

Corrección no Supervisada de Dependencias Sintácticas de Aposición mediante Clases Semánticas*

Unsupervised Correction of Syntactic Dependencies of Apposition through Semantic Classes

Bernardo Cabaleiro, Anselmo Peñas

NLP & IR Research Group
ETSI Informática, UNED, Spain
{bcabaleiro,anselmo}@lsi.uned.es

Resumen: En este artículo exploramos de qué manera el conocimiento semántico adquirido de manera automática a partir de grandes colecciones de texto permite mejorar el resultado de un analizador sintáctico. Para ello proponemos un método no supervisado que corrige dependencias de aposición buscando en su contexto candidatos con mayor compatibilidad semántica que el proporcionado por el analizador sintáctico en primera instancia.

Palabras clave: Analizadores sintácticos, Adquisición de clases semánticas

Abstract: In this article we explore how automatic acquired semantic knowledge from large text collections allows to improve a syntactic parser. We propose an unsupervised method that corrects appositions dependencies by searching candidates in its context more semantically suitable than the one proposed by the syntactic parser in the first place.

Keywords: Syntactic parsers, Semantic classes acquisition

1. Introducción

El análisis de dependencias es un proceso clave para el procesamiento de lenguaje natural. Sin embargo, encontrar la estructura sintáctica correcta depende en muchas ocasiones de conocer las relaciones semánticas entre las palabras involucradas.

En este trabajo mostramos cómo un analizador sintáctico podría aprovechar información semántica extraída automáticamente para mejorar la identificación y corrección de errores en las dependencias de aposición.

Para ello presentamos un método no supervisado de corrección de aposiciones basado en clases semánticas. Consideramos como aposición aquellas construcciones gramaticales con dos sintagmas nominales adyacentes, donde uno define o modifica al otro (House, 2005). Cuando uno de los sintagmas nominales está gobernado por un sustantivo y el otro por una entidad nombrada, generalmente el sustantivo se utiliza para denotar una clase

semántica de la entidad. Por ejemplo:

*David meets his wife, **Julia**.*¹

Los analizadores sintácticos suelen determinar que existe una aposición entre el nombre común “wife” y la entidad nombrada “Julia”, dando lugar a interpretar que “Julia” es de la clase “wife”.

Cuando la entidad no admite el sustantivo como clase se puede inferir que la dependencia dada por el analizador podría ser incorrecta. Por ejemplo:

*David supports the team of his wife, **The Vikings**.*

Un analizador habitualmente determina una dependencia entre “wife” y “Vikings” por lo que se interpreta que Vikings es de la clase wife. Sin embargo, existe una incompatibilidad entre el tipo de la entidad nombrada y la clase, ya que “Vikings” es una organización y “wife” una clase característica de personas, lo que supone una evidencia de que tanto la dependencia como la interpretación que de ella se deriva son incorrectas. En cambio, en el contexto próximo se encuentra la

* This work has been partially supported by the Spanish Ministry of Science and Innovation, through the project Holopedia (TIN2010-21128-C02), and the Regional Government of Madrid, through the project MA2VICMR (S2009/TIC1542)

¹En negrita la entidad nombrada, y subrayado los sustantivos candidatos a gobernante de la aposición.

clase “team”, que sí es válida para organizaciones.

Otro motivo de incompatibilidad es el género. Por ejemplo:

David meets a friend of his wife, Peter.

Aquí el analizador podría identificar una dependencia entre “wife” y “Peter”. Saber que “wife” es una clase exclusivamente femenina y “Peter” una entidad nombrada masculina nos hace considerar más adecuada la clase “friend”.

De acuerdo con el escenario descrito, nos formulamos las siguientes preguntas de investigación:

- ¿Puede detectarse en qué casos el analizador sintáctico no asigna correctamente el gobernante de una aposición debido a una incompatibilidad semántica?
- En caso de encontrar un error, ¿se puede determinar la estructura correcta y corregirlo?

Para responder a estas preguntas tratamos de identificar posibles errores buscando aposiciones con varios sustantivos candidatos a gobernar la relación. Después decidimos cuál de ellos es más apropiado utilizando como soporte el conocimiento antecedente sobre las clases.

