

**ASPECTOS
METODOLÓXICOS DO
DESEÑO E ELABORACIÓN
DO *TESOURO DO LÉXICO
PATRIMONIAL GALEGO E
PORTUGUÉS* (TLPGP)**

Ernesto González Seoane

Instituto da Lingua Galega, Universidade
de Santiago de Compostela

doi: [10.17075/tucmeg.2015.003](https://doi.org/10.17075/tucmeg.2015.003)

O *Tesouro do léxico patrimonial galego e portugués* (TLPGP) é unha iniciativa posta en marcha polo Instituto da Lingua Galega (ILG) co obxecto de rexistrar de maneira sistemática e integrar nun único banco de datos os materiais lexicográficos con referenciación xeográfica relativos aos dominios galego, portugués europeo e portugués brasileiro.²⁵ Os materiais compilados consisten fundamentalmente en vocabularios e glosarios de variedades dialectais galegas e portuguesas, aínda que inclúen tamén estudos léxicos de orientación etnolingüística, materiais dispersos contidos en monografías dialectais, atlas lingüísticos e traballos de campo, etc. Unha boa parte destes materiais procede de teses de doutoramento, memorias de licenciatura e outras modalidades de traballos académicos ata o de agora inéditos ou de moi difícil acceso. Trátase, en suma, de fontes textuais certamente heteroxéneas, pero que comparten un trazo común básico: a de recolleren materiais léxicos cunha localización xeográfica precisa.

O proxecto, desenvolvido baixo a coordinación xeral de Rosario Álvarez Blanco, conta con tres equipos de traballo correspondentes aos tres dominios considerados. Neles participan profesores e membros de centros e unidades de investigación de universidades galegas (ILG-USC), coordinados por Xulio Sousa, portuguesas (CLUL-Universidade de Lisboa, CELGA-Universidade de Coimbra), baixo a coordinación de João Saramago, e brasileiras (Federal de Mato Grosso, Federal de Mato Grosso do Sul, Estadual de Londrina, Federal do Ceará, Federal da Paraíba, Estadual da Paraíba, Estadual do Rio Grande do Norte, Federal do Piauí, Federal do Maranhão, Federal de Santa Catarina, Federal do Pará, Federal do Rio de Janeiro, Bandeirante de São Paulo, São Paulo, Instituto Federal de Educación, Ciencia e Tecnología de São Paulo, Faculdade Taboão da Serra) baixo a coordinación de Vanderci Aguilera e Sílvia Brandão.

O *Tesouro* inscríbese dentro dunha estratexia investigadora pola que o Instituto da Lingua Galega vén apostando desde hai xa varios anos, e que ten por obxecto poñer ao dispor da comunidade investigadora unha serie de

²⁵ Na actualidade unha primeira versión do *Tesouro* atópase dispoñible na páxina web do Instituto da Lingua Galega en: <http://ilg.usc.es/Tesouro/>

ferramentas de libre acceso para a consulta e a explotación de datos lingüísticos de diverso tipo. Deste xeito, tal e como se pode ver na Táboa 1, a día de hoxe están elaborados e dispoñibles en liña ou en fase avanzada de elaboración, por unha banda, catro *corpora* textuais, que compilan textos galegos, escritos nuns casos, orais noutros, producidos en época moderna ou durante a Idade Media, e por outra, tres *corpora* lexicográficos que rexistran e sistematizan a información contida en vocabularios, dicionarios, glosarios e outros instrumentos lexicográficos centrados tanto na lingua oral coma na lingua escrita, xa sexa moderna, xa medieval.

		Textual	Lexicográfico
Medieval		Tesouro Medieval Informatizado da Lingua Galega	Dicionario de Dicionarios do Galego Medieval
Moderno	Oral	Arquivo do Galego Oral / Corpus Oral Informatizado da Lingua Galega	Tesouro do Léxico Patrimonial Galego e Portugués
	Escrito	Tesouro Informatizado da Lingua Galega	Dicionario de Dicionarios da Lingua Galega

Táboa 1. *Corpora* textuais e lexicográficos do ILG

Ademais de polo tipo de materiais que compila, o *Tesouro* singularízase, fronte aos demais *corpora* elaborados polo ILG, por introducir dúas novidades ben visibles. A primeira delas supón unha innovación tecnolóxica, que radica na integración nunha única plataforma de consulta da información de carácter lexicográfico que conteñen os textos compilados xunto cunha representación cartográfica dos datos xerada de maneira automática. Por outra banda, fronte aos demais instrumentos elaborados no ILG, que recollen exclusivamente datos pertencentes ao dominio lingüístico galego, o *Tesouro* incorpora ademais información sobre os dominios do portugués europeo e do portugués brasileiro.

