



This is the final version of the following article:

Buendía, Miriam. 2010. Anotación semántica en el dominio especializado de la Meteorología. In *Modos y formas de la comunicación humana / Ways and modes of human communication*, ed. Rosario Caballero and María Jesús Pinar Sanz, 923-934. Cuenca: Ediciones de la Universidad de Castilla- La Mancha / AESLA.

You can find more articles authored by LexiCon Research Group members at <http://lexicon.ugr.es>.

## **ANOTACIÓN SEMÁNTICA EN EL DOMINIO ESPECIALIZADO DE LA METEOROLOGÍA**

MIRIAM BUENDÍA CASTRO  
*Universidad de Granada*

### RESUMEN

*En el presente estudio hemos llevado a cabo la anotación semántica de un subcorpus del dominio de la Meteorología, extraído del corpus creado para el proyecto de investigación Ecosistema. Hemos utilizado un conjunto de textos pequeños de manera que sirva de entrenamiento (bootstrapping) para la posterior generación de herramientas semi-automáticas de anotación basadas en ontologías. Para llevar a cabo la anotación, nos hemos basado en una ontología incipiente que estamos construyendo, de la cual hemos extraído las clases de entidades que vamos a codificar en el texto, a las que hemos asignado un identificador distintivo. Con este análisis lingüístico preliminar mostramos cómo las distintas entidades se relacionan en una jerarquía, a la vez que damos unas pinceladas acerca de lo que es posible conseguir con la anotación semántica: facilitar la búsqueda, recuperación y extracción de información, lo que podría ser de gran ayuda para la traducción automática, el aprendizaje de ontologías y la minería de textos.*

Palabras clave: anotación semántica, lingüística de corpus, ontología

#### ABSTRACT

*In the current study, we have created a system of semantic tagging for a subcorpus of texts within the domain of Meteorology, which were extracted from the larger corpus created for the research project Ecosistema. This small group of texts has been used for bootstrapping for the subsequent implementation of semi-automatic tagging tools based on ontologies. We have created this tagging system on the basis of an incipient ontology we are constructing, which has been the primary source for the semantic tags in the text. This preliminary linguistic analysis shows how these entities are related to each other in a hierarchy. Semantic annotation facilitates the search, recovery and extraction of information which could be of a great value for automatic translation, the learning of ontologies and web mining.*

Keywords: *semantic annotation, corpus linguistics, ontology*

## 1. INTRODUCCIÓN

La web semántica supone una extensión de la WWW, en la que el significado de los contenidos de los documentos de la red se explicita formalmente para conseguir que los contenidos comprensibles por las personas sean comprensibles por el ordenador. En este sentido, la anotación semántica de los textos, que hace explícito el significado para un ordenador, se ha convertido en uno de los pilares fundamentales en el desarrollo de la web semántica.

La Inteligencia Artificial ha encontrado en las ontologías, el modelo de conocimiento ideal para describir formalmente los recursos web y su vocabulario y, por ende, para hacer explícito, de algún modo, el significado subyacente de los términos incluidos en las páginas web. La semántica ontológica (Niremburg y Raskin 2001) es una teoría que estudia el significado y procesamiento del lenguaje natural, que utiliza un modelo abstracto del mundo —la ontología— como recurso central para extraer y representar el significado de los textos a través del conocimiento que se deriva de ellos (Aguado de Cea *et al.* 2002).

El presente estudio se enmarca dentro del proyecto de investigación *Ecosistema*<sup>1</sup>, cuyo principal objetivo es la representación de la estructura conceptual del dominio de la Gestión Integrada de Zonas Costeras, en forma de un tesoro visual de conceptos costeros organizados en marcos dinámicos de conocimiento especializado (<http://ecolexicon.ugr.es/>). Para ello, hemos creado un corpus que ha sido, por un lado, la fuente de información para establecer la estructura conceptual del dominio a través de la elaboración de jerarquías terminográficas y, por otro, la base para la elaboración de las definiciones que incluyen la información contextual pertinente a cada término.

En esta investigación hemos llevado a cabo la anotación semántica de un subcorpus del dominio de la Meteorología, extraído de nuestro corpus. Hemos utilizado un cuerpo de textos pequeños de manera que sirva de entrenamiento (bootstrapping) para la posterior generación de herramientas semi-automáticas de anotación basadas en ontologías.

Ante la falta de estandarización y herramientas que realicen de forma automática la anotación en el nivel semántico<sup>2</sup>, primeramente hemos creado el mapa conceptual del dominio (cf. 4.2), necesario para la posterior implementación de la ontología incipiente que estamos construyendo, de la que hemos extraído las clases de entidades que vamos a codificar en el texto a las que hemos asignado un identificador distintivo. Posteriormente, mediante la utilización del software SMORE<sup>3</sup>, hemos enriquecido semánticamente el corpus.

