

## INFORME

### PREPARACIÓN DEL PILOTO IV "PATRIMONIO LÉXICO DA GALLAECIA"

Este informe presenta con fecha de 30 de junio de 2019 los objetivos del piloto IV, su metodología y los resultados esperados hasta el fin del año 2019.

#### 1. OBJETIVOS

El objetivo principal del Piloto IV *Patrimonio léxico de la Gallaecia* en estos meses es el desarrollo de una aplicación de acceso abierto que permita realizar consultas por conceptos y conocer las diferentes formas léxicas que sirven para expresar determinado concepto en la Eurorregión Galicia-Norte de Portugal, así como su distribución territorial, representada cartográficamente. Para que estas búsquedas ofrezcan resultados más ricos, incrementaremos continuamente el número de registros correspondientes a dicha Eurorregión en el *Tesouro do léxico patrimonial galego e portugués* (TLPGP), base de datos del que la nueva aplicación que se va a presentar toma los datos léxicos. Este trabajo se enmarca en el objetivo más general de preservar, visibilizar, difundir y poner el valor el léxico patrimonial de la Eurorregión Galicia-Norte de Portugal, como un elemento fundamental de su patrimonio cultural. De este modo, la aplicación, de acceso abierto, tal y como indicamos, y fácil consulta, servirá no solo para facilitar estudios de diferente tipo a los investigadores, sino para mostrar a la sociedad la enorme riqueza léxica que posee y su relevancia, para poner en valor el vocabulario dialectal, potenciando así su uso y con este su preservación y para mostrar así mismo el continuum léxicode la Eurorregión y su carácter de comunidad cultural.

Para el desarrollo de esta nueva aplicación de visualización sobre una base cartográfica fue preciso previamente crear otras dos, de uso interno, cuyas características ofrecemos a continuación.

#### 2. METODOLOGÍA

##### 2.1 Aplicaciones para la vinculación de geosinónimos y para la traducción de conceptos

###### 2.1.1 Objetivo de la aplicación y versiones

Tal y como acabamos de indicar, el punto de partida para la obtención de la información léxica es la base de datos del *Tesouro do léxico patrimonial galego e portugués* (TLPGP) (<http://ilg.usc.es/Tesouro/gl>), que recopila vocabulario gallego y portugués presente en diferentes obras dialectales, publicadas o inéditas. En este recurso se le asigna a cada elemento léxico un clasificador semántico, lo que permite realizar una clasificación onomasiológica básica.

Con el objetivo de vincular los diferentes geosinónimos gallegos y portugueses incluidos en la base de datos (paso previo necesario para permitir búsquedas por conceptos y obtener la distribución geográfica cartografiada de tales geosinónimos) se encargó a la empresa NLPgo Technologies, S.L. una aplicación que permitiese la asignación de hiperlemas o conceptos a las diferentes variantes y lemas de la base de datos del TLPGP. La primera versión de esta aplicación fue entregada por la empresa citada en el verano de 2018 y permite asignar

---

conceptos exclusivamente a las formas gallegas. Al mismo tiempo, la aplicación permite filtrar las búsquedas por variante, lema, clasificador semántico y concepto.

En la imagen que figura a continuación se ofrece una captura de pantalla de la aplicación en esta versión inicial.

**Vinculación de xeosinónimos**  
Tesouro de léxico patrimonial

Variante	Lema	Concepto
<input type="text"/>	<input type="text"/>	<input type="text"/>
Idioma	Clasificador semántico	
gl		
<input type="button" value="Limpiar selección"/>	<input type="button" value="Limpiar todo"/>	<input type="button" value="Buscar"/>
		<input type="text" value="Concepto"/> <input type="button" value="Aplicar aos seleccionados"/>

