

Galnet: WordNet 3.0 do galego

Xavier Gómez Guinovart, Xosé María Gómez Clemente,
Andrea González Pereira, Verónica Taboada Lorenzo

Grupo TALG, Universidade de Vigo
sli@uvigo.es

Resumo

Neste artigo presentamos o proxecto Galnet do Grupo TALG da Universidade de Vigo, dirixido á construción da versión galega do WordNet 3.0. Trátase dun proxecto que se atopa na súa fase inicial de desenvolvemento, mais do que xa se obtiveron uns primeiros resultados que están dispoñibles para a consulta. Describiremos os trazos xerais do proxecto, a metodoloxía e as ferramentas utilizadas, algúns aspectos lingüísticos do traballo e os resultados obtidos nesta primeira etapa.

1. *Introdución*¹

WordNet (Fellbaum, 1998; Miller et al., 1990) é unha base de coñecementos léxicos estruturada en forma de rede semántica. Nesta rede léxico semántica, cada nó é un concepto, e os fíos que conectan estes nós son as relacións semánticas (hiponimia, meronimia...) que se establecen entre os conceptos. Cada concepto na rede está representado polo grupo de lemas sinónimos que poden expresar ese concepto. Na terminoloxía asociada a WordNet, cada grupo de sinónimos é un *synset*, e cada sinónimo que forma parte dese grupo é unha *variant* (ou variante léxica dun mesmo concepto). WordNet inclúe, ao carón de cada *synset*, unha breve definición distintiva (ou *glosa*) do significado compartido por todas as variantes do *synset* e, en certos casos, exemplos de uso das variantes en contexto.

WordNet foi orixinalmente concibido para a lingua inglesa e, aínda que hoxe existen versións do WordNet en moitas linguas, o WordNet do inglés segue sendo arestora a máis desenvolvida e a de referencia. Os traballos do WordNet para esta lingua lévanse a cabo desde 1985 na Universidade de Princeton baixo a dirección do profesor George A. Miller. Na súa versión actual, o WordNet 3.0 do inglés contén 155.287 lemas (variantes) agrupadas en 117.659 grupos de sinónimos (*synsets*)².

WordNet constitúe, sen dúbida, o recurso de semántica léxica computacional máis importante

¹Este traballo foi financiado polo Ministerio de Ciencia e Innovación, dentro do proxecto *Multilingual Central Repository 2.0: TALG* do Subprograma de Acciones Complementarias, 2009 (ref. FFI2009-08317-E/FILO), concedido ao Grupo TALG da Universidade de Vigo.

²<http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html>

na actualidade, especialmente, no ámbito do procesamento da linguaxe natural (PLN), onde é utilizada, por exemplo, en tarefas de desambiguación semántica automática (Agirre y Edmonds, 2006), de recuperación da información (Varelas et al., 2005), de clasificación automática de textos (Elberichi, Rahmoun, e Bentaallah, 2008) ou de resumo automático (Barzilay y Elhadad, 1997).

Na actualidade existen versións do WordNet en diversas fases de desenvolvemento para moi diversas linguas³, incluídas o hebreo (Ordan et al., 2007), o italiano⁴ (Pianta, Benvivogli, e Girardi, 2002), o xaponés⁵ (Isahara et al., 2008), o castelán (Fernández y Vázquez, 2010; Fernández-Montraveta, Vázquez, e Fellbaum, 2008), o catalán (Benítez et al., 1998) e o euskera⁶ (Pociello, Agirre, e Aldezabal, 2011).

