Modelos Computacionales
Actividad 2

José Alberto Benítez Andrades

71454586A

Modelos Computacionales
Máster en Lenguajes y Sistemas Informáticos - Tecnologías del Lenguaje en la Web
UNED
29/01/2011

29 de enero de 2011

0. Enunciado

Contenidos:
o En el Tema 2 se presentan diferentes modelos semánticos, y diversos criterios comúnmente utilizados en la bibliografía para estudiar la cobertura y adecuación de los mismos.
o Bibliografía básica: (Jurafski&Martin 2009, cap 17, 19 y 20). Del capítulo 20 es suficiente estudiar los apartados 20.1, 20.2, 20.3, 2.6 (hasta el tercer párrafo de la pag 668), 20.8 y 20.9.

Ejercicio T2.1: Responder a la preguntas que siguen.

2.1.1 ¿Tiene conocimientos previos sobre representación en lógica de primer orden? Qué libro y en qué asignatura lo ha estudiado? Pase a la pregunta 2.1.3.

Si la respuesta ha sido negativa, pase a la siguiente pregunta, 2.1.2

2.1.2 Si sus conocimientos sobre la lógica de primer orden se corresponden básicamente solo con lo aprendido con los apartados 17.2, 17.3 y 17.4 del capítulo 17 del texto básico, entonces el enunciado del ejercicio es el que sigue. Resumir con mucho detalle la sección 17.4.1 (representing time) sin necesidad de tener muy en cuenta la representación formal y enumerar al menos 6 formas diferentes de denotar tiempo, puntos temporales etc (por ejemplo: con una fecha, con un intervalo de fechas, con adverbios temporales etc) y para cada uno cinco ejemplos variados.

2.1.3 Resumir con mucho detalle la sección 17.4.5 sin necesidad de tener muy en cuenta la representación formal y construir un ejemplo como el del apartado para la temática de la web semántica. Para ello se pueden usar los links encontrados para la práctica, los links (http://es.wikipedia.org/wiki/Web_sem%C3%A1ntica y http://www.w3c.es/Divulgacion/Guiasbreves/WebSemantica) o la información aportada en el tema 3.

2.1.4 Definir (tema 19) roles temáticos, diátesis y describir, con ejemplos, los recursos EuroWordNet, PropBank y FrameNet, elaborando una lista de links relevantes.

1. Resolución.

2.1.1 ¿Tiene conocimientos previos sobre representación en lógica de primer orden? Qué libro y en qué asignatura lo ha estudiado? Pase a la pregunta 2.1.3. Si la respuesta ha sido negativa, pase a la siguiente pregunta, 2.1.2

Sí, tengo conocimientos previos sobre la representación en lógica de primer orden, lo he estudiado a lo largo de la carrera en distintas asignaturas. Siendo un poco más concreto, las asignaturas en las que he abordado este tema con gran profundidad, han sido las siguientes:

- Álgebra (1º Ingeniería Informática, Curso 2006/2007, Universidad de León).
- Teoría de Autómatas y Lenguajes Formales (2º Ingeniería Informática, Curso 2007/2008, Universidad de León).
- Lógica Computacional (2º Ingeniería Informática, Curso 2007/2008, Universidad de León.

A pesar de proporcionarnos siempre una bibliografía, la mayoría de lo estudiado, en lo que a lógica computacional se refiere, ha sido de apuntes propios realizados asistiendo a las distintas clases. No obstante, en algunos trabajos he tenido que realizar la lectura de distintos apartados de los siguientes libros:

- Ben-Ari, Mordechai. Mathematica Logic for Computer Science. Springuer. 2001. ISBN: 1-85233-319-7
- Cori, R., Lascar, D. Logique Mathématique. Masson. 1993. ISBN: 2-225-84079-2, (vol 1º) y 2-225-84080-6, (vol 2º)
- Galton, A. Logic for Information Technology. John Wiley & Sons. ISBN: 0-471-92777-5
- Huth, M., Ryan, M. Logic in Computer Science. Cambridge University Press. 2004. ISBN: 0-521-54310-X
- Lassaigne, R., Rougemont, M. de. Logique et fondements de l informatique. Hermes. 1993. ISBN: 2-86601-380-8
- Nerode, A., Shore, R. Logic for Applications. Springer. 1997. ISBN:0-387-94893-7
- Zhongwan, L. Mathematical Logic for Somputer Science. World Scientific. 1998. ISBN: 981-02-3091-5

2.1.2 ¿ Si sus conocimientos sobre la lógica de primer orden se corresponden básicamente solo con lo aprendido con los apartados 17.2, 17.3 y 17.4 del capítulo 17 del texto básico, entonces el enunciado del ejercicio es el que sigue. Resumir con mucho detalle la sección 17.4.1 (representing time) sin necesidad de tener muy en cuenta la representación formal y enumerar al menos 6 formas diferentes de denotar tiempo, puntos temporales etc (por ejemplo: con una fecha, con un intervalo de fechas, con adverbios temporales etc) y para cada uno cinco ejemplos variados.

En mi caso la respuesta fue positiva, con lo cual he pasado directamente a la siguiente pregunta.

