

CONVENIO PLURIANUAL DE COLABORACIÓN ENTRE A CONSELLERÍA DE CULTURA, LINGUA E XUVENTUDE E A UNIVERSIDADE DE VIGO PARA A REALIZACIÓN DUN PROGRAMA DE ACCIÓNS EN MATERIA DE INVESTIGACIÓN A TRAVÉS DO CENTRO RAMÓN PIÑEIRO PARA A INVESTIGACIÓN EN HUMANIDADES

En Santiago de Compostela,

REUNIDOS:

Dunha parte, Don José López Campos, conselleiro de Cultura, Lingua e Xuventude que, en virtude do seu nomeamento a través do Decreto 44/2024, do 14 de abril (DOG núm. 73, do 14 de abril) actúa de conformidade coas facultades que lle atribúe o artigo 34 da Lei 1/1983, do 22 de febreiro, de normas reguladoras da Xunta e da súa Presidencia, modificada pola Lei 11/1988, do 20 de outubro e pola Lei 12/2007, do 27 de xullo.

Doutra, Don Manuel Joaquín Reigosa Rogel, reitor magnífico da Universidade de Vigo, en virtude do establecido nos seus estatutos e do nomeamento como reitor polo Decreto 59/2018, do 31 de maio, (DOG nº 109, do 8 de xuño), actuando en nome e representación da devandita entidade.

As partes actúan no exercicio dos seus respectivos cargos e coa representación competencial que ostentan, e recoñecéndose mutuamente capacidade para se obrigaren nos termos do presente convenio,

EXPOÑEN:

1. Que para potenciar o emprego do galego, segundo establecen os artigos 5.3 do Estatuto de Autonomía de Galicia e 6.3 da Lei de normalización lingüística, cómpre dispor de estudos científicos sobre diversos eidos da lingua galega, que contribúan a protexela, promovela e modernizala para que sexa un medio de comunicación eficiente en calquera ámbito.

2. Que no Decreto 146/2024, do 20 de maio, polo que se establece a estrutura orgánica da Consellería de Cultura, Lingua e Xuventude (DOG 101, de 17 de maio), atribúeselle á Secretaría Xeral da Lingua (en diante, SXL), 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 da terminoloxía, así como, promover a cooperación e a colaboración con outras institucións competentes en materia de política lingüística, contribuíndo así á súa difusión.



3. Que a Universidade de Vigo (en diante, UdeVigo) é unha entidade que leva a cabo actividades de investigación, docencia e desenvolvemento científico e tecnolóxico, e ten como un dos seus fins o de procurar o seu enraizamento na sociedade galega que a sustenta e, en xeral, contribuír ao seu desenvolvemento social, económico e cultural.

4. Que estes fins son apoiados e compartidos pola Consellería de Cultura, Lingua e Xuventude (en diante consellería). Así, co fin de xestionar o desenvolvemento de proxectos e programas de investigación e conservación das fontes lingüísticas, literarias e antropolóxicas que permitan un estudo profundo da nosa historia e que garantan a utilidade da lingua galega na sociedade actual, con data 11 de febreiro de 1993, foi creado o Centro de Investigacións lingüísticas e literarias Ramón Piñeiro, posteriormente denominado Centro Ramón Piñeiro para a Investigación en Humanidades (en diante CRPIH), polo Decreto 330/1997, do 13 de novembro, dependente organicamente da SXL.

5. Que a dixitalización do coñecemento e a busca da innovación a través da tecnoloxía e da apertura de datos están volvéndose eixos claves da axenda dixital europea e que un dos piares da súa proposta estratéxica de investigación é a transferencia á sociedade e que, neste sentido, o CRPIH xa hai máis dunha década que vén impulsando o acceso aberto aos contidos de investigación, a través de presentacións sinxelas, accesibles, gratuítas e actualizadas.