1. FONTES

Polo que se refire ás fontes compiladas no *Tesouro*, cómpre salientar en primeiro lugar, como xa se indicou, a súa marcada heteroxeneidade tipolóxica²⁶. Entre elas, cabe identificar un primeiro conxunto de obras constituído por monografías e estudos de variedades dialectais galegas, portuguesas e brasileiras. Entre elas ocupan un lugar destacado os vocabularios e glosarios que, con independencia do seu grao de complexidade microestrutural, e atendendo só ao criterio de ordenación primario que adoptan, responden a dous modelos básicos. Deste xeito, temos por unha banda os vocabularios alfabéticos, que presentan os materiais léxicos ordenados do A ao Z, e por outra os temáticos, que agrupan o léxico de acordo con algún sistema máis ou menos estandarizado de clasificación semántica ou temática²⁷, e logo, dentro de cada división ou epígrafe, o ofrecen ordenado alfabeticamente.

Por máis que os glosarios e vocabularios dialectais constitúen o modelo prototípico e, en todo caso, o máis significativo en termos cuantitativos do corpus textual do *Tesouro*, contamos tamén cun conxunto de monografías dialectais que non adoptan o formato de vocabulario, senón que presentan e caracterizan semanticamente o léxico dialectal que recollen inseríndoo en textos en que se describen procesos ou tarefas propios da cultura e civilización tradicional de base agraria e mariñeira.

Ao lado dos vocabularios e monografías dialectais, os atlas lingüísticos ocupan tamén un lugar relevante entre o noso corpus. Con todo, dado que o *Tesouro* ofrece unha representación cartográfica propia de todos os materiais léxicos compilados, non son os mapas lingüísticos incluídos nestes atlas o que máis interesa para os nosos propósitos, senón a relación de formas documentadas neles, xunto co seu significado, no caso de que se indique, e a súa localización xeográfica precisa.

Por último, a base de datos do *Tesouro* aliméntase tamén coas informacións de carácter léxico presentes en monografías e estudos de interese primariamente etnográfico ou antropolóxico. En todo caso, para poderen ser integrados na nosa base estes datos deben estar, como o resto dos materiais, referenciados xeograficamente.

²⁶ Para unha caracterización máis detallada das fontes portuguesas e brasileiras, véxase o traballo de Marta Negro Romero neste mesmo volume.

²⁷ Polo xeral, inspirados de maneira inmediata ou ben a través dunha fonte intermediaria no sistema de clasificación conceptual ideado por Hallig e Wartburg (1952).

2. METODOLOXÍA

2.1. Análise estrutural

A metodoloxía de traballo do *Tesouro* toma como punto de partida a análise estrutural das obras que pretendemos integrar na base. Tendo presente a considerable diversidade tipolóxica que presentan, o obxectivo desta análise non é outro que o de detectar elementos estruturais comúns e diferenciais co fin de telos en conta á hora de acometer a tarefa de deseñar a estrutura da nosa base de datos.

Polo que se refire ás obras de carácter lexicográfico en sentido estrito, é dicir, aquelas formadas por materiais léxicos organizados en artigos lexicográficos, as diferenzas macro e microestruturais entre as obras compiladas son, como era de prever, moi acusadas, e derivan da actuación de tres factores principais.

