrar ou avaliar sistemas ASR. Para este fin, o material sonoro debe segmentarse e aliñarse co texto correspondente, tarefa que pode realizarse de xeito manual (aliñamento manual) ou automático (aliñamento forza-

do). As características dos dous corpus desenvolvidos son as seguintes:

- Nos_ParlaSpeech-GL²⁴. Corpus ASR de máis de 1.670 horas de voz e texto aliñados automaticamente, creado a partir das 316 sesións plenarias celebradas no Parlamento de Galicia entre os anos 2015 e 2022. O corpus divídese en dous subcorpus, “clean” e “other”, que constan de 1.196,92 horas

(667.308 segmentos) e 477,71 horas (130.332 segmentos) respectivamente. Os segmentos incluídos en “clean” foron filtrados de acordo con varios criterios de calidade, mentres que o subcorpus “other” recolle os segmentos descartados neste filtrado. Este corpus está enriquecido cos metadatos asociados ás mencionadas sesións plenarias e ás persoas que interviñeron nos Plenos do Parlamento no período indicado (máis de 200 falantes). Ademais, cada segmento de audio está asociado co identificador do falante correspondente. Os distintos metadatos, recollidos dentro do proxecto ParlaMint²⁵, pódense consultar nesta ligazón²⁶. Para levar a cabo as tarefas de aliñamento dos datos de texto e audio, necesaria para a elaboración do corpus, contouse coa colaboración do centro tecnolóxico Vi-comtech²⁷.

- Nos_TranscriSpeech-GL²⁸. Corpus ASR de 53 horas, de dominio variado, obtido mediante transcripción e aliñamento manuais. Ao igual que no corpus anterior, cada segmento de audio está asociado co falante correspondente, e o corpus acompáñase dos metadatos de cada locutor e da guía empregada na realización das transcripcións. O corpus divídese en catro subcorpus temáticos: Conferencias, Entrevistas, Debates e Discursos.

Alén destes dous corpus, grazas á co-

laboración doutras institucións públicas e proxectos de investigación da USC, na actualidade o proxecto Nós dispón do seguinte material sonoro en forma de datos brutos:

- CRTVG. Telexornais Mediodía da TVG entre 2019-2022, sumando aproximadamente 1.200 horas de material audio-vídeo. Este material acompáñase das súas correspondentes transcripcións (guións).
- AGO (lingua popular). Máis de 8.000 gravacións de lingua popular levadas a cabo dentro do proxecto Arquivo do Galego Oral (AGO).

Corpus TTS

O corpus para síntese de voz, denominado Nos Celtia-GL²⁹, foi creado a partir da voz da actriz Consuelo Díaz Isorna, e gravouse nos estudos Cinemar Films, en Ames (A Coruña).

A voz foi elixida a partir dun proceso de selección no que participaron catro actrices que cumprían as seguintes características: voz feminina, profesional, de entre 30 e 50 anos e bo nivel de galego. Para a escolla, na que participaron expertos en lingüística e membros do proxecto, empregouse un test perceptivo realizado a través da ferramenta FOLERPA, Ferramenta On-line para Experimentación Perceptiva, deseñada polo Instituto da Lingua Galega (ILG).

Para a elaboración do corpus de voz foi precisa a compilación previa dun corpus de frases. Obtívose un corpus de 20.000 frases, máis 3.000 frases de reserva, procedentes de diferentes ámbitos:

- Corpus GTM³⁰ (textos escritos) (10.000 frases)
- Corpus proxecto Nós (textos orais e escritos) (13.000 frases)
- 34% transcricións ortográficas do AGO (Arquivo do Galego Oral)
- 18% definicións do Dicionario RAG
- 20% Diarios de Sesións transcritas do Parlamento de Galicia entre 2015 e 2020
- 27% guións dos Telexornais Medio-día da TVG de 2021
- 1% corpus de entoación

Corpus paralelos

Foron obtidos, mellorados e transformados diferentes corpus paralelos GL-ES e GL-EN para poder ser utilizados por sistemas de tradución automática neuronal. Podemos falar de dous tipos de corpus: auténticos e sintéticos. Os auténticos son aqueles que foron construídos desde corpus paralelos ES-GL e EN-GL, sen ningún tipo de transformación sobre a lingua dos textos. Os sintéticos son aqueles transformados para galego desde corpus ES-PT e EN-PT mediante técnicas de transliteración e tradución PT-GL (Ortega et al., 2022).

Debido á falta de corpus libres entre pares ES-GL e, especialmente, entre EN-GL de tamaño suficiente para adestrar estes sistemas, foi preciso acudir aos corpus da

variante lingüística máis próxima: o portugués. Sobre estes corpus foron realizadas diferentes transformacións para seren convertidos en corpus de galego ILG-RAG válidos: transliteración ortográfica e transformación lexicográfica. Posteriormente, con estes corpus transformados foron realizadas distintas experiencias de adestramento de sistemas NMT con arquitecturas LSTM e Transformer. Como resultado destas conversións obtivemos corpus de EN-GL e ES-GL desde o portugués, para finalmente adestrar os primeiros tradutores neuronais libres con tecnoloxías LSTM e *Transformer* mediante o uso da biblioteca de código aberto OpenNMT³¹.

Corpus conversacional

Obtivéronse dous *datasets* de máis de 3.000 pares de datos tabulares aliñados con comentarios descritivos, un para o castelán e outro para o galego. O corpus en galego é o primeiro corpus para sistemas *data-to-text* en galego coñecido.

Estes *datasets* foron obtidos a partir da base de datos de MeteoGalicia que conta con datos meteorolóxicos dos últimos 10 anos e textos descritivos bilingües escritos polos expertos meteorólogos pertencentes á institución. Cada un dos *datasets* foi revisado e anotado manualmente para incrementar a súa calidade textual e editado para limitar as descricións a aqueles fenómenos relacionados directamente cos datos numéricos.