0 Registros

Variante	Lema	Código	Clasificador semántico	Concepto
No Rows To Show				

Imagen 1. Primera versión de la aplicación de vinculación de geosinónimos

A partir de la experiencia de trabajo con esta aplicación, solicitamos mejoras para facilitar la labor de vinculación de geosinónimos, incorporando la posibilidad de hacer búsquedas por palabras de las definiciones y filtraje por obra. Al mismo tiempo se incluyó el acceso a las variantes y lemas portugueses. Además, con la finalidad de conectar el lemario gallego y portugués, se incluyó en la aplicación un sistema de vinculación, lo que permite que en la aplicación de visualización cartográfica que se desarrollará posteriormente, los lemas gallegos y portugueses cognados se representen con el mismo símbolo. Se ofrece a continuación una captura de pantalla de la aplicación en su versión actual, que requiere de credenciales de acceso para poder ser utilizada (<http://ilg-pw.usc.es/tesouro>). En esta nueva versión también se modificó el sistema técnico de incorporación de conceptos, tras detectar un problema de envergadura, que se detalla en 2.1.3.

## Vinculación de xeosinónimos

### Tesouro de léxico patrimonial

Variante	Lema	Concepto	Idioma
<input type="text"/>	<input type="text"/>	<input type="text"/>	gl <input type="button" value="x"/> pt-eu <input type="button" value="x"/>
Definición	Obra	Clasificador semántico	
<input type="text"/>	<input type="text"/>	<input type="text"/>	
Lema vínculo	<input type="button" value="Aplicar aos seleccionados"/>	Concepto	<input type="button" value="Aplicar aos seleccionados"/>
<input type="button" value="Limpar selección"/>	<input type="button" value="Limpar todo"/>	<input type="button" value="Buscar"/>	

0 Rexistros

Variante	Lema	Clase	C...	Clasificador semánt...	Lema vínculo	Concepto
No Rows To Show						

Imagen 2. Versión actual de la aplicación para vincular geosinónimos

#### 2.1.2 Funcionamiento de la aplicación

El método de trabajo para la asignación de conceptos y la vinculación de lemas gallegos y portugueses es el siguiente:

- 1) En primer lugar, tras establecer el concepto que se quiere incorporar, se realiza un filtrado a partir de la clasificación semántica establecida en el TLPGP. Dentro del resultado obtenido, se realiza una búsqueda de las diferentes expresiones existentes en la base de datos para ese concepto. Es de gran auxilio en la realización de esta tarea la posibilidad de buscar por palabras de las definiciones.
- 2) En segundo lugar, se le asigna el concepto a las variantes y lemas encontrados (para establecer la denominación del concepto y su sistema de aplicación elaboramos unos criterios metodológicos, que figuran como anexo 1). A tal fin se seleccionan los lemas / variantes que interesa, se escribe en la caja correspondiente el concepto que interesa y se aplica a los seleccionados. El concepto se puede escribir también directamente en la celda de la tabla (vid. imagen 3).
- 3) En tercer lugar, y con el objetivo de vincular los lemas gallegos y portugueses cognados establecemos un lema vínculo que incorporamos en el campo correspondiente. Si no hay cognados gallegos y portugueses que vincular, el campo queda vacío. Para introducir el lema vínculo se seleccionan los lemas que interesa conectar, se escribe en la caja correspondiente el lema vínculo que interesa y se aplica a los seleccionados. El lema vínculo se puede escribir también directamente en la celda de la tabla (vid. imagen 3). El lema vínculo solo se le puede asignar a lemas que tienen atribuido previamente un concepto.

Así, por ejemplo, en la siguiente captura de pantalla, puede observarse como los editores le asignaron a algunos de los lemas y variantes situados bajo el clasificador semántico "Animais silvestres (denominacións)" el concepto 'donicela'. Al mismo tiempo, le asignaron al lema portugués *doninha* y al lema gallego *doniña* un lema vínculo (*doniña*) que permite establecer una conexión entre ellos. Todos los lemas que en este campo tienen asignado idéntico lema vínculo serán representados con el mismo símbolo en la aplicación cartográfica que comentaremos en 2.2 (el lema vínculo *denociña* que figura también en el mismo campo sirve para unir el lema portugués *denezinha* y el lema gallego *denociña*, que no figura en la captura; al lema gallego *donicela* no se le asigna lema vínculo, pues no existe cognado portugués).

