

AnCora-Net: Integración multilingüe de recursos lingüísticos semánticos

AnCora-Net: Multilingual integration of linguistic semantic resources

Mariona Taulé, Oriol Borrega, M. Antònia Martí

CLiC-UB, Univesitat de Barcelona

Gran Via de Les Corts. 585, 08007, Barcelona

{mtaule;amarti;oriol.borrega}@.ub.edu

Resumen: AnCora-Net es un léxico verbal multilingüe creado a partir de la integración de los léxicos verbales AnCora-Verb, del catalán y español, en el Unified Verb Index del inglés. El Unified Verb Index aúna diferentes fuentes de conocimiento del inglés de amplia cobertura que son sin duda un referente en representación semántica. La integración de nuestros recursos con los del inglés nos permite enriquecer el contenido de los léxicos AnCora-Verb con información semántica codificada para el inglés. Asimismo, el Unified Verb Index también se enriquece con la incorporación de los léxicos AnCora-Verb, del catalán y español, dando lugar a un recurso multilingüe que puede ser útil para estudios comparativos.

Palabras clave: recurso léxico, recurso multilingüe, semántica, estructura argumental

Abstract: AnCora-Net is a multilingual verbal lexicon built from the mapping of the Catalan and Spanish AnCora-Verb verbal lexicons into the English Unified Verb Index. The Unified Verb Index combines different sources of knowledge for English of wide coverage, which are a referent in semantic representation. The integration of our resources to the Unified Verb Index will enrich the contents of AnCora-Verb lexicons with semantic information coded for English. In the same way, the Unified Verb Index will be also related to equivalent verbs in another languages giving rise a multilingual resource that can be useful for comparative studies.

Keywords: Lexical resource, multilingual resource, semantics, argument structure

1 Introducción

En este artículo se presenta un nuevo recurso de tecnología lingüística, AnCora-Net¹, un léxico verbal multilingüe –catalán, español e inglés– en el que se combina información sintáctico-semántica y conceptual procedente de distintas fuentes. AnCora-Net se ha creado a partir de la integración de los léxicos verbales AnCora-Verb² del catalán y español (Aparicio et al., 2008) en el Unified Verb Index³ (UVI) del inglés. El UVI aúna diferentes fuentes de

conocimiento del inglés –VerbNet⁴, PropBank⁵, FrameNet⁶, WordNet⁷ y OntoNotes⁸– de amplia cobertura que son sin duda un referente en representación semántica. La integración de nuestros recursos al UVI enriquece el contenido de los diferentes recursos léxicos implicados mediante la incorporación de la información semántica codificada en cada uno de ellos. Se parte de la hipótesis que estos recursos aunque representan y clasifican los predicados verbales

⁴<http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>

⁵<http://verbs.colorado.edu/~mpalmer/projects/ace.html>

⁶<http://framenet.icsi.berkeley.edu/>

⁷<http://wordnet.princeton.edu/>

⁸<http://verbs.colorado.edu/VSAP/>

¹ <http://clic.ub.edu/corpus/ancoranet>

² <http://clic.ub.edu/corpus/ancora-lexics>

³ <http://verbs.colorado.edu/verb-index/>

de manera distinta son compatibles y complementarios (Taulé et al., 2010). La conexión de estos recursos permite integrar diferentes aspectos del significado verbal, además de disponer de las relaciones multilingües entre verbos de tres lenguas distintas. Estos léxicos integrados y enriquecidos con información semántica y conceptual podrán ser utilizados tanto en aplicaciones de procesamiento del lenguaje natural (traducción automática y recuperación y extracción de información multilingüe), como en el análisis lingüístico.

A continuación se describen las características de los recursos lingüísticos integrados en AnCora-Net (sección 2) y la metodología seguida para su obtención (sección 3). Finalmente, se presentan las conclusiones (sección 4).

2 Recursos lingüísticos de AnCora-Net

En esta sección se describen las características principales de los recursos lingüísticos que se integran en AnCora-Net.

2.1 AnCora-Verb Lexicons

AnCora-Verb incluye dos léxicos verbales utilizados para la anotación semántica con argumentos, papeles temáticos y clase semántica de los corpus AnCora⁹ (Taulé et al., 2008): AnCora-Verb-CA para el catalán (que contiene un total de 2.248 verbos correspondientes a 3.118 sentidos) y AnCora-Verb-ES para el castellano (con un total de 2.820 verbos correspondientes a 3.938 sentidos).