6. Que as partes son conscientes da necesidade actual de promover as Humanidades dixitais (HD), co propósito de acadar unha mellor infraestrutura para a xestión de datos, recursos dixitais e ferramentas, que permitan a súa compartición e reutilización, aumentando a súa canalización, difusión, proxección e recoñecemento. Así, consideran que o acceso e intercambio de recursos entre elas - toda vez que existen complementariedades e sinerxías en proxectos de interese recíproco- pode ser explotado para acadar unha maior dimensión, mediante a interacción e colaboración científica entre os axentes participantes, o que supón unha optimización dos recursos destinados para o mesmo fin e redundará nun beneficio de ben común a prol do interese público.

7. Que, posto que os seus intereses son concorrentes e perseguen a obtención dunha finalidade común, a UdeVigo o CRPIH veñen colaborando ininterrompidamente desde o ano 1995 (sinatura do primeiro convenio) no desenvolvemento conxunto de actuacións de investigación relacionadas coas áreas de coñecemento mencionadas. Isto amosa un destacado compromiso en continuar cooperando nestas materias de estudo e a UdeVigo -en clave de visión xeral sobre o nivel actual do estado das investigacións- é quen dispón dos coñecementos necesarios para propiciar as contornas de colaboración que favorezan unha axeitada continuidade dos proxectos de investigación en curso que se desenvolven no centro.



8. Que, co propósito de que as tarefas encomendadas ao centro acaden os resultados axeitados, a consellería desexa seguir contando coa colaboración dos investigadores de maior reputación en cada un destes ámbitos e, dadas as particularidades e características específicas dos proxectos de investigación de que se trata, tendo en conta que nas áreas de coñecemento en Tecnoloxía das comunicacións, da Escola de Enxeñaría de telecomunicacións -en concreto no departamentos de Teoría do sinal e comunicacións- e na Escola Superior de Enxeñaría Informática da UdeVigo contan con recursos tecnolóxicos axeitados e con destacados investigadores de recoñecida capacidade científica -en relación á natureza dos proxectos incluídos no anexo deste convenio- os obxectivos perseguidos non poderían acadarse co mesmo nivel de excelencia nun escenario de concorrencia pública.

9. Que é de marcado interese xeral e das partes asinantes, no marco da colaboración establecida, incorporar novas perspectivas -na liña da *Estratexia Galicia Dixital 2030*, (particularmente no relacionado co capítulo 9.1 "Proxección dos sinais de identidade de Galicia")- que potencien a innovación dixital aplicada á dinamización e salvagarda do patrimonio e a produción cultural da Comunidade, como recurso accesible e compoñente fundamental da nosa tradición, identidade e recoñecemento colectivo, incluíndo a lingua galega como creadora de cultura e vehículo de posibilidades representativas, co fin de ampliar o alcance e a accesibilidade dos traballos e resultados da execución destes proxectos, facendo deles unhas ferramentas modernas e eficaces de referencia científica mediante a introdución e mellora no manexo de novos usos asociados ás tecnoloxías da información e da comunicación (TIC).

10. Que a UdeVigo dispón de investigadores de valía profesional e prestixio recoñecido que se dedican sectorialmente á realización deste tipo de actividades e cos que se pode establecer un espazo común de intereses no eido da investigación, que permiten a utilización da figura do convenio de colaboración como fórmula comercial apropiada para articular as súas relacións, dado que non se manifesta nunha contraposición de intereses, senón que se trata de establecer unha colaboración institucional para levar a cabo unha actuación en resposta de obxectivos compartidos e os intereses de ambas as dúas partes teñen por causa a consecución dun fin común en beneficio da comunidade.

Así mesmo, ao abeiro do establecido no artigo 6 -"Convenios e encomendas de xestión"-, da Lei 9/2017 de 8 de novembro de contratos do sector público, este caso considérase excluído e plenamente aplicable á figura do convenio de colaboración debido ás prestacións de investigación e desenvolvemento que, con respecto aos proxectos incluídos no anexo do presente convenio, van ser levadas a cabo por parte da Universidade e que son susceptibles de se incluír no punto 1, apartados b) e c) do citado artigo.



