

CONVENIO DE COLABORACIÓN ENTRE A CONSELLERÍA DE CULTURA, EDUCACIÓN E UNIVERSIDADE, A AXENCIA PARA A MODERNIZACIÓN TECNOLÓXICA DE GALICIA E A UNIVERSIDADE DE SANTIAGO DE COMPOSTELA PARA A XERACIÓN DE RECURSOS LINGÜÍSTICOS E TECNOLÓXICOS EN LINGUA GALEGA

Proxecto Nós: O galego na sociedade e economía da intelixencia artificial.

Santiago de Compostela, 27 de xullo de 2021

REÚNENSE

Don Román Rodríguez González, conselleiro de Cultura, Educación e Universidade que actúa en virtude das facultades atribuídas no artigo 34 da Lei 1/1983, do 22 de febreiro, reguladora da Xunta e da súa Presidencia, modificada pola Lei 11/1988, do 20 de outubro e pola Lei 12/2007, do 27 de xullo, de acordo co seu nomeamento polo Decreto 112/2020, do 6 de setembro e no uso das atribucións que lle confire o Decreto 198/2020, do 20 de novembro, publicado no DOG número 242, do 1 de decembro, no que se establece a estrutura orgánica da Consellería de Cultura, Educación e Universidade e do acordo do Consello da Xunta de Galicia do 27 de maio de 1991, sobre convenios de cooperación con outros entes públicos e de colaboración con particulares, publicado pola Resolución da extinta Consellería de Economía e Facenda do 8 de abril de 1991 (DOG núm 82, do 30 de abril).

Dona Mar Pereira Álvarez, directora da Axencia para a Modernización Tecnolóxica de Galicia, (en diante Amtega) cargo para o que foi nomeada polo Consello da Xunta na súa reunión do 15 de marzo de 2012, en función das atribucións que lle confire o Decreto 252/2011, do 15 de decembro, polo que se crea a Axencia para a Modernización Tecnolóxica de Galicia e se aproban os seus estatutos, modificado polo Decreto 9/2021, do 21 de xaneiro, polo que se suprime o Centro Informático para a Xestión Tributaria, Económi-

co-Financeira e Contable (Cixtec) e se modifica o Decreto 252/2011, do 15 de decembro, polo que se crea a Amtega e se aproban os seus estatutos.

, reitor magnífico da Universidade de Santiago de Compostela, (en diante USC) en virtude do establecido nos seus estatutos e do nomeamento como reitor polo Decreto 57/2018, do 31 de maio, (DOG número 106 do 5 de xuño de 2018), que actúa en nome e representación da devandita entidade en virtude das facultades que lle confire o artigo 20.1 da Lei orgánica 6/2001, do 21 de decembro, de universidades.

Interveñen en función dos seus respectivos cargos e no exercicio das facultades que teñen conferidas e,

EXPOÑEN

Primeiro.- Que o artigo 27 do Decreto 198/2020, atribúelle á Consellería de Cultura, Educación e Universidade a través da Secretaría Xeral de Política Lingüística, entre outras, as funcións de promover, coordinar e desenvolver a política de investigación que favoreza a normalización lingüística nas súas diferentes manifestacións, especialmente no eido do léxico e a terminoloxía, así como, promover o diálogo, a cooperación e a colaboración con outras institucións competentes en materia de política lingüística e desenvolver medidas para o fomento do uso do idioma nos diferentes eidos sociais e económicos de Galicia, co fin de incrementar a dispoñibilidade de produtos e servizos en galego.

Segundo.- Que a Amtega é unha axencia pública autonómica encadrada nas entidades instrumentais do sector público autonómico reguladas no título III da Lei 16/2010, do 17 de decembro, de organización e funcionamento da Administración xeral e do sector público autonómico de Galicia. Está adscrita á Presidencia da Xunta de Galicia e conta con personalidade xurídica diferenciada respecto da Administración xeral da Comunidade Autónoma de Galicia, patrimonio e tesouraría propios e autonomía de xestión nos termos establecidos pola normativa vixente e ten como obxectivos básicos a definición, o desenvolvemento e a execución dos instrumentos da política da Xunta no ámbito das tecnoloxías da información e comunicacións, a innovación e o desenvolvemento tecnolóxico. Os seus estatutos atribúenlle, entre

outras, as seguintes competencias e funcións: a dirección e xestión de todas as actuacións da Xunta de Galicia en materia de tecnoloxías da información e as comunicacións e o deseño e liderado de iniciativas que contribúan a impulsar o desenvolvemento da Sociedade da Información en Galicia, garantindo os dereitos dos cidadáns e eliminando as barreiras que se opoñan á expansión e ao uso das novas tecnoloxías.

Terceiro.- Que a USC é unha entidade que leva a cabo actividades de investigación, docencia e desenvolvemento científico e tecnolóxico, que ten entre os seus fins o de procurar o seu enraizamento na sociedade galega que a sustenta e, en xeral, contribuír ao seu progreso social, económico, cultural e tecnolóxico, e é unha peza clave no proceso de normalización lingüística tanto polos diferentes campos culturais que abrangue como pola súa condición de referente social de prestixio o que a configura como un dos principais axentes xeradores de coñecemento de Galicia. A referida institución universitaria conta entre os seus centros de investigación co Instituto da Lingua Galega (ILG), referente de recoñecido prestixio na súa contribución ao desenvolvemento da investigación científica sobre a lingua galega. Dende o ano 1975, o centro de investigación está traballando na recollida de material dialectal no dominio lingüístico galego e, froito desta recollida, é un ficheiro lexicográfico que contén o vocabulario galego extraído de obras editadas ou inéditas que apareceron en forma distinta á de dicionario. Son tamén resultado do seu labor os materiais do Atlas Lingüístico Galego, do que se levan publicado varios volumes. O ficheiro lexicográfico como os cadernos do ALGa, están dispoñibles para os investigadores na sede do ILG, e son hoxe unha fonte de información imprescindible para calquera traballo sobre gramática e léxico, o centro dispón do Arquivo do Galego Oral, así como o Corpus Oral Informatizado da Lingua Galega (CORILGA), que permitirá investigar sobre a variación e o cambio no galego falado actualmente, e no que se incorporan ferramentas de tecnoloxías da fala.