Adquirimos el conocimiento antecedente a partir de un corpus de 1.5 millones de documentos, empleando patrones sintácticos muy sencillos para la identificación de clases semánticas.

El resto del artículo se organiza de la siguiente manera. Primero explicamos los dos pasos que sigue el sistema: adquisición de conocimiento antecedente (sección 2) e identificación y corrección de errores (sección 3). En la sección 4 mostramos la evaluación, hacemos un breve repaso a trabajos similares en la sección 5. Terminamos con las conclusiones y el trabajo futuro en la sección 6.

2. Adquisición de conocimiento antecedente

El reconocimiento de entidades es un problema muy estudiado en el procesamiento de lenguaje natural y los sistemas alcanzan buenos resultados (Finkel, Grenager, y Manning, 2005), especialmente para los tipos básicos (persona, localización y organización).

Sin embargo, conocer las clases semánticas asociadas a entidades (por ejemplo, una per-

sona puede ser “bombero”, “hijo”, “dueño”, etc) es un problema abierto que resulta interesante para múltiples tareas de NLP. El objetivo de esta fase es extraer relaciones entre clases e instancias con una probabilidad asociada.

Para adquirir esta información representamos en forma de grafo los documentos de una colección. Empleamos como base el paquete Stanford CoreNLP, que incluye uno de los analizadores sintácticos más populares (Klein y Manning, 2003), basado en una gramática libre de contexto probabilística. También utilizamos el etiquetador de partes de la oración (Toutanova y Manning, 2000), el reconocedor de entidades nombradas (Finkel, Grenager, y Manning, 2005) y el sistema de resolución de correferencia (Lee et al., 2011). Además, colapsamos los nodos en referentes de discurso. La descripción completa del método de obtención de la representación se puede ver en (Cabaleiro y Peñas, 2012).

Para obtener un conjunto de clases a partir de los grafos utilizamos una serie de patrones extremadamente sencillos basados en dependencias sintácticas «*governor, dependency, dependant*» en los que participa un sustantivo que será la clase y una entidad nombrada, que a su vez tiene un tipo de entidad asociado (persona, organización o localidad). La lista completa de patrones sintácticos está detallada en la tabla 1. En caso de coincidencia, asignamos el nombre común como clase semántica del nombre propio y obtenemos una instancia *EN – Clase – Tipo*.

Patrón Sintáctico		
EN	nn	NN
EN	appos	NN
EN	abbrev	NN
NN	appos	EN
NN	abbrev	EN
EN	such_as	NN
EN	like	NN

Tabla 1: Patrones para la asignación de clases semánticas. Cada entrada se corresponde con una tripleta «*governor, dependency, dependant*» donde *NN* es un nombre común y *EN* es una entidad nombrada.

Con estos patrones no se pretende obtener todas las relaciones clase-instancia expresadas en la colección, sino adquirir un número

representativo y suficiente de clases con las que evaluar los sustantivos candidatos.

Tras obtener las asignaciones de clases semánticas, agregamos la información de toda la colección para obtener las frecuencias de las coocurrencias entre entidades nombradas, clases y tipos. La Tabla 2 contiene las 5 clases más comunes asociadas al tipo de entidad *person*.

Clase	Tipo de EN	Frecuencia
spokesman	person	140229
president	person	102877
director	person	98182
leader	person	79716
coach	person	55511

Tabla 2: Clases más comunes asociadas al tipo de entidad *person*.

En la fase de identificación y corrección de aposiciones necesitaremos la probabilidad conjunta entre clases y entidades nombradas, así como entre clases y tipos de entidad nombrada (ver sección 3). Para obtenerlas utilizaremos un estimador de máxima verosimilitud:

$$p(c, en) = \sum_t p(en, c, t) \quad (1)$$

$$p(c, t) = \sum_{en} p(en, c, t) \quad (2)$$

Donde c es la clase, en la entidad nombrada y t el tipo de la entidad nombrada.

3. Identificación y corrección de errores en dependencias de aposición

En esta fase se consideran sustantivos candidatos a gobernar la relación de aposición. Para ello, primero identificamos dependencias formadas por dos sintagmas nominales que cumplan dos premisas: el primer sintagma nominal debe contener más de un sustantivo y el núcleo del segundo debe ser una entidad nombrada. Se considerarán candidatos todos los sustantivos contenidos en el primer sintagma nominal.