Por todo isto, a consellería e mais a UdeVigo establecen o presente convenio de colaboración que se rexerá polas seguintes

ESTIPULACIÓNS:

Primeira. Obxecto do convenio

Mediante este convenio establécense as condicións polas que se rexerá a colaboración entre a consellería e a UdeVigo, a través da SXL, para proseguir desenvolvendo o programa de accións en materia de investigación, especialmente no ámbito lingüístico, que se está levando a cabo no CRPIH e que se describe no anexo citado na estipulación segunda.

O fin último desta cooperación é desenvolver proxectos de investigación, fundamentais para a comunidade científica e para toda a sociedade, capaces de garantir a utilidade da lingua galega en todos os ámbitos da vida actual e de contribuír, asemade, ao acceso igualitario do coñecemento no ámbito das Humanidades, en xeral, e das HD, en particular.

Segunda. Obrigas da UdeVigo

As actividades que ha desenvolver a UdeVigo en colaboración co centro, recóllense no ANEXO I ao presente convenio onde se especifican os traballos que se realizarán en cada unha das accións que conforman o programa de investigación.

A UdeVigo cooperará coa consellería e coa SXPL na realización das actividades do programa de actuacións en materia de investigación científica do centro e comprométese a realizar -a través dos seus investigadores- os distintos proxectos que nel se desenvolven:

- PROXECTO RECURSOS PARA O DESENVOLVEMENTO DE TECNOLOXÍAS DA FALA
- PROXECTO ETIQUETADOR-LEMATIZADOR PARA O GALEGO ACTUAL

A UdeVigo colaborará na realización das distintas accións que integran este programa de investigación cunha contribución en especie, consistente en horas de dedicación á investigación, que serán as estimadas para a consecución dunha axeitada execución das actuacións.

A estes efectos, a valoración económica da devandita contribución equivalerá ó resultado do cálculo -en horas- sobre a retribución dun docente investigador -catedrático de universidade (T.C.)-. En consecuencia, a Universidade asumirá, pola súa conta, a cantidade de **dez mil euros (10.000,00 €)** derivada dos custos do seu persoal docente investigador, distribuída en catro anualidades (2026, 2027, 2028 e 2029), por un importe de **2.500,00 €** cada unha.



Ademais, a UdeVigo adoptará as actuacións de soporte técnico necesarias que permitan un intercambio de datos, co fin de contribuír a un mellor desenvolvemento das investigacións, en termos de alcance.

A UdeVigo deberá cumprir coas estipulacións xurdidas do vínculo xerado coa celebración do presente convenio no xeito e termos establecidos.

Terceira. Obrigas da Consellería de Cultura, Lingua e Xuventude

A Consellería ha de propiciar a participación dos profesionais especialistas do CRPIH e a súa contribución activa nos proxectos obxecto de investigacións conxuntas.

Contribuirá á difusión e divulgación dos resultados das investigacións, co fin de canalizar e proxectar os estudos resultantes, con efectos sobre o conxunto da sociedade en xeral e dos profesionais especializados, en particular, non só no ámbito autonómico, senón tamén no nacional e mesmo no internacional.

Facilitará a transferencia de coñecemento científico propio do CRPIH nas áreas temáticas obxecto das investigacións.

Velará polo cumprimento das disposicións, que han de ser executadas no xeito e termos establecidos, en favor do patrimonio público.

A consellería, a través da Secretaría Xeral da Lingua colaborará na realización destas accións cunha achega económica de trinta e catro mil euros **(34.000,00 €)**, repartidos en catro anualidades coa seguinte distribución:

Exercicio 2026: 8.500,00 € con cargo á aplicación orzamentaria 13.02.151A.640.1

Exercicio 2027: 8.500,00 € con cargo á aplicación orzamentaria 13.02.151A.640.1

Exercicio 2028: 8.500,00 € con cargo á aplicación orzamentaria 13.02.151A.640.1

Exercicio 2029: 8.500,00 € con cargo á aplicación orzamentaria 13.02.151A.640.1

Estas cantidades só cubrirán o custo da realización das actividades, polo que non levan asociado ningún beneficio económico para UdeVigo

Este convenio tramítase como expediente anticipado de gasto, posto que se prevé a existencia de crédito axeitado e suficiente no proxecto de orzamentos da Comunidade autónoma de Galicia para o exercicio 2026. Non obstante, estará sometido ao disposto na Orde da Consellería de Economía e Facenda de 11 de febreiro de 1998, modificada polas ordes de 27 de novembro de 2000 e do 25 de outubro de 2001.