58 Registros

Variante	Lema	Clase	C...	Clasificador semán...	Lema vínculo	Concepto
donicela	donicela	sf	6.1	Animais silvestres (d...		donicela
dónina	doninha	sf	6.1	Animais silvestres (d...	doniña	donicela
doninha	doninha	sf	6.1	Animais silvestres (d...	doniña	donicela
doninha	doninha	sf	6.1	Animais silvestres (d...	doniña	donicela
doninha	doninha	sf	6.1	Animais silvestres (d...	doniña	donicela
doninha	doninha	sf	6.1	Animais silvestres (d...	doniña	donicela
dóninha	doninha	sf	6.1	Animais silvestres (d...	doniña	donicela
doniña	doniña	sf	6.1	Animais silvestres (d...	doniña	donicela
doniña	doniña	sf	6.1	Animais silvestres (d...	doniña	donicela
dónozinha	denezinha	sf	6.1	Animais silvestres (d...	denociña	donicela
doniña	doniña	sf	6.1	Animais silvestres (d...	doniña	donicela
dunicela	donicela	sf	6.1	Animais silvestres (d...		donicela
dunicela	donicela	sf	6.1	Animais silvestres (d...		donicela

Imagen 3. Introducción de conceptos y lemas vínculo

En esta aplicación de trabajo interno, los conceptos se asignan siempre en gallego. Con todo, nuestro objetivo es que en la aplicación de visualización que se comentará en 2.2, los conceptos pueden buscarse tanto en gallego como en portugués. Por tal motivo NLPGo desarrolló también una aplicación de traducción de los conceptos gallegos al portugués (<http://ilg-pw.usc.es/tesouro/hiperlemas>) de la que obtendrá la información en la aplicación cartográfica. Esta aplicación también requiere de credenciales de acceso para su uso.

Hiperlema gl ↑	Hiperlema pt
centeo	centeo
cerdeira	cerejeira
chícharo	ervilha
chuvia	chuva
codia	côdea
coello	coelho
debullar	debulhar
donicela	doninha
entroido	entrudo

Imagen 4. Aplicación de traducción de los conceptos

### 2.1.3 Características técnicas

El modelo de base de datos de partida es el que se muestra a continuación:





que las tuplas se eliminan y se crean otras con el mismo contenido, pero con diferente clave primaria).

Este segundo punto no afectó sólo a que la relación tuviese que implementarse como N:M, sino también a que la relación entre los hiperlemas y las entradas (entradas\_hiperlemas) sea un tanto especial. En lugar de referenciar a la clave primaria de las entradas, que son eliminadas y regeneradas por otros procesos relacionados con la carga y modificación de datos, se hizo una referencia por valor, es decir, que cada vez que se establece una relación, se copian los valores de la entrada relacionada en la tabla entradas\_hiperlemas.

De este modo, aunque la tabla de entradas se regenere, los hiperlemas siguen estando conectados con las entradas, siempre y cuando alguno de los valores que determinan la relación no cambie. Estos valores son lema , codigo , variante y entrada\_definicion y, por lo tanto, si alguno de ellos cambia para una entrada, y tenía anteriormente asignado un hiperlema y, opcionalmente, también un lema vínculo, esas dos informaciones se perderían, y habría que volver a asignarle el concepto (o el lema vínculo).

Para evitar problemas, las tuplas de entradas\_hiperlemas que quedan huérfanas no se eliminan, por lo que para saber si una tupla de entradas\_hiperlemas está vigente en un momento dado es necesario ver si está asociada con una entrada y un hiperlema. Las tuplas que no cumplen esto son valores antiguos que ya no están vigentes.

El problema al que se hacía referencia en el párrafo anterior sería que se intentasen borrar las entradas\_hiperlemas no conectadas al hacer cualquier operación, y que justo en ese momento se estuviesen regenerando las tablas de la base de datos. Podría darse el caso de que en el momento de hacer la eliminación, muchas entradas no estuviesen aún en la tabla de entradas, y el resultado podría ser que se eliminase mucha información de conceptos ya introducidos.

La aplicación está desarrollada utilizando Ruby on Rails. Por otro lado, la herramienta tiene un sistema de autenticación muy básico que hace uso del fichero .htpasswd situado en el directorio base del directorio de instalación de la aplicación. Para editar las entradas de ese fichero puede utilizarse el comando htpasswd incluido en apache.

