

Extracción dun vocabulario xurídico-administrativo galego-castelán a partir dun corpus paralelo

Xavier Gómez Guinovart

Ánxeles Torres Padín

Seminario de Lingüística Informática, Universidade de Vigo

1. Introducción

Os córpora paralelos, constituídos por textos na súa versión orixinal e traducida, ofrecen diversas posibilidades de explotación no ámbito do procesamento da linguaxe natural. Os resultados máis importantes da explotación destes córpora obtéñense nas aplicacións nos eidos da tradución e da lexicoloxía, no ámbito da tradución automática estatística, das memorias de tradución e da tradución automática baseada en exemplos, da extracción léxica para recuperación de información multilingüe, e da extracción de léxico bilingüe. Nesta comunicación imos presentar unha investigación en curso do Seminario de Lingüística Informática (SLI) da Universidade de Vigo que se enmarca dentro deste último campo de traballo, a extracción de léxico bilingüe especializado. O obxectivo desta investigación é a constitución dun repertorio terminolóxico bilingüe galego-castelán baseado nos datos tirados das traducións especializadas deste xénero recompiladas no Corpus Lingüístico da Universidade de Vigo (CLUVI).

2. Vocabulario e corpus de baleirado

Concretamente, o corpus utilizado no baleirado deste vocabulario é o Corpus LEGA, un corpus paralelo aliñado a nivel de frase de bitextos pertencentes aos ámbitos da administración e do dereito. Este corpus forma parte do CLUVI, un proxecto que se desenvolve no SLI encamiñado a fornecer un amplo conxunto de córpora de rexistros especializados de lingua galega contemporánea, tanto monolingües coma plurilingües, con diversas aplicacións nos eidos da didáctica, da lexicografía, da terminoloxía e da tradución (Gómez Guinovart e Sacau Fontenla 2004). A sección de córpora paralelos do CLUVI, dunha extensión actual total superior aos 12 millóns de palabras, conta nestes momentos con catro córpora paralelos principais pertencentes a tres rexistros especializados (dos ámbitos xurídico-administrativo, literario e de divulgación científica) e a catro combinacións lingüísticas diferentes (bilingüe galego-español, bilingüe inglés-galego, bilingüe francés-galego e tetralingüe inglés-galego-francés-español). Estes catro córpora son o Corpus LEGA de textos xurídico-administrativos galego-español (5.110.622 palabras), o Corpus UNESCO de divulgación científica inglés-galego-francés-español (2.387.177 palabras), o Corpus TECTRA de textos literarios inglés-galego (1.476.020 palabras), e o Corpus FEGA de textos literarios francés-galego (1.267.119 palabras). Alén diso, estase a traballar na elaboración doutros cinco córpora paralelos do CLUVI: o Corpus EGAL de economía galego-español (718.642 palabras), o Corpus PALOP de literatura poscolonial portugués-español (566.590 palabras), o Corpus VEIGA de subtitulación inglés-galego (71.618 palabras), e nos correlatos inglés-portugués (182.248 palabras) e inglés-español (122.251 palabras) do Corpus TECTRA.

Dentro deste marco de traballo, o Corpus LEGA (de libre consulta no enderezo web <http://sli.uvigo.es/CLUVI/>), de 117.443 unidades de tradución e 5.110.622 palabras (2.452.123 en galego e 2.658.499 en español), constitúe a fonte principal para a extracción terminolóxica na súa especialidade lingüística, e xa está dispoñible unha

primeira achega aos resultados do seu procesamento nun repertorio terminolóxico publicado recentemente pola Área de Normalización Lingüística da Universidade de Vigo, titulado *Vocabulario xurídico-administrativo galego-castelán* (Gómez Guinovart e Torres Padín 2005), que se pode consultar tamén en liña no enderezo <http://webs.uvigo.es/sli/dilega>. Este primeiro achegamento tomou como base unha versión previa do Corpus LEGA de ao redor de 2 millóns de palabras, e para a extracción terminolóxica utilizouse tanto a consulta de vocabularios de referencia, coma a extracción automática.

A orientación fundamental do noso traballo é empírica e descritiva: tratamos de ilustrar o uso real da linguaxe neste ámbito específico de comunicación, e mais de amosar cales son as equivalencias entre o galego e o castelán empregadas efectivamente nesta especialidade da tradución. A descrición deste uso real dos termos no seu contexto converten o glosario nun recurso adecuado para a tradución profesional, un sector ao que consideramos o destinatario principal dos resultados prácticos do noso labor. De acordo con esta orientación empírica, por unha banda, incluimos no glosario só os termos bilingües que aparecían documentados no corpus de baleirado, e por outra banda, ilustramos o uso dos termos postos en equivalencia no glosario con frases de exemplo reais (orixinal e tradución) tiradas do corpus.