Así, ao abeiro da dita normativa, o gasto proxectado queda sometido á condición suspensiva da existencia de crédito adecuado e suficiente para al fin nos orzamentos



xerais da Comunidade Autónoma de Galicia para o ano 2026, no momento da resolución de concesión.

O pagamento, en recibindo a certificación correspondente dos traballos de investigación realizados, efectuarase en oito entregas repartidas do seguinte xeito:

1ª.- 4.250,00 €, que se realizará unha vez executadas, aproximadamente, o 50 % das actividades previstas para o ano 2026 (Data límite, 31 de maio)

2ª.- 4.250,00 €, que se realizará unha vez executadas, aproximadamente, o 100 % das actividades previstas para para o ano 2026 (Data límite, 10 de decembro).

3ª.- 4.250,00 €, que se realizará unha vez executadas, aproximadamente, o 50 % das actividades previstas para o ano 2027 (Data límite, 31 de maio).

4ª.- 4.250,00 €, que se realizará unha vez executadas, aproximadamente, o 100 % das actividades previstas para para o ano 2027 (Data límite, 10 de decembro).

5ª.- 4.250,00 €, que se realizará unha vez executadas, aproximadamente, o 50 % das actividades previstas para o ano 2028 (Data límite, 31 de maio).

6ª.- 4.250,00 €, que se realizará unha vez executadas, aproximadamente, o 100 % das actividades previstas para o ano 2028 (Data límite, 31 de maio).

7ª.- 4.250,00 €, que se realizará unha vez executadas, aproximadamente, o 50 % das actividades previstas para o ano 2029 (Data límite, 31 de maio).

8ª.- 4.250,00 €, que se realizará unha vez executadas, aproximadamente, o 100 % das actividades previstas para o ano 2029 (Data límite, 31 de maio).

Cuarta. Custo total do convenio

Tendo en conta as obrigas da UdeVigo relacionadas na cláusula 2ª, e as da consellería, estipuladas na cláusula 3ª, o custo total do convenio ascende a unha cantidade de corenta e catro mil euros (44.000,00 €), distribuídos anualmente tal e como se recolle no Anexo II

Quinta. Libramento

Para recibir as achegas da Consellería que se indican anteriormente, a UdeVigo entregará na Secretaría Xeral da Lingua, antes das datas indicadas, a seguinte documentación:

- Certificación da persoa xerente da universidade que acredite a execución das actividades do convenio ao longo do exercicio correspondente, xunto coas pertinentes notas de cargo emitidas polo servizo que lle corresponda á UdeVigo para a súa tramitación, así como unha relación valorada das horas dedicadas a cada unha das accións do programa de investigación.



- Unha vez executadas as actividades do convenio, e antes do segundo libramiento de cada ano, os investigadores responsables destas emitirán unha memoria detallada das actuacións realizadas.

No caso de que a acreditación dos custos sexa inferior ao importe do convenio, a achega da consellería diminuírase na mesma proporción.

A Universidade facilitará toda a información que lle sexa requirida pola consellería, así como pola Intervención Xeral da Comunidade Autónoma, Consello de Contas e o Tribunal de Contas no exercicio das súas funcións de fiscalización e control.

Antes de proceder á sinatura do presente convenio e do seu pagamento, a UdeVigo deberá presentar un certificado de non ter contraído débedas, por ningún concepto, coa Administración pública da Comunidade autónoma galega, así como de estar ao día das obrigas tributarias e ao corrente das obrigas sociais.

Sexta. Propiedade ou titularidade dos traballos e publicidade

A propiedade, titularidade e uso dos traballos será común polas partes e só poderán usalos para os fins de cada una delas sempre de xeito gratuíto.