En dichos léxicos, cada predicado puede tener diferentes sentidos ('01, 02... n') y cada sentido está relacionado con uno o más esquemas sintáctico-semánticos (*transitive*, *ditransitive*, *unaccusative*, *inergative*, etc.). Cada uno de estos esquemas tiene asociada la clase semántica, la estructura argumental, con papeles temáticos asignados a cada argumento, y su proyección a la estructura sintáctica (Véase figura 1). Los esquemas sintáctico-semánticos se distinguen de acuerdo a las cuatro clases ontológicas eventivas (*accomplishment*, *achievement*, *states*, y *activities*), y a las alternancias de diátesis en qué aparecen. Sólo se

han considerado las alternancias más productivas: causativa/incoativa, activa/pasiva, resultativa, transitiva/intransitiva, objeto cognado, extensión de objeto y alternancia benefactiva (Vázquez et al., 2000).

En la figura 1 se muestra la información asociada con el verbo *acompañar* en AnCora-Verb-Es: el lema ("acompañar"); la categoría ("verb"); los diferentes sentidos asociados a las clases semánticas correspondientes —en este caso, el primer sentido con la clase *A21.transitive-agentive-patient*; la correspondencia entre la función sintáctica, argumento y rol semántico (p.e., en el primer sentido el sujeto "suj" corresponde al primer argumento "arg0" con el rol semántico de agente "agt"); y las alternancias de diátesis en que puede aparecer el verbo (en este sentido, *acompañar* aparece en pasiva "B22.unaccusative-passive-transitive"). Como se puede observar, la expresión de la alternancia de pasiva implica un cruce de argumentos: el objeto afectado aparece como objeto directo ("cd") en la estructura activa y como sujeto ("suj") en la estructura pasiva, tratándose en ambos casos del argumento ("arg1") con el rol semántico de paciente ("pat"). Finalmente, se incluyen para cada sentido los ejemplos del corpus.

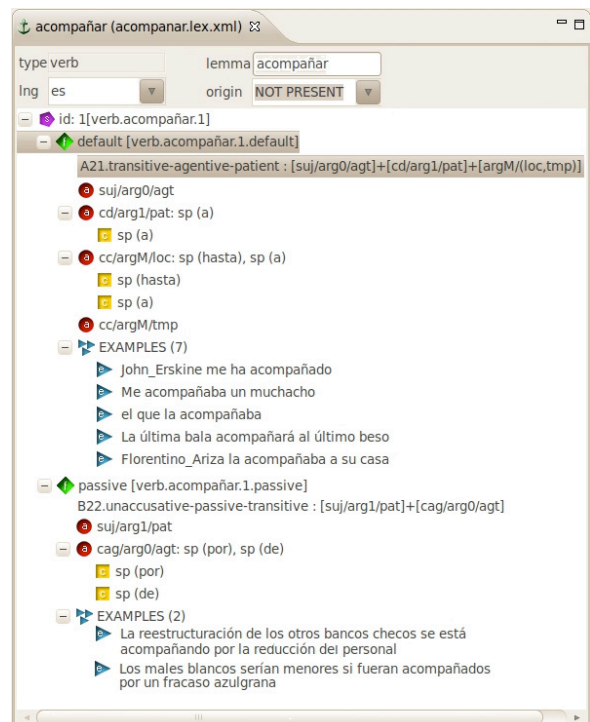


Figura 1: Entrada léxica de “acompañar” en AnCora-Verb-ES

⁹ <http://clic.ub.edu/ancora>.

RETURN HOME | BACK | SEARCH VerbNet v3.1 [VIEW OR MANAGE ALL COMMENTS](#) | UNIVERSITY OF COLORADO

accompany-51.7
Members: 7, Frames: 2

[POST COMMENT](#)

CLASS HIERARCHY
 ACCOMPANY-51.7
 NO SUBCLASSES

MEMBERS KEY

ACCOMPANY (FN 1; WN 2, 4; G 1)	SHEPHERD (FN 1; WN 1)
CONDUCT (FN 1; WN 4; G 4)	STEER (WN 1, 3; G 1)
ESCORT (FN 1; WN 1, 2; G 1)	
GUIDE (FN 1; WN 1, 2, 5; G 1)	
LEAD (FN 1; WN 1; G 5)	

ROLES REF

- AGENT [+ANIMATE]
- THEME [+ANIMATE]
- DESTINATION

FRAMES REF KEY

NP V NP

EXAMPLE "Jackie accompanied Rose."