30-<https://zenodo.org/record/8027725>

31- <https://opennmt.net/>

Recoñecedor automático da fala

Melloras e avances:

- Primeiro recoñecedor en galego totalmente libre e de código aberto
- Interface e API públicas
- Recoñecedor de propósito xeral
- Modelo adestrado cun número reducido de datos (audio-texto aliñado)

Para a elaboración do recoñecedor, empregouse o modelo wav2vec2-large-xlsr-53³², un modelo preadestrado con datos de 53 linguas, para o recoñecemento

automático da fala. Este modelo pódese adaptar a unha nova lingua mediante técnicas de *transfer-learning*, usando *data-sets* relativamente pequenos, ofrecendo resultados moi prometedores.

O punto de partida para este primeiro prototipo foi o modelo wav2vec2-large-xlsr-53-gl³³, obtido mediante axuste fino (*fine-tuning*) do modelo preadestrado usando o corpus OpenSLR77³⁴. Co obxectivo de mellorar o rendemento deste sistema *baseline*, probouse a integrar distintos modelos de linguaxe (LMs) no modelo wav2vec2.

Figura 2. Vista da interface do ASR.



Fonte: Elaboración propia.

32- <https://ai.facebook.com/blog/wav2vec-20-learning-the-structure-of-speech-from-raw-audio/>

33- <https://huggingface.co/diego-fustes/wav2vec2-large-xlsr-gl>

34- <http://www.openslr.org/77>

Para avaliar os distintos sistemas resultantes, fixéronse probas empregando dous conxuntos de datos distintos: subconxuntos de test dos corpus OpenSLR77³⁵ e Common Voice Corpus 7.0. Estes dous corpus presentan claras diferenzas en canto á calidade, control e formato das gravacións.

As medidas empregadas para a avaliación dos modelos foron: a WER (Word Error Rate) e o RTF (Real Time Factor). Coas diferentes configuracións probadas, a WER sitúase entre o 9% (base-line) e o 7% (mellor das configuracións con LM) para o subconxunto de test de OpenSLR, mentres que se reduce, respectivamente, do 22% ao 15% para o subconxunto de CommonVoice. En canto ao RTF, aínda que aumenta cando se integra un LM, este segue estando moi por debaixo de 1 en todos os casos.

As funcionalidades do demostrador ASR pódense probar a través da interface creada a tal efecto, a cal permite tanto gravar voz como subir un ficheiro de audio, devolvendo nos dous casos a transcripción textual correspondente (Figura 2).

Sistema de conversión texto-voz

Melloras e avances:

- Primeiro modelo neuronal libre
- Interface e API públicas
- Corpus de voz libre
- Voz sintética de alta calidade e naturalidade

Na súa versión actual, o demostrador TTS integra dous modelos de voz de distintas características:

- Celtia. Modelo obtido a partir dun subconxunto do corpus Nos_Celtia-GL, gravado por unha actriz de dobraxe profesional, desenvolto no proxecto Nós. Concretamente, para o adestramento do modelo empregáronse 13.000 frases, que se corresponden con 15,5 horas de voz. O modelo Celtia foi adestrado a partir de grafemas e non inclúe módulo de preprocesamento.
- Sabela. Modelo obtido a partir dun corpus de 10.000 frases gravadas por unha locutora de radio profesional. Este corpus, incluído no *dataset* CRPIH UVi-go-GL-Voces³⁶, suma un total de 14 horas e foi desenvolto nun marco de colaboración entre o Centro Ramón Piñeiro para a Investigación en Humanidades (CRPIH) e a Universidade de Vigo. O modelo Sabela foi adestrado a partir de transcripcións fonéticas e inclúe un módulo de preprocesamento baseado no *front-end* de Cotovía³⁷.

Ambos os dous modelos están baseados na arquitectura VITS (*Variational Inference with adversarial learning for end-to-end Text-to-Speech*)³⁸ e foron adestrados desde cero empregando a librería Coqui-TTS³⁹.

Co obxectivo de determinar que modelos TTS ofrecían os mellores resultados, exploráronse e avaliáronse distintas ar-

35- <https://commonvoice.mozilla.org/gl/datasets>

36- <https://zenodo.org/record/8027725>

37- <https://sourceforge.net/projects/cotovia/>

38- <https://arxiv.org/abs/2106.06103>

39- <https://github.com/coqui-ai/TTS>

quitecturas e configuracións. Durante estas probas experimentais empregouse o corpus de Sabela, xa que o corpus Nos_Celtia-GL aínda non estaba dispoñible. Entre as probas realizadas, destacamos as seguintes:

- Avaliáronse dúas opcións para os datos lingüísticos:
 - Adestramento a partir de grafemas (texto plano).
 - Adestramento a partir de fonemas (transcrición fonética de Cotovía):
 - Tra2 (transcrición básica)

- Tra3 (transcrición avanzada con marcas de acento)
- Experimentouse con tres arquitecturas distintas:
 - GlowTTS⁴⁰ + HiFiGAN⁴¹
 - VITS
 - FastSpeech2⁴²
- Por último, analizouse tamén a integración dun módulo de preprocesamento (front-end de Cotovía):
 - Normalización do texto (extensión de números, datas, siglas, abreviaturas etc.).
 - Transcrición fonética.

Figura 3. Vista da interface do TTS.



Fonte: Elaboración propia.

40- <https://paperswithcode.com/method/glow-tts>

41- <https://paperswithcode.com/method/hifi-gan>

42- <https://paperswithcode.com/method/fast-speech-2>

Tradutor neuronal español-galego, inglés galego

As principais melloras e avances ao respecto do estado da arte son as seguintes:

- Primeiro tradutor neuronal feito na Galiza
- Libre e de código aberto
- Calidade similar a Google translator
- Neste momento xa é potencialmente útil para terceiros

Foi desenvolvido un prototipo de probas dun tradutor neuronal para os pares ES-GL e EN-GL, nas dúas direccións, empregando para isto corpus de traducións de inglés-portugués e español-portugués, EN-PT e ES-PT como materia

prima para ser transformada nos pares ES-GL e EN-GL. Para a realización dos experimentos foron empregadas dúas arquitecturas de código aberto: OpenNMT (principalmente) e Fairseq (de forma secundaria). No que respecta aos nosos sistemas utilizados en OpenNMT, estas foron as características (Ortega et al. 2022):

- Para a arquitectura LSTM: dúas capas ocultas, 500 unidades LSTM ocultas por capa, 13 *epochs*, *batch size* de 64. Alteramos tamén os parámetros por defecto do *learning rate* para 100.000 *training steps* e 10.000 *validation steps*.
- No caso do *Transformer*, foi configurado con parámetros de formación por defecto: 6 capas para codificación

Figura 3. Vista da interface do TTS.