## **2.2 Aplicación para la consulta por conceptos y para la visualización cartográfica de los geosinónimos**

La introducción de conceptos así como la vinculación de los lemas gallegos y portugueses es un paso previo imprescindible para poder realizar búsquedas por conceptos y para poder visualizar cartográficamente la distribución de los geosinónimos que expresan un mismo concepto. Una vez diseñadas las aplicaciones anteriores e introducida información en ellas, se encargó a la empresa Enxenio S.L. el diseño de una nueva aplicación que permitiese realizar búsquedas por conceptos, así como visualizar sobre un mapa la distribución de los diferentes lemas que expresan un concepto. En estos momentos se está diseñando la aplicación, de la que existe una versión piloto inicial sobre la que se han pedido cambios y mejoras. En las imágenes 7 y 8 se ofrecen dos capturas de pantalla que sirven para mostrar, a grandes rasgos, como sería la visualización final.

---

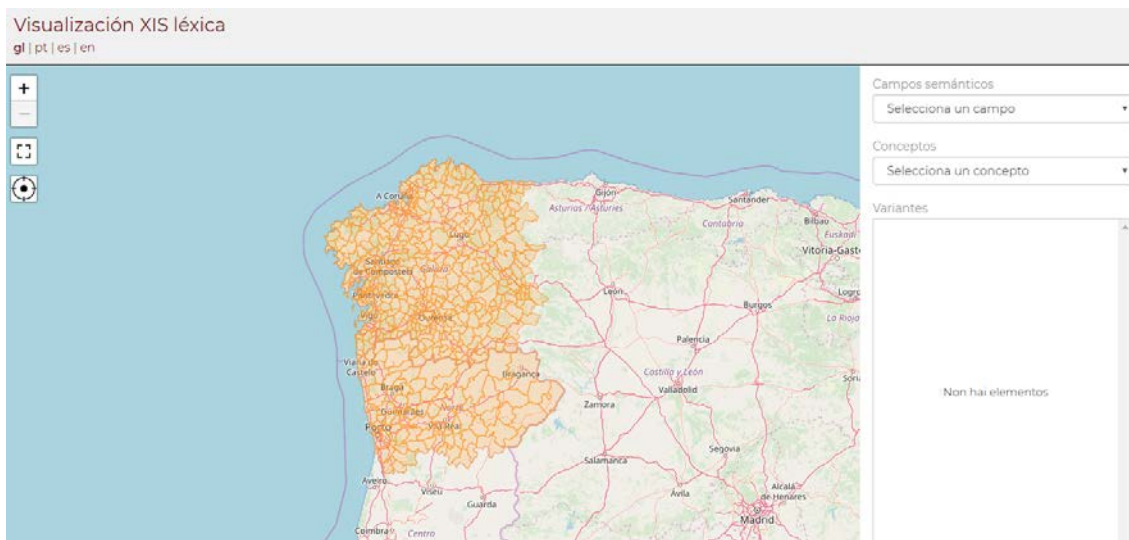


Imagen 7. Versión piloto de la aplicación. Mapa y área de consulta de los geosinónimos