A USC, a través do Centro Singular de Investigación en Tecnoloxías Intelixentes, (CiTIUS) dispón, así mesmo, de recursos materiais e persoais no ámbito tecnolóxico e de investigación que complementan as necesidades para o desenvolvemento e execución do Proxecto Nós. Dito centro conta cunha dilatada experiencia en investigación, desenvolvemento e innovación no ámbito da intelixencia artificial e das tecnoloxías da linguaxe natural, nomeadamente nas linguas ibéricas, como o galego, portugués e

castelán, e colabora con empresas do sector para dinamizar e poñer en valor emprendementos vinculados a proxectos relacionados coa Intelixencia Artificial (IA) e as Tecnoloxías Lingüísticas (TL)

Cuarto.- No ámbito normativo, a Lei 6/2013, do 13 de xuño, do sistema universitario de Galicia establece no seu artigo 111 que a transferencia de coñecemento é un dos mecanismos polos que o valor creado na universidade a través da investigación retorna á sociedade. Así pois, pola súa contribución ao desenvolvemento e ao crecemento económico, a transferencia constitúe unha actividade fundamental para a universidade.

Quinto.- Nos últimos anos, o impulso da presenza da lingua galega no desenvolvemento dixital xa quedou recollido na estratexia tecnolóxica da Xunta de Galicia, tanto na Axenda Dixital 2014.gal, como na posterior Axenda Dixital de Galicia 2020. Neste ámbito realizáronse importantes esforzos para dotar o galego de ferramentas tecnolóxicas (tradutores automáticos, dicionarios en liña, plataformas en liña para a difusión do coñecemento da lingua galega etc.) e poñelas á disposición do conxunto da cidadanía a través de iniciativas como os recursos lingüísticos do portal web <https://www.lingua.gal> da Secretaría Xeral de Política Lingüística da Xunta de Galicia e, outros, promovidos por diversas institucións e entidades públicas ou privadas.

Porén, a situación do galego no contexto dixital non é equiparable a outras 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 e para acadar que o galego dea un paso decidido cara adiante no eido das tecnoloxías intelixentes, o “Proxecto Nós” e a súa estratexia de intelixencia artificial ao servizo da lingua galega, pretende ter visibilidade a través da acción conxunta das partes deste convenio e, nesa dimensión colaborativa, aliñarse tamén cos principios inspiradores da nova Estratexia de Especialización Intelixente de Galicia RIS3 2021-2027, para lles dar respostas ás demandas e ás necesidades da sociedade.

Sexto.- Nós é un proxecto inédito, que leva a denominación dunha xeración de intelectuais galegos que foron quen de impulsar a cultura galega a través da súa lingua e, igual que fixera a intelectualidade galega do século XIX, o proxecto nace, na actualidade, coa virtualidade de configurarse como a pedra alicerce do ecosistema dixital que permitirá xerar recursos dixitais e lingüísticos transversais que lle proporcionaran ao galego a posibilidade de se incorporar a plataformas tecnolóxicas, das cales, ata o de

agora, estivera excluído: plataformas audiovisuais, videoxogos e noutros recursos tecnolóxicos e dispositivos dixitais presentes e futuros.

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 baixo licenzas libres e a través de ferramentas con códigos abertos para ser empregadas, libre e gratuitamente 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.

A infraestrutura de recursos tecnolóxicos e lingüísticos que se xeren co desenvolvemento do proxecto permitirá que o galego experimente un salto cualitativo dentro do universo dixital, fito histórico que favorecerá o proceso de transformación dixital da acción da Xunta de Galicia e doutras administracións públicas e garantirá que a cidadanía en xeral poida participar no mundo dixital sen renunciar a súa lingua ao mesmo nivel que un falante dunha lingua global, como pode ser o inglés ou o castelán. Os recursos resultantes do proxecto clasificaranse en tres categorías: recursos lingüísticos, tecnoloxías lingüísticas e demostradores. Nomeadamente desenvolverase un sintetizador de voz, que vai permitir a terceiros incluír nos seus produtos voces artificiais de alta calidade en galego, podendo escoller xénero e idade; un recoñecedor de fala, que posibilite a interacción persoa-máquina; un xerador de lingua natural que converta datos en oracións en galego; sistemas de diálogo xenérico, así como un elenco de respostas e preguntas, tanto a través de texto como de voz; un tradutor automático a outras linguas e viceversa; e un corrector lingüístico automático, tanto ortográfico como léxico e gramatical.

Sétimo.- O proxecto na súa dimensión global encádrase dentro das estratexias galega, española e europea orientadas ao desenvolvemento de tecnoloxías, ferramentas e recursos para a proxección das linguas nos ámbitos de aplicación da intelixencia artificial. Neste contexto, a Xunta de Galicia vén de aprobar a Estratexia Galicia Dixital 2030, que identifica a innovación tecnolóxica como instrumento para a proxección dos sinais de identidade de Galicia e, en particular, da lingua galega para promover o coñecemento e o seu uso en todo tipo de soportes e formatos dixitais. Participará tamén do proceso participativo da Xunta de Galicia a través da Axencia Galega de Innovación, para a elaboración da Estratexia de Especialización Intelixente RIS3 para o período 2021-2027, que definirá os investimentos para o dito período no ámbito da investigación e a

innovación e que reforzará especialmente, tal como esixe a Comisión Europea, unha maior implicación dos axentes sociais e da cidadanía en xeral.

A primeira fase para o desenvolvemento do proxecto será definir, deseñar e elaborar a planificación pormenorizada da súa execución durante as seguintes catro anualidades a través dunha proposta inicial da estratexia de acción colaborativa conxunta baseada nos principios inspiradores da Intelixencia Artificial e das Tecnoloxías da Linguaxe. A revisión do estado da arte como a catalogación dos recursos dispoñibles para o galego permitirán levar a cabo un deseño máis realista dos diferentes subproxectos de Nós, e que se inserirán en dúas grandes áreas: procesamento de voz e procesamento de texto. Para cada subproxecto definiranse tamén equipos de traballo, estratexias de coordinación entre eles e metodoloxías de seguimento e avaliación da calidade, tanto dos recursos como das ferramentas resultantes.