2.1.3. Resumir con mucho detalle la sección 17.4.5 sin necesidad de tener muy en cuenta la representación formal y construir un ejemplo como el del apartado para la temática de la web semántica. Para ello se pueden usar los links encontrados para la práctica, los links (http://es.wikipedia.org/wiki/Web_sem%C3%A1ntica y http://www.w3c.es/Divulgacion/Guiasbreves/WebSemantica) o la información aportada en el tema 3.

En el punto 17.5 de este tema, se explica de manera detallada cómo funcionan las redes semánticas y su transformación al lenguaje de lógicas descriptivas.

La red semántica es una forma de representación de conocimiento lingüístico en la que los conceptos y sus interrelaciones se representan mediante un grafo. En caso de que no existan ciclos, estas redes pueden ser visualizadas como árboles. Las redes semánticas son usadas, entre otras cosas, para representar mapas conceptuales y mentales. La dificultad de esta manera de representar el conocimiento, radica en que el sistema debe interpretar el significado de lo que se realiza en el propio sistema.

Sin embargo, las lógicas descriptivas representan las redes semánticas de manera más comprensible para las personas, mostrando un mapa conceptual que es especialmente adecuados para ciertos tipos de modelos de dominio. Las lógicas descriptivas están basadas en enfoques lógicos que corresponden a diferentes subconjuntos de FOL (First-Order Logic Web Language).

Cuando utilizamos las lógicas descriptivas, hacemos hincapié en la representación del conocimiento sobre las categorías, los individuos que pertenecen a esas categorías y las relaciones que pueden mantener entre estos individuos. El conjunto de categorías o conceptos que conforman un dominio de aplicación particular, se llama terminología. La parte de una base de conocimientos que contiene los términos se llama *TBox*, que contrasta con la ABox que contiene los datos sobre los individuos. La terminología se organiza mediante una

jerarquía llamada ontología que captura las relaciones subconjunto / superconjunto entre las categorías.

Por ejemplo, para representar un concepto, mediante FOL se utilizan predicados de la forma *Web (x)* para representar un objeto de tipo Web, mientras que, mediante Lógicas Descriptivas sólo se usaría la palabra *Web*.

Para determinar la relación entre las categorías, hay 2 métodos: el primero consiste en definir de forma completa los conceptos y el segundo método consiste en concretar todas las relaciones existentes entre las propias categorías. Entre otras, las distintas relaciones que existen, son las siguientes:

Constructor	DL Syntax	Example
intersectionOf	$C_1 \sqcap \ldots \sqcap C_n$	Human ⊓ Male
unionOf	$C_1 \sqcup \ldots \sqcup C_n$	Doctor ⊔ Lawyer
complementOf	$\neg C$	¬Male
oneOf	$\{x_1 \ldots x_n\}$	{john, mary}
toClass	$\forall P.C$	∀hasChild.Doctor
hasClass	$\exists r.C$	∃hasChild.Lawyer
hasValue	$\exists r.\{x\}$	∃citizenOf.{USA}
minCardinalityQ	$(\geqslant n\ r.C)$	$(\geqslant 2$ hasChild.Lawyer)
maxCardinalityQ	$(\leqslant n\ r.C)$	$(\leqslant 1$ hasChild.Male)
inverseOf	r^-	hasChild$^-$

Axiom	DL Syntax	Example
subClassOf	$C_1 \sqsubseteq C_2$	Human ⊑ Animal ⊓ Biped
sameClassAs	$C_1 \equiv C_2$	Man ≡ Human ⊓ Male
subPropertyOf	$P_1 \sqsubseteq P_2$	hasDaughter ⊑ hasChild
samePropertyAs	$P_1 \equiv P_2$	cost ≡ price
disjointWith	$C_1 \sqsubseteq \neg C_2$	Male ⊑ ¬Female
sameIndividualAs	$\{x_1\} \equiv \{x_2\}$	{President_Bush} ≡ {G_W_Bush}
differentIndividualFrom	$\{x_1\} \sqsubseteq \neg\{x_2\}$	{john} ⊑ ¬{peter}
transitiveProperty	$P \in \mathbf{R}_+$	hasAncestor$^+$ ∈ $\mathbf{R}_+$
uniqueProperty	$\top \sqsubseteq (\leqslant 1\ P.\top)$	$\top \sqsubseteq (\leqslant 1$ hasMother.$\top)$
unambiguousProperty	$\top \sqsubseteq (\leqslant 1\ P^-.\top)$	$\top \sqsubseteq (\leqslant 1$ isMotherOf$^-$.$\top)$

Una de las razones por las cuales se comenzaron a utilizar lógicas descriptivas, fue debido a la poca información que daban los diagramas de red en algunas ocasiones, ya que, en muchos casos, no se podía concretar si un conjunto de categorías era exhaustivo y disjunto.

Voy a intentar realizar un ejemplo para explicar las distintas relaciones, mediante la temática de la web semántica. Supongamos que tenemos las siguientes relaciones:

$$Web \sqsubseteq Servicio\ de\ Internet$$

$$VOIP \sqsubseteq Servicio\ de\ Internet$$

29 de enero de 2011

$$Streaming \sqsubseteq Servicio\ de\ Internet$$

$$Web\ Semántica \sqsubseteq Web$$

$$Web\ 2.0 \sqsubseteq Web$$

$$Web\ 1.0 \sqsubseteq Web$$

Con estas relaciones declaradas anteriormente, dejamos patente que las *Webs,* los servicios de *VOIP* y *Streaming* se encuentran dentro del conjunto de Servicios de internet. A su vez, los Blogs y la web Semántica, se encuentran dentro del conjunto Web. También podríamos realizar la siguiente relación, obteniendo más información sobre los objetos definidos:

$$Web\ 1.0 \sqsubseteq not\ Web\ Semántica$$

Esta relación deja patente que las webs pertenecientes al grupo de Webs 1.0 no pueden estar dentro del conjunto de Web Semánticas.