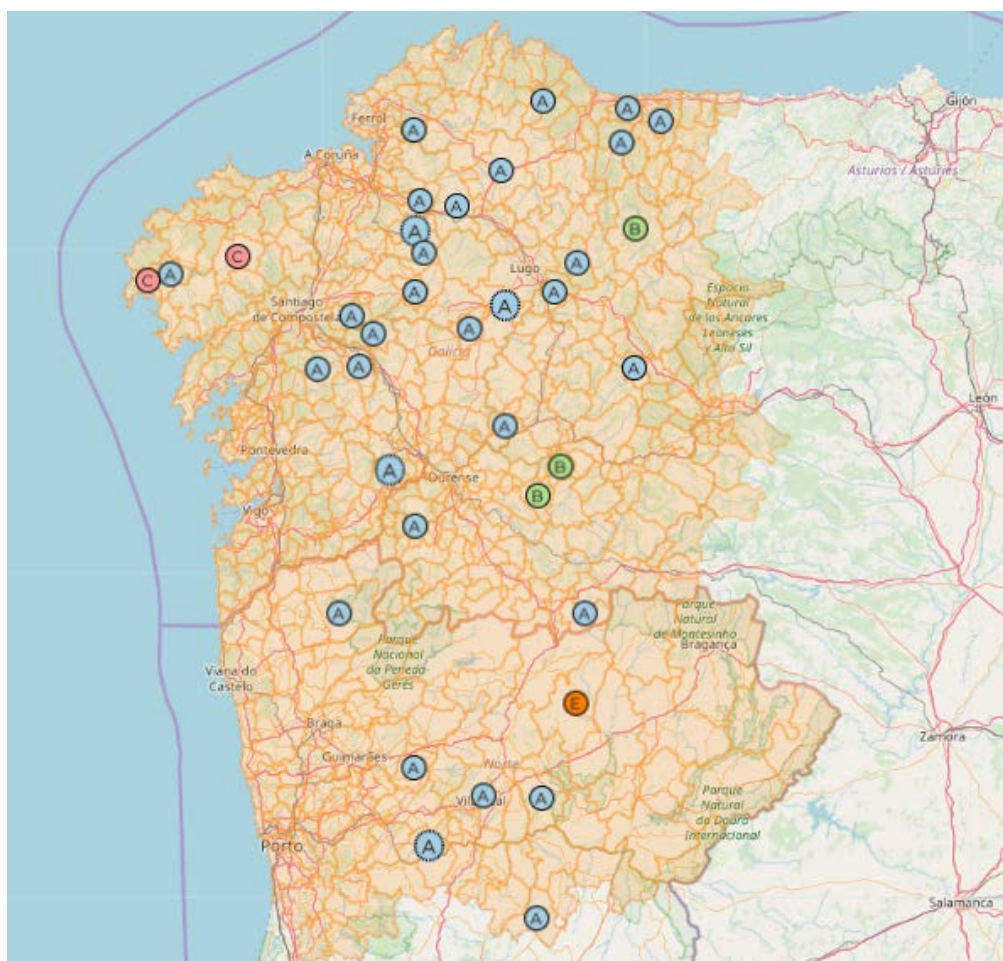


Imagen 8. Versión piloto de la aplicación. Visualización de la distribución de los geosinónimos sobre el mapa



### 2.2.1 Utilidades de la aplicación

Se pretende que la aplicación, de acceso libre, tenga el siguiente funcionamiento y las siguientes utilidades:

- 3) El usuario seleccionará el campo semántico que le interese para realizar las búsquedas y obtendrá el listado de todos los conceptos disponibles para ese ámbito temático. A continuación seleccionará entre ellos el concepto que desee buscar y obtendrá el listado de todos los lemas que expresan ese concepto, así como la visualización de su distribución en el mapa (por municipios).
- 4) El usuario podrá deseleccionar, si así lo desea, uno o varios lemas, de modo que en la cartografía solo se visualicen aquel o aquellos que le interesen.
- 5) Podrá también, si así lo desea, modificar el símbolo que representa el lema sobre el mapa (forma y color).
- 6) A partir del listado de lemas, podrá acceder al listado de variantes que expresan cada uno de esos lemas.
- 7) Además, clicando en el símbolo que representa sobre el mapa cada lema obtendrá un desplegable con el nombre del municipio, el lema o lemas correspondientes, así como con las variantes concretas que expresan ese lema en ese lugar recogidas en la base de datos del TLPGP.
- 8) A partir tanto de lemas como de variantes se podrá acceder directamente al TLPGP para obtener más información (definición, imágenes, categoría gramatical, etc.).
- 9) Los conceptos se podrán consultar en gallego y en portugués, gracias a la aplicación de traducción de conceptos a que aludimos más arriba.
- 10) Se podrá guardar el mapa en formato .png y descargar los datos en formato excel.