Oitavo.- Que as partes son conscientes da responsabilidade que teñen de colaborar no proceso normalizador, así como á hora de cumprir cos obxectivos previstos na Lei 3/1983, de normalización lingüística, no Plan xeral de normalización da lingua galega que aprobou, por unanimidade, o Pleno da Cámara do Parlamento de Galicia na súa sesión do día 21 de setembro de 2004, xa que logo, a comunidade autónoma necesita dispor de traballos de investigación e dunha estratexia común para o desenvolvemento dun proxecto, como Nós, que permitirá xerar os recursos necesarios para colocar o galego na vangarda das tecnoloxías intelixentes, nomeadamente naquelas áreas relacionadas coas tecnoloxías lingüísticas e, de modo xeral, na sociedade e economía da intelixencia artificial.

Noveno.- Que a sinatura deste convenio ten por finalidade a prestación dun servizo de carácter público e social que a través da transferencia do coñecemento incide na sociedade en xeral, e que se executará mediante a acción coordinada e colaborativa pública a través de diferentes departamentos da Xunta de Galicia e a USC, por medio dos organismos especializados dela dependentes, como son, O ILG e o CiTIUS, por ser estes, referencias indiscutibles no eido da investigación lingüística e das tecnoloxías intelixentes. Son as únicas entidades públicas que no ámbito da Comunidade Autónoma Galega contan cos coñecementos técnicos, traballos de investigación avanzada en materia de intelixencia artificial e tecnoloxía lingüística, así

como dos medios materiais e humanos máis especializados para realizar as actuacións derivadas da realización do obxecto do convenio. Pola singularidade das accións que se van desenvolver ao abeiro deste convenio non é posible promover a concorrencia en convocatoria pública.

Por todo isto, as partes asinantes do convenio coinciden en salientar a necesidade da referida colaboración para o desenvolvemento do Proxecto Nós, nas diferentes fases para o seu desenvolvemento global coa finalidade de darlle un pulo decidido a lingua galega no ecosistema dixital e na vangarda das tecnoloxías intelixentes.

CLÁUSULAS

Primeira.- Obxecto do convenio

1 O obxecto deste convenio é establecer as condicións polas que se rexerá a colaboración entre 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 Amtega e a USC a través do ILG e CiTIUS, para definir, deseñar e elaborar a planificación pormenorizada do Proxecto Nós para o seu desenvolvemento durante as seguintes catro anualidades posteriores ao remate da vixencia deste convenio a través dunha proposta inicial da estratexia de acción colaborativa conxunta das partes asinantes baseada nos principios inspiradores da intelixencia artificial e das tecnoloxías da linguaxe coa finalidade de dotar a lingua galega de ferramentas e recursos lingüísticos e tecnolóxicos que permitan o acceso a produtos dixitais presentes e futuros nesa lingua. Atenderase, especialmente a seis ámbitos de actuación:

Actuación 1.- Garantir o protagonismo e asegurar o desenvolvemento e a presenza dixital da lingua galega no mundo.

Actuación 2.- Dotar a lingua galega de ferramentas e recursos suficientes para o seu desenvolvemento na economía e na sociedade dixital

Actuación 3.- Garantir a presenza e uso da lingua galega no proceso de transformación dixital das administracións públicas galegas

Actuación 4.- Favorecer a integración de persoas estranxeiras

Actuación 5.- Favorecer o uso, divulgación e incorporación da lingua galega no desenvolvemento dixital das empresas, en calquera sector de actividade.

Actuación 6.- Favorecer a formación, retención, captación e retorno de talento no eido das tecnoloxías e, en particular, das tecnoloxías da linguaxe natural aplicadas ao galego.

2 O resumo executivo do proxecto Nós, o traballo asociado, así como as liñas de acción que se van desenvolver na fase inicial de deseño, elaboración e definición da proposta colaborativa durante a súa vixencia xúntanse como anexo deste convenio e formarán parte integrante del.

3 O resultado dos traballos que se realicen ao abeiro da proposta inicial da estratexia de acción colaborativa conxunta dende a sinatura do convenio e ata o **15 de decembro do ano 2021** farán posible levar a cabo o maior investimento directo e colaborativo na promoción do galego de toda a historia e, para iso, as partes asinantes do convenio buscarán nas seguintes anualidades achegas económicas dos fondos europeos Next Generation-UE para o seu total desenvolvemento, na medida que o proxecto Nós entronca á perfección co sentido destes fondos europeos e doutras liñas de actuación da Unión Europea en materia tecnolóxica e lingüística.

Segunda.- Partes do convenio

1 A Xunta de Galicia coa implicación dos seguintes departamentos:

1.1 A Consellería de Cultura, Educación e Universidades, a través da SXPL e do CRPIH, dela dependente.

1.2 A Amtega

2 A USC a través do ILG e o CiTIUS, por ser centros de referencia en investigación lingüística e de tecnoloxías intelixentes.

Terceira.- Obxectivo xeral e específicos do convenio

Obxectivo xeral: **colocar o galego na vangarda da intelixencia artificial e das tecnoloxías lingüísticas** para posibilitar deste xeito o uso natural da nosa lingua nas interaccións dixitais entre as persoas e os dispositivos tecnolóxicos.

Obxectivos específicos:

a) **Deseño e elaboración da proposta inicial da estratexia de acción colaborativa e conxunta do plan de traballo do Proxecto Nós**, no que se recollerán, entre outros aspectos: o alcance, obxectivos e resultados ou produtos concretos esperados; as actuacións que se van desenvolver; os medios necesarios para a súa execución o cronograma obxectivo; o modelo tecnolóxico, de sostibilidade e de difusión dos produtos resultantes; así como as mecanismos de validación, seguimento e control. Poderán establecerse plans específicos, ou priorizar, en base aos recursos ou ferramentas previstas:

- Corpus e modelos de linguaxe.
- Resumidores automáticos.
- Correctores lingüísticos automáticos.
- Sistemas de síntese de voz.
- Sistemas de recoñecemento de fala.
- Sistemas de xeración automática de texto.
- Sistemas de diálogo ou axentes conversacionais (combinando os tres anteriores con ferramentas de análise semántica).
- Sistemas de resposta a preguntas.
- Sistemas de tradución automática.
- Sistemas de extracción de información e minaría de textos.