Profundizando un poco más en la web semántica, sabemos que para obtener una definición adecuada de los datos, la Web Semántica utiliza esencialmente RDF, SPARQL, y OWL, mecanismos que ayudan a convertir la Web en una infraestructura global en la que es posible compartir, y reutilizar datos y documentos entre diferentes tipos de usuarios.

- RDF proporciona información descriptiva simple sobre los recursos que se encuentran en la Web y que se utiliza, por ejemplo, en catálogos de libros, directorios, colecciones personales de música, fotos, eventos, etc.
- SPARQL es lenguaje de consulta sobre RDF, que permite hacer búsquedas sobre los recursos de la Web Semántica utilizando distintas fuentes datos.
- OWL es un mecanismo para desarrollar temas o vocabularios específicos en los que asociar esos recursos. Lo que hace OWL es proporcionar un lenguaje para definir ontologías estructuradas que pueden ser utilizadas a través de diferentes sistemas. Las ontologías, que se encargan de definir los términos utilizados para describir y representar un área de conocimiento, son utilizadas por los usuarios, las bases de datos y las aplicaciones que necesitan compartir información específica, es decir, en un campo determinado como puede ser el de las finanzas, medicina, deporte, etc. Las ontologías incluyen definiciones de conceptos básicos en un campo determinado y la relación entre ellos.

Teniendo en cuenta esto, podemos realizar las siguientes relaciones:

$$ReglasLógicas \sqsubseteq Reglas$$

$$SPARQL \sqsubseteq Herramientas$$

$$RSSReader \sqsubseteq Herramientas$$

Conociendo estas relaciones, se puede concretar un poco más la definición de la web semántica y de la web 2.0, del siguiente modo:

$$Web\ Semántica \sqsubseteq Web \sqcap \exists hasHerramientas.SPARQL \sqcap \exists hasReglas.ReglasLógicas$$

$$Web\ 2.0 \sqsubseteq Web \sqcap \nexists hasHerramientas.SPARQL \sqcap \nexists hasReglas.ReglasLógicas$$

Se podría decir también, que la web semántica es de la siguiente manera:

$$Web\ Semántica \equiv Web \sqcap ReglasLógicas \sqcap Herramientas\ de\ consulta$$

$$Web\ Semántica \equiv Web\ 1.0 \sqcup Web\ 2.0$$

$$Web\ Semántica \equiv \forall\ hasReglas.ReglasLogicas$$

$$Web\ Semántica \equiv \forall\ hasHerramientasConsulta.SPARQL$$

2.1.4 Definir (tema 19) roles temáticos, diátesis y describir, con ejemplos, los recursos EuroWordNet, PropBank y FrameNet, elaborando una lista de links relevantes.

Definir:

- Roles temáticos: Son un tipo rol semántico. Un rol semántico es la relación entre un constituyente sintáctico (generalmente, aunque no siempre, argumento del verbo) y un predicado (generalmente, aunque no siempre, un verbo). Ejemplos de roles semánticos son agente, paciente, beneficiario, etc. o también adjuntos como causa, manera, lugar, etc.

  Por lo general, los roles temáticos son un intento de capturar la semántica común entre dos objetos de un mismo tipo. Supongamos que tenemos el siguiente predicado:

  $$\exists e, x, y\ Conduciendo(e) \wedge Conductor\ (e, Pepe) \wedge CosaConducida(e, y) \wedge Coche(y)$$
  $$\exists e, x, y\ Arrastrando(e) \wedge Arrastrador(e, Luis) \wedge CosaArrastrada(e, y) \wedge Toalla(y)$$

  En estos 2 casos, *Conduciendo* y Arrastrando son los verbos, los que poseen la acción de hacer algo, son los AGENTES, ese es su rol temático. Mientras que, *Conductor* y *Arrastrador* son los objetos que realizan la acción de Romper y Abrir, es decir, son los EXPERIMENTADORES del evento. Los objetos directos que reciben la acción del verbo, son *CosaConducida* y *CosaArrastrada*, con lo cual, son los participantes más directos afectados por el evento, su rol temático es TEMA.

  Los roles temáticos son uno de los modelos lingüísticos más antiguos, propuestos inicialmente por el gramático indio Panini entre el 7º y 4º siglo a. C.

- Diátesis: Las alternancias de diátesis, se refiere al hecho de que los verbos se pueden utilizar en diferentes marcos de subcategorización en el que cambian ligeramente su significado semántico.

  Por ejemplo:

  "Pepe comió una pizza" vs "Pepe comió"

  "Pepe rompió la ventana" vs "La ventana se rompió"

  "María le dio unas flores a Pepe" vs "María le dio a Pepe unas flores"

  En realidad el objetivo al que tiene que llegar el verbo, es el mismo, pero la manera en la que se utilizan, es distinta, eso es una diátesis.