A maioría das versións en linguas distintas do inglés seguen o modelo de deseño de EuroWordNet (Vossen, 2002), na que os *synsets* que forman parte do WordNet da lingua propia están vinculados cos *synsets* do resto das linguas a través dun ILI (*interlingual index* ou índice interlingüístico) que é único para cada concepto e que principalmente está baseado nos *synsets* do WordNet inglés de referencia. Deste xeito, o conxunto de léxicos WordNet nos distintos idiomas permiten

³The Global WordNet Association mantén unha listaxe dos léxicos WordNet en distintas linguas na súa páxina web <<http://www.globalwordnet.org/gwa/wordnet-table.htm>>

⁴Estas dúas consultables en liña en <<http://multiwordnet.fbk.eu/english/home.php>>

⁵Consultable en liña en <<http://nlpwww.nict.go.jp/wn-ja/index.en.html>>

⁶Estas tres consultables en liña en <<http://adimen.si.ehu.es/cgi-bin/wei/public/wei.consult.perl>>

a conexión entre os synsets de calquera par de linguas a través do ILI, constituíndo así un recurso de gran utilidade en aplicacións das tecnoloxías lingüísticas que precisan o procesamento pluringüe da linguaxe, como a tradución automática ou a recuperación interlingüística da información. Cómpre salientar tamén que os conceptos que forman parte do ILI están catalogados en xerarquías de dominios e ontoloxías, como a xerarquía de dominios IRST (Bentivogli et al., 2004) ou as ontoloxías SUMO (Pease, Niles, e Li, 2002) e Top Concept Ontology (Álvarez et al., 2008), o que permite un mellor aproveitamento do recurso en diversas aplicacións.

Neste artigo presentamos o proxecto Galnet do Grupo TALG (Tecnoloxías e Aplicacións da Lingua Galega) da Universidade de Vigo, dirixido á construción da versión galega do WordNet 3.0. Trátase dun proxecto que se atopa na súa fase inicial de desenvolvemento, mais do que xa se obtiveron uns primeiros resultados que están dispoñibles para a consulta. Nos seguintes apartados describiremos os trazos xerais do proxecto, a metodoloxía e as ferramentas utilizadas, algúns aspectos lingüísticos do traballo e os resultados obtidos nesta primeira etapa.

2. O proxecto Galnet

O obxectivo do proxecto Galnet non é outro que a construción dun WordNet para o galego (isto é, dun Galnet) aliñado co ILI xerado a partir do WordNet 3.0 do inglés. Este proxecto está incorporado nun proxecto máis amplo encamiñado á integración coordinada das versións castelá, catalá, galega e vasca do WordNet 3.0, no que participan os grupos de investigación do IXA (da Universidade do País Vasco), TALP (Universitat Politècnica de Catalunya), GRIAL (Universitat Autònoma de Barcelona, Universitat de Barcelona, Universitat de Lleida e Universitat Oberta de Catalunya), IULATERM (Universitat Pompeu Fabra) e TALG (Universidade de Vigo).

O marco de desenvolvemento no que se integra o Galnet é o do Multilingual Central Repository (MCR) (Atserias et al., 2004), unha plataforma de libre consulta⁷ desenvolvida ao abeiro do proxecto europeo Meaning (IST-2001-34460) e dos proxectos de financiamento estatal KNOW (TIN2006-15049-C03) e KNOW2 (TIN2009-14715-C04-01). O MCR abrangue na actualidade os léxicos WordNet de cinco linguas (inglés, español, catalán, vasco e galego) enlazados interlingüísticamente polo ILI correspondente ao WordNet 3.0 e cos ILI categorizados

na xerarquía de dominios IRST e nas ontoloxías SUMO e Top Concept Ontology.

A seguir describiremos a metodoloxía e as ferramentas empregadas na construción do Galnet na súa primeira etapa de desenvolvemento.

2.1. Metodoloxía e ferramentas

Os obxectivos desta primeira fase na construción do Galnet foron, en primeiro lugar, elaborar un conxunto de synsets básicos para a operatividade do recurso na lingua galega e, en segundo lugar, fornecer un conxunto suficiente de entradas que servise para ilustrar a utilidade do recurso e ampliar a súa operatividade.