Nestes momentos, xa rematamos a ampliación do Corpus LEGA a máis de 5 millóns de palabras, cunha cantidade equivalente de material textual bilingüe tirado do Diario Oficial de Galicia (DOG), que suma un total de 2.394.407 palabras, e de lexislación diversa de ámbito estatal (publicada no BOE) e supraestatal (Constitución Europea) cun total de 2.716.215 palabras. Polas súas características, daquela, podemos considerar que os textos compoñentes do corpus de baleirado proveñen de fontes fiables e autorizadas, e pertencen na súa meirande parte a dous rexistros especializados da linguaxe xurídica moi próximos mais ben diferenciados: o ámbito administrativo, representado polo DOG, e o ámbito lexislativo, representado polas leis e regulamentos publicados no BOE. Estas leis e regulamentos pertencen, á súa vez, a distintos eidos dentro do ámbito lexislativo: o eido xudicial (representado no corpus, por exemplo, pola *Lei de Axuízamento Civil*), o ámbito da Constitución e dos Estatutos (representado, por exemplo, polo proxecto de *Constitución Europea*), o eido económico (representado, por exemplo, pola *Lei Xeral Tributaria*), o eido social (representado, por exemplo, pola *Lei de protección integral contra a violencia de xénero*) e o ámbito relacionado con sectores específicos (como a *Lei de portos*, o *Regulamento do sector ferroviario* ou a *Lei Orgánica de Universidades*). Ao final deste traballo pódese consultar a lista completa dos textos incluídos no Corpus LEGA.

Tras esta presentación do corpus de baleirado empregado no noso traballo, imos presentar algúns aspectos específicos do seu procesamento relacionados coa identificación no corpus das unidades terminolóxicas e coa elaboración dunha ontoloxía operativa axeitada ao campo de coñecemento representado polo corpus.

3. Identificación das unidades terminolóxicas

A identificación das unidades terminolóxicas no Corpus LEGA de baleirado pasa por dúas fases: unha primeira fase de extracción terminolóxica automática e revisión dos resultados, e unha segunda fase de buscas directas no corpus a partir de listas de termos recompiladas de diversos repertorios, que ten o obxectivo de cubrir as eivas relacionadas coa metodoloxía de extracción empregada na primeira fase.

Para a extracción terminolóxica automática utilizamos un sistema estatístico de extracción de candidatos a termos pluriléxicos realizado polo profesor Gaël Dias, da Universidade da Beira Interior (Covilhã, Portugal). Este programa informático, denominado Senta (Software for the Extraction of N-ary Textual Associations) serve para identificar en corpus crus as atraccións léxicas en secuencias de palabras continuas e non continuas. O Senta utiliza un modelo probabilístico coñecido como expectación mutua que consiste en medir o grao de asociación entre as unidades textuais contiguas. A selección destas unidades textuais realízase a través dun algoritmo chamado LocalMaxs, que elixe os vectores cuxo grao de asociación alcanzan un valor mínimo ou "máximo local" (Dias et al. 2000).

Na primeira fase da identificación das unidades terminolóxicas, logo, utilizouse o programa Senta para extraer os candidatos a termos pluriléxicos do Corpus LEGA constituídos por unha secuencia de entre dúas e cinco palabras. A extracción levouse a cabo de xeito independente na parte en galego e na parte en español do corpus paralelo e xerou unha lista resultante dunhas 80.000 secuencias por lingua. Nas secuencias de tres ou máis palabras, o resultado da extracción automática ofrécenos tanto termos pluriléxicos continuos, coma non continuos. En todos os casos, os resultados inclúen para cada candidato o índice de expectación mutua, isto é, a probabilidade que teñen as unidades textuais compoñentes da secuencia de apareceren unha ao lado da outra; e a súa frecuencia, isto é, o número de veces que se encontra a secuencia no corpus, como se amosa na Figura 1 con secuencias de dúas palabras.