Fonte: Elaboración propia.

e descodificación, 8 cabezas de atención e *batch size* de 4096 *tokens*. Tamén modificamos os parámetros do *learning rate* para os mesmos valores que a configuración LSTM. Neste caso, utilizamos a *tokenización* das sub-palabras, realizada con BPE.

En novembro de 2022, obtivéronse uns resultados de 70,2 BLEU para o par español-galego nunha arquitectura *transformer* e só con corpus auténticos ES-GL, igualando ao estado da arte en RBMT español-galego cun BLEU por enriba de 60 (Bayón e Sánchez Gijón, 2019). En relación ao inglés-galego, os resultados da BLEU, resultantes da utilización dunha arquitectura tamén *transformer*, foron 35, o que aumenta significativamente os resultados anteriores e moito máis os do Carvalho SMT, un sistema estatístico previo, que presentaba unha BLEU de 15,59.

Tradutor neuronal español-galego, inglés galego

As principais melloras e avances ao respecto do estado da arte son as seguintes:

- Primeiro tradutor neuronal feito na Galiza
- Libre e de código aberto
- Calidade similar a Google translator
- Neste momento xa é potencialmente útil para terceiros

Foi desenvolvido un prototipo de probas dun tradutor neuronal para os pares ES-GL e EN-GL, nas dúas direccións, empregando para isto corpus de traducións de inglés-portugués e español-portugués, EN-PT e ES-PT como materia prima para

ser transformada nos pares ES-GL e EN-GL. Para a realización dos experimentos foron empregadas dúas arquitecturas de código aberto: OpenNMT (principalmente) e Fairseq (de forma secundaria). No que respecta aos nosos sistemas utilizados en OpenNMT, estas foron as características (Ortega et al. 2022):

- Para a arquitectura LSTM: dúas capas ocultas, 500 unidades LSTM ocultas por capa, 13 *epochs*, *batch size* de 64. Alteramos tamén os parámetros por defecto do *learning rate* para 100.000 *training steps* e 10.000 *validation steps*.
- No caso do *Transformer*, foi configurado con parámetros de formación por defecto: 6 capas para codificación e descodificación, 8 cabezas de atención e *batch size* de 4096 *tokens*. Tamén modificamos os parámetros do *learning rate* para os mesmos valores que a configuración LSTM. Neste caso, utilizamos a *tokenización* das sub-palabras, realizada con BPE.

En novembro de 2022, obtivéronse uns resultados de 70,2 BLEU para o par español-galego nunha arquitectura *transformer* e só con corpus auténticos ES-GL, igualando ao estado da arte en RBMT español-galego cun BLEU por enriba de 60 (Bayón e Sánchez Gijón, 2019). En relación ao inglés-galego, os resultados da BLEU, resultantes da utilización dunha arquitectura tamén *transformer*, foron 35, o que aumenta significativamente os resultados anteriores e moito máis os do Carvalho SMT, un sistema estatístico previo, que presentaba unha BLEU de 15,59.

Figura 4. Vista da interface do tradutor.



Fonte: Elaboración propia.

Xerador automático de descrições textuais de datos en tempo real

As principais melloras e avances deste sistema de xeración ao respecto do estado da arte son as seguintes:

- Primeiro modelo de xeración data-to-text con tecnoloxía neuronal en galego e castelán.
- Primeira proba de concepto exitosa de fine-tuning dun modelo neuronal text-to-text libre nos dous idiomas.
- Primeiro corpus data-to-text para galego e castelán.

Para a elaboración deste prototipo, en

primeiro lugar realizouse unha revisión do estado da arte sobre os modelos que se empregan actualmente para a xeración de textos, concluíndo que a tendencia actual en xeración *text-to-text* é a de empregar grandes modelos da linguaxe pre-adestrados con millóns de parámetros e aplicar un proceso de *fine-tuning* sobre estes para ensinalles unha nova tarefa. Por este motivo, decidiuse aplicar esta estratexia tamén para o caso do prototipo de xerador automático de textos a partir de táboas de datos (data-to-text).

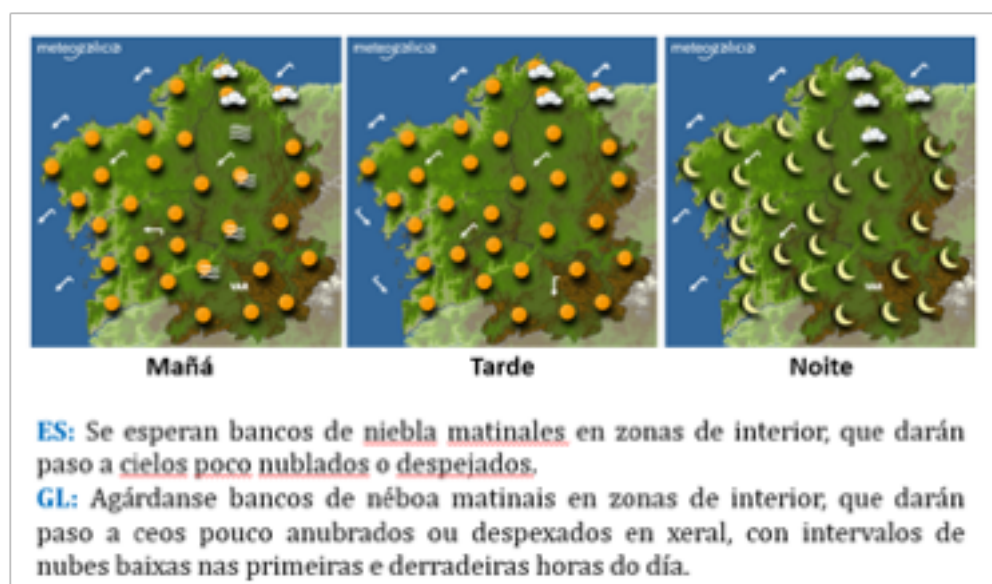
Para aplicar o procedemento de fine-tuning, fíxose unha selección de modelos que tivo en conta os grandes modelos de linguaxe que inclúen nos seus idiomas

adestrados o castelán e o galego. Os modelos seleccionados para as probas foron os seguintes: MT5, FLAN-T5 e MBART.