A metodoloxía utilizada para levar a cabo o primeiro obxectivo consistiu na creación da versión galega dos synsets nominais e verbais pertencentes aos Basic Level Concepts (BLC). Para o segundo obxectivo, elaboramos as entradas galegas para os ficheiros lexicográficos do WordNet correspondentes aos nomes relacionados coas partes do corpo e coas substancias, e para unha parte dos correspondentes aos adxectivos de tipo xeral.

Os Basic Level Concepts (Izquierdo, Suárez, e Rigau, 2007) son un conxunto seleccionado de conceptos do WordNet que representan un compromiso entre dous principios de caracterización contraditorios: representar o maior número posible de conceptos (ser conceptos abstractos) e representar o maior número posible de trazos distintivos (ser conceptos concretos). Así, os BLC aparecen tipicamente na parte media das relacións semánticas xerárquicas de WordNet, sendo deste modo frecuentes e destacados, nin claramente xerais nin demasiado específicos. A primeira tarefa do proxecto Galnet consistiu en elaborar manualmente a versión galega dos BLC (649 synsets nominais e 616 synsets verbais) recollidos no apartado `freqmin20/all` da distribución oficial⁸ dos BLC do WordNet 3.0, sen incluír na adaptación nin as glosas nin os exemplos incluídos nos synsets correspondentes da lingua inglesa.

Unha vez elaborado o núcleo inicial de synsets do Galnet, continuamos a ampliación do recurso a partir da tradución asistida dos ficheiros lexicográficos do WordNet para os nomes relacionados coas partes do corpo e coas substancias, e para unha parte dos adxectivos de tipo xeral. A ferramenta empregada nesta tarefa foi Google Translator Toolkit⁹, unha ferramenta colaborativa en liña que nos permitiu a postedición asistida das propostas de tradución automática do tradu-

⁷<http://adimen.si.ehu.es/web/MCR/>

⁸<http://adimen.si.ehu.es/web/BLC/>

⁹<http://translate.google.com/toolkit/>

tor de Google.

A selección dos ficheiros lexicográficos relacionados coas partes do corpo e coas substancias veu motivada pola nosa vontade de aproveitar o material textual e terminolóxico elaborado en traballos previos do grupo e recollidos no Corpus Técnico do Galego¹⁰ e na base de datos terminolóxica da Termoteca¹¹. A incorporación dos adxectivos xustificouse en virtude dunha maior cobertura ingüística dos resultados nesta fase inicial do traballo. No apartado seguinte, describiremos algúns dos problemas lexicográficos máis relevantes con que nos atopamos nesta xeira.

3. Tratamento da microestrutura

A microestrutura das entradas de WordNet inclúen, para cada concepto ou nó da rede, o grupo de sinónimos ou synset que lexicalizan o concepto, e unha glosa ou definición que pode ir acompañada dun exemplo. En EuroWordNet, ademais, a cada concepto lle corresponde un ILI (índice interlingüístico), un identificador único do concepto que permite relacionar os conceptos entre os léxicos en formato WordNet das distintas linguas.

3.1. Synsets

O grupo de sinónimos ou synset pode conter unha única palabra ou varias. Igualmente, os sinónimos poden ser unidades monoléxicas ou pluriléxicas. A serie de sinónimos dun synset está composta por varias palabras que entran en relación de sinonimia parcial ou total, coa premisa básica de que nalgún contexto poden ser intercambiáveis. WordNet non diferencia os graos da relación de semellanza, nin ten en conta, polo de agora, a connotación do rexistro. Isto provoca que non se poida diferenciar cunha etiqueta de rexistro os usos coloquiais ou vulgares que se recollen nas series sinonímicas no WordNet orixinal inglés e que foron trasladados ao Galnet en versión galega. Por exemplo, no synset correspondente ao concepto con código ILI 05514410-n, a serie sinonímica en inglés {female genitalia, female genitals, female genital organ, fanny} non inclúe na serie ningunha indicación de rexistro vulgar para o último dos sinónimos, o mesmo que sucede na súa adaptación galega {xenitais femininos, órgano xenital feminino, cona}.