Nas actuacións informativas e divulgativas, nas publicacións que se fagan relacionadas coas actividades deste convenio ou cando a UdeVigo ou calquera dos investigadores que participe nas actividades obxecto deste convenio desexe utilizar os resultados acadados para a súa publicación, deberán facer constar, nun lugar destacado, a colaboración da Consellería, a través do CRPIH, e da UdeVigo para o que se usarán os anagramas aprobados pola normativa de imaxe corporativa de ambas as institucións.

Sexa cal for o medio de difusión, cada unha das partes comprométese a facer mención da outra parte e deste convenio.

Sétima. Regulación da participación dos investigadores

A participación dos investigadores da UdeVigo nas actividades recollidas no presente convenio, efectuarase conforme á regulamentación propia da UdeVigo recollida nos seus estatutos e no Regulamento de cursos de especialización e convenios, e de acordo coa Lei de incompatibilidades do persoal ao servizo das administracións públicas e coa Lei orgánica 6/2001, do 21 de decembro de universidades.

Oitava. Utilización da lingua galega

Todas as accións e actividades que se realicen ao abeiro deste convenio deberanse levar a cabo de acordo co contemplado 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 do respecto rigoroso da toponimia oficial nos termos previstos no artigo 10 da citada lei.

Novena. Vixencia

Este convenio entrará en vigor a partir do 1 de xaneiro de 2026 ou no momento da súa sinatura, se é posterior ao 1 de xaneiro de 2026, e terá vixencia durante os anos 2026, 2027, 2028 e 2029, sen posibilidade de prórroga, e poderá ser denunciado por calquera das partes asinantes por incumprimento do que nel se establece, cunha antelación dun mes polo menos, e a consellería poderá, se fose o caso, esixir o reintegro das cantidades achegadas para a súa realización.

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- presidida polo secretario xeral de Política Lingüística e integrada por un representante da UdeVigo, nomeado polo seu reitor, así como polo secretario do Consello científico e executivo do CRPIH, que actuará como secretario da mesma.

Décimo primeira. Causas de resolución do convenio

Serán causa de resolución do convenio:

- a) O mutuo acordo das partes.
- b) O incumprimento das estipulacións do mesmo, previa reunión da Comisión de Seguimento.

A denuncia do convenio poderá ser realizada por calquera das partes, debendo comunicalo por escrito coa antelación establecida na cláusula novena.

Décimo segunda. Normas de aplicación supletorias

Ao presente convenio seralle aplicable subsidiariamente a seguinte normativa:

- I. Lei 40/2015, do 1 de outubro, de Réxime Xurídico do Sector Público (artigos 47 e seguintes), na medida que queda excluído do ámbito de aplicación da Lei 9/2017, de 8 de novembro, de Contratos do Sector Público (artigo 6).

Décimo terceira. Protección de datos de carácter persoal e garantía dos dereitos dixitais.

As partes comprométese a cumprir o establecido no *Regulamento (UE) 679/2016 do Parlamento Europeo e do Consello, de 27 de abril de 2016, relativo á protección de datos das persoas físicas no que respecta ao tratamento de datos persoais e á libre*



circulación destes datos (RXPd) e na Lei orgánica 3/2018, de 5 de decembro, de protección de datos persoais e garantía dos dereitos dixitais (LOPDGDD) e no resto da lexislación española aplicable nesta materia.

O tratamento de datos faise de acordo co regulamento e coa coa lei orgánica citados, así como co resto da lexislación española aplicable, e este tratamento responde aos principios de licitude, lealdade, transparencia, limitación da finalidade, minimización dos datos, exactitude, limitación do prazo de conservación, integridade, confidencialidade e responsabilidade proactiva.

A información que as partes puidesen revelar para a consecución deste convenio e referida ás súas actividades, así como as que se revelen como consecuencia da súa execución terá a consideración de confidencial, debendo gardar segredo sobre toda a información á que poidan ter acceso, coa salvidade de que a mesma sexa de dominio público o se coñecese legal ou lexitimamente pola outra

Décimo cuarta. Responsabilidades das actuacións

O asinamento deste convenio non implica relación laboral, contractual ou de calquera outro tipo entre os profesionais que vaian desenvolver as actividades previstas e a Xunta, de tal xeito que a esta non se lle pode esixir responsabilidade ningunha, nin directa, nin indirecta, nin subsidiaria, polos actos ou feitos acaecidos no desenvolvemento do convenio.