SYNTAX AGENT V THEME

SEMANTICS MOTION(DURING(E), AGENT) MOTION(DURING(E), THEME)

NP V NP PP.DESTINATION

EXAMPLE "Jackie accompanied Rose to the store."

SYNTAX AGENT V THEME {+SPATIAL} DESTINATION

SEMANTICS MOTION(DURING(E), AGENT) MOTION(DURING(E), THEME) PREP(E, THEME, DESTINATION) PREP(E, AGENT, DESTINATION)

This page generated on 2010.4.29 at 8:15 PM. [REFERENCE](#) | [CLASS HIERARCHY](#) [CONTACT](#) | [VERBNET DOWNLOAD & LICENSE](#)

Figura 2: Entrada léxica de 'accompany' en VerbNet

2.2 Unified Verb Index (UVI)

El UVI es un índice que integra diferentes fuentes de conocimiento del inglés –VerbNet, PropBank, FrameNet, WordNet y OntoNotes. Se trata de recursos de amplia cobertura que son sin duda un referente en representación semántica. El UVI tiene como elemento vertebrador el léxico VerbNet (Kipper-Schuler, 2006; Kipper et al., 2002), donde para cada sentido verbal se especifica la clase semántica (Levin, 1993), la estructura léxico-semántica, los esquemas sintácticos, los roles semánticos y las restricciones selectivas (Véase figura 2). Además, los sentidos de VerbNet están asociados a:

- los verbos de PropBank (Palmer et al., 2005), un banco de árboles sintácticos del inglés anotado semánticamente con argumentos y roles semánticos;
- los sentidos verbales de WordNet (Miller, 1990; Fellbaum, 1998);
- los sentidos verbales de la ontología de OntoNotes (Hovy et al., 2006), donde la agrupación de sentidos es más fina que en WordNet y,
- los esquemas conceptuales de FrameNet (Baker et al. 1998, Baker et al. 2003).

En la figura 2 se ejemplifica el contenido básico de VerbNet con la entrada de *accompany*

(‘acompañar’), que incluye la clase semántica de Levin ("accompany-51.7"), los verbos que pertenecen a esta clase ("accompany", "conduct", "escort", etc.), los vínculos con FrameNet, WordNet y Ontonotes ("fn 1; wn 2, 4; gr 1", respectivamente), los roles semánticos ("agent", "theme", "destination") con sus correspondientes restricciones selectivas ("animate") y la estructura sintáctico-semántica (frames).

En UVI hay representados actualmente un total de 5.879 predicados verbales, con uno o más sentidos de WordNet asociados, con 5.726 conexiones al léxico computacional VerbNet, con 4.592 relaciones a PropBank, con 4.186 asociaciones a esquemas conceptuales de FrameNet y con 2.195 conexiones a la ontología de OntoNotes.

3 Metodología

La metodología aplicada para la elaboración de AnCora-Net ha permitido realizar este recurso de manera semiautomática, es decir combinando procesos automáticos y manuales. El vínculo de los léxicos AnCora-Verb al UVI se ha realizado a través de PropBank y VerbNet por tratarse de los recursos más parecidos a AnCora-Verb. De hecho, los léxicos del catalán y español siguen el mismo esquema de anotación que PropBank en la anotación de la

estructura argumental, y están organizados en clases semánticas al igual que VerbNet. Esto ha permitido establecer la conexión entre recursos tanto a nivel de sentido, como a nivel de argumento. Además, cada uno de los sentidos verbales representados en PropBank incluye también información sobre la clase semántica de VerbNet correspondiente.