0.00000539481380	00133	horas extraordinarias
0.00000512455199	00099	transporte ferroviario
0.00000509983374	00101	bens inmobles
0.00000505879871	00024	Benito Corbal
0.00000505879871	00024	IVA incluído
0.00000504752416	00067	organizaciós sindicais
0.00000489731929	00120	entidade xestora
0.00000481603774	00034	departamentos ministeriais
0.00000480948393	00131	actividades económicas
0.00000474644276	00104	pesca marítima
0.00000468632106	00052	estrada convencional
0.00000466883466	00041	consellos reguladores

Figura 1

Na avaliación do resultado da extracción terminolóxica automática consideramos os factores de cobertura e de precisión: a cobertura sería a capacidade do sistema para extraer todas as unidades terminolóxicas pluriléxicas dun texto; e a precisión a capacidade de discriminar entre as unidades pluriléxicas que son termos e as que non o son. Cómpre ter en conta que o Senta extrae todas as secuencias pluriléxicas nas que se detecta un certo grao de atracción mutua, tanto se son termos como se non o son. Entre as secuencias pluriléxicas que non son termos atopamos, por unha banda, entidades con nome (antropónimos, topónimos, nomes de entidades, datas); por outra, locucións adverbiais, prepositivas, conxuntivas e de todo tipo; e, por outra, finalmente, secuencias de palabras sen unidade lingüística. Así, con respecto a estas medidas de avaliación, os resultados da extracción realizada polo programa Senta presenta moito "ruído" e pouco "silencio", é dicir, identifícanse moitas unidades pluriléxicas que non se poden considerar termos e hai algúns termos presentes no corpus que non aparecen na extracción léxica. En consecuencia, tras a extracción automática resulta absolutamente precisa unha comprobación manual dos resultados para validar os termos candidatos.

Pódese observar este tipo de revisión dos resultados na Figura 2, cun exemplo dos resultados da extracción automática con secuencias de dúas unidades.

```
0.00000139033284 00093 xunta xeral
0.00000054518119 00268 ou do
0.00001022415927 00552 os seguintes
0.00004427589738 00468 Estado membro
0.00006029348879 00488 Unión Europea
0.00000639727887 00696 sobre o
0.00000028243534 00030 normas vixentes
0.00002980128738 00557 nos artigos
0.00000847715364 00030 Concepción Arenal
```

Figura 2

Nas secuencias de tres ou máis palabras, o sistema extrae termos pluriléxicos continuos e non continuos. Os termos continuos enténdese que son secuencias ininterrompidas de termos, mentres que os non continuos son secuencias fixadas que se ven interrompidas por outras palabras que soen ser sinónimas ou intercambiabes. Un exemplo diso sería a secuencia *Tratado de ...* onde poderíamos atopar unidades como *Tratado de Niza*, *Tratado de Roma*, etc. No caso das unidades pluriléxicas non continuas a precisión do sistema descende, como se pode observar cos exemplos da Figura 3.

```
0.00000185475176 00014 . <gap> Comisión <gap> de
0.00000063788161 00002 que <gap> atender <gap> necesidades
0.00000248929405 00020 , <gap> carácter <gap> ou
0.00000127576322 00002 Título <gap> Do <gap> Galego
0.00000088045630 00007 que <gap> poida <gap> ,
0.00000063788161 00002 contra <gap> poderanse <gap> os
0.00000063788161 00002 de <gap> continuarán <gap> os
0.00000063788161 00002 as <gap> federacións <gap> de
0.00000095682242 00002 real decreto legislativo <gap> que
0.00000264964524 00004 tribunal <gap> publica <gap> relación
```

Figura 3

Malia a baixa precisión do conxunto de unidades pluriléxicas non continuas identificadas polo programa, cómpre salientar a súa utilidade no "sinalamento" de termos, isto é, para recoñecer unidades terminolóxicas que non daríamos encontrado coa extracción automática partindo só de unidades continuas, como é o caso do termo *real decreto legislativo* (na Figura 3)

Por outro lado, nas unidades pluriléxicas continuas de tres ou máis elementos, atopámonos cunha ampla variedade de secuencias identificadas. No exemplo da Figura 4, constatamos a presenza de antropónimos (*María José Cancelo Baquero*), topónimos (*Santiago de Compostela*), nomes de entidades (*Academia Galega de Seguridade*, *Instituto Galego de Consumo*), datas (*xaneiro de 2005*), locucións de todo tipo (*con cargo á, se procede ou non, baixo a súa responsabilidade*), secuencias fragmentarias (*cada unha das*), e mesmo de unidades terminolóxicas doutros ámbitos, debido á diversidade temática dos textos xurídico-administrativo (*UGM por hectárea, especie bovina, infraestrutura de gas*).

