Creado el 27/01/2014
Última versión: 27/01/2014
Antoni Oliver, Universidat Oberta de Catalunya
Comunidad moocTLH
Esta obra está bajo una licencia de Creative Commons Reconocimiento-CompartirIgual 3.0 Unported.

Enfoques y recursos básicos de las TLH/APROXIMACIONES, RECURSOS Y HERRAMIENTAS DE LAS TLH

Basado en corpus y basado en conocimiento (2/2)

Visto qué es una aproximación basada en conocimiento y qué es una aproximación basada en corpus y los recursos más habituales que se utilizan en PLN, pasamos ahora a matizar y concretarlo en una tarea concreta.

Ejemplo práctico

En este apartado vamos a ver como se puede aplicar todo esto que hemos visto (tanto las técnicas basadas en corpus como las técnicas basadas en conocimiento) en una tarea denominada **desambiguación de sentidos** (*Word Sense Disambiguation*).

Consideremos estas dos oraciones:

A. Juan quería comprase unos zapatos nuevos y fue al **banco** a retirar dinero de su cuenta.

B. Después de pasear por muchas tiendas se encontraba muy fatigado y se fue a sentar en un **banco** para descansar un rato.

En estas dos oraciones la palabra **banco** se refiere a dos significados diferentes: entidad bancaria o mueble urbano o de jardín que sirve para sentarse (en realidad la palabra *banco* puede tener muchos más sentidos). Vamos a ver algunas de las técnicas que podemos aplicar para determinar.

Para empezar veamos cómo podemos aplicar **corpus** para resolver esta tarea. Si disponemos únicamente de **corpus sin anotar** la resolución será complicada, ya que tampoco sabremos a qué sentido se refieren las ocurrencias de esta palabra en el corpus. Si el corpus es **temático** sobre un tema relacionado con la banca, podremos presuponer que muchas de las apariciones se referirán al sentido bancario.

Si tomamos el corpus del Banco Central Europeo en castellano, que tiene 116.113 oraciones, podemos constatar que en:

- en 101 oraciones aparecen tanto las palabras **banco** como **dinero**
- en 504 oraciones aparecen tanto las palabras **banco** como **cuenta**
- en 9 oraciones aparecen tanto las palabras **banco** como **retirar**

en cambio

- en 0 oraciones aparecen tanto **banco** como **madera**
- en 0 oraciones aparece tanto **banco** como **descansar**

De aquí podemos deducir que que en la oración A el sentido de la palabra banco es el de entidad bancaria.

Si disponemos de un corpus etiquetado semánticamente podremos realizar deducciones similares. Desgraciadamente, existen pocos corpus etiquetados semánticamente para el español. Simularemos que disponemos de un corpus de este estilo, que contiene 5 oraciones correspondientes al sentido bancario y 5 al sentido de mueble (subrayamos las palabras que también se encuentran el la frase A y B

<u>Sentido de entidad bancaria:</u>
Aunque la cuenta esté vacía, el **banco** sigue estando obligado a pagar la factura a final de mes.
Si el **banco**, la cooperativa de ahorros y crédito paga intereses en una ETA, éste se reserva el derecho de solicitar un aviso escrito con siete (7) días de anticipación a la fecha de retiro de los fondos.
Antes de trabajar en mercadeo en red, Shawn Herrick fue un cajero de **banco** y un estudiante universitario.
Pero cuando hay un riesgo moral, el **banco** puede elegir un tipo de interés demasiado elevado.
Esto ayuda a facilitar el acceso a pensiones y <u>cuentas</u> en el **banco** de la persona desaparecida.