En primeiro lugar, mesmo entre obras conceptual e estruturalmente máis afíns, existe unha notable diversidade no tocante á riqueza e complexidade da información ofrecida por cada unha. Deste xeito, os elementos estruturais recorrentes límitanse en realidade a dous campos, os correspondentes á variante léxica e á súa localización xeográfica, case sempre acompañados dun terceiro: a glosa, definición ou equivalencia noutro idioma ou variedade da palabra rexistrada. A partir de aquí, os restantes campos que conforman a microestrutura das obras varían dunha a outra. Ocorren con frecuencia variable, mais en ningún caso con regularidade, os correspondentes á información fonética (que algúns textos ofrecen de xeito sistemático, para todas as entradas que rexistran, e outros só con carácter ocasional, por exemplo, para indicar o timbre das vogais medias ou determinados fenómenos ou procesos fonéticos, como a gheada, o rotacismo, o yeísmo, etc.), á clasificación semántica ou temática, á información categorial básica (que se ofrece de maneira sistemática nalgunhas obras e ocasionalmente noutras, por exemplo, con finalidade contrastiva, para marcar as diferenzas entre a variedade estudada e a lingua de referencia, ou ben como recurso auxiliar da desambiguación), aos exemplos, á información de carácter enciclopédico (polo xeral de carácter etnográfico ou antropolóxico), ás remisións e referencias cruzadas, ás fotografías e ilustracións, etc.

Por outra banda, unha segunda fonte de heteroxeneidade radica na adopción por parte dos distintos autores, e en particular no caso de obras presentadas en forma de glosario ou vocabulario, de criterios lexicográficos e gramaticais non coincidentes. Así, unha ollada mesmo superficial aos

textos compilados abonda para decatármonos das ostensibles diferenzas existentes no tocante aos criterios de lematización empregados. Estas diferenzas poden afectar á mesma forma de cita (por exemplo, nos substantivos e adxectivos con flexión de xénero), mais tamén ao tratamento das variantes, que unhas obras consideran por separado en tanto que outras as lematizan conxuntamente, ben baixo un lema único, ben baixo un lema múltiple formado por un lema primario (que é o que serve como elemento ordenador) e outro secundario. A diversidade é tamén a norma no relativo ao tratamento das unidades pluriverbais. Polo demais, rexístranse tamén discrepancias moi acusadas na caracterización gramatical das palabras repertoriadas. As diferenzas afectan, nun plano máis superficial, á táboa de abreviaturas e, nun nivel máis profundo, ao sistema de clasificación categorial adoptado e ao grao de profundidade da propia caracterización, que pode ser moi elemental e limitarse só á indicación da categoría (verbo, substantivo, adverbio...) ou ben introducir en todos ou nalgúns casos precisións adicionais (verbo transitivo, substantivo masculino, adverbio de tempo...).

Finalmente, a análise das obras do corpus permitiu apreciar na súa verdadeira dimensión a heteroxeneidade, xa intuída, derivada das súas diferenzas de natureza lingüística. As máis evidentes radican no feito de as obras estaren redactadas en linguas diferentes, nomeadamente galego, portugués e español, ou mesmo, dentro dalgunha delas, en variedades normativas diferentes. Así, no caso do portugués, contamos con textos escritos tanto na variedade europea coma na brasileira. Para o galego, en cambio, as diferenzas lingüísticas están relacionadas sobre todo coa inexistencia dunha variedade estándar de referencia (nos textos máis antigos) e cunha certa inseguridade no seu manexo (nalgúns dos máis recentes). De por parte, e con independencia da lingua en que estean escritas, rexistramos tamén diferenzas á hora de representar graficamente procesos fonéticos característicos da lingua oral de determinadas variedades xeográficas. Tal é o caso, por citar só algún exemplo rechamante, da diversidade que se rexistra no tratamento gráfico da gheada galega, e que abrangue desde a súa non representación (neste caso a súa existencia reflectiríase só no campo correspondente á transcripción fonética), ata a súa representación mediante signos gráficos diferentes (case sempre <j> e <gh>).

Se no caso das obras de natureza lexicográfica se rexistra, segundo acabamos de ver, unha notable diversidade, os textos de carácter non lexicográfico supoñen un desafío adicional na medida en que a súa estrutura (un conxunto de textos redactados en que se insiren as palabras

que se pretende glosar ou, no caso dos atlas, unha serie de mapas dialectolóxicos coas súas correspondentes lendas) está tan afastada da dunha obra lexicográfica, que require unha adaptación e unha intervención moito máis profunda, como teremos oportunidade de comprobar.