Neste primeiro prototipo, impleméntouse un modelo baseado en MT5 adaptando os datos de entrada ao formato requirido, xa que MT5 é un modelo *text-*

to-text. Os resultados obtidos neste prototipo de *fine-tuning* son textos similares aos de referencia en canto aos termos empregados nas frases e á estrutura destas. Este prototipo é o primeiro xerador *data-to-text* de tipo *end-to-end* coñecido en idioma galego.

Figura 5. Exemplo de uso do modelo de xeración *data-to-text*.



Fonte: Elaboración propia.

4.4 A IMPORTANCIA DOS DATOS

Para acadar o obxectivo de situar o galego no ámbito dixital en igualdade de condicións con outras linguas, é requisito indispensable para o proxecto Nós dispor dunha gran cantidade e variedade de recursos lingüísticos (escritos e orais) de alta calidade (corpus anotados de referencia, corpus a escala web, corpus espe-

cializados por tarefas e dominios, corpus paralelos, bases de coñecemento, dicionarios etc.) que permitan desenvolver tecnoloxías de vangarda para o galego. Estes recursos están a ser compilados a partir de recursos existentes e/ou creados desde cero.

A compilación de recursos existentes, tanto corpus como dicionarios ou lexicóns, é de particular interese para o

proxecto Nós e en ocasións require dunha estreita cooperación coas entidades que dispoñen destes recursos. Tanto a compilación de recursos lingüísticos como o seu tratamento, serán realizados tendo en conta diversos aspectos éticos, que terán impacto directo no funcionamento dos sistemas producidos.

No primeiro ano de proxecto contactáronse máis de 40 entidades e foron asi-

nados 17 acordos de cesión entre os que se atopan: Xunta de Galicia, Parlamento de Galicia, Consello da Cultura Galega, Real Academia Galega, Centro Ramón Piñeiro, CRTVG, Deputación da Coruña, Deputación de Lugo, Editorial Galaxia, Editorial Hugin e Munin, Editorial Laiovento, Praza Pública, Sermos Galiza, Agrocomunicación, Pasa na costa, Atlántica de información, IGADI e Asociación Irimia.

Gráfico 19. Distribución da tipoloxía de entidades cedentes de recursos.



Fonte: Elaboración propia.

4.5 PRÓXIMOS PASOS

Como foi adiantado no inicio deste texto, na actualidade o proxecto Nós está sendo financiado a través do PERTE Nova Economía da Lingua. Neste marco, o

proxecto conta con fondos ata finais do ano 2025. En todo caso, está en elaboración un novo convenio coa Xunta de Galicia, que debería entrar en vigor no 2024, aínda que a duración, contía e obxectivos están por determinar.

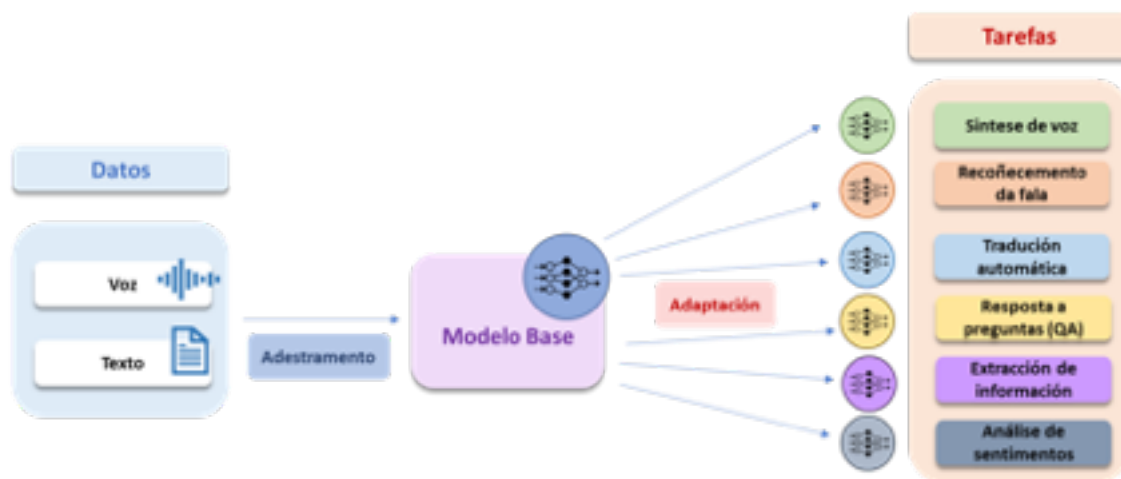
Os obxectivos para o proxecto Nós dentro do PERTE son dobres. Por un lado, o proxecto comprometeuse a:

- Aumentar a cantidade e calidade dos corpus actuais, garantindo a heteroxeidade e representatividade dos mesmos.
- Xerar datos anotados de calidade que permitan adestrar e avaliar os modelos de voz e texto a desenvolver.
- Identificar casos de uso de alto impacto coa participación da industria, fomentando a transferencia tecnolóxica.

Por outra banda, o proxecto está colaborado cos proxectos AINA (Cataluña),

GAITU (País Vasco) e Vives (Valencia) na creación de corpus nas diferentes linguas, así como no adestramento de grandes modelos multilingües cunha visión ibérica. O adestramento destes grandes modelos base (tamén chamados fundacionais) é un requisito indispensable para a creación de novas tecnoloxías lingüísticas en múltiples tarefas. Hoxe en día, a maior parte das tarefas lingüísticas, tal e como se mostra na figura 7, poden abordarse mediante a adaptación e o axuste de parámetros dos modelos base ou fundacionais. Esta adaptación ou fine-tuning do modelo base a tarefas específicas lévase a cabo con datos anotados de calidade de cada unha das tarefas

Figura 6. Arquitectura dos grandes modelos base ou fundacionais e a súa adaptación a tarefas lingüísticas



Fonte: Elaboración propia.