Décimo quinta. Rexistro de convenios. Transparencia e bo goberno

A sinatura do presente 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 de Galicia e co Decreto 126/2006, do 20 de xullo, polo que se regula o rexistro de convenios da Xunta de Galicia.

Así mesmo, coa súa firma, as partes manifestan o seu consentimento para que os datos persoais que aparecen nel e máis o resto das especificacións contidas no mesmo poidan ser publicados no Portal da Transparencia e Goberno Aberto.

Décimo sexta. 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.

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 elas, o interesado deberá contar cun certificado electrónico asociado ao NIF que figure como destinatario da notificación (persoa física ou xurídica).

Décimo sétima. Administración electrónica

As partes asinantes 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 repectar 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. Normativa aplicable e xurisdición

Este convenio ten carácter administrativo e os seus efectos rexeranse polo establecido nas súas estipulacións, e quedará suxeito ao disposto na disposición adicional segunda da Lei 1/2016, do 18 de xaneiro, de transparencia e bo goberno de Galicia.

Aínda que é un dos excluídos do ámbito de aplicación da Lei 9/2017, do 8 de novembro, de contratos do sector público (BOE 9-11-17) segundo especifica o seu artigo 6, ao non previsto, e de acordo co artigo 4 da citada lei, aplicaráselle os principios recollidos nela, co fin de resolver dúbidas e omisións que puidesen presentarse. Así mesmo, supletoriamente terase en conta o Decreto legislativo 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 cos termos deste convenio, as partes comparecentes asinan, no lugar e data mencionados.

Pola Consellería de Cultura, Lingua e Xuventude

Pola Universidade de Vigo

D. José López Campos

D. Manuel Joaquín Reigosa Rogel

Conselleiro de Cultura, Lingua e Xuventude

Reitor da UdeVigo



ANEXO I

As actividades de investigación que ha desenvolver a UdeVigo que se fai referencia na cláusula segunda do presente convenio de colaboración, relaciónanse a continuación:

ÁREA DE LINGÜÍSTICA

LINGÜÍSTICA COMPUTACIONAL

ACCIÓN 1

1. RECURSOS PARA O DESENVOLVEMENTO DAS TECNOLOXÍAS DA FALA

Trátase dunha das áreas máis dinámicas dentro da intelixencia artificial e da linguaxe computacional. O seu desenvolvemento require de importantes recursos lingüísticos, tecnolóxicos e humanos e a investigación neste ámbito constitúe a base sobre a que se constrúe o futuro destas aplicacións. Sen recursos lingüísticos ben desenvolvidos, as linguas que non contan con gran presenza nos mercados globais corren o risco de seren excluídas dos avances tecnolóxicos.

Neste contexto, a situación das linguas minorizadas, require unha atención especial. A súa integración nas tecnoloxías da fala non só é unha cuestión técnica, senón tamén unha cuestión de dereitos lingüísticos e de preservación da diversidade cultural. Promover a investigación en galego neste campo significa garantir a súa presenza nos dispositivos móbiles, nas aulas dixitais, nas plataformas de intelixencia artificial e, en definitiva, na sociedade do futuro.

No marco do presente convenio, o Grupo de Tecnoloxías Multimedia (GTM) da Universidade de Vigo e o CRPIH seguirá centrando as súas actuacións na elaboración e potenciación de recursos para o desenvolvemento das tecnoloxías da fala en lingua galega co fin de garantir a súa presenza e funcionalidade no mundo dixital actual.

Durante este novo período 2026-2029 continuarase coa recollida, análise e tratamento de rexistros sonoros e de vídeo, coa súa correspondente anotación en distintos niveis.

Contéplase seguir adquirindo novos rexistros que conteñan tamén información en lingua de signos, o que potenciará o desenvolvemento de novas liñas de investigación.

Ademais, considerarase a recollida de corpus textuais.