Utilizaremos dos evidencias distintas para determinar cuál de los sustantivos candidatos es más adecuado:

1. Considerar la información mutua normalizada entre la entidad nombrada y el sustantivo candidato $npmi(en; c)$.

2. Considerar la información mutua normalizada entre el tipo de la entidad nombrada y el sustantivo candidato $npmi(t; c)$.

La fórmula que describe la información mutua normalizada es:

$$npmi(x; y) = \frac{pmi(x; y)}{-\log(p(x, y))} \quad (3)$$

Donde:

$$pmi(x; y) = \log \frac{p(x, y)}{p(x)p(y)} \quad (4)$$

$p(x, y)$ es la probabilidad estimada en la fase anterior, y $p(x)$ y $p(y)$ el resultado de marginalizar. Utilizar la información mutua puntual normalizada sitúa los resultados en un rango de $(-1, 1)$, siendo -1 el valor cuando no se ha observado ninguna ocurrencia conjunta y 1 el valor cuando siempre se observan conjuntamente.

4. Evaluación

En esta sección describiremos los datos utilizados en la experimentación, presentaremos los resultados de la misma y haremos un análisis de los diferentes casos encontrados.

4.1. Datos

Para esta tarea se ha utilizado la colección perteneciente a la tarea TAC KBP (Ji, Grishman, y Dang, 2011), compuesta por alrededor de 1.5 millones de documentos pertenecientes a diferentes categorías, incluyendo noticias, blogs y transcripciones de conversaciones telefónicas.

Este corpus se utiliza para generar una colección con todas las oraciones que contienen una asignación de dependencia de aposición candidata a ser corregida (ver sección 3). Encontramos un total de 284845 oraciones, para las cuales alguna de las dos evidencias sugiere otra alternativa en el 47,7% de las ocasiones.

Para evaluar el rendimiento del sistema tomamos una muestra de 300 ejemplos que etiquetamos manualmente indicando si los gobernantes escogidos por el analizador sintáctico y por las evidencias son correctos o no.

4.2. Definición del experimento

Utilizamos como línea base los resultados que obtiene el analizador sintáctico en primera instancia. Para poder comparar la aportación

a la corrección según las diferentes configuraciones del método realizaremos las siguientes pruebas:

Configuración 1: En la primera prueba medimos cuantas dependencias son correctas si se escoge el resultado de aplicar la evidencia 1 ($npmi(ne; c)$). Así comprobaremos en qué medida la entidad nombrada es indicativa de la clase.

Configuración 2: La segunda prueba medirá el acierto del sistema si se escoge el resultado de aplicar la evidencia 2 ($npmi(t; c)$). En esta prueba se comprobará en qué medida el tipo de la entidad nombrada es indicativo de la clase.

Configuración 3: En esta prueba mostramos cuantas dependencias serían correctas si combinamos las configuraciones 1 y 2 escogiendo siempre el resultado correcto: El objetivo de esta prueba es comprobar si existe la posibilidad de agregar la evidencia de las entidades nombradas y sus tipos para mejorar los resultados.

4.3. Resultados

En los 300 ejemplos encontramos 84 casos (28 %) que descartamos porque no se tratan de verdaderas aposiciones (ver sección 4.5). Trabajaremos por tanto con un conjunto de 216 ejemplos, en 150 (69,4 %) de los cuales ninguna de las configuraciones propone un sustantivo diferente al dado inicialmente por el analizador, mientras que en 66 (30,4 %) sí se propone alguna alternativa.

La línea base que hemos definido es el número de dependencias de aposición correctas del analizador sintáctico. En esta muestra la tasa de acierto es de 84,2 %, lo que supone 34 ejemplos clasificados incorrectamente.

En la configuración 1 probamos la utilidad de las entidades nombradas. Nos hemos encontrado el problema de que algunas entidades nombradas no coocurren ninguna vez con sus clases candidatas. En este caso, hemos supuesto que no se corregía la dependencia original por lo que el resultado es correcto si originalmente ya lo era. Podemos comprobar que con esta estrategia se empeoran los resultados un 0,9 % con respecto a la prueba base.