0.00000264964524	00004	María José Cancelo Baquero
0.00000092782778	00002	suxeito pasivo o obrigado
0.00000063788161	00002	a normativa do tributo
0.00000214865395	00004	alleamento dos bens embargados
0.00000198723410	00003	1,4 UGM por hectárea
0.00000397446820	00006	da especie bovina .
0.00000105985816	00002	infraestrutura de gas :
0.00004238407200	00083	cada unha das
0.00000453604707	00008	Instituto Galego de Consumo
0.00000255152645	00004	salario base do convenio
0.00000157017018	00004	de incapacidade permanente total
0.00000271793033	00007	do permiso de conducir
0.00000431119997	00007	o xulgado do contencioso-administrativo
0.00000177497498	00004	cabe recurso de suplicación
0.00000806408389	00016	Xunta de Goberno Local
0.00001123293987	00019	imposto sobre bens inmobles
0.00000446517151	00007	Academia Galega de Seguridade
0.00006219507486	00098	Boletín Oficial del Estado
0.00065255473601	00996	Santiago de Compostela
0.00012663612142	00307	de xaneiro de 2005
0.00001457477902	00053	con cargo á
0.00000081527548	00002	se procede ou non
0.00000464850064	00010	baixo a súa responsabilidade

Figura 4

Tras revisar os resultados da extracción automática, na segunda fase da identificación das unidades terminolóxicas, procedemos á extracción manual. Para esta tarefa adoptamos como guieiros as listas de termos recollidas en diversos glosarios e dicionarios especializados do eido xurídico-administrativo do galego, do español e doutras linguas (Docampo Pereira 1998, Gómez Méixome et al. 2003, *Le droit de A à Z* 1998, Muñiz 1990, Ribó Durán 1995), identificando manualmente os termos relevantes das listas que non apareceran na primeira fase do proceso, e realizando en cada caso a busca destes termos no Corpus LEGA. A razón principal para estes termos non apareceren na extracción automática é a súa baixa frecuencia de aparición no corpus. Algúns dos termos así identificados (*anticrese*, *declaración de vontade* e *dereito transitorio*) son os que presentamos na Figura 5, no formato de táboa que ofrece a interface web ao CLUVI para os resultados dunha busca.

<u>CNC(935)</u>	2.º Os créditos garantidos con anticrese , sobre os froitos do inmovible gravado.	2.º Los créditos garantizados con anticresis , sobre los frutos del inmueble gravado.
<u>CIV(4920)</u>	Emitida a declaración, o executante poderá pedir que se libre, con testemuño do auto, mandamento de anotación ou inscrición no rexistro ou rexistros que correspondan, segundo o contido e obxecto da declaración de vontade .	Emitida la declaración, el ejecutante podrá pedir que se libre, con testimonio del auto, mandamiento de anotación o inscrición en el Registro o Registros que correspondan, según el contenido y objeto de la declaración de voluntad .
<u>CIV(570)</u>	Salvo que se estableza outra cousa en disposicións legais de dereito transitorio , os asuntos que correspondan ós tribunais civís serán substanciados sempre por estes conforme as normas procesuais vixentes, que nunca serán retroactivas.	Salvo que otra cosa se establezca en disposiciones legales de Derecho transitorio , los asuntos que correspondan a los tribunales civiles se sustanciarán siempre por éstos con arreglo a las normas procesales vigentes, que nunca serán retroactivas.

Figura 5

4. Elaboración das entradas terminolóxicas

4.1. Establecemento das equivalencias

Coincidimos coa profesora Teresa Cabré (1992) en que na aplicación práctica da terminoloxía non se debe confundir terminoloxía con tradución e en que o labor terminolóxico non pode consistir en traducir as formas dunha lingua a outra. No noso traballo terminolóxico, por mor da nosa orientación empírica e descritiva, para a elaboración das equivalencias terminolóxicas bilingües limitámonos a recoller, do xeito máis obxectivo posible, as denominacións documentadas no corpus de baleirado que as usuarias e os usuarios dunha lingua utilizan para referirse a un concepto.

Con esta finalidade, en primeiro lugar, analizamos os candidatos a termos extraídos do corpus en galego e decidimos se incluílos ou non no vocabulario. Logo procedemos a buscar a súa equivalencia en castelán no corpus aliñado. Por último, procedemos a fusionar as terminoloxías en galego e castelán, comprobando que os termos extraídos do castelán estivesen xa neste momento seleccionados. Este proceso serviunos para descubrir termos do galego que non aparecían na extracción e viceversa.

Na versión previa publicada do vocabulario, a presentación das entradas inclúe en primeiro lugar a forma lexemática do termo en galego xunto coa súa categoría gramatical. De seguido presentamos a súa equivalencia en castelán. Logo incluimos un contexto real bilingüe de uso extraído do Corpus LEGA, e xunto a el a referencia do texto ao que pertence. Tentamos sempre incluír como contexto ilustrativo algún exemplo de uso do termo que fose o máis definitorio posible. Con todo, isto non foi sempre posible debido á baixa frecuencia de aparición dalgúns dos termos. Na Figura 6 amosamos un exemplo da presentación das entradas na versión publicada do glosario.