A proposta inicial da estratexia de acción colaborativa e conxunta do plan de traballo do Proxecto Nós deberá aprobarse por todas as partes asinantes.

b) **Posta en marcha das accións** que permitan o inicio do desenvolvemento dos recursos e das ferramentas identificadas anteriormente.

Todos os recursos lingüísticos en galego que resulten da realización do proxecto en calquera das súas fases serán publicados en formato dixital baixo licenzas libres e a través de ferramentas tecnolóxicas con código aberto e poderán ser empregadas, libre e gratuitamente, polas persoas destinatarias. Por tal motivo, as partes do convenio comprométese, xa na fase inicial de desenvolvemento da proposta colaborativa e do seu plan de traballo a poñer os devanditos recursos a disposición das entidades ou institucións públicas ou privadas que así o soliciten e da sociedade en xeral.

Cuarta.- Responsabilidade do proxecto

1) A Consellería de Cultura, Educación e Universidade, a través da SXPL e o CRPIH, a Amtega, e a USC a través do ILG e o CiTIUS, serán as encargadas de realizar as actuacións necesarias para desenvolver as actividades obxecto do convenio, nun proceso de colaboración continuo. As entidades asinantes, como coorganizadoras, terán unha participación activa e esencial na realización do obxecto do convenio e compartirán, nos termos que se indiquen neste convenio, a titularidade dos resultados dos traballos realizados.

2) A participación do equipo de traballo ou equipos de traballo dos investigadores da USC nas actividades recollidas neste convenio efectuarase conforme a regulamentación propia da USC recollida nos seus estatutos e no Regulamento para a realización de actividades de investigación aprobado polo Consello de Goberno desta universidade o 20 de xuño de 2005 ao abeiro do establecido no artigo 83 da Lei orgánica 6/2001, do 21 de decembro, de universidades, modificada pola Lei orgánica 4/2007, do 12 de abril.

3) Calquera das institucións e entidades que participan no proxecto, cada unha no seu ámbito competencial de actuación, dispón de persoal asignado aos servizos xurídicos, persoal docente e/ou investigador, lingüistas e a infraestrutura técnica e a tecnoloxía máis adecuada para garantir a consecución dos obxectivos de interese público perseguidos coa sinatura deste convenio. Se fose necesario contratar medios humanos, técnicos ou materiais complementarios para o axeitado desenvolvemento do obxecto do convenio, a devandita contratación realizarase de acordo coa normativa específica que resulte de aplicación ao abeiro da Lei 9/2017, de 8 de novembro, de contratos do sector público, pola que se traspoñen ao ordenamento xurídico español

as directivas do Parlamento Europeo e do Consello 2014/23/UE e 2014/24/UE, de 26 de febreiro de 2014.

4) A sinatura deste convenio non implica relación laboral, contractual ou de calquera outro tipo entre os profesionais que vaian desenvolver as actividades previstas nel e mais a Xunta de Galicia, de tal xeito que a esta non se lle poderá esixir responsabilidade ningunha –nin directa, nin indirecta, nin subsidiaria– polos actos ou feitos acaecidos no desenvolvemento do convenio.

Quinta.- Compromisos das partes

a) A Consellería de Cultura, Educación e Universidade a través da SXPL e o CRPIH comprométese:

A participar no deseño e na elaboración da proposta inicial da estratexia de acción colaborativa e plan de traballo do Proxecto Nós, en especial nos aspectos vinculados ao seu ámbito competencial.

A contribuir ao financiamento do obxecto deste convenio nos termos previstos na cláusula sexta.

b) A Axencia para a Modernización Tecnolóxica de Galicia comprométese:

A participar no deseño e na elaboración da proposta inicial da estratexia de acción colaborativa e plan de traballo do Proxecto Nós, fundamentalmente nos aspectos vinculados ao seu ámbito competencial e, nomeadamente, no relativo ao modelo tecnolóxico, de sostibilidade e de difusión dos produtos resultantes, así como ao aliñamento cos obxectivos da Estratexia Galicia Dixital 2030.

A poñer a disposición do proxecto a infraestrutura técnica dixital para a implementación do obxecto do convenio, e que se concretará nas liñas de acción do plan de traballo, o que non comportará a asunción de obrigas económicas por parte da AMTEGA..

c) A USC, a través do ILG e o CiTIUS comprométese:

- A participar no deseño e na elaboración da proposta inicial da estratexia de acción colaborativa e plan de traballo do Proxecto Nós, en especial nos aspectos vinculados

ao seu ámbito competencial en materia de investigación en lingua galega, Tecnoloxías Lingüísticas (TL) e Intelixencia Artificial (IA)

- Desenvolver a dirección técnica das actuacións recollidas na cláusula terceira deste convenio, para o seu desenvolvemento e xustificar a súa realización e o cumprimento da súa finalidade. Para tales efectos, levará a cabo, directamente ou en colaboración con outras entidades, as actuacións establecidas na cláusula terceira deste convenio debendo proceder, en consecuencia, á asignación dos medios humanos e materiais que sexan necesarios e, de ser o caso, á contratación das colaboracións externas que se precisen para a súa execución
- Achegar coa xustificación os informes e documentos que acrediten fidedignamente o desenvolvemento do obxecto do presente convenio, así como os informes e demais documentos requiridos pola Xunta de Galicia.
- Facer figurar expresamente o financiamento da Xunta de Galicia na totalidade das actuacións asociadas a este Convenio.

Sexta.- Orzamento e financiamento do convenio

O orzamento previsto para a realización do obxecto do presente convenio ascende a **150.617,92 €**

A Consellería de Cultura e Turismo, a través da Secretaría Xeral de Política Lingüística, colaborará na realización do obxecto do convenio cunha achega de **60.000,00 €**, con cargo á aplicación orzamentaria **10.50.151A.640.1 (código de proxecto 2001-07023)** dos orzamentos xerais da Comunidade Autónoma de Galicia para o ano 2021 e destinará o devandito importe ao pagamento dunha parte dos custos das horas de traballo do persoal de dirección e execución dos traballos para definir, deseñar e elaborar a planificación pormenorizada do Proxecto Nós e o seu desenvolvemento posterior a través da proposta inicial da estratexia de acción colaborativa conxunta baseada nos principios inspiradores da intelixencia artificial e das tecnoloxías da linguaxe.