### 2.2.2 Características técnicas

#### 2.2.2.1 Tecnologías de desarrollo

La aplicación web que da soporte al portal está basada en una arquitectura cliente-servidor en capas, que proporciona independencia entre el acceso a datos, la lógica de la aplicación y la vista, tal y como se muestra en la siguiente figura:

---

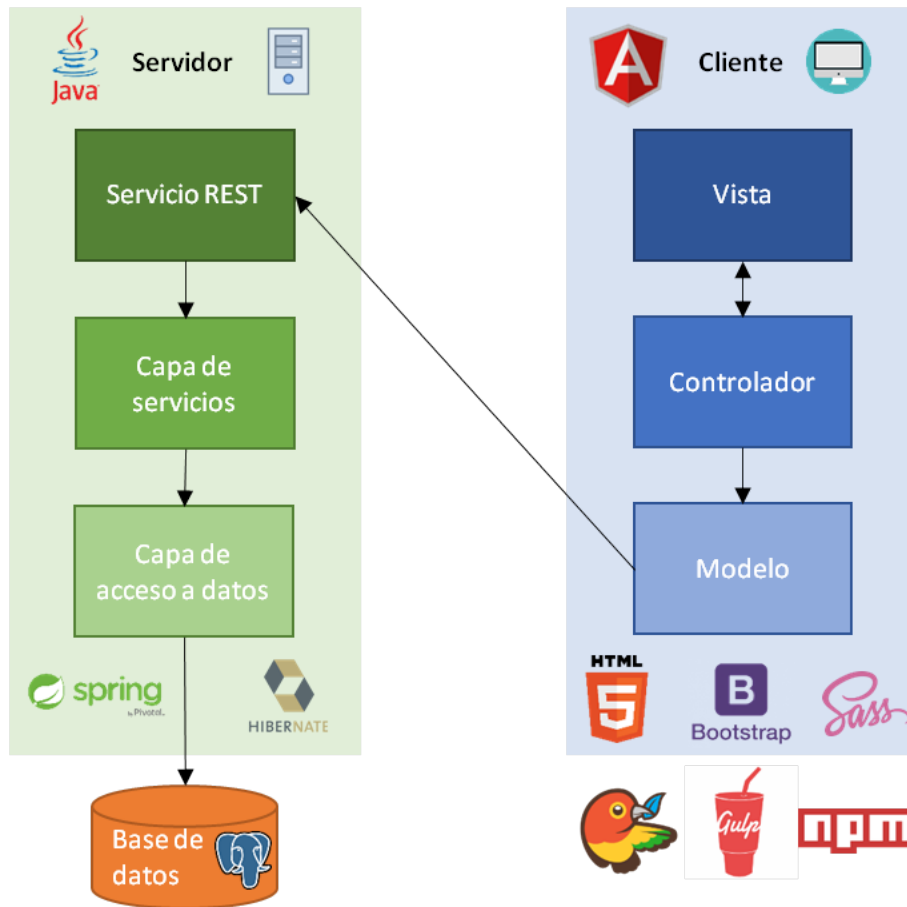


Imagen 9. Arquitectura general del sistema

Esto, unido al uso sistemático de patrones de diseño en su construcción, garantiza la robustez y escalabilidad de la aplicación construida, permitiendo que en el futuro crezca en funcionalidades.

#### 2.2.2.2 Servidor

Es el encargado de persistir los datos en base de datos y de servirlos al lado cliente a través de una serie de operaciones definidas con una API REST (*Representational State Transfer*).

En concreto, para la programación se usa Java 2 Enterprise Edition (J2EE, versión 8), que está diseñado con estándares abiertos que permiten la utilización de diferentes tecnologías en la implementación de los componentes de las distintas capas. En concreto, las tecnologías que se han utilizado en la implementación de cada capa son las siguientes:

- Servicio REST y capa de servicios: se ha utilizado *Spring Framework* (versión 4.3), en particular, las librerías de *Spring Security*, que facilita la autenticación y control de permisos en la aplicación, *Spring Boot* (versión 1.5), que agiliza la creación de aplicaciones, o el módulo de inversión de control que permite reducir el acoplamiento entre los distintos componentes, para mejorar su integración.

- Capa de acceso a datos: el Sistema Gestor de Bases de Datos (SGBD) elegido es PostgreSQL (versión 9.5), y el framework para el acceso a los datos Hibernate (versión 5.4).