1324 - **suxeito xurídico** [s. m.]

ES *sujeto jurídico*

GL En caso de non se formularen aclaracións e precisións, o tribunal só decretará o sobreseimento do preito se non é en absoluto posible determinar en qué consisten as pretensións do demandante ou, se é o caso, do demandado na reconvencción, ou fronte a qué **suxeitos xurídicos** se formulan as pretensións.

ES En caso de no formularse aclaraciones y precisiones, el tribunal sólo decretará el sobreseimiento del pleito si no fuese en absoluto posible determinar en qué consisten las pretensiones del actor o, en caso, del demandado en la reconvencción, o frente a qué **sujetos jurídicos** se formulan las pretensiones.

[Fonte: CIV (3117)]

ES *justiciable*

GL Pártese aquí de dous criterios inspiradores: por un lado, a necesidade de seguridade xurídica e, por outro, a escasa xustificación de somete-los mesmos **suxeitos xurídicos** a diferentes procesos e de provoca-la correspondente actividade dos órganos xurisdiccionais, cando a cuestión ou asunto litixioso pode resolverse razoablemente nun só.

ES Se parte aquí de dos criterios inspiradores: por un lado, la necesidad de seguridad jurídica y, por otro, la escasa justificación de someter a los mismos **justiciales** a diferentes procesos y de provocar la correspondiente actividad de los órganos jurisdiccionales, cuando la cuestión o asunto litigioso razonablemente puede zanjarse en uno solo.

[Fonte: CIV (127)]

203 - **ben de interese cultural** [s. m.]

ES *bien de interés cultural*

GL Resolución do 24 de novembro de 1999, da Dirección Xeral de Patrimonio Cultural, pola que se deixa sen efecto a incoación da declaración de **ben de interese cultural** coa categoría de conxunto histórico a favor da vila de Ribadeo, provincia de Lugo.

ES Resolución de 24 de noviembre de 1999, de la Dirección General de Patrimonio Cultural, por la que se deja sin efecto la incoación de la declaración de **bien de interés cultural** con la categoría de conjunto histórico a favor de la villa de Ribadeo, provincia de Lugo.

[Fonte: D04 (242)]

Figura 6

4.2. Estruturação do campo conceptual

Outra das informacións que pretendemos incorporar ao vocabulario é a adscripción de cada termo á un ámbito temático localizado na árbore conceptual do eido xurídico-administrativo. Para a definición desta ontoloxía, estamos a ter en conta diferentes clasificacións e índices temáticos referidos ao mundo do dereito, da economía e da administración que se poden consultar na Internet (*Wikipedia*, Mestre i Ribera 2004). Consultamos tamén as clasificacións temáticas levadas a cabo por diversos dicionarios, enciclopedias e vocabularios xurídico-administrativos en varias linguas (*Le droit de A à Z* 1998, Muñiz 1990, Ribó Durán 1995). Por último, comprobamos os tesauros da UNESCO (2004) e do CINDOC (1998). A clasificación temática do dicionario de Ribó Durán (1995), coas revisións que se coidaron necesarias, foi o alicerce principal seleccionado para a definición e estruturação do campo conceptual do vocabulario.

A partir destas clasificacións, realizamos unha fusión das subáreas que escollemos para o noso campo conceptual, e comprobamos qué termos do vocabulario se adscribían a cada unha delas. Presentamos na Figura 7 unha primeira versión de traballo do campo conceptual definido para o vocabulario, tentando na medida do posible axustar as subáreas aos datos fornecidos polo corpus, e acompañando cada ámbito temático resultante de exemplos de termos bilingües documentados no corpus de baleirado.

1. Dereito civil

1.1 Parte xeral

GL bens inmuebles - ES bienes inmuebles

GL boa fe - ES buena fe

1.2 Dereitos reais

GL deslinde e marcaxe - ES deslinde y amojonamiento

GL anticrese - ES anticresis

1.3 Obrigas

GL mora - ES mora

GL garantías personais - ES garantías personales

1.4 Contrato en xeral

GL obxecto do contrato - ES objeto del contrato

GL nulidade parcial - ES nulidad parcial

1.5 Contrato en particular

GL arrendador - ES arrendador

GL contrato de obras - ES contrato de obras

1.6 Dereito de danos

GL danos e perdas - ES daños y perjuicios

GL negligencia - ES negligencia

1.7 Dereito de familia

GL patria potestade - ES patria potestad
GL bens gananciais - ES bienes gananciales