A fraseoloxía no Wordnet 3.0 do inglés está tratada dun modo moi irregular, aparecendo entre as variantes nalgún synset illado. Porén, no Galnet desexaríamos que a inclusión de fraseoloxía fose algo común, como un tipo

máis de expresións lexicalizadas que poden entrar en relación sinonímica con outras formas léxicas. Asemade, pensamos que Galnet pode contribuir á xeneralización do uso destas expresións pluriléxicas unidas claramente a aspectos socio-culturais da nosa lingua. A inclusión de unidades fraseolóxicas no Galnet é, ata certo punto, independente da súa aparición no WordNet do inglés, como se pode apreciar nos seguintes catro exemplos (onde tampouco se marca na versión galega a connotación de rexistro coloquial que sería necesaria nalgúns casos):

01041061-v: {close up, clam up, dummy up, shut up, belt up, button up, be quiet, keep mum} >{calar, calar coma un peto, calar coma unha estoa, coser a boca, pechar o bico, non dar un chío, estar en silencio}

01823149-v: {care a hang, give a hoot, give a hang, give a damn} >{importar un pataco, importar un farrapo de gaita, pesarlle tanto a alguén un ombro coma outro, non lle importar a alguén algo unha chisca}

01825125-v: {begrudge, resent} >{dar de mala gana, dar a contragusto, dar torcendo o fociño, ter envexa}

01983597-v: {bristle, uprising, stand up} >{arrepia as carnes, poñer os pelos de espeto, pór os pelos de punta, pór os pelos dereitos}

Canto ao tratamento da terminoloxía, no WordNet recóllense todos os termos asociados a un determinado concepto, tanto se pertencen a niveis de especialización altos como formas populares ou vulgarizadas, coa limitación xa indicada para o léxico xeral e fraseolóxico da imposibilidade de marcar as formas con etiquetas de rexistro:

14848642-n: {colutorio, enxaugadura} ‘solución médica usada para facer gargarexos e para lavar a boca’

5573895-n: {fémur, óso da coxa} ‘óso máis longo e grosso do esqueleto humano. Esténdese dende a pelve ata o xeonllo’

Con todo, o problema principal que atopamos á hora de facer un tratamento axeitado da terminoloxía foi a aparición de erros na taxonomía da zooloxía incluída nos BLC nominais da WordNet do inglés:

01352574-n: {bacteria genus} >{xénero das bacterias} (Porén, as bacterias son un dominio, non un xénero zoolóxico)

01342529-n: {animal order} >{orde dos animais} (Animalia é un reino)

01862557-n: {mammal family} >{familia dos mamíferos} (Os mamíferos son unha clase)

¹⁰<http://sli.uvigo.es/CTG/>

¹¹<http://sli.uvigo.es/termoteca/>

3.2. Glosas

Nas definicións de WordNet 3.0 non hai un patrón homoxéneo que sexa respectado na totalidade dos synsets, senón que existen diferentes esquemas entre os que se poden salientar os seguintes tipos:

a) Definición intensional: defínese o termo enumerando as características ou propiedades que fan que se sitúe dentro dun concepto determinado. Por exemplo:

14631295-n: {berilio, Be, glucinio, número atómico 4} ‘elemento metálico, bivalente, tóxico, gris, crebadizo, duro e luminoso’

b) Definición negativa: defínese o termo polo seu contrario.

00023383-a: {impreciso, inexacto, inxusto} ‘non exacto’. Ex: *unha tradución inexacta; o termómetro é impreciso*

c) Definición operacional: defínese o termo indicando a súa composición e outros elementos relativos á súa magnitude, como a masa, o tempo, a temperatura, etc.

15103911-n: {metal de Wood, aliaxe de Wood} ‘aliaxe que contén o 50 % de bismuto máis chumbo, estaño e cadmio; fúndese a preto de 160 graos Fahrenheit’

d) Definición por xénero e diferenza: no inicio da definición clasifícase o termo polo xénero ao que pertence e despois engádense características propias que axudan a delimitar o concepto.