Recursos Semánticos:

El primero de ellos, FrameNet, sirvió de base para el primer gran trabajo sobre etiquetado automático de roles semánticos, precursor de los posteriores sistemas y del actual interés de la comunidad del Procesamiento del Lenguaje Natural en esta problemática. El segundo de ellos, PropBank, es responsable de la sensible mejora en los resultados de los sistemas actuales, principalmente debido a su clara vocación de corpus enfocado a la construcción de sistemas de etiquetado de roles semánticos.

FrameNet

FrameNet es un proyecto que pretende identificar y describir los aspectos lexicográficos de las palabras de un gran corpus de texto en inglés (esencialmente extraído del British National Corpus), tratando de reflejar con ello la relación entre las propiedades sintácticas y semánticas existentes en el idioma. El proyecto FrameNet contiene, entre otras cosas, un conjunto de oraciones que intentan abarcar exhaustivamente toda la casuística que se da en el inglés en relación a las realizaciones sintácticas de todos los posibles contenidos semánticos, proporcionando para dichas frases un etiquetado parecido al que hemos descrito como la tarea de etiquetado de roles semánticos.

Un ejemplo de las relaciones entre marcos semánticos en FrameNet

El nombre de FrameNet refleja precisamente la relación con esta teoría, así como con el hecho de que se establecen relaciones de herencia y composición entre estos marcos semánticos, formándose las redes de significado en las que participan las palabras. En realidad, los conceptos de marco semántico y elementos de un marco considerados en la teoría de marcos semánticos y por ende en FrameNet son más ambiciosos y tienen mayor complejidad conceptual que los conceptos correspondientes de la teoría de roles semánticos. En FrameNet

se identifican y describen los posibles marcos semánticos existentes en el inglés, y se analizan los significados de las palabras directamente refiriéndose al marco semántico en el que aparecen, estudiando las propiedades sintácticas de las palabras y cómo las propiedades semánticas se plasman en una realización sintáctica concreta.

Un conjunto determinado de palabras, que pueden constituir una proposición o ser simplemente un sintagma de algún tipo, estarán enmarcadas en términos semánticos en un marco o frame concreto. El significado particular de alguna de las palabras participantes es el que determina cuál es el marco semántico correcto. Este par formado por una palabra y un significado concreto para la misma se conoce en FrameNet como unidad léxica (lexical unit), y se dice entonces que una unidad léxica evoca un marco semántico. Así es como toma forma la idea base de la semántica basada en marcos, según la cuál el significado de las palabras debe ser explicado en términos del marco semántico en el que se enmarcan.

Además de los conceptos de marco semántico y unidad léxica, es necesario definir también el concepto de elemento de un marco o frame element antes de citar algunos ejemplos que resultaran muy clarificadores. Los elementos de un marco o frame elements son los distintos tipos de entidades que participan en un marco semántico determinado. En los términos empleados en el contexto del etiquetado de roles semánticos, un frame element viene a ser un rol semántico para una clase semántica determinada. Los elementos que participan en un marco semántico concreto son específicos de dicho marco, por lo que existen multitud de elementos distintos, siendo ésta una diferencia fundamental con la visión más generalista utilizada en la mayoría de los etiquetadores de los roles semánticos y en otros recursos lingüísticos como PropBank, en la que se utilizan un conjunto de roles más reducido y compartido por las distintas clases semánticas. Puede hacerse una analogía entre los elementos de un marco y los argumentos de un predicado de lógica de primer orden, o simplemente con los argumentos de algún tipo de función. Así pues, dada una secuencia de palabras que evocan un marco semántico determinado, hay que decidir cuáles de esas palabras instancian cada uno de los elementos requeridos por el marco semántico.

Se presenta a continuación un ejemplo de marco semántico y los elementos que participan en el mismo:

Frame : Transfer

Frame Elements : DONOR, THEME, RECIPIENT

Descripción : Alguien (DONOR) está en posesión de algo (THEME) y en tonces hace que alguien más (RECIPIENT) esté en posesión del THEME, quizás ocasionando que el THEME se mueva al RECIPIENT.

Los nombres que se utilizan para identificar los elementos del marco no deben entenderse literalmente. Por ejemplo, DONOR no significa necesariamente “donante", como indica la definición de la palabra, sino que debe entenderse en los términos expuestos en la descripción del marco semántico. Los nombres utilizados cumplen simplemente un objetivo mnemónico.

Veamos ahora dos realizaciones sintácticas del marco semántico anterior:

1. The teacher gave the student a book.
2. The teacher gave a book to the student.

Según la filosofía de FrameNet, el significado de los constituyentes de la oración debe ser entendido en términos de los roles semánticos y gramaticales que desempeñan con respecto al verbo give. El verbo es en este caso la palabra.

give	FEs:	Donor	Theme	Recipient
	PTs:	NP	NP	NP
	GFs:	Ext	Comp	Obj
give	FEs:	Donor	Theme	Recipient
	PTs:	NP	NP	PP-to
	GFs:	Ext	Obj	Comp