A referida achega é unha porcentaxe do importe total do proxecto (39.836%) e financiará exclusivamente unha parte dos custos do convenio, sen que a USC obteña beneficio ou lucro ningún.

A USC, a través do ILG e o CiTIUS contribuirá, así mesmo, cunha achega en especie (horas de traballo do equipo ou equipos de traballo) e que se destinará a financiar unha parte dos custos da dirección técnica e de execución do obxecto deste convenio e, así mesmo, outros custos necesarios para o bo desenvolvemento do proxecto. A dita achega é equivalente á porcentaxe do 60,164% sobre o importe total .

O custo referido ao **valor/hora** do equipo ou equipos de traballo que participen na dirección técnica e execución da proposta inicial da estratexia de acción colaborativa conxunta do proxecto Nós durante a vixencia do convenio será a seguinte :

Táboa 1: Recursos propios da USC (dedicación do persoal investigador do ILG e do CITIUS):

Persoal investigador do ILG e do CiTIUS (Período 1.07.2021-15.12.2021)	Custo/Hora	Núm. de horas dedicadas ao proxecto duran- te a vixencia do convenio	Importe
Persoal investigador ILG			
Investigadora principal do proxecto (CDr)	45,54	209	9.517,86
Co-investigador principal do proxecto (CU)	54,27	209	11.342,43
Persoal investigador (1 CU, 2 TU)	47,93	267	12.797,31
Persoal investigador CITIUS (Período 1.07.2021-15.12.2021)			
Investigador principal (TU)	45,60	264	12.038,40
Co-investigador principal (CU)	47,16	264	12.450,24
Persoal investigador (1 CU, 2 TU, 2 CDr, 2 RyC, 1 JCI)	36,48	616	22.471,68
TOTAL USC DEDICACIÓN PERSO- AL PROPIO			80.617,92

Custos do material bibliográfico e documental, recursos informáticos, técnicos e tecnolóxicos, e outros necesarios para o desenvolvemento do proxecto			10.007,45
TOTAL USC RECURSOS PROPIOS			90.625,37

Táboa 2: Persoal investigador de nova contratación:

Novas contratacións (Período 1.09.2021-15.12.2021)	Custo/Hora	Núm. de horas dedicadas ao proxecto durante a vixencia do convenio	Importe
Persoal investigador contratado (3 Inv. Asociados)	26,3986497	1533	40.469,13
Persoal investigador contratado (2 Técnico Superior)	19,10315068	1022	19.523,42
Total novas contratacións			59.992,55

A USC asumirá con cargo aos seus orzamentos propios a diferenza entre a achega económica da Consellería de Cultura, Educación e Universidade a través da Secretaría Xeral de Política Lingüística e o orzamento total do convenio, que será equivalente a **90.617,92 €**, importe no que se inclúen, ademais de parte dos custos de dirección técnica e da execución do proxecto, de ser o caso, os custos do material bibliográfico e documental necesario, os recursos e medios informáticos, técnicos e tecnolóxicos, e, en xeral, os que sexan necesarios para o bo desenvolvemento do proxecto nas sedes dos centros investigadores.

Sétima.- Forma de pagamento e xustificación do convenio

A documentación requirida na fase xustificativa presentarase electronicamente accedendo ao expediente na Carpeta do cidadán da persoa solicitante/representante, <https://sede.xunta.gal/carpeta-do-cidadan>. Código de procedemento PR004A.

A USC para recibir a achega económica da Consellería de Cultura, Educación e Universidade a través da SXPL para facer fronte a unha parte dos custos da dirección técnica e de execución do obxecto deste convenio, presentará electronicamente, coa data límite do **3 de decembro** de 2021, a seguinte documentación:

- Un certificado da persoa responsable da Xerencia da USC ou do órgano de control que corresponda, que acredite a toma de razón en contabilidade de que se executou o obxecto do convenio e que as achegas das partes cofinanciadoras foron destinadas ao pagamento dos custos por hora do persoal investigador propio e, de ser o caso, os custos do persoal externo contratado, que irá acompañado dunha relación valorada das horas dedicadas ao proxecto polo equipo ou equipos de traballo da USC e, tamén, dos colaboradores externos.
- Unha memoria asinada pola persoa investigadora principal do equipo de traballo (ou dos equipos de traballo, de configurarse máis dun) na que se fará constar de forma detallada as actuacións realizadas.

No caso de que a acreditación dos custos do convenio e, máis concretamente, a súa equivalencia en horas de dedicación ao proxecto, sexa inferior ao orzamento total do convenio, a achega das partes reducirase na mesma proporción.

A Consellería de Cultura, Educación e Universidade a través da Secretaría Xeral de Política Lingüística, asume as tarefas de verificación e comprobación da xustificación que presente a USC para o cumprimento do convenio.

O pagamento, realizarase na conta bancaria que sinale a USC mediante certificado acreditativo da súa titularidade.

Antes de proceder á sinatura deste convenio e para o seu pagamento, a USC deberá presentar:

- Un certificado de non ter pendente de pagamento ningunha débeda coa Administración pública da Comunidade Autónoma de Galicia, expedido pola Consellería de Facenda.
- Un certificado de estar ao día das obrigas tributarias, expedido pola correspondente delegación ou axencia do Ministerio de Economía e Facenda.
- Un certificado de estar ao corrente das obrigas sociais, expedido pola Tesourería da Seguridade Social.

A sinatura deste convenio comportará a autorización ao órgano xestor (SXPL) para solicitar as certificacións que deban emitir a Axencia Estatal da Administración Tributaria, a Tesourería da Seguridade Social e a Consellería de Facenda da Xunta de Galicia.

A USC facilitará toda a información que lle requira a Consellería de Cultura, Educación e Universidade, así como a Intervención Xeral da comunidade autónoma, a Intervención Xeral do Parlamento, o Consello de Contas e o Tribunal de Contas no exercicio das súas funcións de fiscalización e control.

Oitava.- Resultados conxuntos

1) Os resultados obtidos como consecuencia da colaboración obxecto deste convenio corresponderán de xeito conxunto ás partes asinantes e, cada unha delas, poderán facer uso gratuíto destes co consentimento unánime das outras partes.