<u>Sentido de mueble:</u>
De vez en cuando, todos necesitamos encontrar algo que nos sirva de alto en el camino, como un **banco** en el que podamos <u>sentarnos</u> un rato y pararnos a <u>descansar</u>.
El área cuenta con rejas para mayor privacidad y un **banco** para <u>sentarse</u> a <u>descansar</u> mientras se reflexiona.
Quizás toma el transporte público, ¿tiene la parada de autobús un **banco** donde una persona mayor pueda <u>sentarse</u> mientras espera?
En el camino se van encontrando **bancos** que invitan a <u>sentarse</u> y contemplar el entorno.
Un café especial, comprarse una flor o simplemente <u>sentarse</u> en un **banco** con los ojos cerrados a escuchar el canto de los pájaros puede ser una excelente forma de desconectar.

Estas coocurrencias nos pueden dar pistas para determinar el sentido de la palabra **banco** en las oraciones A i B.

Veamos ahora la entrada de **banco** en el diccionario de la Real Academia de la Lengua Española:

banco.

(Del fr. ant. *bank,* y este del germ. **banki*).

1. m. Asiento, con respaldo o sin él, en que pueden sentarse varias personas.

2. m. Madero grueso escuadrado que se coloca horizontalmente sobre cuatro pies y sirve como de mesa para muchas labores de los carpinteros, cerrajeros, herradores y otros artesanos.

3. m. **cama** (‖ del freno). U. m. en pl.

4. m. En los mares, ríos y lagos navegables, bajo que se prolonga en una gran extensión.

5. m. Conjunto de peces que van juntos en gran número.

6. m. Establecimiento público de crédito, constituido en sociedad por acciones.

7. m. Establecimiento médico donde se conservan y almacenan órganos, tejidos o líquidos fisiológicos humanos para cubrir necesidades quirúrgicas, de investigación, etc.*Banco de ojos, de sangre*

8. m. *Arq.* **sotabanco** (‖ piso habitable).

9. m. *Geol.* Estrato de gran espesor.

10. m. *Ingen.* Macizo de mineral que presenta dos caras descubiertas, una horizontal superior y otra vertical.

11. m. *Ven.* Extensión de terreno con vegetación arbórea que sobresale en la llanura.

12. m. p. us. Persona que cambia moneda.

Vemos que nos ofrece hasta 12 posibles significados. Si nos centramos en los dos significados de este ejemplo veremos que las definiciones son:

1. m. Asiento, con respaldo o sin él, en que pueden sentarse varias personas.

6. m. Establecimiento público de crédito, constituido en sociedad por acciones.

En el primer significado tenemos también una palabra coincidente con nuestra frase de ejemplo, lo que nos permetiría decidir que se utiliza en la oración con este significado. Para el banco como entidad financiera no tenemos ninguna palabra coincidente.

Veamos ahora qué nos cuenta la Wikipedia sobre bancos.

Banco

Para otros usos de este término, véase Banco (desambiguación).

Este artículo o sección necesita **referencias** que aparezcan en una **publicación acreditada**, como revistas especializadas, monografías, prensa diaria o páginas de Internet fidedignas.
Puedes añadirlas **así** o avisar al autor principal del artículo en su página de discusión pegando: `{{subst:Aviso referencias|Banco}} ~~~~`

Un **banco** es una empresa financiera que se encarga de captar recursos en la forma de depósitos, y prestar dinero, así como la prestación de servicios financieros. La **banca**, o el **sistema bancario**, es el conjunto de entidades o instituciones que, dentro de una economía determinada, prestan el servicio de banco. La internalización y la globalización promueven la creación de una Banca universal. Al igual que la palabra española "banco", la palabra griega (trá·pe·za) que se traduce como banco significa literalmente "mesa". En un contexto financiero, como por ejemplo el trabajo de los cambistas, se refiere al mostrador para la transacción de dinero.[1]

Sede del Banco Mundial en Washington D. C.