2.2. Deseño da base de datos

Os datos e informacións obtidos da análise dos textos do corpus, ou máis ben dunha mostra representativa deles, permiten acometer con garantías a seguinte fase do proxecto, que consiste en deseñar unha base de datos cuxa estrutura sexa capaz de satisfacer diversos requirimentos. Así, esta estrutura ha de poder acomodarse a un corpus acusadamente heteroxéneo, preservando en todo o posible a integridade e mesmo a literalidade das obras compiladas, mais, ao mesmo tempo, debe ser quen de superar esa heteroxeneidade tipolóxica ou lingüística a través da habilitación de novas vías para a recuperación da información contida nas obras. Os campos que conforman a base de datos son os seguintes:

CAMPO	DENOMINACIÓN
1	Variante
2	Fonética
3	Categoría gramatical
4	Definición
5	Exemplos e refráns
7	Páxina
8	Sección
10	Cita bibliográfica
11	Código xeográfico
11b	Táboa localidades
12	Imaxes
13	Remisións cruzadas
15	Remisión a textos
20	Rexistro creado
9	Clasificación semántica
16	Lema (na lingua correspondente)
17	Categoría
6	Comentarios procesador
18	Sinónimo de referencia
19	Lema común

Táboa 2. Campos da base de datos

Os catorce primeiros campos, que na táboa figuran en cor verdosa, están concibidos para recoller a información que figura nos textos orixinais. Os seis restantes, pola contra, son da exclusiva responsabilidade do equipo de elaboración do *Tesouro*, e non dos autores dos textos compilados. Destes campos, os catro que figuran en cor avermellada son procesados na presente fase de elaboración do proxecto, en tanto que os dous últimos, que aparecen en cor azul escura, corresponden a unha fase posterior, aínda non abordada, na que se pretende crear vínculos entre os lemarios das distintas linguas (para deste xeito posibilitar, por poñer un exemplo, a recuperación conxunta do gal. *muiño* e do port. *moínho*) e, ademais, entre as distintas unidades léxicas que presentan un significado coincidente (para así poder recuperar de maneira conxunta, por exemplo, os rexistros correspondentes a *píntega*, *salamántiga*, *sacabeira*, *sacarrancha*, *pinchorra*, etc.).

As denominacións dos diferentes campos resultan abondo transparentes, polo que non será preciso entrar nunha descrición detallada. Cómpre advertir, en todo caso, que dos catorce primeiros só dous son obrigatorios, o que nós denominamos «variante» (campo 1), e que corresponde á forma lingüística que na obra orixinal figura na cabeceira da entrada; e o código xeográfico (campo 11), que vincula esa variante co lugar ou lugares en que a obra en cuestión a documenta (concellos no caso de Galicia e Portugal; mesorreixións no caso do Brasil), e que por outra banda permitirá logo xerar os mapas de maneira automática. Como se dixo antes, o criterio que rexe o tratamento dos textos é marcadamente conservador, polo que a forma consignada no campo 1 reproduce de maneira literal a que figura no texto orixinal. Cómpre advertir, así e todo, que naqueles casos en que atopamos entradas complexas (que conteñen máis dunha acepción ou que agrupan nun único artigo lexicográfico información relativa a máis dunha variante) optamos por desdobralas tantas veces sexa necesario para dispoñermos dun rexistro para cada variante e para cada acepción. Deste xeito, unha entrada como

bochicar *v.i.* **1** Falar polo baixo, rosmar. «¿Que andas bochicando por aí?». **2** Contestar, responder de mal modo. «¡Non te atrevas a bochicarme!».

desdóbrase para crear dous rexistros, un para cada acepción:

bochicar *v.i.* **1** Falar polo baixo, rosmar. «¿Que andas bochicando por aí?».

bochicar *v.i.* **2** Contestar, responder de mal modo. «¡Non te atrevas a bochicarme!».

Os restantes doce campos son opcionais, e cóbrense só se a obra orixinal ofrece algún tipo de información ao respecto, e sempre reproducindo fielmente a literalidade do texto. A través deles vértense na base as informacións de diferente tipo que cada obra ofrece sobre as «variantes» que rexistra e, ademais, cando é pertinente, vincúlase o rexistro cos arquivos de imaxes que ilustran a entrada (campo 12) e que logo se poderán recuperar na aplicación de consulta.