Os dereitos incorporais de propiedade intelectual de carácter patrimonial respecto dos recursos dixitais ou tecnolóxicos que, de ser o caso, se xeren nas diferentes fases de desenvolvemento do proxecto, de titularidade común das partes asinantes, pola sinatura do presente convenio cederanse de xeito gratuíto á Xunta de Galicia para o seu destino ao uso público con formatos dixitais baixo licenzas libres e con códigos abertos para ser empregados, libre e gratuitamente, polas persoas destinatarias.

2) Exceptúase do disposto no parágrafo anterior o resultado da actividade propia do labor investigadora dos traballos de investigación que se utilicen como medios para o

desenvolvemento do obxecto do convenio e cuxa propiedade intelectual lle corresponderá á USC.

Novena.- Difusión da colaboración das partes

2) Nas actuacións informativas e divulgativas, así como nas publicacións que se fagan relacionadas coas actividades deste convenio, farase constar nun lugar destacado a colaboración das partes asinantes e utilizaranse, en igual tamaño e consideración, os anagramas ou logos que correspondan, segundo o previsto pola normativa de identidade corporativa da Xunta de Galicia e da USC.

Sexa cal sexa o medio de difusión, cada unha das partes comprométese a facer mención das outras partes e deste convenio.

Décima.- Comisión de seguimento

Crearase unha comisión de seguimento das accións obxecto do presente convenio, facultada para proporlles ás partes as resolucións de cantas dúbidas xurdan na súa execución e presidida polo representante da Consellería de Cultura, Educación e Universidade, ou persoa na que delegue, e integrada ademais por:

- Unha persoa representante da USC, nomeada polo seu titular.
- Unha persoa representante da Amtega, nomeada polo seu titular.

A comisión de seguimento manterá reunións periódicas de coordinación, seguimento e control das actividades realizadas e xestionará calquera modificación que xurda no desenvolvemento do proxecto nun proceso de colaboración mutua entre as partes asinantes. Ás reunións da comisión asistirán os seus membros ou as persoas en quen deleguen.

A comisión poderá adoptar as decisións e ditar as instrucións necesarias para asegurar o cumprimento do obxecto do presente convenio. Os acordos da comisión serán inmediatamente executivos.

Décimo primeira.- Vixencia

Este convenio entrará en vigor o mesmo día da súa sinatura dixital pola persoa asinante que o faga en último lugar e terá vixencia ata o **15 de decembro** do ano 2021, agás que sexa denunciado por escrito por algunha das partes asinantes polo incumprimento dos compromisos que nel se recollen, cun mes de antelación polo menos.

Antes da finalización do prazo de vixencia, os asinantes poderán acordar unanimemente a súa prórroga por un máximo dun ano máis. De prorrogarse o convenio, na correspondente addenda estableceríanse as obrigas concretas das partes asinantes.

Décimo segunda.- Utilización da lingua galega

Todas as accións e actividades que se realicen ao abeiro deste convenio deberán levarse a cabo de acordo co indicado nas vixentes normas ortográficas e morfolóxicas do idioma galego, aprobadas pola Real Academia Galega no ano 2003, tal e como se establece na disposición adicional da Lei 3/1983, do 15 de xuño, de normalización lingüística. Así mesmo, coidarase especialmente o respecto rigoroso á toponimia oficial nos termos previstos no artigo 10 da citada lei.

O incumprimento desta estipulación será considerada causa específica de resolución deste convenio.

Décimo terceira.- Rexistro de Convenios. Transparencia e bo goberno

A sinatura deste convenio suporá o consentimento expreso do interesado á Administración para incluír e facer públicos os datos referidos ao convenio, de conformidade co artigo 15 da Lei 1/2016, do 18 de xaneiro, de transparencia e bo goberno, co artigo 8 da Lei 19/2013, do 9 de decembro (BOE 10-12-2013), de transparencia, acceso á información pública e bo goberno e co Decreto 126/2006, do 20 de xullo, polo que se regula o Rexistro de Convenios da Xunta de Galicia

Décimo cuarta.- Causas de resolución

Este convenio resolverase:

- Por incumprimento total ou parcial dalgunha das estipulacións que o regulan.
- Polo transcurso do tempo estipulado sen que se culminen as actividades acordadas.

De se resolver o convenio, as partes asinantes poderán, se é o caso, esixir a devolución das cantidades achegadas.

Décimo quinta.- Notificación electrónica

Segundo o artigo 14.2 da Lei 39/2015, do 1 de outubro, do procedemento administrativo común das administracións públicas, os colectivos que están obrigados a relacionarse a través de medios electrónicos coas administracións públicas e, polo tanto, a recibir notificacións por canles electrónicas son: as persoas xurídicas, as entidades sen personalidade xurídica, os colexios profesionais, os que representen un interesado que estea obrigado a relacionarse electronicamente coa Administración e os empregados das administracións públicas para os trámites que realicen con elas por razón da súa condición de empregado público. Neste punto e conforme a disposición adicional quinta da Lei 39/2015, o Parlamento rexerese polo previsto na súa normativa específica, no marco dos principios que inspiran a actuación administrativa de acordo coa citada lei.

As notificacións realizaranse só por medios electrónicos a través do Sistema de Notificación Electrónica de Galicia Notific@, <https://notifica.xunta.gal>. Para poder acceder a unha notificación electrónica, o interesado deberá contar cun certificado electrónico asociado ao NIF que figure como destinatario da notificación (persoa física ou xurídica). Así mesmo, o interesado pode autorizar a calquera outra persoa a acceder ao contido das súas notificacións.

Décimo sexta.- Protección de datos de carácter persoal

As partes asinantes comprométese a cumprir, en todo momento, as disposicións contidas na normativa básica estatal e autonómica que xustifica a lexitimidade do tratamento dos datos persoais e dos dereitos dixitais.

Na aplicación deste convenio respectaranse cantas esixencias establece o Regulamento UE 679/2016 xeral de protección de datos e a Lei orgánica 3/2018, do 5 de decembro

(BOE 06/12/2018) de protección de datos persoais e garantía dos dereitos dixitais, adoptando cantas medidas resulten necesarias para cumprir as súas previsións e, en particular, para garantir a seguridade e integridade dos datos persoais e dos dereitos dixitais e, a súa protección fronte a alteracións, perdas, tratamentos ou accesos non autorizados.