#### 2.2.2.3 Cliente

Está formado por una *Single Page Application*, que, a su vez, utiliza el patrón de arquitectura Modelo Vista Controlador (*Model View Controller*).

Para el desarrollo de esta parte se utiliza, principalmente, AngularJS (versión 1.6). Angular es un *framework* MVC de JavaScript para el desarrollo web *front-end*. Además, para complementar el uso de este framework se utilizaron otros lenguajes como HTML5, CSS3, JavaScript, JQuery o Bootstrap.

Por último, para la visualización geográfica de la información en un mapa se utiliza la librería de código abierto *Leaflet* (versión 1.5), que ofrece la posibilidad de incluir *plugins* gestionados por una amplia comunidad, permitiendo añadir nuevas funcionalidades y dando como resultado un mapa más completo.

### 3 RESULTADOS

Se prevé que la aplicación que estamos desarrollando esté finalizada en verano de 2019 y se pueda abrir para consulta libre unos meses después, una vez elaborada la página web desde la que se podrá acceder y que incorporará, además de la aplicación, instrucciones y normas de uso, créditos, reconocimientos institucionales, etc. Se prevé que en ese momento, se puedan consultar alrededor de 100 conceptos. Esta cifra se ira ampliando paulatinamente. Además, se incrementará durante estos meses del número de obras y registros léxicos de la Euroregión Galicia-Norte de Portugal (alrededor de 50.000 nuevos registros), mediante la incorporación de diversos trabajos que están solo pendientes de subir a la base o en fase de revisión, así como de al menos parte de los resultados obtenidos por diferentes centros de enseñanza en el marco del proyecto *Lingua, patrimonio e coñecemento tradicional* (Plan Proxecta de la Xunta de Galicia), donde el alumnado es el encargado de recoger el léxico de su entorno.

---

## ANEXO 1

### DENOMINACIÓN DOS CONCEPTOS

#### 1) Criterios básicos

Segundo os casos, a denominación do concepto pode ser unha unidade léxica ou ben unha descrición breve do concepto. En principio a primeira vía que debemos explorar é a de optar por unha unidade léxica (vid. con todo o que se indica ao final deste apartado). As dúas características básicas que se deben ter en conta á hora de seleccionar a dita unidade léxica son as que a seguir se indican.

- a. Que sexa unha forma normativa
- b. Que sexa unha forma estándar, frecuente

Así, por exemplo:

*Concepto*

‘Xogo infantil que consiste en andar á pata coxa empuxando ao mesmo tempo co pé o peletre, seguindo certas regras, a través dun percorrido establecido formado por un conxunto de casas numeradas, pintadas ou raiadas no chan’

- a) Aplicando a) unicamente teríamos dúas posibilidades para escoller, *mariola* e *peletre*, posto que son as únicas formas recollidas no DRAG.
- b) Aplicando b) debemos seleccionar *mariola*, porque é a forma máis habitual na variedade estándar (ademais de que, de acordo cos datos do TLPGP, ocupa tamén unha maior extensión xeográfica).

No caso de que exista unha expresión normativa claramente representativa do concepto e claramente estándar, pero que resulta ambigua no campo semántico en que se sitúa, empregaremos un desambiguador, de acordo cos criterios expostos en 3. É preferible optar por unha forma ambigua (desambiguada) se é normativa e frecuente antes ca por unha forma unívoca pero pouco frecuente.

Se non encontramos unha expresión normativa claramente representativa do concepto e claramente estándar, pero si existe a posibilidade de expresar o concepto mediante unha descrición breve, optaremos por esta vía.

*folla do piñeiro*

Ademais, algúns conceptos que se definen basicamente a partir da súa relación con outros conceptos, deben recibir sistematicamente un mesmo tipo de denominación a través dunha descrición breve, aínda que nalgún caso poida existir unha unidade léxica que cumpre as condicións arriba sinaladas:

---

'cría de': *cría da vaca, cría do porco* etc.