En la configuración 2 probamos la utilidad de los tipos de entidad nombrada. Con esta estrategia mejoramos ligeramente (1 %) los resultados de la prueba base.

En la configuración 3 mostramos cuantas

dependencias serían correctas si combinamos las configuraciones 1 y 2 de manera óptima. Con este sistema ideal mejoraríamos el sistema base en un 9,8 %. Este resultado motiva el desarrollo de un modelo conjunto.

En la Tabla 3 mostramos los resultados de cada configuración. Las filas se corresponden con el número de dependencias de aposición en las que se mantiene el sustantivo original (no corregidas) y el número para el que se propone un candidato alternativo (corregidas). En la configuración 1 también se muestran los casos donde no se han encontrado coocurrencias entre la entidad nombrada y ninguna de las clases (Sin información). Consideramos correctas las aposiciones que coinciden con las etiquetas manuales.

La tabla 4 muestra ejemplos de oraciones correctamente corregidas. En el primer ejemplo, los dos métodos propuestos escogen el gobernador mejor que el analizador sintáctico. En el segundo no se encuentran coocurrencias en el conocimiento antecedente entre la entidad y las clases, la información que aporta el tipo de la entidad sirve para corregir la dependencia.

En la tabla 5 podemos ver ejemplos en los que alguno de los métodos ha fallado. En el primer ejemplo el tipo de la entidad reconocido es incorrecto, lo que induce a error a la configuración 2. Es interesante que las puntuaciones de todos los candidatos son negativas, lo que nos hace pensar que detectando casos en los que las clases candidatas no son compatibles con el tipo de la entidad podemos corregir errores en la clasificación de entidades nombradas (ver sección 6)

En el segundo ejemplo la evaluación de los sustantivos según la configuración 2 falla. En oraciones así, considerar el género gramatical ayudaría a mejorar el resultado.

4.4. Casuística

En este apartado analizamos los diferentes casos encontrados que afectan a las dependencias de aposición: sintagmas nominales con múltiples sustantivos, sintagmas nominales compuestos por conjunciones, sintagmas nominales ambiguos semánticamente y casos de no aposición.

4.4.1. Sintagmas nominales con múltiples sustantivos

Generalmente, el núcleo del primer sintagma nominal debería ser el gobernante de la aposición, sin embargo, determinarlo no es un

		Correctas	Incorrectas	Total
Línea Base	Total	186 (84,2 %)	34 (15,8 %)	216
Configuración 1 <i>npmi(en; c)</i>	No corregidas	118 (54,6 %)	7 (3,2 %)	125 (57,9 %)
	No corregidas (Sin información)	46 (21,3 %)	11 (5,1 %)	57 (26,4 %)
	Corregidas	16 (7,4 %)	18 (8,3 %)	34 (15,7 %)
	Total	180 (83,3 %)	36 (16,7 %)	216
Configuración 2 <i>npmi(t; c)</i>	No corregidas	150 (69,4 %)	0 (0 %)	150 (69,4 %)
	Corregidas	34 (15,8 %)	32 (14,8 %)	66 (30,6 %)
	Total	184 (85,2 %)	32 (14,8 %)	216
Configuración 3 <i>npmi(en; c) & npm_i(t; c)</i>	No corregidas	169 (78,2 %)	0 (0 %)	169 (78,2 %)
	Corregidas	34 (15,8 %)	13 (6 %)	47 (21,8 %)
	Total	203 (94 %)	13 (6 %)	216

Tabla 3: Línea base y resultados del sistema para las tres configuraciones.

Oración	Tipo de EN	Clase	Inicial		Alternativa		
			Conf 1	Conf 2	Clase	Conf 1	Conf 2
... the <u>chief</u> of the Shin Bet security <u>agency</u> , Yuval Diskin , ...	person	agency	0.13	-0.25	chief	0.28	0.09
... the <u>head</u> of the Pellervo economic research <u>institute</u> , Paula Horne	person	institute	-1	-0.13	head	-1	0.08

Tabla 4: Aposiciones correctamente corregidas.

problema trivial. En ocasiones contiene varios sustantivos que son candidatos válidos a ser el gobernante de la relación.