Con este análisis lingüístico preliminar mostramos cómo las distintas entidades se relacionan en una jerarquía, a la vez que damos unas pinceladas acerca de lo que es posible conseguir con la anotación semántica: facilitar la búsqueda, recuperación y extracción de información, lo que podría ser de gran ayuda para la traducción automática, el aprendizaje de ontologías y la minería de textos.

## 2. LA NECESIDAD DE ANOTAR EL CORPUS

Los corpus planos ofrecen con frecuencia ciertas limitaciones en comparación con los corpus codificados. Si, por ejemplo, estamos estudiando un determinado fenómeno lingüístico en un corpus plano, tendremos que buscar a través de palabras o expresiones específicas (exact matching), no podemos realizar la búsqueda por categorías de elementos lingüísticos como son los elementos gramaticales o las funciones sintácticas, como ocurre en los corpus anotados.

Tal y como señala Leech (1997: 4): “The fact is that to extract information from a corpus, we often have to begin by building information in”. Y esto precisamente es lo que se pretende con la codificación (encoding/markup) de los textos: enriquecer los textos de un corpus con información acerca del corpus haciendo que las características implícitas de los elementos textuales se expliciten.

### *2.1 La anotación: definición, tipos y aplicaciones*

La anotación puede definirse como: “[...] the practice of adding interpretative linguistic information to a corpus” (Leech 2004).

Leech (1997) considera en una primera fase los siguientes tipos de anotación: ortográfica, fonética, prosódica, gramatical, sintáctica, semántica, del discurso y pragmática o estilística. En una publicación posterior (2004) aúna los distintos tipos en: gramatical, fonética, semántica, pragmática, del discurso, estilística y léxica.

Basándonos en Leech (2004 [1997]), consideramos seis tipos de anotación teniendo en cuenta los elementos lingüísticos que se codifican:

- La etiquetación morfológica o gramatical (*part-of-speech*, POS): identifica las partes de una oración, como son el nombre, verbo, artículo.
- La anotación léxica, anotación de lemas o lematización: consiste en añadir la identidad del lema de cada palabra de un texto. En inglés, la lematización puede considerarse redundante pero, en lenguas más flexivas como el español o el alemán, puede resultar de una gran utilidad para la extracción de información.
- Anotación sintáctica: es la práctica de añadir información sintáctica a un corpus a través de la incorporación en el texto de indicadores de la estructura sintáctica.
- La anotación semántica: consiste en añadir información acerca de la categoría semántica de las palabras.
- La anotación del discurso: consiste en añadir información sobre elementos anafóricos de un texto.
- La anotación pragmática: consiste en añadir información acerca de los tipos de actos de habla que ocurren un diálogo.

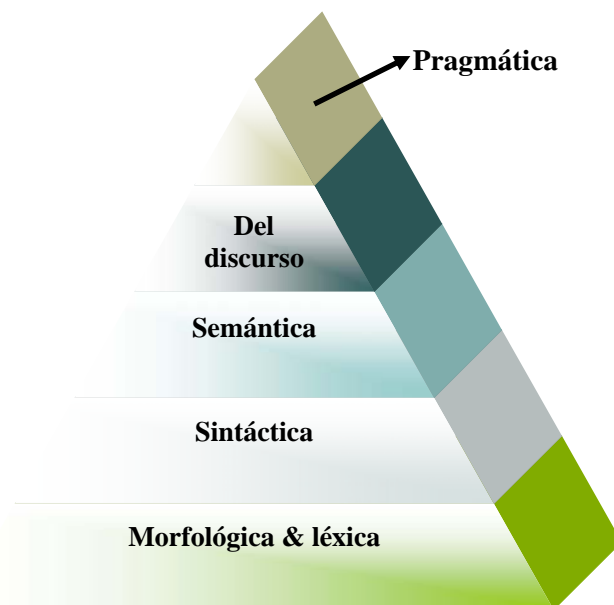


Figura 1. Niveles de anotación lingüística

La anotación tiene objetivos y aplicaciones muy variadas: facilita la extracción y recuperación de la información, permite la reutilización de los corpus, a la vez que es multifuncional (Leech 2004 [1997]). Por ejemplo, en la etiquetación morfológica, la información sobre la clase gramatical de las palabras nos puede servir para la lexicografía, el análisis sintáctico, para elaborar listas de frecuencia y otras muchas aplicaciones.

#### 2.1.1 Normas para la anotación

Leech (2004 [1997]) enumera una serie de normas o estándares de buena práctica necesarias a tener en cuenta para la anotación de un corpus:

1. Las anotaciones tienen que estar separadas del texto. Siempre tiene que ser fácil poder separar las anotaciones del *raw corpus*, el corpus puro propiamente dicho sin tratar, de forma que se pueda recuperar el corpus de la misma forma en la que estaba antes de que se añadieran las anotaciones.