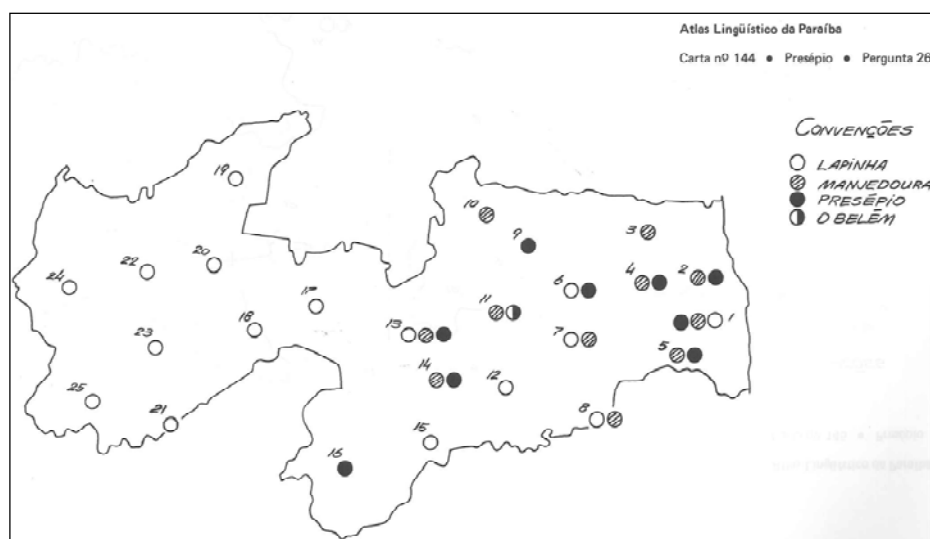
Por outra banda, os textos non lexicográficos non poden ser trasladados directamente á base de datos, senón que requiren, segundo se indicou máis arriba, un maior grao de intervención. Así, nestes casos as formas correspondentes ao campo 1 (Variante) son extraídas directamente dos textos (ou das lendas dos mapas, no caso dos atlas) e ilustradas no campo 4 (Definición) con un ou varios fragmentos dese mesmo texto que permitan comprender o significado da palabra e o seu contexto de uso (no caso dos atlas, o campo 4 é ocupado por regra xeral polo título do mapa). Deste modo, a partir dun fragmento como o que reproducimos a seguir creamos os rexistros que figuran na Táboa 3:

Unha *coutada* é un monte *coutado*, un *couteiro*: o dono págalle a un home, o *couteiro*, que fai de vixiante e coida que ninguén toque esas terras (Álvarez 1974: 13).

1	4	7	8	10
coutada	Unha <i>coutada</i> é un monte <i>coutado</i> , un <i>couteiro</i> : o dono págalle a un home, o <i>couteiro</i> , que fai de vixiante e coida que ninguén toque esas terras.	13	Os terreos	Álvarez 1974
coutar	Unha <i>coutada</i> é un monte <i>coutado</i> , un <i>couteiro</i> : o dono págalle a un home, o <i>couteiro</i> , que fai de vixiante e coida que ninguén toque esas terras.	13	Os terreos	Álvarez 1974
couteiro	Unha <i>coutada</i> é un monte <i>coutado</i> , un <i>couteiro</i> : o dono págalle a un home, o <i>couteiro</i> , que fai de vixiante e coida que ninguén toque esas terras.	13	Os terreos	Álvarez 1974
couteiro	Unha <i>coutada</i> é un monte <i>coutado</i> , un <i>couteiro</i> : o dono págalle a un home, o <i>couteiro</i> , que fai de vixiante e coida que ninguén toque esas terras.	13	Os terreos	Álvarez 1974

Táboa 3. Mostra da adaptación dun texto redactado

Polo que se refire á integración na base dos datos léxicos contidos nos atlas, na mostra que se ofrece a seguir, tirada do *Atlas Lingüístico da Paraíba*, pode verse un exemplo do procedemento seguido:



1	4		10	11
lapinha	Presépio [cena reproduzindo o nascimento de Cristo].	44	ALPB	92501; 92502; 92503; 92504
manjedoura	Presépio [cena reproduzindo o nascimento de Cristo].	44	ALPB	92501; 92502; 92503; 92504
presépio	Presépio [cena reproduzindo o nascimento de Cristo].	44	ALPB	92501; 92502; 92503; 92504
belém	Presépio [cena reproduzindo o nascimento de Cristo].	44	ALPB	92501; 92502; 92503; 92504

Táboa 4. Mostra da adaptación dun atlas lingüístico

2.3. Procesamento dos materiais compilados

O tratamento dos materiais contidos nas obras que se pretende integrar no *Tesouro* comeza polo vertido dos datos na nosa base seguindo o manual de procesamento, en que se recollen todas as posibles variables detectadas no proceso de análise macro e microestrutural antes sumariamente descrito. Unha vez rematado este proceso, ponse en marcha a segunda fase, na que se realiza a lematización, caracterización categorial básica e asignación dun clasificador semántico.