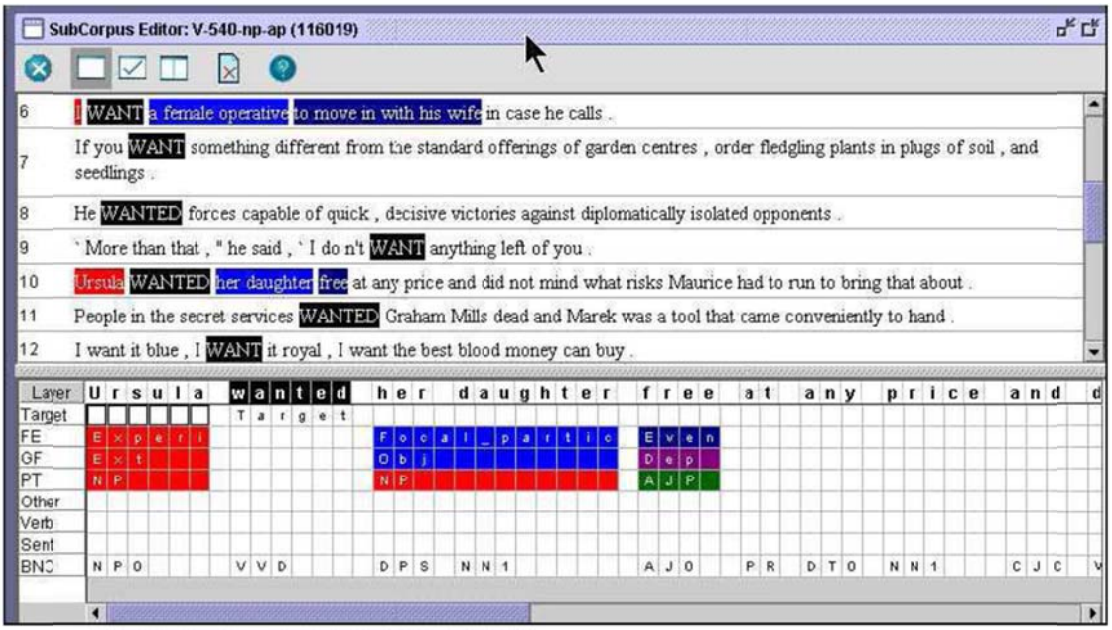
Tabla 3.1: Patrones de valencia para el verbo give en FrameNet. Para cada combinación de frame elements, se expresan las funciones sintácticas y gramaticales de las posibles realizaciones sintácticas de cada frame element que evoca el marco semántico Transfer, siendo la palabra que juega ese papel conocida como target en la terminología usada por FrameNet (se puede traducir por objetivo o, utilizando la terminología utilizada generalmente en los etiquetadores de roles semánticos y en recursos como PropBank, predicado). Los roles semánticos que participan en la oración serán los elementos del marco.

En FrameNet, al requisito por el cual una palabra debe combinarse con tipos particulares de sintagmas en una oración se le conoce como la valencia de una palabra (valence), por analogía con el término utilizado en química para referirse a las posibilidades de combinación de los átomos. La valencia puede entenderse en términos sintácticos y semánticos.

La valencia semántica vendría de una palabra especificada por los elementos del marco que evoca la palabra. Por ejemplo, la palabra give en el marco semántico Transfer debe ir acompañada de los elementos DONOR, THEME, y RECIPIENT. Para describir la valencia semántica de la palabra en toda su extensión será necesario por tanto especificar todos los marcos semánticos que puede evocar, y qué conjuntos de elementos deben acompañarla en cada caso.

Por otro lado, la valencia sintáctica de una palabra debe expresarse en términos de cuáles son las funciones sintácticas y gramaticales de los elementos semánticos de cada uno de los marcos semánticos asociados a la palabra. En el ejemplo anterior, las dos oraciones que aparecen reflejan parte de las propiedades de valencia sintáctica de la palabra give en el marco semántico Transfer. En ambos casos, el rol DONOR está formado por un sintagma nominal (the teacher), y el rol THEME está expresado por otro sintagma nominal (a book). El tercer elemento semántico en discordia, RECIPIENT está constituido en la primera oración por un sintagma nominal (the student) y en la segunda por un sintagma preposicional (to the student). Además, gramaticalmente, the teacher es el sujeto de gave en ambas oraciones (en FrameNet, el sujeto es denominado argumento externo, o de forma abreviada Ext). En la primera oración, the student es el copmlemento directo (Obj) del verbo, y a book funciona como complemento indirecto (Comp). En la segunda, sin embargo, a book funciona como complemento directo y to the student como indirecto. Toda esta información que caracteriza las propiedades de valencia de un predicado se encuentran anotadas en FrameNet mediante “patrones de valencia".

El proyecto FrameNet incluye la creación de una serie de herramientas que facilitan las tareas de búsqueda de ejemplos y etiquetado.