05445389-n: {mitocondria} ‘orgánulo que contén as encimas responsables da produción de enerxía’

e) Definición teórica: defínese o termo seguindo unha determinada teoría ou disciplina.

14842703-n: {aire} ‘un dos catro elementos que segundo Empédocles compoñen o universo’

Outro tipo de definicións máis problemáticas son as que recolleemos a continuación:

f) Definicións circulares: definicións que non dan máis información da que xa se extrae do propio termo. Este tipo de definicións son moi abundantes no WordNet 3.0 do inglés. Por exemplo:

15075298-n: {toilet roll} ‘a roll of toilet paper’ >{rolo de papel hixiénico} ‘rolo de papel fino e estreito que serve para usos hixiénicos’ (Carballeira Anllo, 2009).

g) Definicións pouco específicas. Dentro deste tipo de definicións atopamos pouca especificación en diferentes elementos:

g1) Existen casos nos que a información que chega a definición non aclara o concepto, posto que non se dá información esencial (ou complementaria) que axude a identificar o termo:

15090742-n: {B-complex vitamin, B complex, vitamin B complex, vitamin B, B vitamin, B} ‘originally thought to be a single vitamin but now separated into several B vitamins’ >{complexo vitamínico B, complexo B, vitamina do complexo B, vitamina B, B} ‘grupo de vitaminas hidrosolubles cada unha coa súa propia función metabólica diferente das demais’ (Daviña Facal, 2000).

g2) Imprecisión na descrición dalgunhas características:

15056827-n: {spindrift, spoondrift} ‘spray blown up from the surface of the sea’ >{escuma de mar, espuma de mar} ‘conxunto de burbullas que se forman na tona da auga do mar cando esta bate contra as rochas ou contra un corpo que se move no seu interior’ (Carballeira Anllo, 2009).

g3) Nalgúns casos só se describen os usos ou a orixe do termo pero non as súas características esenciais:

15062468-n: {tallow} ‘obtained from suet and used in making soap, candles and lubricants’ >{sebo} ‘graxa sólida dos animais, constituída principalmente por estearato e palmitato de glicerilo. Soluble en etanol e éter; emprégase na fabricación de xabóns e candeas, para o curtume do coiro e como produto intermedio’ (Real Academia de Ciencias Exactas, Físicas y Naturales, 1996).

g4) En oposición ao punto anterior, existen tamén definicións que describen características pero non usos:

14673032-n: {crocolite} ‘a rare lead chromite mineral that forms bright orange crystals’ >{crocoíta} ‘mineral raro de cromato de chumbo que forma cristais laranxas brillantes e que se emprega ás veces como pigmento’ (definición propia).

h) Confusión conceptual da definición con respecto ao termo. Este tipo de problema apréciase no seguinte synset, no que se fala do lume como un combustible cando, en realidade, é o resultado dunha combustión:

14686186-n: {fire} ‘fuel that is burning and is used as a means for cooking’ Ex: *put the kettle on the fire; barbecue over an open fire* >{lume} ‘chama ou labarada dunha substancia en combustión; materia en combustión’ Ex: *no alto do monte había lume; no inverno acendemos o lume na lareira* (Carballeira Anllo, 2009).

i) Existen series de cohipónimos nos que non se establece un patrón de definición; por exemplo, nas veas e arterias debería aparecer sempre a orixe, o final e a función do vaso. Nos seguintes exemplos non se explicita a orixe do conduto sanguíneo:

05349906-n: {laryngeal artery, arteria laryngea} ‘either of two arteries that supply blood to the larynx’ >{arteria larínxea, laryngea arteria} ‘unha das dúas arterias que fornecen sangue á larínxe’

05357366-n: {anastomotic vein, vena anastomotica} ‘either of two communicating veins serving the brain’ >{vea anastomótica, vena anastomotica} ‘unha das dúas veas comunicantes que serve o cerebro’

3.3. Exemplos

Con respecto aos exemplos que poden complementar a glosa, non existe no WordNet 3.0 do inglés unha estrutura recorrente, isto é, non aparecen exemplos en todos os synsets e, no caso de apareceren, non hai unha pauta que unifique a cantidade dos mesmos. Nalgúns casos, como no dos adxectivos, os exemplos son, predominantemente, citas textuais.