El proceso de conexión de los recursos lingüísticos AnCora-Verb del español y del catalán a PropBank y VerbNet se ha llevado a cabo en cuatro etapas:

1) Primero se ha realizado la vinculación automática de los léxicos AnCora-Verb y UVI a través de WordNet 3.0 y PropBank-1.75. Se ha utilizado WordNet 3.0 como un léxico bilingüe para establecer los posibles vínculos entre las distintas lenguas (catalán-inglés, español-inglés) y PropBank para vincular los argumentos. El alineamiento automático de los argumentos ha sido posible gracias a que ambos recursos, AnCora-Verb y PropBank, utilizan el mismo esquema de anotación. En el proceso de vinculación automática de los léxicos se realizaron 25.007 vínculos para el castellano (una media de 5,41 vínculos por esquema sintáctico-semántico verbal) y 21.519 para el catalán (media de 5,84 vínculos por esquema sintáctico-semántico verbal). En este primer momento, no fue posible establecer vínculo alguno para 890 esquemas sintáctico-semánticos en castellano (16,1% del total) y 961 esquemas sintáctico-semánticos en catalán (20,7%). De acuerdo con los resultados finales (sección 3.1), teniendo en cuenta tanto el número de vínculos válidos como la media de vínculos por sentido, se estima que este preproceso ha permitido reducir en al menos un 33% el tiempo total invertido para la anotación.

2) En segundo lugar, se ha realizado la validación manual del resultado de la vinculación automática, y la conexión de aquellos sentidos que no han obtenido ninguna correspondencia automática. La vinculación entre los léxicos AnCora-Verb y PropBank se establece a dos niveles: la asignación a nivel de sentido verbal y a nivel de estructura argumental —posición argumental (arg0, arg1,... argn) y papel temático (agente, tema, manera, etc.). Las correspondencias que se establecen son de n a n sentidos, considerando como equivalentes tantos verbos del inglés como se crea necesario y viceversa. En el análisis de cada equivalencia, se prioriza la

estructura argumental de los verbos. Es decir, si un sentido verbal de AnCora-Verb comparte la misma estructura argumental que un sentido verbal de PropBank es un candidato con más posibilidades para establecer la correspondencia. No obstante, la diferencia en la estructura argumental de los verbos no será nunca un criterio excluyente para considerar como válida una equivalencia verbal. En general, para las diátesis de un mismo sentido en AnCora-Verb se utilizan las mismas equivalencias. Por ejemplo, los esquemas sintáctico-semánticos correspondientes al sentido acompañar.1, es decir, la alternancia transitiva ("verb.acompañar.1.default") y pasiva ("verb.acompañar.1.passive"), se vinculan a los mismos sentidos verbales de PropBank ("accompany.01" y "escort.01"). Mientras que las alternancias de diátesis no afectan a las equivalencias verbales, las diferencias de sentido sí lo hacen. Así, por lo general, cada sentido de un verbo de AnCora-Verb se traduce por verbos distintos al inglés. No obstante, puede ocurrir que un mismo sentido verbal de PropBank albergue sentidos distintos de AnCora-Verb. Por ejemplo: "adoptar.1" (*adoptar una resolución, adoptar sanciones...*) y "adoptar.2" (*adoptar un niño*) se corresponden ambos al mismo sentido verbal de PropBank ("adopt.01"). En este caso los dos sentidos de AnCora-Verb se asocian al mismo sentido de PropBank. Puede darse también el caso contrario, es decir, que AnCora-Verb tenga un único sentido para dos o más sentidos en PropBank. En estos casos, si existe en inglés un verbo que tenga una misma traducción al castellano para todos sus sentidos, entonces se vincula a éste. En caso de que la diferencia que existe entre los sentidos del inglés y el del español sea muy acusada y su traducción sea completamente distinta, se le asocian todas las equivalencias posibles. Por ejemplo: "adquirir.1" (*adquirir un coche, adquirir renombre...*) se vincula a los sentidos "acquire.01", "buy.01" y "assume.01" de PropBank. En Fernández et al. (2010) se presentan de manera detallada los criterios seguidos en la vinculación de AnCora-Verb y PropBank.