Dadas as particularidades do proxecto, todos os desenvolvementos estanse a levar a cabo no marco do PLN fiable

(trustworthy), que promove especialmente a UE, e que contempla os seguintes aspectos: acción e supervisión humanas,

solidez técnica e seguridade, así como xestión da privacidade e dos datos. Seguindo ese mesmo marco, procuraremos compilar recursos e adestrar modelos coa máxima transparencia posible con base na trazabilidade e na explicabilidade. Asemade, buscaremos maximizar a diversidade, minimizando a discriminación e fortalecendo a equidade e o benestar social e ambiental, e sempre rendendo contas á administración e á sociedade. Deste xeito, e con isto concluímos, os produtos

derivados dos proxectos (corpus, modelos e demostradores) encóntranse en xeral a libre disposición de terceiros en webs e repositorios abertos:

- En GitHub: <https://github.com/proxectonos>
- En Zenodo: <https://zenodo.org/communities/proxecto-nos>
- En HuggingFace: <https://huggingface.co/proxectonos/>

5 EPÍLOGO

As tecnoloxías lingüísticas (TL) xogan un papel crítico nunha sociedade interconectada e multicultural, permitindo ás computadoras comprender, interpretar e xerar linguaxe humana de maneira moi eficiente. As TL conforman un campo interdisciplinario da intelixencia artificial e a lingüística. Engloba diversos campos de coñecemento, como o recoñecemento e síntese de voz, clasificación e análise de texto e a corrección gramatical, entre outras.

A súa capacidade para interpretar e xerar linguaxe natural está a abrir novas oportunidades e a transformar a maneira de traballar, aprender, comunicarse e vivir. Entre os principais ámbitos no que ten un impacto significativo destacan a formación e saúde personalizadas, o soporte técnico, a análise de información valiosa, a automatización de procesos e a análise de sentimentos.

A industria das tecnoloxías da linguaxe se conforma en torno a empresas de software, empresas de servizos lingüísticos e institucións académicas, principalmente. O conxunto de solucións que proporcionan abranguen un amplo conxunto de técnicas e ferramentas, que permiten facer tradución automática, interacción conversacional, procesamento da linguaxe natural e recoñecemento e síntese de voz.

O principal beneficio das solucións das TL para as empresas é utilizar a automatización para reducir custos e mellorar as operacións comerciais para maximizar a produtividade e a rendibilidade. Neste sentido, as empresas que levan varios

anos empregando solucións de procesamento da linguaxe natural argumentan que esta tecnoloxía mellora a eficiencia, xera vantaxes competitivas, reduce o tempo de produción e riscos, aforra custos e permite acceder a novas oportunidades de negocio.

Estamos ante unha actividade cun potencial de negocio extraordinario, segundo se desprende dos principais estudos feitos a nivel mundial que prognostican un crecemento exponencial. Na actualidade o mercado a nivel global oscila entre os 15.000 e os 20.000 millóns, dependendo da metodoloxía utilizada, e se espera un incremento, no escenario máis optimista, de ata 160.000 millóns a finais da década. O marxe da proxección mais ou menos fiable, estamos ante un sector que se expandirá e no que é preciso posicionarse, tanto a nivel nacional coma rexional. A cota de mercado esperada para Europa a nivel global se estima no 20%, onde España ocupa o quinto posto cun 9% do mercado europeo, sendo unha fortaleza o feito de que coexistan unha ampla variedade de linguas que precisan de desenvolvemento para a aplicación das tecnoloxías lingüísticas.

O mercado está moi concentrado, e unhas poucas empresas repártense entre o 55% e o 60% do mercado mundial, quedando a restante cota en mans de outros moitos operadores. Os grandes líderes a nivel mundial son grandes corporacións, pero tamén hai que destacar a presenza de moitos operadores rexionais e locais que poden contribuir a crear actividades de alto valor engadido e emprego moi

cualificado. Despois de depurar os datos de diversas fontes o número de axentes atopados no sector das tecnoloxías da linguaxe en España é de 214, dos cales, 115 son empresas. A maior parte se sitúan en Madrid, Cataluña, País Vasco e Valencia, mentres que Galicia ocupa a quinta posición.

Con datos da Enquisa do uso das TIC e o comercio electrónico nas empresas, en 2022 a nivel español o uso máis frecuente tivo que ver coa de automatización de fluxos de traballo ou a axuda na toma de decisións. Galicia destacou pola IA que converte a linguaxe falada en formato lexible por unha máquina. En termos xerais, segundo se desprende de dita enquisa, as empresas galegas superan a media nacional, tanto no emprego da IA para a análise da linguaxe escrita coma no emprego da IA para converter a linguaxe falada en formato lexible por unha máquina. Neste posicionamento pode estar a influir o impulso das tecnoloxías da linguaxe aplicadas ó galego.

O feito de ter unha lingua propia pode erixirse nunha vantaxe competitiva que permita impulsar as tecnoloxías da linguaxe natural en Galicia. A singularidade lingüística dunha comunidade presenta desafíos específicos para os desenvolvedores de tecnoloxías da linguaxe. Ter unha lingua propia abre a porta á creación de solucións específicas adaptadas ás necesidades lingüísticas locais, permitindo unha mellor precisión do procesamento da linguaxe natural. Ademais, o desenvolvemento de aplicacións específicas favorece a adopción de tecnoloxías e fomenta a diversidade e inclusión na era dixital. De aí que o galego sexa unha acti-

vo estratéxico que está servindo para potenciar o desenvolvemento e a innovación no ámbito das tecnoloxías da linguaxe.