Índice [ocultar]

1 Historia de la actividad bancaria
2 Operaciones bancarias típicas
2.1 Operaciones pasivas
2.2 Operaciones activas
2.2.1 Encaje bancario
2.2.2 Préstamos y créditos
2.2.2.1 Clasificaciones de préstamos
2.2.3 Cuentas de crédito
2.2.4 El descuento de efectos
2.2.5 Cartera de valores
2.2.6 Cesiones temporales de activos
2.2.7 Margen de intermediación
2.2.8 Prestación de servicios
3 Clases de banco
4 Banca por país
5 Institución financiera
6 Véase también
7 Referencias
8 Enlaces externos

Como significado principal nos ofrece el de entidad financiera, aunque en la primera línea ya nos advierte de la existencia de otros usos y nos dirige a la página de desambiguación.

Banco (desambiguación)

Banco puede hacer referencia a:

- **Banco**, mueble largo, de estructura sencilla, en el que pueden sentarse varias personas a la vez.
  - **Banco azul**, conjunto de escaños o bancada que en la tradición parlamentaria española suele reservarse al Gobierno
- **Banco de pesas**, estructura destinada al levantamiento de pesas.
- **Banco de tiempo**, sistema de intercambio de servicios por tiempo.
- **Banco** o *predela*, en arquitectura, es la parte inferior de un retablo.
- **El Banco (centro clandestino de detención)** Centro clandestino de detención que funcionó entre fines de 1977 y mediados de 1978 en la actual Brigada Femenina XIV de la Policía de la Provincia de Buenos Aires.

Índice [ocultar]

1 Naturaleza y Botánica
2 Finanzas
3 Organizaciones de almacenamiento y distribución
3.1 Medicina y Veterinaria
4 Geografía
4.1 Arrecifes
5 Varios
6 Enlaces externos

Naturaleza y Botánica [editar · editar código]

- **Banco**, conjunto de peces.
- **Banco**,tierra junto a un cuerpo de agua.
- **Banco**, gran banco de pesca de la plataforma continental norteamericana en el océano Atlántico
- **Banco**, por acumulación a lo largo del litoral o en el lecho de un río.
- **Banco**, es un arrecife, en su mayoría sumergido, localizado en el sector noreste del Canal de Mozambique.

Finanzas [editar · editar código]

- **Banco**, institución financiera.
  - **Banco de desarrollo**, entidad que financia proyectos para promover el desarrollo económico.
  - **Banco de inversión**, organización especializada en obtenerel dinero o los instrumentos para realizar determinada inversión.
  - **Banco malo**, entidad financiera que ayuda a limpiar los activos tóxicos.
  - **Banco comunal**, organización sin ánimo de lucro otorgue préstamos a un grupo de personas.
  - **Banco zombi**, institución financiera con un valor económico neto menor que cero.

Veamos ahora cuantas veces aparecen las palabras que acompañan a banco en nuestra frase A en las páginas de la Wikipedia dedicadas a banco como entidad bancaria y banco como mueble:

A. Juan quería comprase unos zapatos nuevos y fue al **banco** a retirar dinero de su cuenta.

	entrada de banco como entidad financiera en la Wikipedia	entrada de banco como mueble en la Wikipedia
comprar	0	0
zapatos	0	0
retirar	0	0
dinero	20	0
cuenta	12	0

Vemos que algunas palabras como comprar o zapatos están en la frase de manera accidental y nada tienen que ver con la palabra banco. En la tabla se observa que las palabras que co-ocurren con banco en la oración A y aparecen con mayor frecuencia en la Wikipedia tienen que ver con el sentido de entidad financiera (dinero aparece 20 veces y cuenta 12). En cambio ninguna palabra de la frase A aparece en la entrada de la Wikipedia correspondiente a banco como mueble. Podemos deducir de manera clara que en la frase A banco tiene el significado de entidad financiera.