2. Se tiene que proporcionar documentación detallada y explícita. Burnard (2004) destaca la necesidad de proporcionar documentación adecuada sobre el corpus y los textos que lo forman. De igual modo, es importante proporcionar información detallada y explícita sobre las anotaciones en un corpus anotado, como la respuesta a preguntas de cómo, dónde, cuándo o por quién fueron hechas las anotaciones; qué esquema de anotaciones (annotations scheme) y qué sistema de codificación (coding scheme) se han aplicado. Por esquema de anotaciones se entiende un sistema explicativo que proporcione información acerca de las prácticas de anotación que se han seguido. Por sistema de codificación se entiende las convenciones simbólicas que se han usado para representar las propias anotaciones.
3. Las prácticas de anotación tienen que estar lingüísticamente consensuadas.
4. Las prácticas de anotación tienen que respetar los estándares/normas *de facto*. Por *de facto* se entiende cualquier tipo de estandarización que esté empezando a darse. Las normas *de facto* resumen lo que la gente ha hallado en el pasado, lo que significa que los investigadores en ciernes en este ámbito deberían adoptarlo para apoyar un consenso creciente en la comunidad.

### 3. LA WEB SEMÁNTICA

En palabras de Berners-Lee, Hendler y Lassila (2001), la web semántica es una extensión de la WWW en la que la información recibe un significado bien definido para que pueda ser procesable por los ordenadores: “The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation”.

El objetivo primordial de la web semántica es conseguir que los ordenadores entiendan la semántica de los textos. A este respecto, la anotación semántica de los textos que hace explícito el significado

para un ordenador se ha convertido en un pilar fundamental para el desarrollo de la web semántica.

### 3.1 La anotación semántica

La anotación semántica consiste en asignar a las entidades de un texto su descripción semántica. Al añadir una etiqueta a cada palabra de un texto que indique el campo semántico<sup>4</sup> al que se refiere nos permite extraer todas las palabras relacionadas del campo semántico en concreto.

En la siguiente ilustración, extraída de Kiryakov *et al.* (2003: 484) podemos observar el proceso:

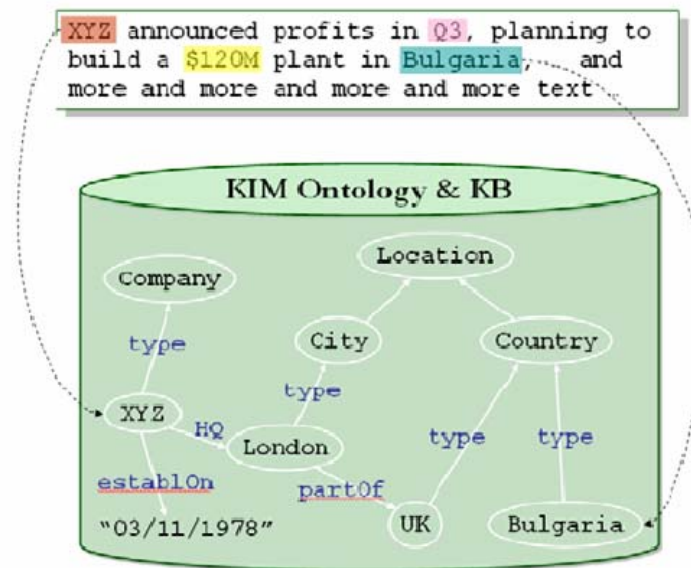


Figura 2. Anotación semántica

Es una cuestión terminológica de si a este tipo de anotación debería llamársela *anotación semántica*, *de entidades* o de cualquier otro modo. Parece no haber un término bien establecido para la anotación semántica ni tampoco un significado concreto de lo que este tipo de anotación conlleva. Lo que sí está claro son las innumerables aplicaciones que permite la anotación semántica como la indexación,



recuperación y categorización de la información o generación de metadatos más avanzados.

Aguado de Cea *et al.* (2002) corroboran la idea de que la anotación lingüística puede ayudar a los ordenadores a comprender la información textual de un documento de la web semántica. Apuestan por encontrar un modo de conjugar e identificar complementariedades entre los modelos de anotación semántica de la Inteligencia Artificial (IA) y las anotaciones propuestas por la Lingüística de Corpus (LC).

Los investigadores de la IA han encontrado en las ontologías el modelo de conocimiento ideal para describir formalmente los recursos web y su vocabulario y, por tanto, para hacer explícito de algún modo el significado subyacente de los términos incluidos en las páginas web (Aguado de Cea *et al.* 2002: 38).

### 3.2 Las ontologías

Una ontología, en el dominio de la Lingüística Computacional, hace referencia, según una definición ya estandarizada, a una especificación formal y explícita de una conceptualización compartida (Gruber 1993: 199): “[...] a formal, explicit specification of a shared conceptualization”.

*Formal* se refiere a que pueda ser procesable por un ordenador; *explícito* hace referencia a la necesidad de especificar de forma consciente los distintos conceptos que conforman una ontología; *conceptualización* describe un modelo abstracto del mundo en el que se identifican los conceptos relevantes y *compartida* se refiere a que una ontología debe dar cuenta de un conocimiento consensuado.