Para la validación de los vínculos obtenidos automáticamente, o la creación de nuevos vínculos en el caso de no haberse obtenido automáticamente, se ha utilizado la herramienta de creación y anotación AnCora-Pipe (Bertran et al., 2010). Esta herramienta incluye un editor

The screenshot shows the AnCora-Pipe interface. At the top, there is a table with columns: ancoralexid, propbankid, Revised, and Valid. The table lists three entries for the verb 'acompañar':

ancoralexid	propbankid	Revised	Valid
verb.acomodar.1.passive	accommodate.01	reviewed	yes
verb.acompañar.1.default	accompany.01	reviewed	yes
verb.acompañar.1.default	escort.01	reviewed	yes

Below the table, there are checkboxes for 'Valid link' and 'Revised', and a 'Comment' field with the text 'link proposal via wordnet senses'. A 'Reset' button is also present.

The main part of the interface is a table titled 'Arguments AncoraLex <-> Propbank'. It has four columns: Details, AncoraLex Arg, Propbank Arg, and Details. The rows show the following mappings:

Details	AncoraLex Arg	Propbank Arg	Details
[subj/arg0/agt]:	arg0	arg0	[subject accompanier: Agent]: "a chart" in The figure was incorrectly shown as a net loss in a chart accompanying Friday's Heard on the Street column.
[cd/arg1/pat]: sp(a)	arg1	arg1	[object accompanied: Theme]: "Friday 's Heard on the Street column" in The figure was incorrectly shown as a net loss in a chart accompanying Friday's Heard on the Street column.
[cc/argM/loc]: sp(hasta) + sp(a)	argM	arg4	[end point:]: "to the fire" in John led the unhappy group of travellers from the frying pan to the fire.
[cc/argM/tmp]:	argM	<none>	<none>
<none>	<none>	arg3	[start point:]: "from the frying pan" in John led the unhappy group of travellers from the frying pan to the fire.

At the bottom, there are buttons for 'Open frame', 'Refresh', 'Open roleset', and 'Find another roleset'. Below these buttons, there is a section for 'default (default) A21.transitive-agentive-patient' and 'EXAMPLES:' with text examples in Spanish. On the right side, there is a section for 'name = go along with' and 'verbnet classes = 51.7', and another 'EXAMPLES:' section with an example in English: '(active) The figure was incorrectly shown as a net loss in a chart'.

Figura 3: Interfaz de anotación AnCora-Pipe

que permite vincular los léxicos AnCora-Verb a PropBank, y acceder y consultar fácilmente a los distintos recursos implicados: AnCora-Verb, PropBank, VerbNet, WordNet, FrameNet y Ontonotes (Véase figura 3).

En la figura 3 se puede observar la interfaz de anotación de AnCora-Pipe. Las correspondencias se establecen a dos niveles: por un lado se vinculan los sentidos (en este caso "acompañar.1" y "accompany.01") y por otro lado se vincula la estructura argumental (en la figura 3, "arg0", "arg1" y "argM" de "acompañar.1" se vinculan a los argumentos "arg0", "arg1" y "arg4" de "accompany.01" respectivamente), teniendo en cuenta, sobre todo, los papeles temáticos ("agt", "pat", "loc" y "tmp" en el caso de la fuente castellana y "agent", "theme", "endpoint" y "startpoint" en la de la inglés). Como se puede observar en la figura 3, un complemento no argumental ("argM") en AnCora-Verb puede vincularse a un complemento argumental ("arg4"). Del mismo modo, puede darse el caso que un

argumento quede sin correspondencia (por ejemplo, el argumento "argM-tmp" del castellano no encuentra equivalencia en el inglés). En la parte inferior de la interfaz se proporcionan ejemplos de AnCora-Verb (a la izquierda) y PropBank (a la derecha) con vínculos a la entrada del lexicón castellano y a la entrada correspondiente al *Roleset* inglés, desde donde se puede acceder al resto de recursos para esta lengua.

3) En tercer lugar, se han enriquecido automáticamente los léxicos AnCora-Verb con la información contenida en las fuentes de información léxica del inglés integradas en UVI gracias a la vinculación entre nuestras entradas léxicas y PropBank. Es decir, toda la información sintáctico-semántica incluida en VerbNet-3.1, el esquema conceptual (*frame*) de FrameNet, el vínculo con los sentidos equivalentes de WordNet 3.0 del inglés y las agrupaciones de sentidos verbales de OntoNotes (Véase figura 4).