1.8 Dereito de sucesións

GL sucesión intestada - ES sucesión intestada
GL testamento - ES testamento

1.9 Propiedade intelectual

GL propiedade intelectual - ES propiedad intelectual

2. Dereito hipotecario

GL anotación preventiva - ES anotación preventiva
GL hipoteca legal tácita - ES hipoteca legal tácita

3. Dereito notarial

GL acta notarial - ES acta notarial
GL dación de fe - ES dación de fe

4. Dereito de augas

GL augas públicas - ES aguas públicas

5. Dereito mercantil

GL rexistro mercantil - ES registro mercantil
GL seguro de responsabilidade civil - ES seguro de responsabilidad civil

5.1 Sociedades

GL capital social - ES capital social
GL sociedade anónima - ES sociedad anónima

5.2 Títulos valores

GL obriga de pagamento - ES pagaré
GL cheque - ES cheque

5.3 Dereito industrial

GL patente - ES patente
GL marca comunitaria - ES marca comunitaria

5.4 Dereito bancario

GL aval bancario - ES aval bancario
GL depósito bancario - ES depósito bancario

5.5 Dereito de navegación

GL fletamento - ES fletamento
GL consignatario de buque - ES consignatario de buque

6. Dereito concursal

GL quitación e espera - ES quita y espera
GL prelación de créditos - ES prelación de créditos

7. Dereito procesual

7.1 Dereito procesual civil

GL carga da proba - ES carga de la prueba
GL xulgado de paz - ES juzgado de paz

7.2 Dereito procesual penal

GL recurso de casación - ES recurso de casación
GL policía xudicial - ES policía judicial

7.3 Dereito procesual laboral

GL auto de sobreseimento - ES auto de sobreseimiento
GL xulgado do social - ES juzgado de lo social

8. Dereito constitucional

GL comisión do goberno - ES comisión del gobierno
GL Defensor do Pobo - ES Defensor del Pueblo

<p>9. Dereito administrativo GL multa coercitiva - ES multa coercitiva GL persoal laboral - ES personal laboral</p> <p>9.1 Contratación pública GL prego de cláusulas - ES pliego de cláusulas GL mesa de contratación - ES mesa de contratación</p> <p>9.2 Dereito urbanístico GL solo urbano - ES suelo urbano GL plans de saneamento - ES planes de saneamiento</p> <p>9.3 Xustiza administrativa GL recurso contencioso-administrativo - ES recurso contencioso-administrativo GL trámite de vista - ES trámite de vista</p> <p>10. Dereito fiscal GL base imponible - ES base imponible GL imposto sobre o valor engadido - ES impuesto sobre el valor añadido</p> <p>10.1 Dereito aduaneiro GL arancel de aduanas - ES arancel de aduanas GL aduanas - ES aduanas</p> <p>11. Dereito laboral GL comité de empresa - ES comité de empresa GL extinción do contrato - ES extinción del contrato</p> <p>12. Dereito da seguridade social GL xubilación parcial - ES jubilación parcial GL invalidez permanente - ES invalidez permanente</p> <p>13. Dereito penal GL falso testemuño - ES falso testimonio GL liberdade condicional - ES libertad condicional</p> <p>14. Dereito ambiental GL residuos sólidos - ES residuos sólidos</p> <p>15. Dereito informático GL comercio electrónico - ES comercio electrónico GL sinatura electrónica - ES firma electrónica</p> <p>16. Dereito internacional 16.1 Dereito comunitario GL Banco Central Europeo - ES Banco Central Europeo GL moeda única - ES moneda única</p>

Figura 7

5. Conclusión

O obxectivo desta comunicación é dar conta do traballo de investigación en curso, que se está a realizar no Seminario de Lingüística Informática (SLI) da Universidade de Vigo, encamiñado á constitución dun repertorio terminolóxico bilingüe xurídico-administrativo galego-castelán baseado nos datos que fornece o Corpus Lingüístico da Universidade de Vigo (CLUVI). A finalidade deste traballo é a creación dunha ferramenta de consulta terminolóxica que poida ser da máxima utilidade para a tradución xurídico-administrativa profesional en galego, e mais para o persoal da administración pública e do campo do dereito que traballa coa nosa lingua.

Entre as liñas futuras de traballo, hai dúas que se albiscan como prioritarias. Por unha banda, a adecuación da árbore conceptual aos contidos dunha realidade ampla e cambiante como é a do mundo do dereito e da administración pública. Por outra, a ampliación con máis textos do Corpus LEGA de baleirado, coa finalidade de acadar un grao significativo de representatividade no maior número de aspectos posibles desta realidade. Na versión actual do corpus, existen eidos mellor representados ca outros. Nos primeiros, a variedade terminolóxica documentada no corpus é moi abundante, por exemplo, nos eido do dereito civil, administrativo, penal, fiscal ou constitucional. Porén, noutros eidos como o dereito internacional, ambiental ou informático, a variedade terminolóxica documentada é moito menor. A curto prazo, imos tentar solucionar esta eiva cunha ampliación do Corpus LEGA centrada nos campos conceptuais nos que o número de textos é, nestes intres, reducido ou inexistente.