Otra definición de ontología más concreta la ofrece Weigand (1997): “An ontology is a database describing the concepts in the World or some domain, some of their properties and how the concepts relate to each other”.

De estas definiciones se desprende que una ontología siempre estará compuesta, como mínimo de conceptos, atributos, relaciones y reglas de restricción.

#### 3.2.1 El lenguaje de ontologías web, OWL

OWL (*Web Ontology Language* - Lenguaje de Ontologías Web) es un lenguaje de ontologías web que extiende RDFS<sup>5</sup> (*Resource Description Format Schema*) para permitir la expresión de relaciones complejas entre diferentes clases RDFS, y mayor precisión en las restricciones de clases y propiedades específicas. Utiliza la conexión proporcionada por RDF para dotar a las ontologías de la capacidad para ser distribuidas a través de varios sistemas, compatibles con los estándares web de accesibilidad e internacionalización, a la vez que hacen que las ontologías sean abiertas y extensibles.

OWL presenta tres variantes —OWL-Lite, OWL-DL, OWL-Full—, que incorporan diferentes funcionalidades y que, en general, van de una versión más simplificada —OWL-Lite—, utilizada en la representación de jerarquías simples, a OWL-Full.

En la web ya podemos encontrar un gran número de ontologías OWL como, por ejemplo, una ontología del cáncer desarrollada por el Centro de Bioinformática del Instituto Nacional del Cáncer de los EEUU<sup>6</sup> o la versión OWL de la ontología médica GALEN, desarrollada por la Universidad de Manchester<sup>7</sup>.

#### 4. EL PROCESO DE ANOTACIÓN SEMÁNTICA

Tal y como se ha comentado anteriormente, ante la falta de herramientas que lleven a cabo de forma totalmente automática la anotación en el nivel semántico, hemos elaborado primeramente el mapa conceptual del dominio a través del análisis de un mini corpus extraído de nuestro corpus. Posteriormente se ha procedido a la implementación de una ontología —todavía en construcción—; y por último, hemos anotado semánticamente el corpus mediante el programa SMORE.

##### 4.1 *El corpus*

El corpus de Ecosistema es un corpus escrito, formado por textos completos, especializado en Ingeniería de Puertos y Costas y Medio Ambiente que ha sido compilado de forma manual. Se trata asimismo de un corpus comparable puesto que alberga textos originales en español, inglés y alemán que no presentan variación horizontal (Cabré

2002), esto es, son temáticamente especializados pero dentro de una misma área de conocimiento; y los textos no son traducciones unos de otros.

Para la lengua española, el caso que nos ocupa, contiene alrededor de seis millones de palabras. Para la presente investigación, hemos extraído un mini corpus de 500.000 palabras enmarcado en el dominio de la Meteorología. Además, hemos utilizado a modo de evaluación un corpus web *ad-hoc* creado automáticamente mediante el programa BootCat (Baroni *et al.* 2006).

#### *4.2. El mapa conceptual*

A raíz del análisis del corpus, se ha procedido al diseño del mapa conceptual mediante el programa CmapTools<sup>8</sup> para la posterior implementación de la ontología. Recordemos que un mapa conceptual es una herramienta para la organización y representación del conocimiento, cuyo principal objetivo es representar las relaciones entre conceptos. Asimismo, contribuye al aprendizaje activo, integrando explícitamente conocimientos nuevos y antiguos.

Como podemos observar en la Figura 3., *viento* ocupa el eje central del mapa y de él salen una serie de relaciones, entre las que se destacan, *tipo de*, *acción*, *efecto*, *resultado* o *instrumento de medida*.