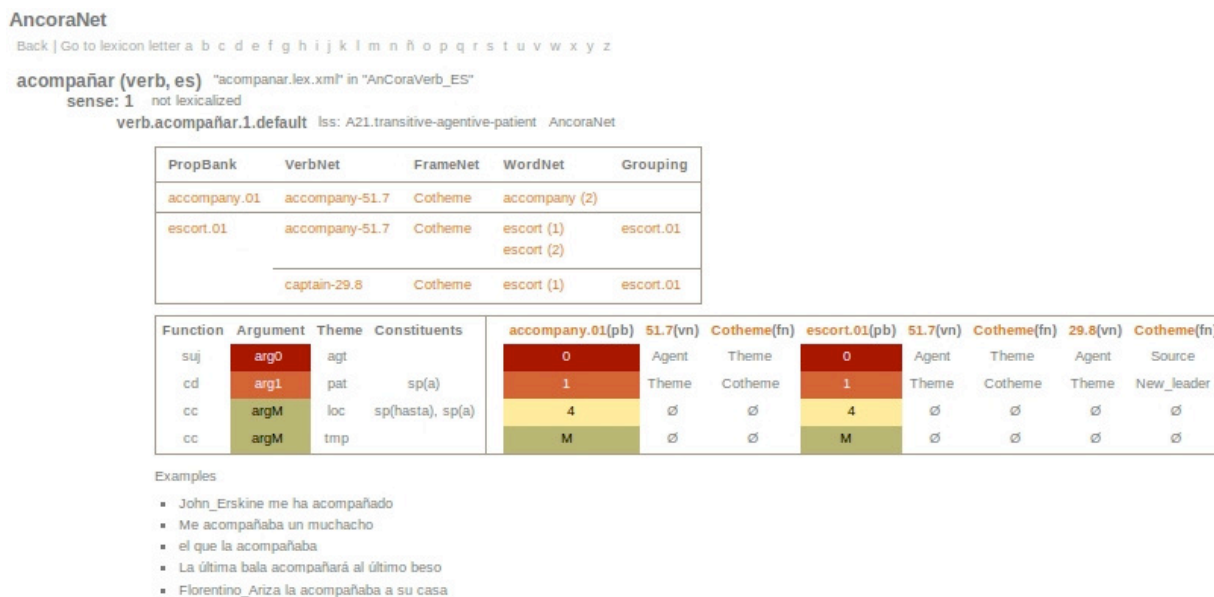


Figura 4: Entrada léxica de ‘acompañar’ en AnCora-Net

En la figura 4, se muestra la entrada léxica parcial de *acompañar.1* en AnCora-Net. En la entrada se establecen las distintas correspondencias entre el sentido verbal del léxico español AnCora-Verb con las distintas fuentes del inglés: el sentido verbal de PropBank ("accompany.01" y "escort.01"), la clase semántica de VerbNet ("accompany-51.7"), así como el marco conceptual de FrameNet ("Cotheme") y el sentido correspondiente en WordNet ("accompany (2)" y "escort (1) y (2)") y en OntoNotes (Grouping, "escort.01") inferidos automáticamente. En la entrada también quedan establecidas las correspondencias a nivel de argumento y papel temático.

4) Finalmente, se ha obtenido de manera automática la conexión entre el léxico AnCora-Verb-CA del catalán y AnCora-Verb-Es del español, a través de las correspondencias establecidas con los sentidos de PropBank. Es decir, se compara si un sentido verbal de AnCora-CA y AnCora-ES comparten el mismo sentido en PropBank. Una vez realizada esta conexión se ha realizado la validación manual para garantizar que los vínculos establecidos son los correctos. Se han obtenido automáticamente un total de 11.009 vínculos entre las dos lenguas de AnCora-Verb, que cubren 2.508 sentidos para el catalán y 2.884 para el castellano, con una media de 3,81 vínculos por sentido en castellano y 4,38 por sentido en catalán. Tras la validación manual,

se obtiene que la conexión automática resulta válida en un 90% de los casos.

3.1 Resultados

En la tabla 1 se presenta el número de vínculos obtenidos entre los léxicos AnCora-Verb del catalán y del español con las distintas fuentes del inglés.