Aplicación para el etiquetado de ejemplos en FrameNet

Los marcos semánticos constituyen un método para caracterizar las relaciones semánticas entre palabras. Consideremos por ejemplo los verbos give y receive. Ambas palabras evocan el marco anterior Transfer. Es evidente para un lector que es al menos parecido decir que el profesor le dio un libro al alumno, o que el alumno recibió un libro del profesor. En FrameNet se considera que ambos verbos aportan una perspectiva distinta del mismo marco semántico. Por supuesto, las realizaciones sintácticas son distintas, y las funciones gramaticales realizadas por cada elemento del marco semántico también son distintas. Pero a nivel semántico, ambas oraciones quedan relacionadas a través del marco semántico. Esto contrasta con la visión clásica usada por los lingüistas para describir las estructuras de argumentos en las teorías del nexo, en la que se utilizan un conjunto más general de roles temáticos. Los roles temáticos tratan de capturar las regularidades existentes en las relaciones entre la semántica y la función gramatical de los constituyentes. Un análisis basado en roles temáticos asignará distintos roles a los participantes de una oración con el verbo give y a una oración con el verbo receive.

give	Agent	Theme	Recipient
receive	Recipient	Theme	Source

Con la vista puesta en la utilización de los datos para construir un etiquetador automático de roles semánticos mediante modelos estadísticos, la utilización de frame elements en lugar de roles temáticos tendrá consecuencias que habrá que considerar a la hora de utilizar FrameNet como corpus de entrenamiento. Por un lado, se gana en generalidad entre distintas palabras que evocan un mismo marco semántico, al tener todas ellas el mismo conjunto de roles temáticos sin importar la perspectiva impuesta por cada palabra. Pero por otro lado, perdemos las generalizaciones relacionadas con las teorías del nexo, que vienen mejor expresadas en términos de roles temáticos, como por ejemplo que el rol Agent suele funcionar gramaticalmente como sujeto.

Los distintos marcos semánticos están además relacionados entre sí en FrameNet. Existen básicamente dos tipos de relaciones: de herencia y de composición. En las relaciones de herencia, un marco semántico se dice que hereda de otro marco si posee todas las propiedades del marco padre y añade algunos detalles específicos. Por ejemplo:

a. The teacher gave the student a message.
b. The teacher mail the student a message.

Si enviamos un correo electrónico, estamos realizando una transferencia en el sentido descrito en el marco Transfer, sólo que ahora, por ejemplo, el donante pasa a ser emisor. La descripción del marco semántico será distinta, más especializada. También los nombres de los elementos del marco serán distintos, aunque el número de estos y las propiedades sintácticas de los verbos que evocan el nuevo marco semántico serán idénticos a los del padre.

Otras veces, la relación es de composición. Por ejemplo , en el marco semántico Commercial Transaction, evocado por verbos como sell o buy, podemos considerar que aparecen dos eventos, cada uno de los cuáles vendría a corresponderse con un marco semántico Transfer: un comprador da al vendedor dinero, y el vendedor le da algo a cambio. De esta manera, las propiedades sintácticas de estos verbos son las mismas que las de los verbos que evocaban el marco semántico Transfer, aunque existen ahora más roles semánticos.

Una vez se tienen claras las palabras que pueden evocar un marco semántico determinado, y se estudian todas las propiedades de valencia de cada una, el proyecto FrameNet busca frases que sirvan de ejemplo de todo esto, y las etiqueta con la información semántica y sintáctica anterior. De esta manera, decimos que el corpus de frases anotadas que nos proporciona FrameNet busca ante todo la exhaustividad, es decir, al menos un ejemplo de todas las combinaciones posibles de cada marco semántico, como si de un diccionario de marcos semánticos se tratara.

PropBank

Este recurso, cuyo nombre completo es realmente Proposition Bank, consiste en una versión enriquecida del corpus Penn Treebank II, que básicamente incluía información de las estructuras sintácticas. A diferencia del enfoque utilizado en el recurso FrameNet, en PropBank se lleva a cabo un acercamiento eminentemente práctico al problema del etiquetado semántico, de forma que los integrantes del grupo de trabajo del proyecto no estaban interesados en realizar un estudio tan pormenorizado como en FrameNet de todas las clases semánticas existentes y de las relaciones de herencia y composición entre ellas, ni tampoco en representar en el corpus fenómenos semánticos globales complejos tales como la correferencia, la cuantificación o la resolución de anáforas. En vez de esto, lo que se busca en PropBank es realizar un análisis superficial de la estructura semántica de cada oración, identificando para cada una de las oraciones del corpus TreeBank los argumentos o roles semánticos que participan en cada una de las proposiciones. Se pretende con ello disponer de un corpus lo suficientemente amplio como para ser relevante desde el punto de vista estadístico, posibilitando su posterior uso en tareas como la que nos ocupa del etiquetado de roles semántico automático.

PropBank es un recurso más reciente, posterior a los primeros trabajos publicados sobre el etiquetado de roles semánticos automático, y esto queda patente en el enfoque práctico escogido. Mientras en FrameNet se intentan analizar todas las posibles realizaciones

sintácticas de todas las clases semánticas existentes en el inglés y aportar un ejemplo para cada una de ellas, constituyendo así una especie de diccionario semántico del idioma, PropBank tiene vocación de corpus anotado con roles semánticos útil para la construcción de modelos de aprendizaje automático. De hecho, así como FrameNet inspiró el primer trabajo importante sobre etiquetado automático de roles semánticos, la aparición de PropBank ha propiciado la explosión de trabajos en este área y la mejora en el rendimiento de los sistemas actuales.

Para cada uno de los verbos que aparecen en el corpus original, se han definido un conjunto de posibles roles semánticos, para posteriormente anotar cada ocurrencia de los mismos en el texto. PropBank se centra exclusivamente en los verbos, estudiando los roles semánticos como argumentos de los verbos. En ningún momento se contempla la posibilidad de que un nombre, adjetivo o adverbio funcionen como núcleos o predicados para un conjunto de roles, tal como ocurría en FrameNet, habiéndose dejado esta tarea para futuras revisiones.