6. Bibliografía

- Cabré, T. (1992). *La terminologia: la teoria, els mètodes, les aplicacions*. Barcelona: Empúries.
- CINDOC (1998). *Tesouro de derecho* [en liña]. Madrid: CSIC. <http://pci204.cindoc.csic.es/tesauros/Derecho/DerTes.htm> [Consulta: 5 de xullo de 2005]
- Dias, G.; Guilloré, S.; Lopes, J.G.P. (2000). "Extraction automatique d'associations textuelles à partir de corpora non traités". En Rajman, M.; Chappelier, J-C. (eds) (2000) *Proceedings of 5èmes journées internationales d'analyse statistique des données textuelles*. Lausanne. 213-221
- Docampo Pereira, M. (1998). *Fraseoloxía xurídico-administrativa. Glosario galego-alemán-inglés-castelán* [en liña]. Vigo: Universidade de Vigo. <http://webs.uvigo.es/traductor/mdocampo/> [Consulta: 5 de xullo de 2005].
- Gómez Guinovart, X.; Sacau Fontenla, E. (2004). "Parallel corpora for the Galician language: building and processing of the CLUVI (Linguistic Corpus of the University of Vigo)". En Lino, T. et al. (ed.) (2004). *Proceedings of the 4th International Conference on Language Resources and Evaluation, LREC 2004*. Lisboa: ELRA. 1179-1182.
- Gómez Guinovart, X.; Torres Padín, A. (2005). *Vocabulario xurídico-administrativo galego-castelán. Baleirado terminolóxico documentado no Corpus Lingüístico da Universidade de Vigo*. Vigo: Universidade de Vigo (Servizo de Publicacións).
- Gómez Méixome, A.; González Montañés, A.; González Rodríguez, B. (2003). *Compilación de terminoloxía xurídica castelán-galego* [en liña]. [Compostela:] Consello da Avogacía Galega. <http://sal.avogacia.org/cdcomforma/docs/CRLC-G.pdf> [Consulta: 5 de xullo de 2005].
- Le droit de A à Z: dictionnaire juridique pratique*. 3ª ed. París: Editions Juridiques Européennes, 1998.
- Mestre i Ribera, E. (2004). *Manual sobre la Classificació Decimal Universal* [en liña]. <http://www.geocities.com/bibliotecaris/> [Consulta: 5 de xullo de 2005].
- Muñiz, E.G. (1990) *Diccionario terminológico de economía, comercio y derecho*. Madrid: Editorial Fontenebro.

Ribó Durán, L. (1995). *Diccionario de derecho*. 2ª ed. Barcelona: Bosch.

UNESCO (2004). *Tesaurus de la UNESCO* [en liña]. [París:] UNESCO. <http://databases.unesco.org/thesp/> [Consulta: 5 de xullo de 2005].

Wikipedia [en liña]. <http://es.wikipedia.org/wiki/Derecho> [Consulta: 5 de xullo de 2005].

7. Corpus de baleirado

7.1. Eido xudicial

Ley de Enjuiciamiento Civil / Lei de Axuizamento Civil, 7/01/2000.

Ley del Poder Judicial / Lei do Poder Xudicial, 23/12/2003.

Estatuto General de la Abogacía Española / Estatuto Xeral da Avogacía Española, 22/6/2001.

7.2. Constitución e estatutos

Constitución Europea / Constitución Europea, 2005.

Estatuto de Galicia / Estatuto de Galicia, 6/4/1981.

Constitución Española / Constitución Española, 27/12/1978.

7.3. Eido económico

Ley de Presupuestos Generales del Estado para el año 2005 / Lei de orzamentos xerais do Estado para o ano 2005, 27/12/2004.

Ley de medidas fiscales, administrativas y del orden social / Lei de medidas fiscais, administrativas e da orde social, 30/12/2003.

Ley de Medidas de Reforma del Sistema Financiero / Lei de medidas de reforma do sistema financeiro, 22/11/2002.

Ley General Tributaria / Lei Xeral Tributaria, 17/12/2003.

Reglamento del Impuesto sobre Sociedades / Regulamento do imposto sobre sociedades, 30/07/2004.