Observemos qué ocurre en la frase B:

B. Después de pasear por muchas tiendas se encontraba muy fatigado y se fue a sentar en un **banco** para descansar un rato.

	entrada de banco como entidad financiera en la Wikipedia	entrada de banco como mueble en la Wikipedia
pasear	0	0
tienda	1	0
encontrar	0	0
fatigado	0	0
sentar	0	1

descansar	0	0
rato	0	0

En este caso la cosa no está tan clara. Para acabarnos de decidir tenemos que tener en cuenta que la entrada de la WIkipedia de banco como entidad financiera es mucho más grande (2810 palabras) que la de banco como mueble (con sólo 331 palabras). En este caso podemos dar un peso mucho mayor a la aparición de sentar en la entrada de la WIkipedia como mueble para decidir que este es el sentido que le corresponde a banco en la frase b.

Observemos ahora qué información nos proporciona WordNet sobre estos dos sentidos de la palabra **banco**.

08420278-n **banco** una institución financiera que acepta depósitos y canaliza el dinero en actividades de préstamo; ese banco tiene la hipoteca de mi casa; cobrar un cheque en el banco
1 hiperónimo: 08054721-n institución financiera, organización financiera
diversos hipónimos: 08234628-n asociación de crédito, cooperativa de crédito; 08422524-n entidad de ahorros; 08419163-n banco comercial, banco de negocios; banco financiero; banco mercantil; etc.
algunos conceptos relacionados como 02310855-v depositar, ingresar; 02343374-v operar en banca

02828884-n **banco** asiento largo para más de una persona
1 hiperónimo 04161981-n asiento
diversos hipónimos 0278838-n banqueta, banquillo, taburete; 04002371-n reclinatorio; 04177920-n escaño

Para utilizar WordNet en la tarea de desambiguación de sentidos podríamos utilizar una estrategia similar a la utilizada con las definiciones del diccionario, utilizando las glosas (o definiciones) de WordNet. Pero WordNet nos proporciona un mecanismo mucho más potente mediante las relaciones que ofrece entre los conceptos. Estas relaciones nos permiten calcular distancias semánticas entre diferentes conceptos. Una de las medidas más sencillas de distancia semántica es la llamada *distancia de caminos* (path similarity) que se basa en el cálculo del camino más corto entre dos sentidos en la taxonomía es_un (is_a) (hiperonimia/hiponimia). Así, por ejemplo, podemos calcular las distancias semánticas entre los conceptos **banco** (como entidad financiera 08420278-n) y **dinero** (13384557-n) y **cuenta** (13929037-n) y de banco (como mueble 02828184-n) estos mismos conceptos. Veamos los resultados en la siguiente tabla:

	banco (entidad financiera) 08420278-n	banco (mueble) 02828184-n
dinero - 13384557-n	0.0833	0.0625
cuenta - 13929037-n	0.0909	0.0666

Estas distancias nos permetirían deducir que el sentido de banco en la oración A es el

correspondiente a entidad financiera.

Veamos por último la información que contiene ConceptNet 5 sobre banco (en español)

banco	TranslationOf	bench
banco	TranslationOf	school_of_fish
banco	TranslationOf	shoal, group of fish
banco	TranslationOf	bank (institution)
banco	TranslationOf	bank
banco	ConceptuallyRelatedTo	bancario
banco	ConceptuallyRelatedTo	bancada
banco	ConceptuallyRelatedTo	banqueta
banco	ConceptuallyRelatedTo	banca
banco	ConceptuallyRelatedTo	banquito
banco	ConceptuallyRelatedTo	banquillo
banco	Synonym	banca

Esta información no está desembiguada por sentidos por lo que no es directamente aplicable a la tarea del ejemplo.

Conclusiones

En este tema hemos visto dos aproximaciones diferenciadas para la resolución de tareas lingüísticas: las basadas en corpus y las basadas en conocimiento. Aunque existe una diferencia de base entre estas aproximaciones, no siempre se puede marcar una frontera clara entre una y otra, ya que la utilización de un corpus ya supone en sí una fuente de conocimiento y si etiquetamos con muchísima información un corpus lingüístico, su uso se aproxima más a las aproximaciones basadas en conocimiento.

En el último apartado hemos esbozado una serie de técnicas para tratar la tarea de desambiguación de sentidos (*Word Sense Disambiguation)*