En algunos casos existen compuestos nominales cuyos sustantivos tienen la misma importancia, como en los siguientes ejemplos:

... used by its domestic subsidiary airline, **Eagle Air**, ...

... by the IOC 's chief Beijing organiser, **Hein Verbruggen**, ...

En otras ocasiones el sintagma nominal incluye una oración subordinada, que también tiene un sustantivo apto para gobernar la aposición, como en el siguiente ejemplo:

Another passenger who gave only his surmane, **Chen** ...

Además también hay casos en los que el núcleo es menos discriminativo como clase semántica que el otro sustantivo, por ejemplo:

Henry is grounded by his illustrator partner, **Rudy** ...

Tanto escoger el gobernador de la dependencia de aposición como asignar la clase a la entidad nombrada depende del objetivo final del sistema. En este trabajo hemos decidido

que a efectos de evaluación el último sustantivo válido del sintagma será el gobernante correcto de la relación de aposición.

4.4.2. Sintagmas nominales compuestos por conjunciones

Este caso se produce cuando el primer sintagma nominal contiene una o varias conjunciones, con lo que hay dos sustantivos válidos para gobernar la relación de aposición. Por ejemplo:

... a prominent Jewish writer and Holocaust survivor, **Ralph Giordano** ...

En este caso consideramos que la solución ideal sería mantener la aposición en el sustantivo en el que se haya encontrado, y añadir una aposición nueva para los demás sustantivos candidatos. A efectos de puntuación del sistema, hemos considerado que escoger cualquiera de los dos sustantivos es una elección correcta. Hemos encontrado frases con esta casuística en cuatro ocasiones (1,85 %).

4.4.3. Sintagmas nominales ambiguos semánticamente

En algunas oraciones existen dos clases referidas a dos entidades diferentes, pero que pre-

Oración	Tipo de EN	Clase	Inicial		Alternativa		
			Conf 1	Conf 2	Clase	Conf 1	Conf 2
The <u>bronze medalist</u> , China's Ma Yuxi , set ...	organization (erróneo)	medalist	-1	-0.23	bronze	-1	-0.07
... leaving with his <u>father</u> and his <u>mother</u> , Linda .	person	mother	0.2	0.06	father	0.01	0.07

Tabla 5: Aposiciones mal corregidas.

cisan de conocimiento extra-lingüístico para decidir cómo están relacionadas. Por ejemplo:

... *at least one brother of another defendant, **Ali Dayeh**.*

En el ejemplo anterior existen dos clases, “brother” y “defendant”, que se refieren a dos entidades distintas, “Ali Dayeh Ali” y una entidad desconocida. Sin conocimiento externo no puede decidirse si Ali Dayeh es de clase “brother” o de clase “defendant”.

4.5. Casos de no aposición

En la inspección manual hemos encontrado múltiples ejemplos de situaciones en las que el sistema detecta una relación de aposición errónea. Corregir estos errores es por sí mismo una línea de trabajo para mejorar el análisis sintáctico, pero en este trabajo nos limitaremos a indicar qué casos hemos identificado.

- Conjunción entre dos oraciones. Por ejemplo:

*Guo Wenjun won the women's 10-meter air pistol, **Guo Jingjing** and Wu Minxia the women's synchronized 3-meter springboard, ...*

- Relación incorrecta de aposición entre un sintagma nominal y un sintagma verbal. En muchos casos, el análisis correcto sería considerar el sintagma nominal como objeto del sintagma verbal. Por ejemplo:

... *will serve as the incoming president's chief of staff, President-elect **Ma Ying-jeou's** office announced.*

- Estructuras que denotan la relación localidad-región. Por ejemplo:

... *ESA mission control in Toulouse, southwestern **France**, ...*

- Enumeraciones. Por ejemplo:

*A spa tucked away in the basement includes a large pool, a **whirlpool**, workout room, saunas, a solarium ...*

- Texto sin sentido. Ejemplo:

... *Anti-Muslim Bigots , V-for-Vendicar, fruitella, Zionism equal Racism, **The Chemical Oil Nazi**, LORD RAMA RANTER , Muslim With Mission, ...*

5. Trabajo relacionado

La relación que existe entre la ambigüedad sintáctica y el análisis semántico de un texto, así como sus aplicaciones a la desambiguación sintáctica es un tema de estudio en PLN desde hace años (Church y Patil, 1982; Resnik, 1993).