O proceso de lematización ten por obxecto asignar a cada un dos rexistros existentes, que están encabezados por un lema determinado polo autor do glosario, e que nós denominamos «variante» (campo 1), un lema normalizado (campo 16), que será o que permitirá logo aos usuarios da base recuperar conxuntamente todas as variantes correspondentes a un único elemento léxico. Este lema normalizado vai acompañado, ademais, da indicación da súa categoría gramatical, tamén unificada e normalizada (campo 17). Deste xeito concíliase o imperativo de respectar e preservar a

literalidade das obras orixinais (da que dan conta os campos 1 e 3) coa vontade de introducir elementos de harmonización que posibiliten a recuperación conxunta de rexistros pertencentes a obras elaboradas con criterios ás veces moi diferentes.

Non é este o lugar para expoñer polo miúdo os criterios que rexen o proceso de lematización e que forman parte do manual de procesamento do *Tesouro*. En calquera caso, si pode ser conveniente chamar a atención sobre algunhas das súas pautas xerais.

En primeiro lugar, segundo queda dito, nesta fase do proxecto optouse por manter diferenciados os lematarios portugués e galego. Deste xeito, por exemplo, para os textos galegos adóptase como lematario de referencia o VOLG, na súa versión electrónica dispoñible na rede.²⁸ Para os textos portugueses e brasileiros, en cambio, tómanse como referencia, respectivamente, os lematarios do dicionario Porto Editora (PEd) e Aurélio e, complementariamente, o doutras obras lexicográficas, como os dicionarios Priberam (DPLP), Aulete e Houaiss. Isto quere dicir que no estado actual do traballo só resulta posible recuperar de maneira conxunta os resultados galegos e portugueses naqueles casos en que existe coincidencia formal plena entre eles. A vinculación de ambos os lematarios, a través do campo 19, constitúe, conforme sinalamos máis arriba, un dos obxectivos da terceira fase do proxecto.

Por outra banda, tampouco convén perder de vista que o proceso de lematización ten por obxecto unicamente agrupar e vincular cun lema canónico todas aquelas formas que se poden considerar variantes fónicas (ou eventualmente variantes gráficas), como é o caso, por exemplo, dos distintos resultados que proceden dun mesmo étimo ou de solucións que son o resultado de procesos fonéticos característicos da lingua oral. Deste xeito, a lematización permitiranos recuperar de maneira conxunta variantes como *fociño* e *fuciño*, *entroido* e *antroido*, *pobre* e *probe*, *boroa* e *broa*, *gando* e *gado*, etc. Pola contra, optouse por lematizar por separado as variantes léxicas e morfolóxicas, é dicir as formas que, compartindo un mesmo significado, presentan raíces léxicas diferentes (*donicela* e *garridiña*, *porco* e *cocho*, *golpe* e *raposo*, *píntega* e *sacabeira*...) ou ben aquelas que, compartindo significado e raíz léxica, presentan diferenzas na súa estrutura morfolóxica (*donicela* e *doniña*, *alustrar* e *relustrar*, *queixo* e *queixelo*...). Tal e como se dixo máis arriba, na terceira fase do proxecto as variantes deste tipo quedarán vinculadas mediante un sinónimo de referencia, a través do campo 18.

²⁸ Dispoñible en: <http://academia.gal/recursos-volg>

Por último, polo que se refire á clasificación semántica cabe sinalar que todos e cada un dos rexistros levan asignado un clasificador de acordo cun esquema conceptual de ata tres niveis de profundidade. Cómpre advertir que, segundo se verá máis abaixo, o clasificador ten como función primordial servir ben como elemento de desambiguación, ben como filtro para seleccionar de entre os resultados das consultas aqueles rexistros que lle verdadeiramente interesen ao usuario. En calquera caso, a aplicación do *Tesouro* ofrece unha modalidade de busca avanzada que permite obter unha listaxe alfabética das variantes ou dos lemas agrupados baixo cada un dos clasificadores.