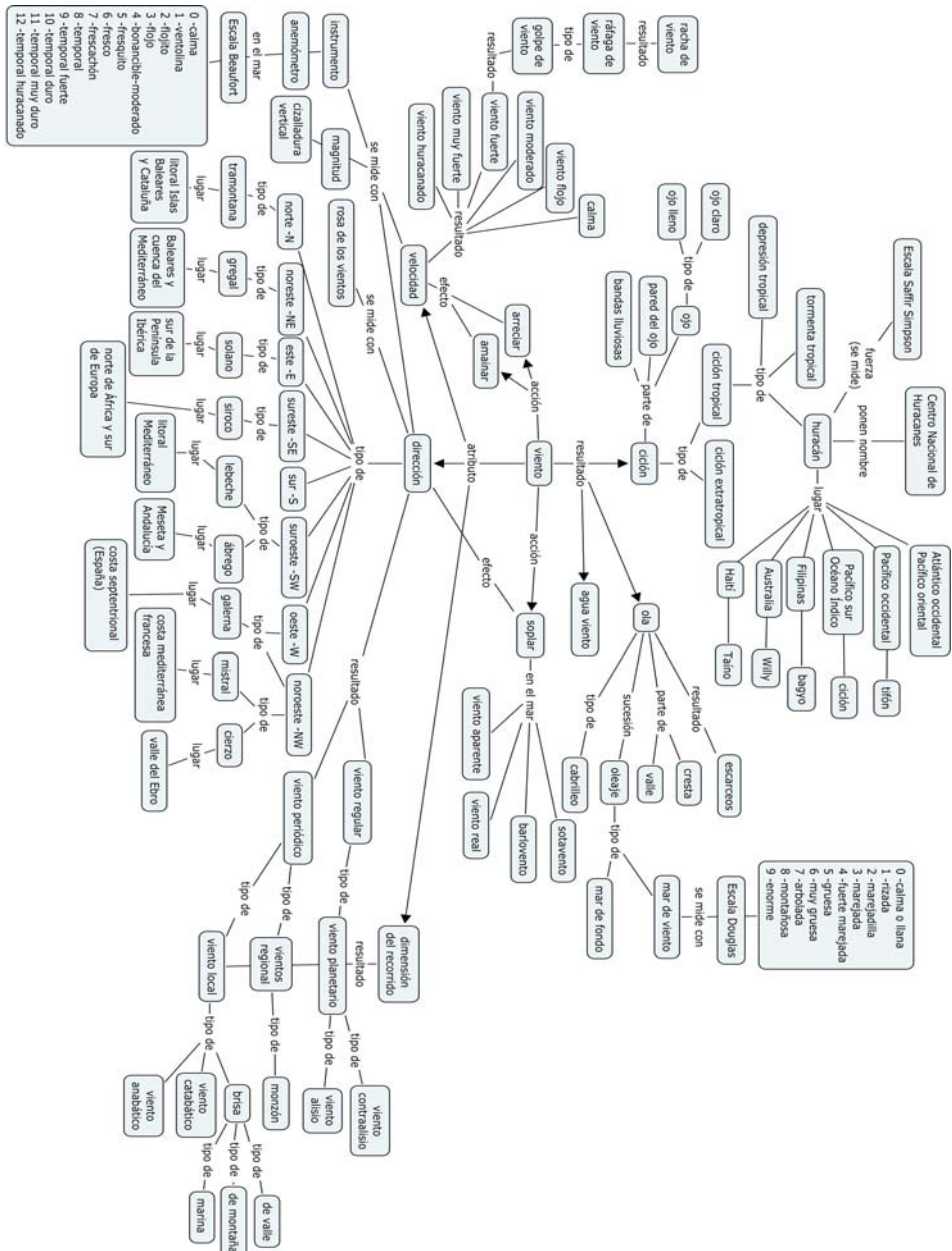


Figura 3. Mapa conceptual de viento

### 4.3. La ontología

La ontología ha sido implementada mediante el editor de ontologías Protégé<sup>9</sup>, después de extraer los conceptos del mapa conceptual y de estudiar las relaciones que se establecen entre ellos.

A modo de ejemplo, ofrecemos la siguiente ilustración, en la que se estudia el concepto *escala de Beaufort*. Como podemos observar, la *escala de Beaufort* es un tipo de *anemómetro*, un *anemómetro* es a su vez un *instrumento* que se encarga de medir la *dirección* del viento. En la parte de la derecha se despliegan todas las instancias que incorpora la *escala de Beaufort* que corresponden a los niveles de la escala y al nombre que recibe cada uno de ellos.

