

CONVENIO PLURIANUAL DE COLABORACIÓN ENTRE A CONSELLERÍA DE CULTURA, EDUCACIÓN E UNIVERSIDADE E A UNIVERSIDADE DA CORUÑA 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, Román Rodríguez 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 12/2007, do 27 de xullo, de acordo co seu nomeamento polo Decreto 112/2020, de 6 de setembro, e no uso das atribucións que lle confire o Decreto 130/2020, de 17 de setembro (DOG nº 190, do 18 de setembro), polo que se establece a estrutura orgánica da Consellería de Cultura, Educación e Universidade.

Doutra, Don Julio Ernesto Abalde Alonso, reitor magnífico da Universidade da Coruña, en virtude do establecido nos seus estatutos e do nomeamento como reitor polo Decreto 5/2020, do 9 de xaneiro, (DOG nº 7 de 13 de xaneiro), actuando en nome e representación da devandita entidade.

As partes actúan no exercicio dos seus respectivos cargos e coa representación competencial que ostentan, e recoñecendo 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 198/2020, do 20 de novembro (DOG do 1 de decembro) establécese a estrutura orgánica da Consellería de Cultura, Educación e Universidade, e atribúeselle á Secretaría Xeral de Política Lingüística (en diante, SXPL), entre outras,



as funcións de promover, coordinar e desenvolver a política de investigación que favoreza a normalización lingüística nas súas diferentes manifestacións, 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 da Coruña (en diante UDC) 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, Educación e Universidade (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 SXPL.

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 redunda 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 UDC e 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 UDC -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 os departamentos de Filoloxía da UDC 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 UDC 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 UDC 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 UDC, a través da SXPL, 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 UDC

As actividades que ha desenvolver a UDC 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 UDC 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 ETIQUETADOR-LEMATIZADOR PARA O GALEGO ACTUAL



A **UDC** 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 dous mil catrocentos euros (**2.400,00 €**) derivada dos custos do seu persoal docente investigador, distribuída en catro anualidades (2022, 2023, 2024 e 2025), por un importe de **600,00 €** cada unha.

Ademais, a UDC 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 UDC deberá cumprir coas estipulacións xurdidas do vínculo xerado coa celebración do presente convenio no xeito e termos establecidos.

Terceira. Obrigas da Consellería de Cultura, Educación e Universidade

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 de Política Lingüística colaborará na realización destas accións cunha achega económica de **catorce mil catrocentos euros (14.400,00 €)**, repartidos en catro anualidades coa seguinte distribución:

Exercicio 2022: 3.600 € con cargo á aplicación orzamentaria 10.50.151A.640.1

Exercicio 2023: 3.600 € con cargo á aplicación orzamentaria 10.50.151A.640.1



Exercicio 2024: 3.600 € con cargo á aplicación orzamentaria 10.50.151A.640.1

Exercicio 2025: 3.600 € con cargo á aplicación orzamentaria 10.50.151A.640.1

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

Para atender os compromisos de gasto derivados das accións establecidas existe crédito axeitado e suficiente na aplicación orzamentaria sinalada, dos presupostos xerais da comunidade autónoma para os citados exercicios orzamentarios.

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

1ª.- 1.800,00 € (1 de xuño de 2022)

2ª.- 1.800,00 € (10 de novembro de 2022)

3ª.- 1.800,00 € (1 de xuño de 2023)

4ª.- 1.800,00 € (10 de novembro de 2023)

5ª.- 1.800,00 € (1 de xuño de 2024)

6ª.- 1.800,00 € (10 de novembro de 2024)

7ª.- 1.800,00 € (10 de xuño de 2025)

8ª.- 1.800,00 € (10 de novembro de 2025)

Cuarta. Custo total do convenio

Tendo en conta as obrigas da UDC, relacionadas na cláusula 2ª, e as da consellería, estipuladas na cláusula 3ª, o custo total do convenio ascende a unha cantidade de dezaseis mil oitocentos **euros (16.800,00 €)**, distribuídos anualmente tal e como se recolle no Anexo II

Quinta. Libramento

Para recibir as achegas da Consellería de Cultura, Educación e Universidade que se indican anteriormente, a UDC entregará na Secretaría Xeral de Política Lingüística, 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 UDC 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 libramento 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 UDC 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 UDC 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 gratuito.

Nas actuacións informativas e divulgativas, nas publicacións que se fagan relacionadas coas actividades deste convenio ou cando a UDC 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 de Cultura, Educación e Universidade, a través do CRPIH, e da UDC, 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 UDC nas actividades recollidas no presente convenio, efectuarase conforme á regulamentación propia da UDC 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 o mesmo día da súa sinatura e terá vixencia durante os anos **2022, 2023, 2024 e 2025**, 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 UDC, 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 comprométense 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 e polo que se derroga a Directiva 95/46/CE (Regulamento xeral de protección de datos)*, así como na *Lei orgánica 3/2018, de 5 de decembro, de protección de datos persoais e garantía dos dereitos dixitais* 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 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. 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 por triplicado, no lugar e data mencionados.

Pola consellería de Cultura, Educación e
Universidade

Pola Universidade da Coruña

D. Román Rodríguez González
Conselleiro de Cultura, Educación e Universidade

D. Julio Ernesto Abalde Alonso
Reitor da UDC



ANEXO I

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

ÁREA DE LINGÜÍSTICA

LINGÜÍSTICA COMPUTACIONAL

ACCIÓN

PROXECTO ETIQUETADOR /LEMATIZADOR DO GALEGO ACTUAL

No marco da edición 2020-2021 desenvóléronse técnicas de aprendizaxe profunda para detectar entidades nomeadas (NER, named entity recognition).