En (Ciaramita, 2007) se experimenta si añadir rasgos semánticos en un analizador sintáctico pueden mejorar su rendimiento. Para ello etiquetan las entidades nombradas de un corpus mediante un reconocedor de entidades y tratan las etiquetas semánticas como partes de la oración. En (Agirre, Baldwin, y Martinez, 2008; Agirre et al., 2011) el foco está en utilizar WordNet para generalizar palabras relacionadas, como, por ejemplo, tomar la clase “herramienta” en vez de las instancias “tijeras” o “cuchillo”, para mejorar analizadores sintácticos.

Más en la línea de nuestro trabajo, (Clark y Harrison, 2009) tratan de explotar la redundancia de instancias adquiridas mediante patrones en un corpus previo. Nosotros además combinamos clases con tipos de entidades nombradas. Así encontramos evidencia de que, por ejemplo, las clases “spokesman” y “leader” se relacionan con más frecuencia con entidades de tipo “person”, y “group” y “company” a “organization”.

Existen múltiples ontologías o diccionarios con información sobre clases semánticas que podríamos utilizar como conocimiento antecedente, ya sea creados a mano como WordNet (Miller et al., 1990) o de manera semisupervisada como DBPedia (Mendes, Jakob, y

Bizer, 2012), pero su cobertura es insuficiente, especialmente para clases abiertas, y no incorporan fácilmente conocimiento nuevo.

Para solucionar este problema se han propuesto múltiples métodos no supervisados de adquisición de clases semánticas, lo que se conoce entre otros nombres “semantic-class learning” o “semantic class induction” (Lin y Pantel, 2001). Una técnica común es procesar los textos mediante un analizador morfosintáctico y seleccionar uno o varios patrones superficiales para extraer un conjunto de clases semánticas, y posteriormente refinar los resultados obtenidos (Hearst, 1992; Kozareva, Riloff, y Hovy, 2008).

En nuestro caso, hemos optado por no restringir el número de clases, escogiendo varios patrones comunes y aplicándolos a un corpus de millones de documentos. De esta manera pretendemos aprovechar la idea de (Schubert, 2002), que sostiene que los textos contienen conocimiento general en forma de aserciones, que puede ser explotado mediante el procesamiento de grandes cantidades de texto, como en KNEXT (Schubert, 2002) o DART (Clark y Harrison, 2009). De esta manera, nos centramos más en encontrar redundancias que en perfeccionar el análisis intra-documento.

6. Conclusiones y trabajo futuro

En este artículo hemos estudiado si considerar clases semánticas podría mejorar el análisis de dependencias de aposición. Para ello evaluamos la compatibilidad semántica de los sintagmas nominales que participan en la dependencia.

Para caracterizar los sintagmas nominales utilizamos conocimiento antecedente de clases semánticas adquirido automáticamente. Este conocimiento está dividido en dos evidencias distintas, una que relaciona clases semánticas con entidades nombradas y otra que relaciona clases semánticas con tipos de entidad nombrada.

El conocimiento antecedente que relaciona las clases con las entidades no mejora los resultados, ya que la dispersión de las instancias clase-entidad hace que la probabilidad estimada de coocurrencia no sea robusta. Esto podría mejorarse sustituyendo los estimadores de máxima verosimilitud por otros como modelos basados en entropía condicional o en información mutua.

Es destacable que en el conjunto de test utilizado no hay ningún ejemplo que sea res-

pondido correctamente únicamente con la evidencia clase-entidad ya que o bien coincide con la evidencia del tipo de la entidad o con la respuesta del analizador sintáctico en primera instancia.

Al considerar la información aportada por los tipos de las entidades nombradas el problema es que algunas clases son muy dominantes (“chief”, “business”) y tienden a sobreasignarse.

La combinación óptima de la información de las entidades nombradas y de sus tipos mejora notablemente los resultados, pasando de un 84 % a un 94 % de dependencias correctas. El siguiente paso es por tanto generar un modelo más robusto que considere conjuntamente esta información, e incluso otra adicional, como el género gramatical de las entidades nombradas.