AnCora-Net	AnCora-Net-ES	AnCora-Net-CA
verbos	2.820	2.248
verbos conectados	2.512	2.029
Vínculos a VerNet	7.284	5.942
Vínculos a PropBank	7.968	6.494
Vínculos a FrameNet	6.685	5.495
Vínculos a WordNet	12.241	10.247
Vínculos a OntoNotes	6.752	5.852

Tabla 1: Resumen de los vínculos establecidos entre las distintas fuentes léxicas de Ancora-Net

La variabilidad en cuanto al número de conexiones con los diferentes recursos del inglés (VerbNet, WordNet, OntoNotes, PropBank y FrameNet) viene dictada por las conexiones ya existentes en esta lengua. Como sucede en nuestro caso, los vínculos no son de 1 a 1 y, además, no son exhaustivos. Esto

provoca, entre otras cosas, que haya sentidos verbales que no queden conectados con todos los recursos ingleses, como hay sentidos verbales ingleses que no tienen todas las conexiones a los recursos disponibles.

Tras la revisión manual y la conexión de los sentidos que no habían recibido una propuesta de conexión automática, han quedado un total de 7.976 vínculos válidos para el castellano, con una media de 1,64 vínculos por sentido verbal, y 6.498 vínculos válidos para el catalán, con una media de 1,56 vínculos por sentido. Los sentidos con más conexiones coinciden en castellano y catalán: *acompañar.2/acompanyar.2* (transitivo-causativo) es en los dos casos el sentido más conectado, con 10 y 11 enlaces respectivamente. Pese a todo, la desviación estándar en las dos lenguas es de 0.99, lo que se traduce en que más del 80% de los sentidos tienen únicamente 1 o 2 vínculos válidos.

Estos resultados son deseables: un índice de vínculos por sentido alto reflejaría una dispersión grande en las clases verbales definidas en AnCora-Verb respecto a las definidas en VerbNet. Un índice bajo apunta a que la hipótesis de partida de que las clases de AnCora-Verb, más generales, podrían definir “superclases” en VerbNet, cuya división es mucho más fina, está bien encaminada.

Los léxicos AnCora-Net están disponibles en la página web del proyecto: <http://clic.ub.edu/corpus/ancoranet>.

4 Conclusiones

En este artículo se presenta AnCora-Net, un léxico verbal multilingüe creado a partir de la integración de los léxicos AnCora-Verb del catalán y español, con los diferentes recursos de base semántica que conforman el UVI del inglés.

La integración de dichos recursos permite enriquecer y complementar los contenidos de los diferentes recursos implicados que, a pesar de seguir criterios distintos de representación, resultan compatibles entre ellos. AnCora-Verb se enriquece con la información semántica proporcionada por las fuentes inglesas: clase semántica de VerbNet, argumentos semánticos de PropBank, esquemas conceptuales de FrameNet, sentidos de WordNet y OntoNotes. A su vez, las distintas fuentes del inglés se

benefician con la incorporación de los léxicos AnCora-Verb, del catalán y español, dando lugar a un recurso multilingüe. AnCora-Net agrupa los sentidos de VerbNet y PropBank en las clases más genéricas de AnCora-Verb reduciendo así la excesiva granularidad de estas fuentes léxicas.

Como los corpus AnCora están anotados con los léxicos AnCora-Verb, la información que se obtiene del inglés puede añadirse automáticamente a los corpus anotados, enriqueciéndolos a su vez con nueva información semántica.

Es objetivo futuro la obtención, a partir de los corpus AnCora y PropBank, de un corpus comparable multilingüe formado por los ejemplos de cada uno de los predicados verbales interconectados en las diferentes lenguas, de modo que se podrá disponer de ejemplos con información equivalente sobre los predicados verbales en las tres lenguas. Por lo tanto, AnCora-Net se puede utilizar para el alineamiento semántico de corpus monolingües y multilingües previamente anotados con cualquiera de las fuentes léxicas integradas.

Cabe señalar también que se está trabajando en la incorporación a AnCora-Net del léxico EPEC-ADI del euskera (Aldezabal, 2010), de modo que se ampliaría el carácter multilingüe del recurso.