## 2) Recursos para o establecemento das denominacións

- Para determinar o carácter normativo ou non dunha forma acudimos ao DRAG. En principio (agás excepcións xustificadas), todas as formas que figuran como denominación de conceptos deben ter entrada neste dicionario como formas recomendadas.
- Para determinar, entre as formas normativas as máis estándares, acudimos por un lado ao noso propio coñecemento da lingua. Se existe calquera dúbida, tentamos comparar a frecuencia da formas:
  - 1) A través dos propios datos do TLPGP
  - 2) A través de corpus
- Para saber se unha posible denominación é ambigua ou non acudimos ao DRAG. Se a palabra ten diferentes acepcións, vemos se existe ambigüidade sempre dentro do campo semántico en que nos situamos (o campo corresponde co nivel xerárquico situado por riba do concepto concreto na clasificación onomasiolóxica establecida).

## 3) Desambiguación / clarificación

Como se indicou en 1, no caso de que a forma normativa máis frecuente sexa ambigua no marco do campo semántico en que se sitúa, é preferible engadir entre parénteses un elemento desambiguador antes que optar por unha denominación pouco frecuente. Así:

bailarete (xogo)

bailarete (xoguete)

Do mesmo xeito, aínda que a palabra escollida non sexa ambigua no campo semántico en cuestión, se se considera que pode resultar difícil determinar a realidade á que fai referencia (porque a realidade en si é pouco coñecida ou porque a palabra seleccionada, aínda non sendo ambigua no campo semántico en cuestión, é unha palabra con moitas acepcións), botaremos man dun elemento desambiguador / clarificador:

ovo (peza do rodicio)

Eses elementos desambiguadores / clasificadores poden ser de diferente tipo. No caso dos nomes de animais e plantas, podemos botar man do seu nome científico se non encontramos outro elemento claro e breve que permita a univocidade da

---

denominación. Poden ser tamén pequenas definicións, como no exemplo anterior, ou indicacións categoriais:

amencer (verbo)

amencer (substantivo)

#### 4) Outros criterios de selección

- a) Entre dúas formas normativas, cun grao de estandarización similar e non ambiguas, é preferible seleccionar como denominación do concepto aquela forma máis transparente. Por exemplo, entre *mestra* e *abella mestra* é preferible a segunda porque a presenza do núcleo *abella* axuda ao usuario a identificar o referente.
- b) Se hai dúas formas normativas que son variantes dun mesmo lema no TLPGP, debemos escoller como denominación do concepto aquela que coincide co lema do TLPGP. Así por exemplo, *amañecer* e *amencer* (ambas formas normativas) agrupáronse baixo o lema *amencer* no TLPGP. Daquela, a denominación do concepto debe ser *amencer* (coincidente co lema) e non *amañecer*.
- c) Hai que ter en conta á hora de establecer a denominación do concepto o tratamento que teñen as correspondentes expresións na base. Por exemplo, hai varias entradas que son *andar á pata coxa* ou *ir á pata coxa*, con dobre lema [lemas *andar* e *á pata coxa*]. Daquela, e dado que imos traballar con lemas, o concepto non debería chamarse *andar á pata coxa* senón simplemente *á pata coxa* ou en todo caso [*andar, ir...*] *á pata coxa*, posto que só lle asignaremos "concepto" ás correspondentes expresións adverbiais (*á pata coxa, á pata galana, á perna coxa...*) (non a *andar* ou a *ir*).

#### ESTABLECIMENTO DOS CONCEPTOS

##### 5) Lemas encadrables baixo un mesmo concepto

Situaremos baixo un mesmo concepto todos aqueles lemas que sexan realización léxica dese concepto aínda que inclúan algún matiz de tamaño, forma, comportamento etc. que restrinxa a súa extensión. Por exemplo baixo o concepto *can*, figuraría tamén a seguinte variante, a pesar da restrición establecida no significado:

**cuzo, -a:** Perro/-a que vive constantemente en el hogar sin salir de él. Su principal utilidad está en defender la casa.

Caso diferente sería que a palabra en cuestión designase a cría do can, porque esta tería o seu propio concepto.

---

**6) Aspectos formais**

- d) Os nomes dos conceptos deben escribirse con minúscula inicial (*mariola*, non *Mariola*)