Serán de aplicación todos os dereitos e obrigas derivados da normativa aplicable en materia de protección de datos e garantía dos dereitos dixitais.

Decimo sétima.- Administración dixital

As partes asinantes deste convenio respectarán cantas esixencias establece a Lei 4/2019, do 17 de xullo, de Administración dixital de Galicia (DOG 26 de xullo de 2019), que será aplicable ao sector público autonómico, integrado, de acordo co artigo 3 da Lei 16/2010, do 17 de decembro, de organización e funcionamento da Administración xeral do sector público autonómico de Galicia, pola Administración xeral e polas entidades instrumentais e, tamén, será aplicable á cidadanía que ten o deber de utilizar os servizos públicos dixitais que se poñan á súa disposición de xeito adecuado, co respecto das normas de uso, xerais ou específicas, que se establezan para cada un dos servizos e, en particular, os cidadáns e as cidadás deberán utilizar os sistemas de identificación e de sinatura electrónicas de que sexan lexítimos titulares, usar a información dispoñible conforme a política de privacidade publicada e respectar a normativa en materia de protección de datos de carácter persoal (artigos 2 a) e b) e 7 apartados 1 e 2 respectivamente da citada lei).

Décimo oitava.- Natureza xurídica e xurisdición

Este convenio ten carácter administrativo e os seus efectos rexeranse polo establecido nas súas estipulacións, e quedará suxeito ao disposto nos artigos 47 e seguintes da Lei 40/2015, do 1 de outubro, do réxime xurídico do sector público.

Así mesmo, será de aplicación á Xunta de Galicia a Lei 14/2013, do 26 de decembro, de racionalización do sector público autonómico, polo que se refire ao contido da colaboración mediante o desempeño conxunto de tarefas de interese público. O artigo 14 prevé que a Administración xeral da comunidade autónoma e as entidades do sector

público, ou estas entre elas, poderán formalizar acordos de colaboración ou cooperación para o servizo aos intereses xerais da súa competencia.

Doutra banda, aínda que é un dos negocios xurídicos excluídos do ámbito de aplicación da Lei 9/2017, de 8 de novembro, de contratos do sector público, pola que se traspoñen ao ordenamento xurídico español as directivas do Parlamento Europeo e do Consello 2014/23/UE e 2014/24/UE, de 26 de febreiro de 2014, no non previsto aplicánselle os principios recollidos na citada lei, co fin de resolver dúbidas e omisións que puidesen presentarse. Así mesmo, supletoriamente terase en conta o Decreto lexislativo 1/1999, do 7 de outubro, pola que se aproba o texto refundido da Lei de réxime financeiro e orzamentario de Galicia e demais normativa vixente na materia.

A xurisdición contencioso-administrativa será a competente para dirimir as cuestións que puidesen xurdir da aplicación deste convenio.

En proba de conformidade co seu contido as partes asinan telematicamente este convenio.

O conselleiro de Cultura, Educación e
Universidade

Román Rodríguez González

A directora da Amtega

Mar Pereira Álvarez

O reitor magnífico da Universidade de
Santiago de Compostela

Anexo:

Resumo executivo do proxecto Nós, plan de traballo asociado ao convenio de inicio

Resumo executivo

O Proxecto Nós permitirá xerar os recursos necesarios para colocar o galego na vangarda das Tecnoloxías Intelixentes, nomeadamente naquelas áreas relacionadas coas Tecnoloxías Lingüísticas (TL) e, de modo xeral, na sociedade e economía da Intelixencia Artificial. Dentro do proxecto compilaranse e desenvolveranse recursos e ferramentas para o procesamento automático do galego, tanto oral como escrito, que serán distribuídas baixo licenzas libres e demostradores completos finais que permitirán amosar as posibilidades dos ditos recursos en proxectos de alto impacto socioeconómico en diferentes contextos e facilitarán o desenvolvemento de produtos e servizos por parte das empresas e outras organizacións, tanto públicas como privadas. O proxecto encádrase dentro das estratexias galega, española e europea orientadas ao desenvolvemento de tecnoloxías, ferramentas e aplicacións para a proxección e uso das linguas nos ámbitos de aplicación da IA.

O Proxecto Nós será dirixido por un equipo investigador da Universidade de Santiago de Compostela formado por persoal do Instituto da Lingua Galega (ILG) e do CiTIUS - Centro Singular de Tecnoloxías Intelixentes da USC, que comporá equipos de investigación coordinados para cada un dos subproxectos que se desenvolverán en Nós, e que se corresponden con algunhas das grandes tarefas do Procesamento das Linguas Naturais:

- **Tecnoloxías da fala**, que permitirán incluír sistemas de voz en dispositivos.
- **Modelos de lingua**: permitirá converter datos estruturados en oracións en galego, desenvolver un axente conversacional de carácter xeral en galego, e crear sistemas de pregunta-resposta, tanto a través de texto como de voz.

- **Tradución automática:** de galego a outras linguas e viceversa, permitindo entrada textual ou por voz.
- **Extracción de información:** consistirá na identificación e extracción das principais entidades, relacións semánticas e eventos dun texto ou conxunto de textos.
- **Verificación de noticias e análise de opinións:** permitirá clasificar noticias en relación á veracidade do contido e á polaridade (desde moi negativa até moi positiva) do propio texto.
- **Resumos automáticos:** xerará resumos automáticos de conxuntos de documentos.
- **Corrección e avaliación lingüística** automática, tanto ortográfica como léxica, gramatical e de estilo.

Durante o período do convenio 2021 vanse definir cales van ser os recursos lingüísticos que se van desenvolver no proxecto e que posteriormente van ser facilitados para o seu uso en todo tipo de aplicacións, servizos ou produtos, tanto pola comunidade científica, como por empresas ou institucións. Identificaranse tamén cales van ser os demostradores específicos baseados neses recursos que se van desenvolver, en función do seu impacto ou interese socioeconómico dentro do denominado *Living Lab*. Estes desenvolvementos serán completos, plenamente funcionais e servirán como demostradores visíbeis e accesíbeis da tecnoloxía desenvolvida que terán un efecto tractor que dea lugar ao desenvolvemento de novos produtos. A título de exemplo, algúns destes produtos finais poden ser: asistentes para a realización de tarefas domésticas coma a compra para persoas maiores e dependentes, asistentes conversacionais para a interacción cos servizos de información das administracións públicas, por texto ou voz, subtítulado e/ou audiodescrición automática de contidos dixitais (TV) ou, en xeral, a inclusión do galego en todos os aparellos con interacción persoa-máquina, tanto en modo texto como a través da voz (elevadores, televisións, GPS, pequenos e grandes electrodomésticos, etc.).