Finalmente, dado que AnCora-Verb está conectado con el léxico de nominalizaciones deverbales AnCora-Nom (Peris y Taulé, 2011), podría vincularse de manera automática estos nombres deverbales con los verbos correspondientes incluidos en AnCora-Net ampliando, por lo tanto, este recurso con predicados nominales.

Un recurso de estas características puede constituir un módulo para cualquier aplicación de procesamiento del lenguaje natural que implique el tratamiento del multilingüismo y la interpretación semántica.

Agradecimientos

Este trabajo ha sido posible gracias a los proyectos TEX-MESS 2.0 (TIN2009-13391-C04-04) y Araknion (FFI2010-114774-E) del Ministerio de Ciencia e Innovación.

Referencias bibliográficas

Aldezabal I. 2010. ‘Basis for the annotation of EPEC-RolSem’, Proceedings of *Interdisciplinary Workshop on Verbs. The*

- identification and Representation of Verb Features*. Scuola Normale Superiore - Laboratori di Linguistica. Università di Pisa. Pisa (Italia).
- Aparicio, J., Taulé, M. y Martí M.A. 2008. AnCora-Verb: A Lexical Resource for the Semantic Annotation of Corpora. En *Proceedings of 6th International Conference on Language Resources and Evaluation*. Marrakesh (Morocco).
- Baker, C. F., Fillmore, Ch. J., Lowe, J. B. 1998. 'The Berkeley FrameNet project', Proceedings of the COLING-ACL, Montreal (Canada).
- Baker, C. F., Ruppenhofer, J. 2002. 'FrameNet's Frames vs. Levin's Verb Classes'. In J. Larson and M. Paster (Eds.) *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*. 27-38.
- Baker, C., Collin F., Charles J. Fillmore, B. Cronin 2003. 'The Structure of the Framenet Database', *International Journal of Lexicography*, Volume 16.3: 281-296.
- Bertran, M., O. Borrega, M.A. Martí, M. Taulé, 2011. *AnCoraPipe: A new tool for corpora annotation*, Working paper 1: TEXT-MESS 2.0 (Text-Knowledge 2.0), Universitat de Barcelona. <http://clic.ub.edu/ca/publicacions>
- Fellbaum, C., editor 1998. 'WordNet: An Electronic Lexical Database'. *Language, Speech and Communications*. MIT Press, Cambridge, Massachusetts.
- Fernández, P., Borrega, O., Taulé, M., Martí, M.A. 2010. *Guía de anotación para vincular Ancora-Verb y PropBank*. Working paper 1: AnCora-Net Project. CLiC-Universitat de Barcelona. Barcelona.
- Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., Weischedel, R. 2006. 'OntoNotes: The 90% Solution', Proceedings of HLT/NAACL, New York.
- Kipper, K. M. Palmer, O. Rambow 2002. 'Extending PropBank with VerbNet Semantic Predicates'. *Workshop on Applied Interlinguas*, held in conjunction with AMTA-2002. Tiburon, CA.
- Levin, B. 1993. *English Verb Classes and Alternations: A preliminary investigation*. University of Chicago Press, Chicago.
- Miller, G. A. 1990. 'WordNet: An on-line lexical database'. *International Journal of Lexicography*.
- Palmer, M., P. Kingsbury, D. Gildea 2005. 'The Proposition Bank: An Annotated Corpus of Semantic Roles'. *Computational Linguistics*, 21 (1), MIT Press, USA.
- Peris A., Taulé, M. 2011. AnCora-Nom: A Spanish Lexicon of Deverbal Nominalizations, *Procesamiento del Lenguaje Natural*, nº46, Alicante (España).
- Taulé, M, Martí, M.A., Recasens, M. (2008). 'Ancora: Multilevel Annotated Corpora for Catalan and Spanish'. Proceedings of 6th International Conference on Language Resources and Evaluation. Marrakesh (Morocco).
- Taulé, M., M.A. Martí, O. Borrega (2010). Proceedings of *Interdisciplinary Workshop on Verbs. The identification and Representation of Verb Features*. Scuola Normale Superiore - Laboratori di Linguistica. Università di Pisa. Pisa, Italia.
- Vázquez, G., Fernández, A.; Martí M.A. 2000. *Clasificación verbal. Alternancias de diátesis*. Lleida: Edicions de la Universitat de Lleida.