A seguir, na Táboa 5, pódese ver representado en forma de táboa o resultado do procesamento dun pequeno grupo de rexistros correspondentes á fala de Vilardevós (Gil 1969).

1	2	3	4	5	7	8	10	11	12	13	20	9	16	17	6
abeacas			Piezas del arado que separan la tierra a ambos lados.		88	LA AGRICULTURA. El arado	Gil 1969	32091	vilardevos_17.jpeg	ameisello; fachón; relladoiro		3.4.2	abeaca	sf	
ameisello			Clavija de madera que une las dos abeacas del arado.		88	LA AGRICULTURA. El arado	Gil 1969	32091	vilardevos_17.jpeg	abeacas		3.4.2	meiselo	sm	
arado de pau			Arado de madera.		88	LA AGRICULTURA. El arado	Gil 1969	32091	vilardevos_17.jpeg	mosquito		3.4.2	arado de pau	loc sm	
cuño			Cuña que levanta o baja el arado para que profundice más o menos.		88	LA AGRICULTURA. El arado	Gil 1969	32091	vilardevos_17.jpeg			3.4.2	cuño	sm	
charrúa			Arado de hierro.		88	LA AGRICULTURA. El arado	Gil 1969	32091				3.4.2	charrúa	sf	
chavella			Clavija que une el yugo a la lanza del arado.		88	LA AGRICULTURA. El arado	Gil 1969	32091	vilardevos_17.jpeg	tamoeiro		3.4.2	chavella	sf	
fachón			Armazón de mimbre que se pone desde las abeacas a la punta de la telera e impide que la tierra caiga en el surco.		88	LA AGRICULTURA. El arado	Gil 1969	32091		abeacas		3.4.2	fachón	sm	
mosquito			Arado de hierro y de madera que tiene la misma forma que el arado de pau.		88	LA AGRICULTURA. El arado	Gil 1969	32091			arado de pau	3.4.2	mosquito	sm	
rabela			Esteva del arado.		88	LA AGRICULTURA. El arado	Gil 1969	32091	vilardevos_17.jpeg			3.4.2	rabela	sf	
rella			Pieza de hierro en la punta del arado.		88	LA AGRICULTURA. El arado	Gil 1969	32091	vilardevos_17.jpeg			3.4.2	rella	sf	
relladoiro			Listón pequeño y fijo entre el cuerpo central del arado y las abeacas.		88	LA AGRICULTURA. El arado	Gil 1969	32091	vilardevos_17.jpeg	abeacas		3.4.2	relladoiro	sm	
tamoeiro			Correa de la chavella.		88	LA AGRICULTURA. El arado	Gil 1969	32091		chavella		3.4.2	tamoeiro	sm	
tamón			Lanza del arado.		88	LA AGRICULTURA. El arado	Gil 1969	32091	vilardevos_17.jpeg			3.4.2	tamón	sm	
telera			Telera.		88	LA AGRICULTURA. El arado	Gil 1969	32091	vilardevos_17.jpeg			3.4.2	teiroa	sf	

Táboa 5. Mostra do procesamento e integración na base dun texto

3. A APLICACIÓN DE CONSULTA

Na actualidade, como xa indicamos anteriormente, está dispoñible a través da páxina web do Instituto da Lingua Galega unha primeira versión do *Tesouro* (véxase nota 2). Por máis que o número de obras vertidas na base e completamente procesadas é aínda limitado (no momento da súa presentación, en febreiro de 2014, o *Tesouro* acolle un total de 115 obras, das cales 11 corresponden a Brasil, 49 a Galicia e 55 a Portugal), consideramos que pode ser abondo para mostrar as súas funcionalidades e para dar unha idea das posibilidades que pode ofrecer.

A aplicación resulta bastante intuitiva, de xeito que non require nin complexas instrucións nin un adestramento previo para poder ser manexada. Como se pode ver na Fig. 1, que se ofrece a continuación, a pantalla divídese en catro áreas ou seccións.