Como activo estratéxico, se pode estimar o valor económico da lingua utilizando metodoloxías que inclúen análises econométricos, métodos input-output, e técnicas baseadas en indicadores de emprego, cálculos limitados pola dispoñibilidade e desagregación de datos estatísticos. Para a dita estimación, identifícanse varios sectores vinculados ao idioma, como as industrias da cultura e os medios de comunicación, o ensino de idiomas, a tradución e interpretación, e as actividades de promoción lingüística e cultural das administracións públicas. Pero tamén é preciso incluír o emerxente sector das tecnoloxías da linguaxe.

O valor total da produción dos sectores vinculados ao galego en 2019 foi de 4.118 millóns de euros. Este valor componse dun 44,7% de impacto directo (1.841,7 millóns de euros), un 9,7% de impacto indirecto (397,5 millóns de euros), e un 45,6% de efecto inducido (1.878,6 millóns de euros). Esta produción vinculada ao impacto do galego representa o 3,4% do total da economía galega. En termos de emprego, o impacto total dos sectores do galego foi de 49.893 postos de traballo equivalentes a tempo completo (4,7% do total da economía galega). Isto inclúe un 55,5% de impacto directo (27.702 empregos), un 8,5% de impacto indirecto (4.242 empregos), e un 36,0% de efecto inducido (17.949 empregos).

Pero manter ou incrementar dito valor dependerá da estratexia da rexión relativa ás tecnoloxías da linguaxe aplicadas ó galego. Nun mundo máis dixitalizado

e automatizado, o desenvolvemento das tecnoloxías da linguaxe precisa da integración do galego para asegurar que non quede ó marxe da revolución tecnolóxica. Isto non só contribúe á inclusión dixital, senón que tamén axuda a preservar a riqueza lingüística e cultural, incrementando o valor económico da propia lingua.

De aí a importancia que ten dispor de proxectos propios que axuden a impulsar o sector das tecnoloxías da linguaxe en Galicia. Nos últimos anos realizáronse importantes esforzos para dotar o galego de ferramentas tecnolóxicas e poñelas á disposición do conxunto da cidadanía. Porén, a situación do galego no contexto dixital non é equiparable ó doutras linguas como o inglés ou o castelán, polo que precisa dun impulso decidido que a sitúe como lingua de uso convencional nos recursos tecnolóxicos de uso común pola sociedade.

Para conseguir que o galego dea un paso decidido cara adiante no eido das tecnoloxías intelixentes, o “proxecto Nós” naceu coa finalidade de configurarse como o alicerce do ecosistema dixital que permitirá xerar recursos dixitais e lingüísticos transversais que proporcionarán ao galego a posibilidade de se incorporar a plataformas tecnolóxicas, das cales, até agora, estivera excluído.

A finalidade que persegue o proxecto Nós co seu desenvolvemento é acadar unha ampla variedade de recursos lingüísticos en galego de alta calidade e que serán publicados en formato dixital, en xeral baixo licenzas libres, e a través de ferramentas con códigos abertos para ser empregadas en calquera tipo de soporte dixital polas diferentes institucións públi-

cas e, mesmo, polos axentes que conforman o tecido económico empresarial e, asemade, pola sociedade en xeral.

En 2023, o proxecto Nós incorporouse ao PERTE Nova Economía da Lingua, formando parte dun proxecto conxunto cos outros proxectos análogos no ámbito do catalán e éuscaro. Este fito situou a Galicia nunha posición privilexiada, debido a captación de recursos que garanten o impulso do sector en Galicia ata 2025.

O proxecto NOS ten oito liñas de traballo que van dende a síntese de voz, o recoñecemento da fala, á tradución automática, a xeración de textos ou mesmo, ós sistemas de diálogo. No primeiro ano desenvóléronse seis corpus lingüísticos, tres demostradores e un prototipo. O desenvolvemento das tecnoloxías lingüísticas precisa de recursos tanto orais como escritos, de xeito que no primeiro ano se construíron tres corpus de recoñecemento da fala, un de síntese de voz e outro corpus galego-español de tradución automática. Tamén se elaborou un recoñecedor automático da fala e un corpus conversacional. Outros recursos moi destacables foron o tradutor neuronal de galego a español e inglés. Tamén se desenvolveu un xerador automático de descrições textuais en tempo real.

Os seguintes pasos, na medida que conta con fondos ata o 2025, consistirán en aumentar o corpus actuais, xerar datos de calidade para adestrar modelos de voz e texto, e identificar casos de uso de alto impacto. Tamén, no marco da colaboración cos proxectos do catalán e o eúskaro, se pretenden adestrar grandes modelos base para a creación de novas tecnoloxías en múltiples tarefas.

En Galicia, o sector das tecnoloxías lingüísticas eríxese como un campo de extraordinario potencial, impulsado pola riqueza de contar cun idioma propio. O inicio deste camiño xa foi trazado, atopándose no ecosistema galego unha converxencia de actores crave, no que conviven empresas de tecnoloxías da información e a comunicación, empresas de tecnoloxía lingüísticas, institucións académicas, a administración pública e os centros lingüísticos especializados. Esta amálgama de recursos proporciona unha base sólida para impulsar o incipiente sector espe-

cializado en tecnoloxías lingüísticas. Con todo, para materializar este potencial, é imperativo formar e nutrir de talento especializado, fortalecer a colaboración coa industria local, apostar por proxectos transversais que aproveiten sinerxias, e continuar impulsando o sector mediante investimentos financeiros e o acceso a recursos lingüísticos desde a administración. Este enfoque integral non só consolidaría a posición de Galicia no mapa das tecnoloxías lingüísticas, senón que tamén contribuiría ao desenvolvemento sostible e á proxección internacional da rexión.



REFERENCIAS BIBLIOGRÁFICAS

ACAP e OESIA (2018). Estudio de Caracterización del Sector de las Tecnologías del Lenguaje en España, Madrid. Disponible en: <https://www.ontsi.es/es/publicaciones/Estudio-de-caracterizacion-del-sector-de-tecnologias-del-lenguaje-en-Espana>

ACAP, OESIA (2018). Estudio de Caracterización del Sector de las Tecnologías del Lenguaje en España, Madrid. Disponible en: <https://www.ontsi.es/es/publicaciones/Estudio-de-caracterizacion-del-sector-de-tecnologias-del-lenguaje-en-Espana>

Bayón, María Do Campo, and Pilar Sánchez-Gijón. 2019. "Evaluating machine translation in a low-resource language combination: Spanish-Galician". In Proceedings of Machine Translation Summit XVII: Translator, Project and User Tracks, pages 30–35, Dublin, Ireland. European Association for Machine Translation.

