

*Tesouro do léxico patrimonial galego e portugués:*  
**análise dos materiais galegos**

*María Beatriz Domínguez Oroña*

Instituto da Lingua Galega

Universidade de Santiago de Compostela

[marybet89@hotmail.com](mailto:marybet89@hotmail.com)

**Abstract.** The *Tesouro do léxico patrimonial galego e portugués* is a database of the Galician, European Portuguese and Brazilian Portuguese dialectal lexicon. This database collects lexical items from various sources with geographical indication. Most of the Galician works forming part of the corpus of the *Tesouro* are unpublished academic papers. These papers offer a lot of lexical information, but also other types as phonetic transcriptions or usage examples. The *Tesouro* has a computer application of query which access is public and free.

Keywords: lexical database, corpus, source, geographical indication, use.

Palavras-chave: base de datos lexicográfica, corpus, fontes, georreferência, utilidade.

**1. O *Tesouro do léxico patrimonial galego e portugués***

O *Tesouro do léxico patrimonial galego e portugués* é un proxecto iniciado no ano 2007 e orientado a integrar nun único banco de datos léxico dialectal do galego, do portugués de Portugal e do portugués do Brasil. Un dos principais obxectivos deste proxecto é conservar parte do patrimonio inmaterial destes países que se encontra ameazado pola abandono do modo de vida tradicional, ademais de pór ao servizo da comunidade científica unha gran cantidade de información lexicográfica espallada en materiais de diverso tipo e de difícil acceso. O material lexicográfico compilado extráese de fontes heteroxéneas (traballos de campo, inquéritos de atlas, traballos académicos inéditos, etc.), con frecuencia descoñecidas do público e/ou de difícil acceso aos investigadores; a condición básica é que o léxico sexa recollido de forma directa dos informantes e que teña referencia xeográfica. O proxecto está coordinado desde o Instituto da Lingua Galega da Universidade de Santiago de Compostela pola profesora Rosario Álvarez Blanco e nel colaboran universidades de Galicia, Brasil e Portugal.

## 2. O corpus do *Tesouro*: os materiais galegos

Como xa indicamos anteriormente, as obras que forman parte do corpus do *Tesouro* son heteroxéneas. Canto ao formato das obras, encontramos libros, capítulos de libros, traballos académicos, recursos web e artigos de publicacións periódicas. No que se refire á presentación da información léxica dentro de cada obra tamén encontramos diferenzas, desde glosarios até descrições etnolingüísticas. Algúns glosarios presentan a información léxica en entradas ordenadas alfabeticamente e outros presentan unha organización semántica. Tamén hai diversidade na localización desa información dentro de cada obra, ás veces esa información constitúe o corpo central do traballo e outras é un apartado do traballo ou un anexo.

O corpus do *Tesouro* non é un corpus pechado, senón que se está incrementando continuamente. Na data de elaboración do póster (o 21 de outubro do 2013), o *Tesouro* estaba conformado por un total de 139.950 rexistros pertencentes a 97 obras: 2 de Brasil, 55 de Portugal e 40 de Galicia. Na data da súa presentación pública, o 20 de febreiro do 2014, as cifras incrementáronse (155.676 rexistros e 112 obras). Como sinalamos anteriormente, cando falamos de obras referímonos a fontes de diverso tipo.

Gran parte destas fontes presentan un glosario do léxico común dun determinado lugar, parroquia ou concello. Ademais das formas dialectais e a súa definición, estes traballos acostuman presentar outro tipo de información: a transcripción fonética, a categoría gramatical, exemplos de uso ou ilustracións das palabras que se recollen. Todo este material é introducido na base de datos, onde a cada rexistro se lle asigna un lema, unha clasificación semántica e unha categoría gramatical seguindo os criterios establecidos previamente polos membros do proxecto.

Das 38 obras galegas introducidas na data en que se elaborou o póster, 31 son traballos académicos, na súa maioría memorias de licenciatura, moitas delas inéditas (29 memorias de licenciatura, 1 tese de doutoramento e 1 traballo académico non incluíble nas epígrafes anteriores). O resto dos materiais galegos son artigos de publicacións periódicas (4), libros (2) ou capítulos de libros (1) e recursos en páxinas web (2).



Na táboa que colocamos a seguir recóllense as cifras actuais:

TIPO DE FONTE	NÚMERO DE OBRAS	%
Memorias de licenciatura	33	67,3%
Teses de doutoramento	1	2,05%
Outros traballos académicos	3	6,1%
Libros	2	4,1%
Capítulos de libros	1	2,05%
Artigos	5	10,2%
Páxinas web	4	8,2%
<b>Total</b>	<b>49</b>	<b>100%</b>

A maioría das 49 obras galegas introducidas actualmente seguen sendo memorias de licenciatura (67,3%), seguidas a moita distancia por artigos (10,2%), páxinas web (8,2%), outros traballos académicos (6,1%), libros (4,1%) e, compartindo último lugar, teses de doutoramento e capítulos de libros (2,05%). Exceptuando dúas obras que están datadas en 1938 e 1949, o resto sitúanse cronoloxicamente entre 1967 e o 2012. A maioría delas concéntranse na década dos 70. Gran parte do léxico que se recolle nestas obras está vinculado á agricultura, á gandería, aos oficios tradicionais e a outros ámbitos da vida cotiá, como o ser humano (partes do corpo e prácticas sociais), o ámbito doméstico ou o tempo meteorolóxico.