Unha vez máis trátase de que os recursos elaborados sexan polivalentes. Con esta finalidade, os distintos niveis de anotación serán discernibles e utilizables por separado, en función das necesidades da aplicación final. Por este motivo serán tamén útiles para investigadores doutros ámbitos, particularmente do campo da lingüística.

ACCIÓN 2

1. PROXECTO "ETIQUETADOR-LEMATIZADOR PARA O GALEGO ACTUAL"

Previsión de tarefas a desenvolver nos anos 2026-2029

No eido do convenio 2022-25, estudouse a viabilidade dos modelos de *aprendizaxe automática* (AA) para a detección de *expresións multipalabra* (EMs), centrándonos nas *entidades nomeadas* (ENs) e *locucións* xa etiquetadas nos corpora Xiada e CORGA. Nesta liña, ensaiáronse diferentes técnicas de *etiquetación de secuencias*. Pretendemos ampliar este estudo no campo da detección e identificación de EMs en galego, a través dunha serie de tarefas relacionadas:

- Xeración automática de recursos de adestramento.
- Detección do carácter idiomático das expresións.
- Identificación de EMs verbais.
- Avaliación de LLMs (*Large Language Models*) en tarefas de detección/identificación de EMs.

Xeración automática de recursos de adestramento.

Unha necesidade imperativa no desenvolvemento de ferramentas informáticas baseadas en AA son os recursos de adestramento. No noso caso, corpora etiquetados para a detección e/ou identificación de EMs. A solución ideal pasaría por expandir a anotación de Xiada e CORGA a novos tipos de EMs, máis alá de ENs e locucións. Ausente esa posibilidade, unha alternativa é a xeración de recursos de xeito sintético. Un xeito de abordar o problema é a xeración de novos casos a partir de textos xa anotados, mediante técnicas específicas. Aquí podemos mencionar a inserción, borrado e intercambio de palabras [Wei and Zou, 2019]; a substitución por sinónimos [Liesting et al., 2021], palabras relacionadas semanticamente [Marivate and Sefara, 2020], ou preditas usando modelos da linguaxe [Wu et al., 2019]. Outra estratexia é a xeración de paráfrases, mediante o método de *backtranslation* [Senrich et al., 2016] ou modelos adestrados a tal obxecto [Kumar et al., 2019]. Finalmente, pódese optar pola interpolación das representacións vectoriais de exemplos existentes para xerar outros novos sen transformar as entradas [Zhang et al., 2017].



Aínda que prometedoras, estas técnicas requiren unha cantidade significativa de texto anotado do que non dispoñemos. Ademais, na súa maior parte, trátase de metodoloxías pensadas para o adestramento de clasificadores de textos, pero pouco axeitadas de etiquetadores de secuencias. Aínda que os primeiros poden ser usados en tarefas como a determinación do carácter idiomático de posibles EMs, son os segundos os que, como no caso dos modelos *transformer*, estamos a utilizar para a detección de EMs en novo texto. Por iso, sen descartar os métodos anteriores, cremos máis prometedor o **uso de LLMs para a obtención de texto anotado sintético**, como ChatGPT [Piedboeuf and Langlais, 2023], GPT-4 [Møller et al., 2024], ou opcións libres como Llama3 [Dubey et al., 2023] ou DeepSeek-V3 [Liu et al., 2024]. Tamén existen alternativas monolingües para o galego, como Carballo-bloom-1.3B e Carballo-cerebras-1.3B [Gamallo et al., 2025].

Estes LLMs están adestrados para producir unha resposta, en forma de texto, a partir das indicacións textuais (*prompts*) dos seus usuarios. A tarefa do usuario é deseñar os *prompts* apropiados para obter o resultado desexado (*prompt engineering*), a xeración automática de texto anotado no noso caso. Esta dinámica abre a posibilidade de estudar as diferencias na calidade dos textos sintéticos resultantes debidas ao uso de diferentes LLMs, ás diversas estratexias para xerar *prompts*, eá incluso nestes duns poucos exemplos de texto anotado (*few-shot learning*) ou ningún (*zero-shot learning*).