Tanto para os recursos lingüísticos que se desenvolverán nos dez subproxectos, coma para os demostradores do *Living Lab* acordarase entre a USC e a Xunta de Galicia (AMTEGA) dentro deste período o modelo tecnolóxico mais adecuado que facilite unha ampla difusión e adopción dos mesmos para a creación de novos servizos que se porán a disposición de empresas, institucións e público en xeral.

Plan de Trabajo asociado ao convenio

O traballo realizado a través deste convenio ten como obxectivo deseñar o proceso de execución do Proxecto Nós, sendo estes os resultados principais a entregar:

1. Documento de **revisión do estado da arte** das tecnoloxías lingüísticas obxecto do proxecto, con especial atención á súa aplicación á lingua galega.
2. **Catálogo dos recursos lingüísticos** actualmente dispoñíbeis para o galego en formato dixital, atendendo ás súas licenzas de uso e dereitos de autor.
3. Documento de planificación pormenorizada e deseño do proceso de execución dos diferentes subproxectos de Nós e dos demostradores a implementar ao longo do proxecto.
4. Documento de **especificación do modelo de provisión** dos recursos desenvolvidos e **do modelo tecnolóxico** para a configuración dos diferentes servizos que presten os demostradores desenvolvidos ao longo do proxecto.
5. **Plan de difusión e diseminación.**

1. Revisión do estado da arte: o primeiro dos obxectivos consistirá en elaborar un documento que conteña unha revisión detallada do estado da arte das tecnoloxías lingüísticas obxecto do proxecto. Por unha parte, analizaranse aqueles traballos máis relevantes do punto de vista do procesamento computacional, que permitirán coñecer as necesidades –tanto de recursos como de infraestruturas– de cada unha das áreas do proxecto. Por outra parte, será estudada a bibliografía máis destacada sobre a compilación, clasificación e anotación de recursos lingüísticos, principalmente en relación coa súa aplicación en sistemas de intelixencia artificial. Esta revisión analizará os estudos máis relevantes das tecnoloxías lingüísticas en xeral, mais terá en conta

aqueles aspectos de especial interese para o deseño e implementación das mesmas en lingua galega. Tamén se dará especial atención ás tecnoloxías máis resilientes en relación á capacidade de adaptación de innovacións así como de adecuación a cambios de dominio de aplicación.

2. Catalogación dos recursos lingüísticos: o segundo obxectivo consiste na catalogación dos recursos lingüísticos, tanto textuais como acústicos, existentes para o galego en formatos dixitais. Tendo en conta que o Proxecto Nós distribuirá recursos e ferramentas lingüísticas con licenzas abertas, nesta catalogación atenderase de maneira especial aos dereitos de autor e ás licenzas de uso e difusión dos recursos dispoñíbeis. Esta catalogación ten como finalidade principal o deseño dun panorama de recursos lingüísticos para o galego que permita planificar o proxecto de maneira máis efectiva, coñecendo previamente aquelas áreas con maiores necesidades de compilación e construción de recursos.

3. Planificación do Proxecto Nós: o obxectivo central do convenio consistirá na planificación pormenorizada da execución do Proxecto Nós durante as seguintes catro anualidades. Tanto a revisión do estado da arte como a catalogación dos recursos dispoñíbeis para o galego permitirán levar a cabo un deseño máis realista dos diferentes subproxectos de Nós, que se inserirán en dúas grandes áreas: procesamento de voz e procesamento de texto.

Esta planificación incluírá unha definición detallada dos diferentes subproxectos e da planificación (obxectivos, tarefas, fitos e entregas de cada un deles). Para cada subproxecto definiranse equipos de traballo e estratexias de coordinación entre eles, así como metodoloxías de seguimento e avaliación da calidade tanto dos recursos como das ferramentas resultantes. En relación á infraestrutura, será realizada unha análise das necesidades computacionais tanto para procesar e almacenar os recursos como para adestrar os modelos. Identificaranse, atendendo ao seu interese e impacto socio-económico, os demostradores que se van desenvolver dentro do proxecto. Os recursos lingüísticos que resulten precisos para o funcionamento dos demostradores definiranse e encadraranse nos subproxectos relacionados.

4. Provisión de recursos e modelo tecnolóxico: por último, será realizada unha análise das diferentes posibilidades para a provisión dos recursos resultantes do

proxecto, clasificados en tres categorías: (i) recursos lingüísticos; (ii) tecnoloxías lingüísticas; (iii) demostradores. En relación aos recursos, avaliaranse os pros e os contras das diferentes alternativas para provelos ao público obxectivo (difusión en repositorios institucionais *versus* repositorios públicos abertos, así como as diferentes posibilidades de licenciamiento). Sobre as tecnoloxías lingüísticas, seleccionaranse aquelas ferramentas de maior difusión e calidade no procesamento das linguas, para as que se implementarán modelos específicos para o galego. Por último, definiranse un conxunto de demostradores tecnolóxicos que tiren proveito dos recursos e das ferramentas desenvolvidas. Tanto para os recursos lingüísticos como para os demostradores deseñarase o modelo tecnolóxico máis adecuado para conseguir unha ampla difusión e adopción dos mesmos para a creación de novos servizos a disposición de empresas, institucións e público en xeral.

5. Plan de difusión e diseminación: definirase un plan de difusión do proxecto e de diseminación dos resultados orientado aos diferentes públicos obxectivo, en función dos tres tipos de recursos desenvolvidos. O plan incluírá accións dirixidas á sociedade, á comunidade científica relacionada (lingüística, tecnoloxías da lingua, intelixencia artificial), aos sectores produtivos e a organismos e institucións públicas. Como parte do plan de difusión incluíranse as especificacións do sitio Web que albergará a información pormenorizada sobre o desenvolvemento do proxecto e os seus resultados.