

Procesamiento de Lenguaje Natural

TEMA 4

Análisis semántico

Enrique Alfonseca

Pilar Rodríguez

Índice

- **Desambiguación de sentidos**
 - **Análisis semántico**
 - **Análisis pragmático**
- Introducción
 - Clasificador Bayes
 - Teoría de la Información
 - WSD con diccionarios
 - WSD con tesauros
 - WSD con densidad conceptual
 - WSD con topic signatures
 - Combinaciones
 - WSD no supervisada
 - Otros métodos
 - **Introducción**
 - **Cálculo lambda**
 - **HPSG**
 - **Otros formalismos**
 - **Introducción**

WSD

Introducción

Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

Introducción

La desambiguación del sentido de las palabras consiste en descubrir con qué sentido se está usando una palabra en un contexto dado.



*Nos sentamos en el banco toda la tarde.
Este banco te da más intereses que aquel.*

WSD

Introducción

Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis

semántico

Pragmática

Introducción (II)

En general, tendremos una palabra w , en un contexto c , y se trata de descubrir, entre una serie de sentidos $\{s_1, \dots, s_k, \dots, s_K\}$, con qué sentido se está utilizando.

Por tanto, es un problema de clasificación.

WSD

Introducción

Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis

semántico

Pragmática

Introducción (III)

Problema: Distintos diccionarios pueden considerar las mismas acepciones de la palabra como un solo sentido, o como dos.

En general, los sentidos de las palabras son difusos.

Por ejemplo:

Entré por la puerta. Cerré la puerta.



Espacio en la pared que permite el acceso a un recinto.

Barrera giratoria o deslizante que cierra el acceso a un recinto.



Entré por la puerta y la cerré después

WSD

Introducción

Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

Introducción (III)

Hay sentidos en los que todo el mundo está de acuerdo en que son diferentes (e.g. **banco**).

En estos casos, puede ser útil desambiguar las palabras, para

- Recuperación de información.
- Traducción automática.
- Generación de resúmenes multi-documento.
- Búsqueda de respuestas.
- etc.

WSD

Introducción
Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

Desambiguación con clasificadores Bayesianos

Utilizan el contexto de la palabra para deducir su sentido.

Dada una palabra w en un contexto $c = \{v_1, \dots, v_j, \dots, v_J\}$, con posibles sentidos $\{s_1, \dots, s_k, \dots, s_K\}$,

- Seleccionar $s' \iff P(s'|c) > P(s_k|c) \forall s_k \neq s'$.

WSD

Introducción
Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

Clasificadores Bayesianos (II)

Estimación de las probabilidades $P(s_k|c)$.

Aplicando el teorema de Bayes,

$$P(s_k|c) = \frac{P(c|s_k)P(s_k)}{P(c)}$$

Por tanto,

$$\begin{aligned} s' &= \operatorname{argmax}_{s_k} P(s_k|c) \\ &= \operatorname{argmax}_{s_k} P(c|s_k)P(s_k) \\ &= \operatorname{argmax}_{s_k} [\log P(c|s_k) + \log P(s_k)] \end{aligned} \quad (1)$$

WSD

Introducción
Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

Clasificadores Bayesianos (III)

Estimación de las probabilidades $P(c|s_k)$.

Hipótesis de Bayes Naïve: las palabras que ocurren en el contexto de w son todas condicionalmente independientes.

$$P(c|s_k) = P(\{v_j|v_j \text{ in } c\}|s_k) = \prod_{v_j \text{ in } c} P(v_j|s_k)$$

Por tanto,

$$\begin{aligned} s' &= \operatorname{argmax}_{s_k} [\log P(c|s_k) + \log P(s_k)] \\ &= \operatorname{argmax}_{s_k} [\log P(s_k) + \sum_{v_j \text{ in } c} \log P(v_j|s_k)] \quad (2) \end{aligned}$$

WSD

Introducción
Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

Clasificadores Bayesianos (IV)

Estimación de las probabilidades $P(v_j|s_k), P(s_k)$.

Utilizando el corpus de entrenamiento (donde la palabra w está desambiguada), el estimador de máxima *likelihood*.

$$P(v_j|s_k) = \frac{C(v_j|s_k)}{C(s_k)}$$

$$P(s_k) = \frac{C(s_k)}{C(w)}$$

WSD

Introducción
Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

Clasificadores Bayesianos (V) (Gale, Church and Yarowsky, 1992)

Entrenamiento:

1. Estimar, utilizando el corpus, $P(s_k) \forall k$
2. Estimar, utilizando el corpus, $P(v_j | s_k) \forall j, k$

Desambiguación:

1. Para todos los posibles sentidos s_k ,
 - (a) $score(s_k) = \log P(s_k)$
 - (b) Para todas las palabras v_j en c ,
$$score(s_k) + \log P(v_j | s_k)$$
2. Escoger $argmax_{s_k} score(s_k)$

Precisión medida con seis palabras difíciles: 90%

WSD

Introducción
Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

Clasificador basado en Teoría de la Información

Se trata de agrupar las características utilizadas para la clasificación de manera que nos den la máxima información posible para distinguir los sentidos.

Dada una serie de posibles características, se trata de descubrir cuales proporcionan la máxima información para desambiguar la palabra.