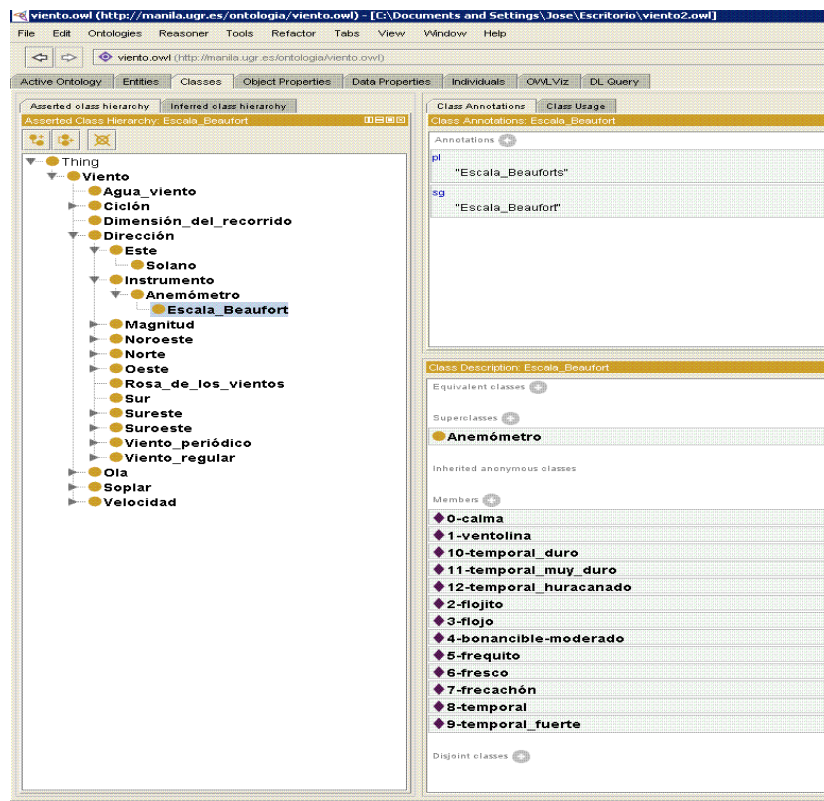


Figura 4. Despliegue de la clase *Escala de Beaufort* dentro de la ontología

#### 4.4. El programa de anotación semántica: SMORE

SMORE<sup>10</sup> está diseñado para permitir a los usuarios etiquetar documentos en HTML con OWL mediante el uso de ontologías web. A este respecto, proporciona la manera de usar clases, propiedades y ejemplos de las ontologías existentes, permite editar dichas ontologías e incluso crear una ontología nueva usando términos extraídos de documentos web.

Como acabamos de ver, SMORE permite crear una ontología desde cero o enriquecer una ontología ya existente. Nosotros nos hemos decantado por la segunda opción, para así disponer de un conocimiento ya adquirido y poder partir de una base.

SMORE nos ha servido fundamentalmente para dos aspectos: i) enriquecer nuestra ontología con conceptos extraídos de textos y documentos web, que no aparecían previamente en la ontología, pero que están relacionados con ella; ii) anotar semánticamente el corpus.

Para enriquecer la ontología es suficiente con seleccionar texto de cualquier sitio web descargado o documento local y hacer clic en el botón correspondiente de la barra de herramientas para crear clases, propiedades y ejemplos OWL. A modo de ejemplo, ofrecemos la Figura 5, en la que observamos la incorporación de los conceptos *rolar*, como subtipo de la acción de arreciar, *rizos* relacionado con *cresta* o *olas gruesas* como un subtipo de ola.

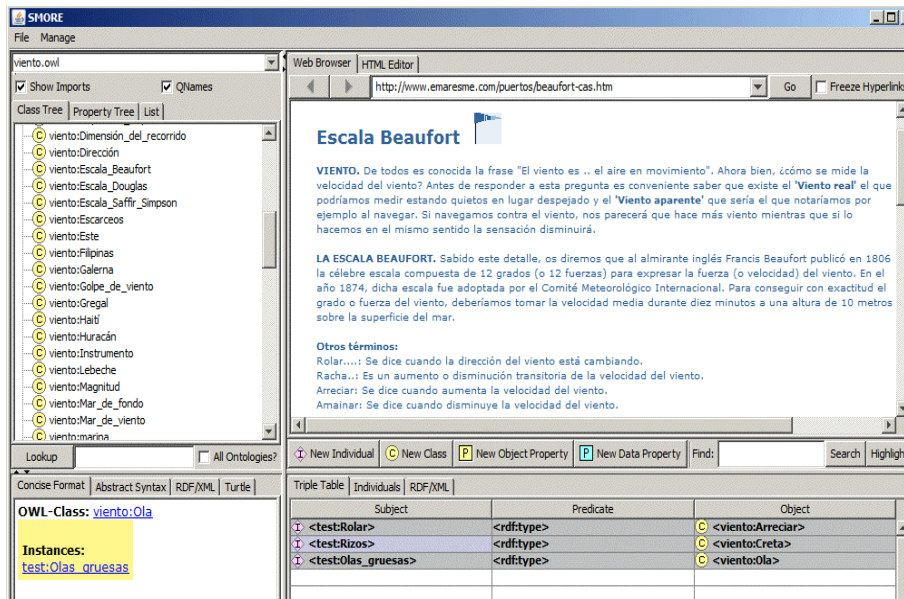


Figura 5. Enriquecimiento de la ontología mediante SMORE

Además de incorporar nuevos conceptos a la ontología, hemos etiquetado los conceptos del texto o documento web que ya aparecían en nuestra ontología, con el objetivo de estudiar posteriormente las relaciones que se establecen entre ellos.

Posteriormente se genera un fichero RDF/XML en el que se ofrecen las anotaciones semánticas. Como podemos observar en la Figura 6, usando los mismos conceptos que en la figura anterior, se establecen las relaciones entre los conceptos de la ontología y el corpus. Lo que se pretende es que cada uno de los términos del texto, en vez de constituir una simple cadena de caracteres, se ubiquen conceptualmente en una parte de la ontología y faciliten, de esta forma, la recuperación de información y la inferencia de conocimiento.