Dada la dificultad de definir un conjunto general de roles semánticos común a todos los predicados posibles, lo cuál sería muy interesante desde el punto de vista de la generalización entre verbos que aportaría, en PropBank se han definido los roles para cada uno de los verbos por separado, pero este proceso se ha realizado tratando de permitir algún grado de generalización entre los distintos verbos, aunque no de manera totalmente estricta. Para cada verbo, los argumentos o roles semánticos son numerados empezando en 0. Por ejemplo, para un verbo en particular, el rol Arg0 será habitualmente aquel argumento del verbo que cumple las funciones de Agente, mientras que el rol etiquetado como Arg1 se reservará siempre que sea posible al argumento que funciona como Paciente. Para los argumentos siguientes no es posible realizar generalizaciones tan claras, aunque en la medida de lo posible se han intentado seguir unos criterios comunes (en concreto, se utiliza la organización de roles que aparece en el recurso VerbNet). Además de los roles numerados específicos de cada verbo, también se definen algunos roles genéricos que pueden ser aplicados a cualquier verbo.

Para cada acepción considerada de un verbo, se definen un conjunto de roles que participan en el predicado en cuestión, recibiendo este conjunto el nombre de roleset. Además, cada roleset se puede asociar con las posibles realizaciones sintácticas del predicado, indicando las funciones sintácticas en las que pueden aparecer cada uno de los roles anteriores. La unión entre un conjunto de roles y las posibles realizaciones sintácticas es conocida en PropBank como frameset.

Un verbo polisémico podrá tener de este modo varios framesets, siempre que las diferencias en el significado sean lo suficientemente profundas como para requerir participantes o roles semánticos distintos. Todos los framesets utilizados en PropBank son definidos en un fichero (Frame File), en el que para cada frameset se incluyen:

- El verbo en cuestión junto a un número que indica la acepción que se está considerando.

- El conjunto de roles numerados, junto a un descriptor para cada uno que indica al menos superfluamente cuál es el papel que juega cada argumento en la acepción actual. Este descriptor debe entenderse sólo como un mnemónico informativo para los anotadores que participan en un proyecto, y no tiene ninguna intención teórica.

- Por último, una serie de oraciones de ejemplo extraídas del corpus etiquetadas convenientemente con los roles anteriores, que tratan de reflejarlas distintas

realizaciones sintácticas en las que puede presentarse el verbo que se está considerando en su acepción actual.

1. Frameset **accept.01** "take willingly"

Arg0: Acceptor

Arg1: Thing accepted

Arg2: Accepted-from

Arg3: Attribute

Ex:[Arg0 He] [ArgM-MOD would][ArgM-NEG n't] accept [Arg1 anything of value] [Arg2 from those he was writing about].

2. Frameset **kick.01** "drive or impel with the foot"

Arg0: Kicker

Arg1: Thing kicked

Arg2: Instrument (defaults to foot)

Ex1: [ArgM-DIS But] [Arg0 two big New York banksi] seem [Arg0 *trace*i] to have kicked [Arg1 those chances] [ArgM-DIR away], [ArgM-TMP for the moment], [Arg2 with the embarrassing failure of Citicorp and Chase Manhattan Corp. to deliver $7.2 billion in bank financing for a leveraged buy-out of United Airlines parent UAL Corp].

Ex2: [Arg0 Johni] tried [Arg0 *trace*i] to kick [Arg1 the football], but Mary pulled it away at the last moment.

Generalmente, como puede verse en los ejemplos, cada frameset consta de dos, tres o hasta cuatro argumentos numerados, aunque existen casos en los que puede haber hasta seis argumentos numerados, especialmente en algunos verbos relacionados con el movimiento como el siguiente:
1. Frameset **edge.01** "move slightly"

Arg0: causer of motion

Arg1: thing in motion

Arg2: distance moved

Arg3: start point

Arg4: end point

Arg5: direction

Ex: [Arg0 Revenue] edged [Arg5 up] [Arg2-EXT 3.4%] [Arg4 to $904 million] [Arg3 from $874 million] [ArgM-TMP in last year's third quarter].

Además de los argumentos numerados, existe uno especial etiquetado como ArgA que se utiliza en situaciones en las que existe más de un argumento funcionando en cierto modo como agente. Por ejemplo, en la frase Mary hustled John to school promptly at 7:30 pm, es John quien lleva a cabo la acción de escaparse de clases antes de tiempo, pero aún así Mary también está ejerciendo de agente de alguna manera incitando a John a llevar a cabo la acción.

Es en estos casos en los que se etiqueta a este segundo participante, Mary en nuestro ejemplo, con la etiqueta ArgA. Por último, también se utilizan etiquetas para roles que son independientes de los verbos, y que en general podemos asociar con el concepto gramatical de adjuntos (aunque esto no es absolutamente preciso en todos los casos). Estos argumentos se conocen en PropBank como Funcional tags. Los argumentos independientes de este tipo que aparecen en PropBank son los siguientes:

Funcional tag	Descripción
ArgM-TMP	Modificador temporal.
ArgM-LOC	Modificador de lugar.
ArgM-DIR	Modificador de dirección.
ArgM-MNR	Modificador de manera o modo.
ArgM-CAU	Indica la causa de algo.
ArgM-ADV	Se utiliza para adverbios de nivel de oración y otros agentes que no queden recogidos en ninguna otra categoría.
ArgM-DIS	Etiquetan a partículas conectivas del discurso.
ArgM-NEG	Partículas de negación.
ArgM-PNC	Indican la motivación (no la causa) de una acción.
ArgM-REC	Indican acciones reflexivas o recíprocas.