O recoñecemento de entidades nomeadas é a tarefa de localizar referencias a entidades en textos e clasificalas en categorías predefinidas. A día de hoxe, NER é un compoñente crucial de calquera aplicación de minaría de textos, como xa se demostrou en moitos dominios diferentes (desde a moda ata a documentación legal) e resulta ser especialmente relevante en aplicacións biomédicas.

Desde o punto de vista da implementación, NER estivo ligado ao paradigma de etiquetaxe de secuencias desde o seu inicio, un tipo de tarefa de clasificación onde a entrada é unha secuencia de valores observados e a saída é unha etiqueta categórica para cada membro da secuencia. En procesamento da linguaxe natural (NLP, *natural language processing*), os valores de entrada adoitan ser palabras, aínda que, dependendo da tarefa, tamén poden ser unidades máis pequenas como caracteres individuais ou unidades máis grandes, como frases.

A etiquetaxe das parte do discurso (PoS, *part-of-speech*) é probablemente o exemplo máis arquetípico aplicado a NLP, porque a tarefa en si defínese como asignar unha etiqueta (unha parte do discurso, como un verbo ou un substantivo) a cada elemento (palabra) en unha secuencia. Así, podería dicirse, por definición, que calquera implementación de etiquetaxe de PoS está a realizar unha forma de etiquetaxe de secuencias, aínda que as implementacións pioneiras fosen específicas para a tarefa (así como para a linguaxe e para o conxunto de etiquetas), xa que estaban baseadas en regras manuscritas. Non obstante, os enfoques máis modernos presentan modelos adestrables que aprenden a correspondencia palabra-etiqueta no contexto como un problema de aprendizaxe supervisada e,



polo tanto, son exemplos de verdadeiros modelos xenéricos de etiquetaxe de secuencias.

A fragmentación, ás veces chamada análise superficial, implica atopar frases relevantes (normalmente verbos, substantivos e/ou frases preposicionais) nun texto. Aínda que os primeiros enfoques tamén foron *ad hoc*, o problema foi reformulado como un problema de etiquetaxe introducindo un esquema de codificación chamado etiquetaxe IOB (*inside, outside, beginning*). Neste enfoque, a cada palabra asígnaselle unha etiqueta "I", "O" ou "B" dependendo de se ocorre "Dentro", "Fóra" ou no "Inicio" dun fragmento. Desde entón, a maioría dos enfoques de fragmentación, incluídos os enfoques de aprendizaxe automática máis recentes e os enfoques de aprendizaxe profunda, utilizaron o esquema de IOB ou variantes deste. O NER está moi relacionado coa fragmentación, no sentido de que o obxectivo é extraer segmentos relevantes de palabras nun texto, e tamén se pode aplicar o concepto de etiquetado IOB. Polo tanto, non é de estrañar que os enfoques de etiquetado de secuencias para NER, baseados en variantes de etiquetado IOB, sexan populares desde os primeiros anos e seguisen a evolución xeral das arquitecturas de etiquetaxe descritas anteriormente.

No contexto deste convenio, o principal problema para a aplicación de estratexias de aprendizaxe automática e aprendizaxe profunda é a falta de textos anotados para NER en galego. Coa colaboración dos lingüistas do CRPIH, pretendemos mellorar (en tamaño e calidade da anotación) os *corpora* dispoñibles (CORGA e Xiada), sen desbotar a opción de converter un corpus que xa estea dispoñible para algunha lingua próxima, como pode ser o portugués ou o español. Unha vez que se teña un corpus anotado de calidade, a plataforma Bertiño [2] parece un bo punto de partida, xa que amosou moi bos resultados nas primeiras probas experimentais.

Por último, é interesante comentar que os enfoques antes descritos para a detección de entidades nomeadas non teñen en conta a estrutura do texto. Polo tanto, a integración de información sintáctica podería supoñer unha fonte de información adicional, aínda que non sempre de utilidade práctica, xa que a miúdo o uso de técnicas baseadas na xerarquización pode resultar moi complexo para o tratamento de fenómenos de estrutura puramente liñal. Con todo, unha alternativa atractiva pode ser a interpretación da análise sintáctica coma unha simple tarefa de etiquetaxe de secuencias, o que debería permitir integrar a información sintáctica sen necesidade de renunciar ás técnicas estándar.



Deste xeito, todo o exposto anteriormente motiva que as tarefas de investigación que temos pensado desenvolver na UDC durante os anos 2022 a 2025 sexan as seguintes:

- Exploración dos corpora existentes e análise de alternativas para desenvolver un corpus para NER en galego a través de mecanismos translingües.
- Definición de estratexias para o recoñecemento de entidades nomeadas en galego.
- Estudo da posibilidade de mellorar o recoñecemento das devanditas entidades mediante o uso de información sintáctica.



ANEXO II

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

Acción	Custo total anual	Horas estimadas anuais	Achega Consellería anual	Contribución UDC anual
2. "Etiquetador-lematizador para o galego actual"				
TOTAL ANUAL	4.200,00 €	140,00 €	3.600,00 €	600,00 €
TOTAL CONVENIO (CUATRIANUAL)	16.800,00 €	560 €	14.400,00 €	2.400,00 €