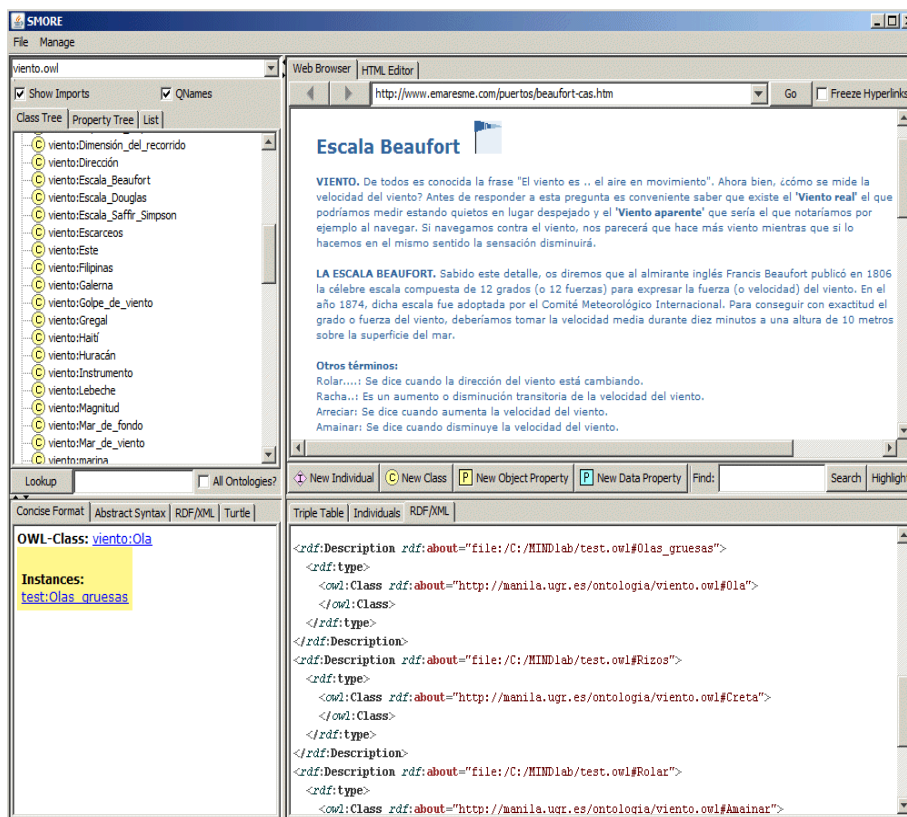


Figura 6. Anotación semántica mediante SMORE

## 5. CONCLUSIONES

En esta investigación hemos llevado a cabo la anotación semántica de un subcorpus del dominio de la Meteorología, extraído de nuestro corpus. Hemos utilizado un cuerpo de textos pequeños de manera que sirva de entrenamiento (bootstrapping) para la posterior generación de herramientas semi-automáticas de anotación basadas en ontologías.

Con este análisis lingüístico preliminar mostramos cómo las distintas entidades se relacionan en una jerarquía, a la vez que damos unas pinceladas acerca de lo que es posible conseguir con la anotación semántica: facilitar la búsqueda, recuperación y extracción de



información, lo que podría ser de gran ayuda para la traducción automática, el aprendizaje de ontologías y la minería de textos.

En una fase posterior, pretendemos aumentar y enriquecer nuestra ontología y el número de textos anotados semánticamente, para poder implementar un programa de anotación semántica semi-automático.

---

## NOTAS

<sup>1</sup>Esta investigación forma parte del proyecto de investigación Ecosistema: Espacio único de Sistemas de Información Ontológica y Tesauros sobre el Medio Ambiente (FFI2008-06080-C03-01/FILO) financiado por el Ministerio de Ciencia e Innovación.

<sup>2</sup>Aunque se han planteado algunas recomendaciones acerca de la codificación léxico-semántica, EAGLES (1999), no se ha publicado todavía ninguna estandarización para la anotación semántica de corpus, al contrario de lo que ocurre con la anotación morfológica (EAGLES, 1996a) o sintáctica (EAGLES, 1996b), que ya disponen de su normativa.

<sup>3</sup>Para más información acerca de los sistemas de anotación semántica que se aplican en los sistemas de ontologías, véase Senso Ruiz y Abel Leiva Mederos (2008).

<sup>4</sup>Recordemos que un campo semántico, campo conceptual, dominio semántico, campo léxico o dominio léxico es una construcción mental que agrupa las palabras que están relacionadas con el mismo concepto mental.

<sup>5</sup>Para más información acerca de RDF, véase <<http://www.w3.org/TR/rdf-schema/>>

<sup>6</sup>Para más información, véase <<http://www.mindswap.org/2003/CancerOntology/>>

<sup>7</sup>Para más información, véase <<http://www.opengalen.org/index.html>>

<sup>8</sup><<http://cmap.ihmc.us/conceptmap.html>>

<sup>9</sup><<http://protege.stanford.edu>>

<sup>10</sup><<http://www.mindswap.org/2005/SMORE/>>

## REFERENCIAS BIBLIOGRÁFICAS

AGUADO DE CEA, G., I. ÁLVAREZ DE MON, y A. PAREJA-LORA. 2003. “Primeras aproximaciones a la anotación lingüístico-ontológica de documentos de Web Semántica: OntoTag”, *Revista Iberoamericana de Inteligencia Artificial*, 1: 37-49.