Hay dos casos particulares de argumentos funcionales que no son independientes de los verbos, sino que aparecen en los framesets como parte de los argumentos participantes en una acepción de un verbo. Son los argumentos EXT, que indica un constituyente numérico o de cantidad, y PRD, que marca una relación predicativa entre dos argumentos. Este último es un poco más difícil de entender. Sea la siguiente oración:

1. Mary called John a doctor

Existe ambigüedad en el significado de la frase, ya que por un lado podemos entender que Mary llamó doctor a John (es decir, dijo que John era un doctor), o bien Mary llamó a un doctor para que viera a John. En el primer caso, se establece una relación predicativa entre John y doctor, y por tanto en PropBank vendrá etiquetado el argumento a doctor con la etiqueta funcional PRD

EuroWordNet

EuroWordNet es una gran base de datos léxico-semántica creada en 1985 por George A. Miller. El objetivo del proyecto es representar la información semántica de las palabras del inglés con la vista puesta en el procesamiento computacional de dicha información. Así como en un diccionario se expresa el significado de las palabras mediante definiciones, en WordNet se establecen grupos de palabras sinónimas o synsets, de forma que una palabra se define por equiparación con otras que significan lo mismo o, al menos, algo muy parecido. Además también se establecen relaciones semánticas entre estos synsets, formando una gran red de palabras que da nombre al recurso.

Si una palabra tiene varios significados, aparece en distintos synsets. Para cada uno de los synsets, se incluye una pequeña definición y unos cuántos ejemplos de su uso dentro del lenguaje. Dentro de un synset podemos encontrar palabras individuales o secuencias de palabras que juntas expresan un significado concreto (collations), como por ejemplo máquina de coser. Para cada palabra se almacena el número de significados en que aparece (en cuántos synsets está incluida), y para cada acepción, existe una estimación de la frecuencia con la que se da. Actualmente en el proyecto existen 150.000 palabras agrupadas en 115000 synsets (versión 1.5 de WordNet).

Las relaciones semánticas dependen del tipo de palabra sobre la que se definan. Entre los nombres se establecen relaciones de hiperonimia e hiponimia (que en términos informáticos definen una relación de generalización y especialización respectivamente), así como de holonimia, meronimia (relaciones de composición) y términos coordinados (hermanos en la jerarquía de herencia, siguiendo con la metáfora informática). Para los verbos se definen la hiperonimia y la troponimia (esta última es similar a la hiponimia en los nombres), así como la implicación y términos coordinados. Para los adjetivos y adverbios se definen relaciones para indicar si están relacionados con algún nombre, verbo o adjetivo.

Todo esto queda resumido en la siguiente enumeración:

- Nombres
  hiperonimias : Y es una hiperonimia(generalización) de X si todo X es un (o algún tipo de) Y
  hiponimias : Y es una hiponimia(especialización) de X si todo Y es un (o algún tipo de) X

  términos coordinados : Y is a término coordinado con X si X e Y comparten una hiperonimia (es una relación conmutativa)

  holonimia : Y es una holonimia de X si X es una parte de Y
  meronimia : Y es una meronimia de X si Y es una parte de X

- Verbos
  hiperonimia : el verbo Y es una hiperonimia del verbo X is la actividad
  X es un (o algún tipo de) Y
  troponimia : el verbo Y es una troponimia del verbo X si la actividad Y implica hacer X de alguna manera
  implicación : el verbo Y está implicado por X si haciendo X debes estar haciendo Y
  términos coordinados : dos verbos que comparten una hiperonimia

- Adjectivos
  nombres relacionados
  participio de verbo

- Adverbios
  adjetivo raíz

También existen relaciones entre palabras directamente, básicamente relaciones entre antónimos y derivados. La relación más importante y en la que más hincapié se hace en Wordnet para los nombres y verbos es la de hypernym (IS A). Según estas relaciones, todos los nombres y verbos están organizados en jerarquías hasta llegar a un conjunto base de categorías generales o primitivas, 25 para los nombres, 15 para los verbos. En la siguiente tabla muestro las categorías base de las que parten todos los nombres por hiperonimia:

act,action,activity	animal,fauna	artifact
attribute,property	body,corpus	cognition,knowledge
communication	event,happening	feeling,emotion
food	groups,collection	location,place
motive	natural object	natural phenomenon
person,human being	plant,flora	possession
process	quantity,amount	relation
shape	state,condition	substance
time		

Los adjetivos están organizados principalmente mediante relaciones de antonimia. Los adverbios se organizan en base a los adjetivos de los que se derivan.

WordNet constituye un precursor de FrameNet, y de ahí por tanto su importancia en el etiquetado de roles semánticos. Además, la información contenida en WordNet podría servir para mejorar los sistemas de etiquetado de roles semánticos, por ejemplo enriqueciendo recursos, o utilizando el synset al que pertenece una palabra para ayudar a decidir sobre el rol semántico que desempeña.