A discusión anterior tamén é pertinente para unha segunda arquitectura de xeración para texto anotado: a **anotación automática por LLMs** de texto extraído dun corpus sen anotar.

Unha terceira vía de xeración automática de recursos parte do **uso de corpora paralelos**. Un corpus deste tipo pódese dividir en segmentos, cada un contendo a tradución a ambos idiomas do mesmo fragmento de texto. Aplicando un proceso de extracción de candidatos a EM nos dous idiomas, mediante métricas de distancia, información mutua ou heurísticas; pódense identificar pares de EMs que aparecen con frecuencia nos mesmos segmentos de texto aliñado. Outra posibilidade é partir dun lexicón de EMs nunha das linguas e obter as EMs correspondentes na outra a partir da presenza simultánea desta últimas coas do lexicón [Semmar et al., 2019]. Actualmente, existen dous corpora paralelos de galego co español e inglés: Nos_ES-GL_aut (<https://zenodo.org/records/7671278>), e Nos_EN-GL_aut (<https://zenodo.org/records/7675110>), que ofrecen a posibilidade de extraer EMs. Tamén aí a posibilidade de usar algún dos lexicóns en inglés ou español listados en [Mititelu et al., 2017], se resultan apropiados. O resultado sería unha lista de EMs en galego, que podería usarse para anotar automaticamente textos extraídos de recursos como Xiada ou CORGA.



DetECCIÓN DO CARÁCTER IDIOMÁTICO DAS EXPRESIÓNS.

Distinguir entre aparicións literais e idiomáticas é unha cuestión básica no procesamento de EMs. e pódese enfocar como un problema de clasificación de textos, concretamente das frases nas que expresións posiblemente idiomáticas aparecen. Actualmente existen dous recursos para galego con potencial para ser usados nesta tarefa: o corpus de proba da competición SemEval-2022 Task 2 ([Madabushi et al. 2022]) e a base de datos de expresión con grados diversos de composicionalidade [Castro and Temerko and García, 2025]. Xa comezamos o seu estudo no último semestre do convenio 22-25, co obxectivo de xerar conxuntos de datos de adestramento.

IDENTIFICACIÓN DE EMs VERBAIS.

Un dos conxuntos de datos máis amplos neste eido é o correspondente á acción PARSEME [Savary et al. 2015]. Na súa versión 1.3 contén corpora de adestramento e proba para a identificación de EMs verbais en 26 linguas. Desgraciadamente, o galego non é un deles. Isto devólvenos a posibilidade da xeración de conxuntos de datos de adestramento e proba, seguindo algunha das técnicas automáticas comentadas previamente, cunha revisión manual posterior dos mesmos por parte de falantes nativos e conforme as guías de anotación PARSEME.

AVALIACIÓN DE LLMs PARA TAREFAS DE DETECCIÓN E IDENTIFICACIÓN DE EMs.

Nas tarefas anteriores contemplamos o uso de LLMs para a xeración automática de corpora anotados, que poden usarse no adestramento de modelos *transformer* máis pequenos, que precisan de menos recursos e que poden executarse localmente, en lugar de seguir o modelo de *software como servizo*. Pero, coas técnicas apropiadas de *prompting*, os LLMs tamén poden usarse directamente para a detección e identificación de EMs en textos sen anotar proporcionados polo usuario. Iso abre a posibilidade de estudar a súa efectividade para o seu procesamento no galego.



ANEXO II

O custo total do convenio, coas achegas correspondentes de ambas as partes asinantes, distribuirase anualmente do seguinte xeito:

Accións	Custo total anual	Horas estimadas anuais	Achega Consellería anual	Contribución UDV anual
1. Recursos para o desenvolvemento de tecnoloxías da fala	6.380,00 €	200,00	4.780,00 €	1.600,00 €
2. Etiquetador – lematizador para o galego actual	4.620,00 €	160,00	3.720,00 €	900,00 €
TOTAL ANUAL	11.000,00 €	360,00	8.500,00 €	2.500,00 €
TOTAL CONVENIO (CUADRIANUAL)	44.000,00 €	1.440,00	34.000,00	10.000,00 €