Na versión galega, optamos por manter esta estrutura e respectar as citas para non alterar o paralelismo con outras linguas, agás nos casos nos que o exemplo identificaba o termo como unha realidade propia doutra cultura allea á galega. Nestes últimos tentamos que os exemplos se aproximasesen, na medida do posible, á nosa cultura, como se pode observar nos seguintes exemplos:

00699651-a: {inflected} ‘showing alteration in form (especially by the addition of affixes)’ Ex: ‘boys’ and ‘swam’ are inflected English words; German is an inflected language >{flexionado, flexivo} ‘que mostra modificacións na forma (especialmente a través da adición de afixos)’ Ex: ‘nenos’ e ‘nadou’ son palabras flexionadas en galego; o alemán é unha lingua flexiva

02297664-a {standard, received} ‘conforming to the established language usage of educated native speakers’ Ex: standard English (American); received standard English is sometimes called the King’s English (British) >{estandarizado, estándar, dentro do estándar} ‘que concorda co uso da linguaxe establecida dos falantes cultos nativos’ Ex: galego estándar

A continuación séguese unha breve descrición do tipo de exemplos que consideramos problemáticos no WordNet 3.0 do inglés e, no seu caso, das solucións achegadas para o GalNet.

a) Existen exemplos que se adaptan só a unha determinada zona, cultura ou lingua. O que intentamos facer, neste caso, foi modificar os exemplos para que fosen máis xerais ou, simplemente, axeitalos ao caso galego:

05514410-n: {female genitalia, female genitals, female genital organ, fanny} ‘external female sex

organs’ Ex: in England ‘fanny’ is vulgar slang for female genitals >{xenitais femininos, órgano xenital feminino, cona} ‘órganos sexuais femininos externos’ Ex: en Galicia ‘conna’ é un termo vulgar para referirse aos órganos xenitais femininos

b) Algúns exemplos presentan características esenciais ou complementarias importantes para explicar o concepto, polo que serían elementos propios da definición. Un exemplo deste tipo atópase no seguinte synset:

14707361-n: {adenosine deaminase, ADA} ‘an enzyme found in mammals that can catalyze the deamination of adenosine into inosine and ammonia’ Ex: ADA deficiency can lead to one form of severe combined immunodeficiency disease; the gene encoding ADA was one of the earlier human genes to be isolated and cloned for study >{adenosina deaminase, ADA} ‘encima que se atopa nos mamíferos que poden catalizar a desaminación da adenosina en inosina e amoníaco’ Ex: a discapacidade da ADA pode levar a unha forma de enfermidade da inmunodeficiencia combinada severa; o xene que codifica a ADA foi un dos xenes humanos que antes de puido illar e clonar para estudalo

Neste caso o primeiro dos exemplos é unha característica que delimita o concepto, polo que o normal sería que aparecese na definición.

c) Hai exemplos que non contribúen esencialmente a aclarar o concepto e que transmiten outro tipo de informacións, algunhas delas de marcado carácter ideolóxico. Na adaptación do inglés ao galego das entradas para nomes de partes do corpo e de substancias, este tipo de exemplos foron modificados, na medida do posible, para facelos máis neutrais:

14699752-n: {gem, gemstone, stone} ‘a crystalline rock that can be cut and polished for jewelry’ Ex: he had the gem set in a ring for his wife; she had jewels made of all the rarest stones >{xema, pedra preciosa} ‘pedra cristalina, que pode ser cortada e pulida para xoiaría’ Ex: os rubís son xemas; tiña xoias feitas de todas as pedras preciosas máis raras