BARONI, M., A. KILGARRIFF, J. POMIKÁLEK, y P. RYCHLÝ. 2006. “WebBootCaT: instant domain-specific corpora to support human translators”. *EAMT 2006 - 11th Annual Conference of the European Association for Machine Translation. Oslo: The Norwegian National LOGON Consortium and The Departments of Computer Science and*

- 
- Linguistics and Nordic Studies at Oslo University (Norway)*: 247-252. [Documento de Internet disponible en [http://trac.sketchengine.co.uk/attachment/wiki/WBC/DocsIndex/webbootcat\\_eamt06.pdf?format=raw](http://trac.sketchengine.co.uk/attachment/wiki/WBC/DocsIndex/webbootcat_eamt06.pdf?format=raw)].
- BERNERS-LEE T., J.A. HENDLER, y O. LASSILA. 2001. "The Semantic Web". *Scientific American*. 284 (5): 34-43, May 2001. [Documento de Internet disponible en <http://www.sciam.com/article.cfm?id=the-semantic-web> 12].
- BURNARD L., 2004, "Metadata for Corpus Work". *Developing Linguistic Corpora: a Guide to Good Practice*. Eds. M. Wynne. Oxbow Books: Oxford. 30-46. [Documento de Internet disponible en <http://ahds.ac.uk/linguistic-corpora>].
- CABRÉ CASTELLVÍ, M.T. 2002. "Textos especializados y unidades de conocimiento: metodología y tipologización". *Texto, terminología y traducción*. Eds. J. García Palacios y M. T. Fuentes. Ediciones Almar: Salamanca. 15-36.
- EAGLES. 1999. "EAGLES LE3-4244 Preliminary Recommendations on Lexical Semantic Encoding Final Report" [Documento de Internet disponible en <http://www.ilc.cnr.it/EAGLES96/EAGLESLE.PDF>].
- EAGLES. 1996a. "EAGLES: Recommendations for the Morphosyntactic Annotation of Corpora". *EAGLES Document EAG--TCWG--MAC/R*. [Documento de Internet disponible en <http://www.ilc.cnr.it/EAGLES/annotate/annotate.html>].
- EAGLES. 1996b. "EAGLES: Recommendations for the Syntactic Annotation of Corpora". *EAGLES Document EAG—TCWG—SASG/1.8*. [Documento de Internet disponible en <http://www.ilc.cnr.it/EAGLES/pub/eagles/corpora/sasg1.ps.gz>].
- GRUBER, T.R. 1993. "A translation Approach to Portable Ontologies". *Journal on Knowledge Acquisition*. 5 (2): 199-220.
- KIRYAKOV A., B. POPOV, D. OGNJANOFF, D. MANOV, A. KIRILOV y M. GORANOV. 2003. "Semantic Annotation, Indexing, and Retrieval" *2nd International Semantic Web Conference (ISWC2003)*, 20-

---

23 October 2003, Florida, USA, Springer-Verlag Berlin Heidelberg 2003. Vol. 2870: 484-499. [Documento de Internet disponible en [http://www.ontotext.com/publications/SemAIR\\_ISWC169.pdf](http://www.ontotext.com/publications/SemAIR_ISWC169.pdf)].

- LEECH, G. 1997. "Introducing Corpus Annotation". *Corpus Annotation: Linguistic Information from Computer Text Corpora*. Eds. R. Garside, G. Leech y T. McEnery. London: Longman. 1-18.
- LEECH G., 2004, "Adding Linguistic Annotation". *Developing Linguistic Corpora: a Guide to Good Practice*. Eds. M. Wynne. Oxbow Books: Oxford. 17-29. [Documento de Internet disponible en <http://ahds.ac.uk/linguistic-corpora/>].
- NIREMBURG S., y V.RASKIN. 2001. "Ontological Semantics". [Documento de Internet disponible en <http://crl.nmsu.edu/Staff/pages/Technical/sergei/book/index-book.html>].
- SENSO RUIZ, J.A. y A. ABEL LEIVA MEDEROS. 2008 (en prensa). *El proceso de anotación semántica: tendencias y retos*. Cursos de Doctorado, Universidad de Granada.
- WEIGAND, H. 1997. "Multilingual Ontology-Based Lexicon for News Filtering –The TREVI Project". *15th International Joint Conferences on Artificial Intelligence (IJCAI-97)*. 160-165. [Documento de Internet disponible en <http://crl.nmsu.edu/Events/IJCAI>].