Reglamento del Impuesto sobre la Renta de las Personas Física / Regulamento do imposto sobre a renda das persoas físicas, 30/07/2004.

Reglamento de planes y fondos de pensiones / Regulamento de plans e fondos de pensións, 20/02/2004.

Ley de ordenación y supervisión de los seguros privados / Lei de ordenación e supervisión dos seguros privados, 29/10/2004..

Ley de Contratos de las Administraciones Públicas / Lei de contratos das administracións públicas, 12/10/2001.

Ley concursal / Lei concursal, 9/7/2003.

7.4. Eido social

Ley de extranjería / Lei de estranxeiría, 30/12/2004.

Ley de Protección Integral contra la Violencia de Género / Lei de protección integral contra a violencia de xénero, 28/12/2004.

Ley de igualdad de oportunidades / Lei de igualdade de oportunidades, 2/02/2003.

Estatuto de los Trabajadores / Estatuto dos Traballadores, 1994.

7.5. Sectores

Ley de Pesca Marítima del Estado / Lei de Pesca Marítima do Estado, 26/03/2001.

Ley de puertos / Lei de portos, 26/11/2003.

Ley de cohesión y calidad del Sistema Nacional de Salud / Lei de cohesión e calidade do Sistema Nacional de Saúde, 28/05/2003.

Reglamento General de Circulación / Regulamento xeral de circulación, 21/11/2003.

Reglamento del Sector Ferroviario / Regulamento do sector ferroviario, 30/12/2004.

Ley Orgánica de Universidade / Lei Orgánica de Universidades, 21/12/2001.

Lei Orgánica de calidad de la educación / Lei Orgánica de calidade da educación, 23/12/2002.

7.6. Diarios oficiais

Diario Oficial de Galicia, ano XXIV, núm. 101, 27 de maio de 2005.

Diario Oficial de Galicia, ano XXIV, núm. 62, 1 de abril de 2005.

Diario Oficial de Galicia, ano XXIV, núm. 49, 11 de marzo de 2005.

Diario Oficial de Galicia, ano XXIV, núm. 43, 3 de marzo de 2005.

Diario Oficial de Galicia, ano XXIV, núm. 39, 25 de febreiro de 2005.

Diario Oficial de Galicia, ano XXIV, núm. 36, 22 de febreiro de 2005.

Diario Oficial de Galicia, ano XXIV, núm. 32, 16 de febreiro de 2005.

Diario Oficial de Galicia, ano XXIV, núm. 23, 3 de febreiro de 2005.

Diario Oficial de Galicia, ano XXIV, núm. 16, 25 de xaneiro de 2005.

Diario Oficial de Galicia, ano XXIII, núm. 177, 10 de setembro de 2004.

Diario Oficial de Galicia, ano XXIII, núm. 176, 9 de setembro de 2004.

Diario Oficial de Galicia, ano XXIII, núm. 174, 7 de setembro de 2004.

Diario Oficial de Galicia, ano XXIII, núm. 166, 26 de agosto de 2004.

Diario Oficial de Galicia, ano XXIII, núm. 162, 20 de agosto de 2004.

Diario Oficial de Galicia, ano XXIII, núm. 157, 13 de agosto de 2004.

Diario Oficial de Galicia, ano XXIII, núm. 154, 10 de agosto de 2004.

Diario Oficial de Galicia, ano XXIII, núm. 141, 22 de xullo de 2004.

Diario Oficial de Galicia, ano XXIII, núm. 135, 14 de xullo de 2004.

Diario Oficial de Galicia, ano XXIII, núm. 133, 12 de xullo de 2004.
Diario Oficial de Galicia, ano XXIII, núm. 129, 6 de xullo de 2004.
Diario Oficial de Galicia, ano XIX, núm. 10, 17 de xaneiro de 2000.
Diario Oficial de Galicia, ano XIX, núm. 9, 14 de xaneiro de 2000.
Diario Oficial de Galicia, ano XIX, núm. 8, 13 de xaneiro de 2000.
Diario Oficial de Galicia, ano XIX, núm. 7, 12 de xaneiro de 2000.
Diario Oficial de Galicia, ano XIX, núm. 6, 11 de xaneiro de 2000.
Diario Oficial de Galicia, ano XIX, núm. 5, 10 de xaneiro de 2000.
Diario Oficial de Galicia, ano XIX, núm. 4, 7 de xaneiro de 2000.
Diario Oficial de Galicia, ano XIX, núm. 3, 5 de xaneiro de 2000.
Diario Oficial de Galicia, ano XIX, núm. 2, 4 de xaneiro de 2000.
Diario Oficial de Galicia, ano XIX, núm. 1, 3 de xaneiro de 2000.