00512261-a: {incompetent, unqualified} ‘legally not qualified or sufficient’ Ex: a wife is usually considered unqualified to testify against her husband; incompetent witnesses >{incapacitado, legalmente incapacitado, legalmente non capacitado} ‘privado de capacidade xurídica pola concorrencia de determinadas causas legais, legalmente non cualificado’ Ex: en xeral non se aceptan nun ruízo as testemuñas legalmente incapacitadas

	WN30		Galnet	
	Vars	Syns	Vars	Syns
N	117798	82115	9183	5646
V	11529	13767	1414	616
Adx	21479	18156	4864	3114
Adv	4481	3621	0	0
TOTAL	155287	117659	15461	9376

Táboa 1: Resultados iniciais.

4. Resultados

Na Táboa 1 preséntanse, agrupados en categorías (nomes, verbos, adxectivos e adverbios) e diferenciando entre synsets e variantes, os resultados acadados desde un punto de vista cuantitativo no estado actual de desenvolvemento do proxecto Galnet¹². Estes resultados corresponden a 649 sysnets (1.333 variantes léxicas) dos BLC de categoría nominal, 616 synsets (1.414 variantes) dos BLC de categoría verbal, 2.014 synsets (3.550 variantes) do ficheiro lexicográfico de nomes relacionados coas partes do corpo, 2.983 synsets (4.300 variantes) do ficheiro lexicográfico de nomes de substancias, e 3.114 synsets (4.864 variantes) do conxunto de adxectivos de tipo xeral incluídos en WordNet 3.0. Tendo en conta os resultados obtidos en todas as categorías, podemos concluír que o crecemento lexicográfico do Galnet nesta primeira fase do proxecto ofrece unha cobertura semántica moi próxima ao 10 % con relación a cobertura de referencia da WordNet 3.0 en lingua inglesa.

De momento, os resultados iniciais do proxecto Galnet poden ser consultados en liña a través da interface WEI (Web EuroWordNet Interface)¹³ do Multilingual Central Repository e forman parte tamén da plataforma RILG de Recursos Integrados da Lingua Galega¹⁴. Porén, está prevista a distribución libre dos recursos léxicos xerados no Galnet para toda a comunidade científica con licenza GPL (GNU General Public License da Free Software Foundation)¹⁵.

5. Conclusións

Neste artigo presentamos o proxecto Galnet do Grupo TALG da Universidade de Vigo, dirixido á construción da versión galega do WordNet 3.0. Aínda que se trata dun proxecto en curso, algúns dos seus resulta-

dos xa se poden consultar libremente en Internet mediante a interface web de consulta do WEI (Web EuroWordNet Interface) accesible en <<http://adimen.si.ehu.es/cgi-bin/wei/public/wei.consult.perl>> ou a través da plataforma RILG de Recursos Integrados da Lingua Galega dispoñible en <<http://sli.uvigo.es/RILG/>>. Así mesmo, está prevista a dispoñibilización libre en Internet con licenza GPL das sucesivas versións do Galnet xeradas ao longo do proxecto.