### 3. A aplicación de consulta

Desde o 2011 o *Tesouro* conta cunha aplicación de consulta accesible unicamente para os membros do proxecto que desde a súa presentación en febreiro de 2014 é de acceso público e gratuito (<http://ilg.usc.es/Tesouro/>).

Esta aplicación permite realizar, por un lado, unha *busca simple* por variantes (as formas concretas rexistradas nos glosarios) e por lemas (formas máis ou menos canónicas, que agrupan variantes fonéticas e gráficas, no posible coincidentes coas formas estándares); e, por outro, unha *busca avanzada* por países, campos semánticos e categorías gramaticais.

O *Tesouro* ofrece toda a información léxica, etnográfica, sociolingüística etc., contida nas fontes, a distribución xeográfica e a cartografía en tres mapas, Galicia, Portugal e Brasil, así como unha clasificación semántica de cada variante. A estrutura da base de datos do *Tesouro* é tan flexible que permite, por un lado, recoller toda a información contida nas fontes e, por outro, darlle un aspecto homoxéneo a esa información facilitando a súa consulta. De seguido presentamos unha breve descrición dos principais campos nos que se organiza a información:

1. **Variante:** é a forma que aparece na obra orixinal e encabeza a entrada de cada un dos rexistros. Cada variante asóciase a un lema.
2. **Transcrición fonética:** algunhas obras presentan transcrición fonética das variantes, a pesar de que o que se prima no *Tesouro* é a información léxica, tamén se recolle a información fonética relevante dando un tratamento unificado aos diversos tipos de transcrición que se encontran nas obras, empregando para iso o Alfabeto Fonético Internacional.
3. **Clase e categoría:** a información que na propia obra se dá sobre a clase de palabra de cada rexistro.
4. **Definición:** a definición literal que o autor da obra asigna a cada rexistro. A definición pode conter información lexicográfica ou sinónimos. En moitas obras galegas anteriores aos oitenta a definición está en castelán.

5. **Exemplos e refráns:** cando os hai inclúense neste campo. Nalgúns casos os exemplos son fragmentos tirados de textos orais transcritos.
6. **Lema:** é a forma baixo a que se recollen as diferentes variantes fónicas ou léxicas, no posible coincidentes coas formas normativas. O lema é particular para cada variedade lingüística.
7. **Clase e categoría:** a clase gramatical deste campo é asignada polo manipulador da obra, atendendo á definición que o autor dá da variante. Esta clasificación correspóndese ao lema, independentemente da categoría que o autor indicase na obra.
8. **Clasificación semántica:** introdúcese un clasificador semántico para cada un dos rexistros, o que se pretende con esta clasificación é facilitar as buscas.
9. **Imaxes:** moitas obras conteñen fotografías ou debuxos que ilustran parte do léxico recollido. Estas imaxes son incorporadas na base de datos e vinculadas aos rexistros aos que fan referencia.

Ademais de todos estes campos hai outros dedicados a incluír a referencia bibliográfica, a páxina e o código xeográfico de cada rexistro. A información xeográfica trasládase ao mapa correspondente iluminándose o concello ao que pertence cada rexistro.

#### 4. As utilidades do *Tesouro*

A utilidade desta ferramenta é notoria tanto no campo da lexicografía coma no da variación lingüística, concretamente na dialectoloxía léxica. A súa aplicación de consulta permite, entre outras cousas, coñecer palabras e significados tradicionalmente ausentes dos dicionarios e obter de maneira directa e inmediata a distribución de formas en cada un dos territorios e tamén conxuntamente en Galicia, Portugal e Brasil. O *Tesouro* tamén pode ser de utilidade noutros campos de investigación lingüística como o da etimoloxía, o da historia da lingua, o da fraseoloxía ou o da fonética. Ademais, pode contribuír á investigación noutros ámbitos científicos non lingüísticos, como a etnografía ou a botánica.

O feito de que a aplicación de consulta sexa dixital e de acceso libre trae consigo moitas vantaxes, como a posibilidade de facer buscas de distinto tipo e acceder de forma sinxela a gran cantidade de material inédito e disperso.

#### Referencias

- Álvarez Blanco, Rosario / Xosé Afonso Álvarez Pérez / João Saramago / Xulio C. Sousa Fernández (2009) Presentación de un corpus digital de léxico tradicional: El Tesouro do léxico patrimonial galego e portugués. *Fonetică și Dialectologie*, XXVIII, 5-19.
- Álvarez Pérez, Xosé Afonso / Xulio C. Sousa Fernández (2013) The Tesouro do Léxico Patrimonial Galego e Portugués. A Galician and Portuguese word bank. *Dialectologia*, Special issue, IV, pp. 5-24.
- González Seoane, Ernesto X. (2012): Aspectos metodolóxicos do deseño e elaboración do Tesouro do léxico patrimonial galego e portugués. In *X Congreso Internacional da Asociación Internacional de Estudos Galegos. "Galiza alén do Arco Atlántico. Unha visión multidisciplinar dos Estudos Galegos"*, Prifysgol Caerdydd / Cardiff University, 12-14 de setembro de 2012.
- Mouzo Villar, Rosa / María López Suárez / Marta Negro Romero (2013) O Tesouro do Léxico Patrimonial Galego e Portugués. Unha ferramenta para o estudo do léxico [Póster]. In *I Simposio Internacional RELEX. Lexicografía románica: Estado da cuestión*, A Coruña, 15-17 de outubro de 2013.
- Osorio Peláez, César (2011) Desenvolvemento dun sistema para a visualización web do TESOURO DO LÉXICO PATRIMONIAL GALEGO E PORTUGUÉS. Tráballo de fin de grao, Universidade de Santiago de Compostela.