Este modelo tendría varias utilidades: (1) mejorar la corrección de aposiciones, (2) mejorar la propia asignación de clases, utilizando aposiciones corregidas para la adquisición de conocimiento antecedente y (3) mejorar el reconocimiento de entidades nombradas, identificando casos en los que las clases candidatas no sean adecuadas al tipo de la entidad nombrada.

Este trabajo está centrado en las dependencias de aposición, pero es interesante estudiar si se puede extrapolar esta técnica para resolver otros tipos, como abreviaturas o verbos copulativos, o incluso resolución de coreferencia.

Bibliografía

- Agirre, Eneko, Timothy Baldwin, y David Martinez. 2008. Improving parsing and PP attachment performance with sense information. En *Proceedings of ACL-08: HLT*, páginas 317–325, Columbus, Ohio, June. Association for Computational Linguistics.
- Agirre, Eneko, Kepa Bengoetxea, Koldo Gojenola, y Joakim Nivre. 2011. Improving dependency parsing with semantic classes. En *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2*, HLT '11, páginas 699–703, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Cabaleiro, Bernardo y Anselmo Peñas. 2012.

- Representación gráfica de documentos para extracción automática de relaciones. *Procesamiento del Lenguaje Natural*, 49(0).
- Church, Kenneth y Ramesh Patil. 1982. Coping with syntactic ambiguity or how to put the block in the box on the table. *Comput. Linguist.*, 8(3-4):139–149, Julio.
- Ciaramita, Massimiliano. 2007. Dependency parsing with second-order feature maps and annotated semantic information. En *Proc. of the 12th International Workshop on Parsing Technologies (IWPT)*.
- Clark, Peter y Phil Harrison. 2009. Large-scale extraction and use of knowledge from text. En *Proceedings of the fifth international conference on Knowledge capture, K-CAP '09*, páginas 153–160, New York, NY, USA. ACM.
- Finkel, Jenny Rose, Trond Grenager, y Christopher Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. *ACL '05*, páginas 363–370, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Hearst, Marti A. 1992. Automatic acquisition of hyponyms from large text corpora. En *Proceedings of the 14th conference on Computational linguistics - Volume 2, COLING '92*, páginas 539–545, Stroudsburg, PA, USA. Association for Computational Linguistics.
- House, Random. 2005. *Random House Kernerman Webster's College Dictionary*.
- Ji, Heng, Ralph Grishman, y Hoa Dang. 2011. Overview of the TAC2011 Knowledge Base Population Track. En *TAC 2011 Proceedings Papers*.
- Klein, Dan y Christopher D. Manning. 2003. Accurate unlexicalized parsing. En *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1, ACL '03*, páginas 423–430, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Kozareva, Zornitsa, Ellen Riloff, y Eduard Hovy. 2008. Semantic class learning from the web with hyponym pattern linkage graphs. En *Proceedings of ACL-08: HLT*, páginas 1048–1056, Columbus, Ohio, June. Association for Computational Linguistics.
- Lee, Heeyoung, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, y Dan Jurafsky. 2011. Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task. *CONLL Shared Task '11*, páginas 28–34, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Lin, Dekang y Patrick Pantel. 2001. Induction of semantic classes from natural language text. En *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '01*, páginas 317–322, New York, NY, USA. ACM.
- Mendes, Pablo N., Max Jakob, y Christian Bizer. 2012. Dbpedia for nlp: A multilingual cross-domain knowledge base. En *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, Mayo.
- Miller, George A, Richard Beckwith, Christiane Fellbaum, Derek Gross, y Katherine J Miller. 1990. Introduction to wordnet: An on-line lexical database*. *International journal of lexicography*, 3(4):235–244.
- Resnik, Philip. 1993. Semantic classes and syntactic ambiguity. En *Proceedings of the workshop on Human Language Technology, HLT '93*, páginas 278–283, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Schubert, Lenhart. 2002. Can we derive general world knowledge from texts? En *IN PROC. HLT 2002*, páginas 24–27.
- Toutanova, Kristina y Christopher D. Manning. 2000. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. *EMNLP '00*, páginas 63–70, Stroudsburg, PA, USA. Association for Computational Linguistics.