Bibliografía

- Agirre, Eneko e Philip Edmonds. 2006. *Word Sense Disambiguation*. Springer, Berlin.
- Álvez, J., J. Atserias, J. Carrera, S. Climent, A. Oliver, e G. Rigau. 2008. Consistent Annotation of EuroWordNet with the Top Concept Ontology. En *Proceedings of the 4th Global WordNet Conference (GWC'08)*.
- Atserias, J., L. Villarejo, G. Rigau, E. Agirre, J. Carroll, B. Magnini, e P. Vossen. 2004. The MEANING Multilingual Central Repository. En *In Proceedings of the Second International WordNet Conference*, páxinas 80–210.
- Barzilay, Regina e Michael Elhadad. 1997. Using lexical chains for text summarization. En *Proceedings of the ACL/EACL 1997 Workshop on Intelligent Scalable Text Summarization*, páxinas 10–17.
- Bentivogli, Luisa, Pamela Forner, Bernardo Magnini, e Emanuele Pianta. 2004. Revising the wordnet domains hierarchy: semantics, coverage and balancing. En *Proceedings of the Workshop on Multilingual Linguistic Resources*, páxinas 101–108.
- Benítez, Laura, Sergi Cervell, Gerard Escudero, Mònica López, German Rigau, e Mariona Taulé. 1998. Methods and Tools for Building the Catalan WordNet. En *In Proceedings of ELRA Workshop on Language Resources for European Minority Languages*.
- Carballeira Anllo, Xosé María, editor. 2009. *Gran diccionario xerais da lingua*. Edicións Xerais da Lingua, Vigo.
- Daviña Facal, Luís. 2000. *Diccionario das ciencias da natureza e da saúde*. Deputación Provincial da Coruña, A Coruña.
- Elberrichi, Zakaria, Abdellatif Rahmoun, e Mohamed Amine Bentaallah. 2008. Using wordnet for text categorization. *Int. Arab J. Inf. Technol.*, 5(1):16–24.

¹²Os datos cuantitativos completos sobre o WordNet 3.0 do inglés poden consultarse en <<http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html>>

¹³<<http://adimen.si.ehu.es/cgi-bin/wei/public/wei.consult.perl>>

¹⁴<<http://sli.uvigo.es/RILG/>>

¹⁵<<http://www.gnu.org/licenses/gpl.html>>

- Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge.
- Fernández, A. e G. Vázquez. 2010. La construcción del wordnet 3.0 en español. En M. A. Castillo e J. M. García, editores, *La lexicografía en su dimensión teórica*, páxinas 201–220, Málaga. Universidad de Málaga.
- Fernández-Montraveta, A., G. Vázquez, e C. Fellbaum. 2008. The Spanish Version of WordNet 3.0. En A. Storrer, editor, *Text Resources and Lexical Knowledge*, páxinas 175–182, Berlín. Mouton de Gruyter.
- Isahara, Hitoshi, Fransis Bond, Kiyotaka Uchimoto, Masao Utiyama, e Kyoko Kanzaki. 2008. Development of the Japanese WordNet. En *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*.
- Izquierdo, R., A. Suárez, e G. Rigau. 2007. Exploring the automatic selection of basic level concepts. En *Proceedings of the International Conference on Recent Advances on Natural Language Processing (RANLP'07)*.
- Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross, e Katherine J. Miller. 1990. Introduction to WordNet: an On-line Lexical Database. *International Journal of Lexicography*, 3(4):235–244.
- Ordan, Noam, Bar Ilan, Noam Ordan, e Shuly Wintner. 2007. Hebrew WordNet: a test case of aligning lexical databases across languages. *International Journal of Translation*, 19:39–58.
- Pease, Adam, Ian Niles, e John Li. 2002. The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. En *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*.
- Pianta, Emanuele, Luisa Bentivogli, e Christian Girardi. 2002. MultiWordNet: developing an aligned multilingual database. En *Proceedings of the First International Conference on Global WordNet*.
- Pociello, Elisabete, Eneko Agirre, e Izaskun Aldezabal. 2011. Methodology and construction of the Basque WordNet. *Language Resources and Evaluation*, 45:121–142.
- Real Academia de Ciencias Exactas, Físicas y Naturales. 1996. *Vocabulario científico y técnico*. Espasa, Madrid.
- Varelas, Giannis, Epimenidis Voutsakis, Paraskevi Raftopoulou, Euripides G.M. Petrakis, e Evangelos E. Milios. 2005. Semantic similarity methods in wordnet and their application to information retrieval on the web. En *Proceedings of the 7th annual ACM international workshop on Web information and data management*, páxinas 10–16.
- Vossen, Piek. 2002. WordNet, EuroWordNet and Global WordNet. *Revue française de linguistique appliquée*, 7:27–38.