Bughin, J., Seong, J., Manyika, J., Chui, M., e Joshi, R., (2018). Notes from the AI frontier: Modeling the impact of AI on the world economy. McKinsey Global Institute. Disponible en: <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy>

Capelouto, J. D., (23 de xaneiro do 2023). Microsoft announces new multi-billion dollar investment in ChatGPT creator OpenAI. Semafor. Disponible en: <https://www.semafor.com/article/01/23/2023/microsoft-announces-new-multi-billion-dollar-investment-in-openai>

de-Dios-Flores, Iria, Carmen Magariños, Adina Ioana Vladu, John E. Ortega, José Ramon Pichel, Marcos García, Pablo Gamallo, Elisa Fernández Rei, Alberto Bugarín-Diz, Manuel González González, Senén Barro, and Xosé Luis Regueira. 2022. "The Nós Project: Opening routes for the Galician language in the field of language technologies". In Proceedings of the Workshop Towards Digital Language Equality within the 13th Language Resources and Evaluation Conference, pages 52–61, Marseille, France. European Language Resources Association.

Domo (2020). Data never sleeps 8. Disponible en: <https://www.domo.com/learn/infographic/data-never-sleeps-8>

Fortune Business Insights (2022). Natural language processing (NLP).

GAIASTECH (s.f.). Mapa de capacidades digitales de Galicia. Disponible en: <https://gaiastech.xunta.gal/es/mapa-de-capacidades-digitales-de-galicia>

INE (s.f.). Enquisa do uso de TIC e Comercio Electrónico nas Empresas.

Meertens, L., Choukri, K., Aguzzi, S e Vasiljevs, A. (2019). Final study report on CEF automated translation value proposition in the context of the European LT market/ecosystem – , Publications Office of the European Union. Dispoñíbe en :<https://data.europa.eu/doi/10.2759/142151>

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, 58(3), 103434.

Ministerio de Asuntos Económicos y Transformación Digital (s.f.). Catálogo de entidades y soluciones. Disponible en: <https://plantl.mineco.gob.es/tecnologias-lenguaje/catalogo-TL/Paginas/catalogo-TL.aspx>

Mishra, A. N., & Pani, A. K. (2021). Business value appropriation roadmap for artificial intelligence. *VINE Journal of Information and Knowledge Management Systems*, 51(3), 353-368.

Mordor Intelligence (2020). Global natural language processing (NLP).

Musto, S., e Maguire, C. (2021). The Total Economic Impact™ Of IBM Watson Natural Language Processing (NLP) Solutions. Forrester Research. Disponible en: <https://www.ibm.com/downloads/cas/XMRMP7XK>

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Available at SSRN 4375283.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Available at SSRN 4375283.

Observatorio para el Análisis y el Desarrollo Económico de Internet (ADEI) (2015). Digitalización y productividad. Disponible en <http://www.observatorioadei.es/publicaciones/NotaTecnica-Digitalizacion-y-productividad.pdf>

Observatorio para el Análisis y el Desarrollo Económico de Internet (ADEI) (2014). Digitalización y desempeño empresarial. Disponible en http://observatorioadei.es/publicaciones/digitalizacion_y_desempeno_empresarial.pdf

Ortega, John, Iria de-Dios-Flores, José Ramon Pichel, and Pablo Gamallo, (2022), "A neural machine translation system for Galician from transliterated Portuguese text". In Proceedings of the Annual Conference of the Spanish Association for Natural Language Processing: Projects and Demonstrations (SEPLN-PD 2022), 21-23 September, A Coruña, CEUR Workshop Proceedings, Vol. 3024, pp. 92-95. ISSN: 1613-0073.

PERTE Nueva economía de la lengua, (2022). Disponible en: <https://planderecuperacion.gob.es/como-acceder-a-los-fondos/pertes/perte-nueva-economia-de-la-lengua>

Rehm, G., & Hegele, S. (2018, May). Language Technology for Multilingual Europe: An Analysis of a Large-Scale Survey regarding Challenges, Demands, Gaps and Needs. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).

Rehm, G., Marheinecke, K., Hegele, S., Piperidis, S., Bontcheva, K., Hajič, J., ... & Yvon, F. (2020). The European language technology landscape in 2020: Language-centric and human-centric AI for cross-cultural communication in multilingual Europe. arXiv preprint arXiv:2003.13833.

Reinsel, D., Gantz, J., e Rydning, J., (2018). The digitization of the world - From Edge to Core. USA. Disponible en: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

Varone, M., (2022). The 2023 expert NLP survey report trends driving NLP investment and innovation. Expert.ai. Disponible en <https://www.expert.ai/wp-content/uploads/2022/12/The-2023-Expert-NLP-Survey-Report-Trends-driving-NLP-Investment-and-Innovation.pdf>

Vasiljevs, A., Choukri, K., Meertens, L., & Aguzzi, S. (2019). Final study report on CEF Automated Translation value proposition in the context of the European LT market/ecosystem. DOI, 10, 142151.

Vasiljevs, A., Skadiņa, I., Sāmīte, I., Kauliņš, K., Ajausks, Ē., Meļņika, J., & Bērziņš, A. (2020, May). The Competitiveness Analysis of the European Language Technology Market. In Proceedings of the Twelfth Language Resources and Evaluation Conference (pp. 3381-3389).

Vladu, Adina Ioana, Iria de-Dios-Flores, Carmen Magariños, Adina Ioana Vladu, John E. Ortega, José Ramon Pichel, Marcos García, Pablo Gamallo, Elisa Fernández Rei, Alberto Bugarín-Diz, Manuel González González, Senén Barro, and Xosé Luis Regueira (2022). "Proxecto Nós: Artificial Intelligence at the Service of the Galician Language". In Proceedings of the Annual Conference of the Spanish Association for Natural Language Processing: Projects and Demonstrations (SEPLN-PD 2022), co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2022). A Coruña, 21-23 de setembro de 2022.