Fig. 1. A aplicación do *Tesouro*. Fonte: web do ILG

A primeira delas, localizada na parte superior, consiste nunha sinxela caixa na que se debe escribir a forma que se pretende buscar. Ao seu carón figuran uns cadros que permiten seleccionar unha das dúas modalidades de procura que se ofrecen, a busca por *lema* (campo 16) ou por *variante* (campo 1).

Na segunda área, situada na zona central da pantalla (marcada na imaxe con sombreado azul), ofrécense os rexistros correspondentes á forma procurada, con indicación da obra que os recolle, do dominio lingüístico a

que corresponde cada un (galego, portugués europeo ou portugués brasileiro) e do clasificador semántico correspondente. Ademais, mediante unha icona advírtese, cando é pertinente, da existencia dun arquivo de imaxe asociado ao rexistro.

A terceira área localízase nunha columna situada á esquerda da pantalla (marcada na imaxe con sombreado amarelo). Nela ofrécese ao consultor a posibilidade de refinar a súa busca activando ou desactivando os filtros correspondentes ás localizacións, aos campos semánticos, aos lemas e ás variantes rexistrados. Deste xeito, á vista do resultado inicial da consulta do lema *cesto* o usuario pode entrecollel, marcando os filtros correspondentes, por exemplo, unicamente os rexistros galegos da variante *cesto* correspondentes ao ámbito das «artes de pesca».

Por último, na área situada na zona central inferior da pantalla (marcada na imaxe con sombreado rosa) ofrécese a cartografía automática dos rexistros mostrados en cada momento na metade superior.

4. CABO

En definitiva, consideramos que o *Tesouro do léxico patrimonial galego e portugués* constitúe unha contribución relevante para a recuperación, preservación e difusión dun legado que se encontra en serio risco de desaparición. Esta ameaza vese agravada pola dispersión das fontes e polas dificultades para acceder a unha boa parte delas, que consisten en traballos académicos que nunca foron publicados e que se encontran, en non poucos casos, en perigo pola mesma deterioración do seu soporte material. Ademais, o *Tesouro* habilita novas ferramentas de acceso sinxelas e áxiles que, respectando a literalidade dos materiais orixinais, permite superar unha excesiva fragmentación da información por mor das diferenzas de todo tipo (lingüísticas, tipolóxicas, metodolóxicas, etc.) que separan as obras do corpus.

5. REFERENCIAS BIBLIOGRÁFICAS

- ÁLVAREZ BLANCO, Rosario (1974): *Notas lingüísticas e etnográficas de Ramirás*, memoria de licenciatura, Universidade de Santiago de Compostela.
- ARAGÃO, Maria do Socorro Silva de / Cleusa Bezerra de MENEZES (1984): *Atlas linguístico da Paraíba. V. Vols. 1 e 2*, Brasília, CNPq/UFPB.
- Aulete = *iDicionário Aulete*. Versión en liña: <http://aulete.uol.com.br/>
- Aurélio = FERREIRA, Aurélio Buarque de Holanda (2009): *Novo Dicionário Aurélio da Língua Portuguesa*, Curitiba, Editora Positivo.
- DPLP = *Dicionário Priberam da Língua Portuguesa*. Versión en liña: <http://www.priberam.pt/DLPO/>
- GIL SUÁREZ, María del Carmen (1969): *El habla de Villardevós*, memoria de licenciatura, Universidade de Santiago de Compostela.
- HALLIG, Rudolf / Walther von WARTBURG (1952): *Begriffssystemals Grundlage für die Lexikographie. Versuch eines Ordnungsschemas*, Berlin, Akademie Verlag.
- HOUAISS, Antônio (2001): *Dicionário Houaiss da Língua Portuguesa*, Rio de Janeiro, Objetiva.
- PEd = *Dicionário da Língua Portuguesa da Porto Editora*. Versión en liña: <http://www.infopedia.pt/lingua-portuguesa/>
- VOLG = Antón SANTAMARINA / Manuel GONZÁLEZ GONZÁLEZ (coords.) (2004): *Vocabulario ortográfico da lingua galega*. A Coruña: Real Academia Galega / Instituto da Lingua Galega. Existe unha versión en liña: <http://www.realacademiagalega.org/recursos-volg>