Wamba-Taguimdje, S. L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects. *Business Process Management Journal*, 26(7), 1893-1924.

Way, A., Rehm, G., Dunne, J., Giagkou, m., Gomez-Perez, J. M., Hajič, J., Hegele, s., Kaltenböck, M., Lynn, T., Marheinecke, K., Resende, N., Skadiņa, I., Skowron, M., Vojtěchová, T., Grützner-Zahn, A. (2022). Report on the state of Language Technology in 2030. D2.18. European Language Equality (ELE). Disponible en https://european-language-equality.eu/wp-content/uploads/2022/05/ELE__Deliverable_D2_18_Report_on_State_of_LT_in_2030_.pdf

Yadav, S., Rake, B., kumar, V., (2022). Natural language processing (NLP) market. Allied Market Research. United States.

Documentos do Foro Económico de Galicia

Un dos obxectivos fundamentais do Foro é poñer á disposición da Sociedade unha colección de documentos onde se conxuga a análise da situación actual con propostas de cambio e mellora. Os documentos pretenden contribuir a enriquecer o debate público en materia económica en Galicia. A responsabilidade de cada documento é das persoas que o asinan. O Foro só asume a responsabilidade sobre os documentos consensuados e asinados polo colectivo. O editor dos documentos é o membro nato do Foro, José Francisco Armesto Pina.

Documentos publicados

40/2023-

Potencial de Galicia no sector das tecnoloxías lingüísticas.

39/2023-

O Sector Cárnico en Galicia. Propostas para o futuro.

38/2023-

O desenvolvemento da Intelixencia Artificial en Galicia: están preparadas as empresas galegas?.

37/2022-

O futuro da Agricultura Ecolóxica galega no marco do Pacto Verde Europeo.

36/2022-

Unha estratexia en IA para a empresa galega.

35/2022-

A reforma que a Administración necesita: propostas para o SXXI.

34/2022-

A reforma do sistema de financiamento: unha perspectiva dende Galicia.

33/2021-

A economía social como modelo de desenvolvemento para Galicia.

32/2021-

Fuga de talento en Galicia ¿mito o realidade?.

31/2020-

O monte galego do Século XXI.

30/2020-

El sector Fintech: propuesta para desarrollar un HUB en Galicia.

29/2020-

Innovación y Tecnologías Inteligentes.

28/2019-

Infraestructuras Viarias: Inversión y Tarificación.

27/2019 -

Planificación e proxección do sector vitivinícola de Galicia:

Realidade, desafíos e retos.

26/2018 -

Os retos para a Galicia do Século XXI.

No sistema europeo de transporte marítimo-terrestre.

25/2018 -

El impacto de la reestructuración bancaria sobre la concentración y la exclusión financiera en España y Galicia.

Documentos descargables en:

<http://www.foroeconomicodegalicia.es>

- 24/2018 -
Turismo Rural: diagnose e propostas.
- 23/2018 -
Diagnóstico actual e propostas de futuro para a dixitalización.
en Galicia.
- 22/2017 -
Unha estratexia para a Galicia rural do século XXI.
Diagnóstico e propostas para o debate.
- 21/2017 -
Unha estratexia industrial para Galicia.
- 20/2017 -
Desafíos dunha sociedade avellentada e en declive: desequilibrios
territoriais e prestación de servizos.
- 19/2017 -
Galicia diante da reforma do sistema de financiamento autonómico.
- 18/2016 -
Novas demandas para o rural galego.
- 17/2016 -
A I+D+i Empresarial en Galicia: Dez mensaxes e unha proposta.
- 16/2016 -
Galicia, a perspectiva internacional.
- 15/2016 -
La formación profesional dual en España.
- 14/2016 -
O capital risco e outras fontes alternativas de financiamento das empresas
galegas.
- 13/2015 -
O Complexo Lácteo Galego nun mercado liberalizado. Extratexias e políticas
ante a desaparición do sistema de cotas na UE.
- 12/2015 -
Sobre o risco de estancamento na economía Europea.
- 11/2015 -
Os cambios no modelo territorial: o seu impacto fiscal en Galicia.
- 10/2015 -
O sector forestal en Galicia: problemática actual e perspectivas futuras.
- 9/2014 -
Desequilibrio exterior, modelo competitivo y recuperación económica.
- 8/2014 -
Galicia: bases para una estrategia marca-país y líneas de trabajo para su
acción exterior.

7/2014 -

¿Qué futuro demográfico nos agarda e como pode incidir no noso benestar?.

6/2013 -

Internacionalización, políticas y crecimiento.

5/2013 -

Mejorando la Eficiencia de la Política de infraestructuras.

4/2013 -

O fondo de compensación interterritorial: Unha proposta desde Galicia.

3/2013 -

O sector alimentario en Galicia: desafíos e oportunidades. Cales deben ser as prioridades?.

2/2013 -

Análisis y propuestas para una estrategia de futuro para el sector naval en Galicia.

1/2013 -

Chaves para o aproveitamento marítimo -Portuario- Loxístico de Galicia.

A large, light blue decorative graphic on the left side of the page, featuring a pattern of dots and stars that fades out towards the right.

www.foroeconomicodegalicia.es

O Foro Económico de Galicia é unha plataforma de transferencia de coñecemento en materia económica dende as empresas e universidades galegas á sociedade e aos espazos de decisión pública. O Foro integra a profesores e investigadores, empresarios e directivos representativos dos diferentes sectores e áreas de Galicia, e xornalistas galegos de referencia.

SOCIOS



POTENCIAL DE GALICIA NO SECTOR DAS TECNOLOXÍAS LINGÜÍSTICAS

DOCUMENTO 40 / 2024

COORDINADORES:
SENÉN BARRO AMENEIRO (USC)
LUIS OTERO GONZÁLEZ (USC)