Ejemplos:

Palabra	Indicador	Ejemplos
<i>prendre</i>	objeto directo	<i>measure</i> → <i>to take</i> <i>décision</i> → <i>to make</i>
<i>cent</i>	palabra previa	<i>per</i> → % <i>número</i> → <i>céntimo</i>

WSD

*Introducción
Bayesiana*

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

Teoría de la Información (II)

Por ejemplo, supongamos que queremos desambiguar una palabra w

(e.g. *prendre*)

según la palabra inglesa en que es traducida en un corpus bilingüe

(*take, make, rise, speak*)

utilizando como característica de la clasificación su objeto directo

(*measure, note, exemple, décision, parole*)

WSD

Introducción
Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

Teoría de la Información (III)

Algoritmo flip-flop. (*Brown et al., 1991*)

Tenemos dos conjuntos de características:

$$T = \{t_1, \dots, t_m\} = \{take, make, rise, speak\}$$

$$X = \{x_1, \dots, x_n\} = \{measure, note, exemple, decision, parole\}$$

1. Comenzar con una partición aleatoria de T ,

$$P = \{P_1, P_2\}$$

2. Mientras se produzca una mejora en I ,

- (a) Encontrar una partición de X , $Q = \{Q_1, Q_2\}$ que maximice $I(P; Q)$

- (b) Encontrar $P = \{P_1, P_2\}$ que maximice $I(P; Q)$

WSD

Introducción

Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis

semántico

Pragmática

Teoría de la Información (IV)

En el ejemplo, se termina (independientemente de la asignación inicial aleatoria) con

$$P_1 = \{take\}$$

$$P_2 = \{make, rise, speak\}$$

$$Q_1 = \{measure, note, exemple\}$$

$$Q_2 = \{decision, parole\}$$

WSD

*Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros*

Análisis
semántico

Pragmática

WSD con diccionarios

Es de esperar que las definiciones de los distintos sentidos de una palabra estén altamente relacionadas semánticamente con cada uno de esos sentidos.

Consideremos el siguiente ejemplo para *banco*:

banco Institución financiera que acepta depósitos y canaliza el dinero a actividades de préstamos.

banco Un asiento largo para más de una persona.

■ *Metí dinero en el banco*

Todos los asientos de esta sala son bancos

WSD

Introducción
Bayesiana
T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

WSD con diccionarios (II)

Algoritmo: medir el solapamiento entre el contexto de w y cada definición del diccionario.

Dado el contexto c ,

1. Para cada sentido s_k de w ,

(a) $\text{score}(s_k) = \text{overlap}(D_k, \cup_{(v_j \text{ in } c)} v_j)$

2. Escoger $\text{argmax}_{s_k} \text{score}(s_k)$

Problema: las definiciones suelen ser cortas y tienen poca evidencia.

Nos sentamos en el banco

WSD

Introducción
Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

WSD con diccionarios (III)

(Lesk, 1986)

Solución: medir el solapamiento entre el contexto de w , y la definición de cada palabra que aparece en las palabras de D_k .

Dado el contexto c ,

1. Para cada sentido s_k de w ,
 - (a) Para cada palabra v_j de D_k , con sentidos

s_{j_1}, \dots, s_{j_L} ,

• $E_{v_j} = \cup_{j_l} D_{j_l}$

(a) $\text{score}(s_k) = \text{overlap}(D_k, \cup_{(v_j \text{ in } c)} E_{v_j})$

2. Escoger $\text{argmax}_{s_k} \text{score}(s_k)$

WSD

*Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros*

Análisis
semántico

Pragmática

WSD con diccionarios (IV)

- Las palabras en las definiciones no se desambiguan.
- El corpus de entrenamiento (las definiciones) sigue siendo más pequeño que con los métodos anteriores.
 - ⇒ Muchas palabras se quedan sin desambiguar
- Precisión medida: 50-70%.
- Posibles mejoras, seleccionando definiciones para v_l , etc.

WSD

*Introducción
Bayesiana
T. I.
Diccionarios
Tesoros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros*

Análisis
semántico

Pragmática

WSD con tesauros

Disponemos de un tesoro, como Roget's, en el cual las palabras están clasificadas en categorías (e.g. *Matemáticas, Leyes, etc.*)

Dada una palabra w ,

- Sus sentidos $s_1, \dots, s_k, \dots, s_K$
- corresponden a las categorías $t_1, \dots, t_k, \dots, t_K$ a las que pertenece w .

WSD

Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros

Análisis
semántico

Pragmática

WSD con tesauros (II)

Algoritmo de (Walker, 1987)

Dado el contexto c

1. Para cada sentido s_k de w ,
 - $\text{score}(s_k) = \sum_{v_j \text{ in } c} \delta(t(s_k), v_j)$
2. Escoger $\text{argmax}_{s_k} \text{score}(s_k)$

$\delta(t(s_k), v_j) = 1$ si el dominio del sentido s_k coincide con alguno de los dominios de v_j .

Problemas:

1. Una clasificación por dominios no es siempre apropiada.
2. Muchas palabras específicas de dominio, y los nombres propios, no aparecerán en el tesoro.

WSD

Introducción
Bayesiana
T. I.
Diccionarios
Tesoros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros

Análisis
semántico

Pragmática

WSD con tesauros (III)

Algoritmo de (Yarowsky, 1992)

Para cada contexto c_i en el corpus

1. Para cada categoría t_l del tesoro,
 - $\text{score}(c_i, t_l) = \log \frac{P(c_i|t_l)P(t_l)}{P(c_i)}$
2. Escoger $\{t_l | \text{score}(c_i, t_l) > \alpha\}$

Para cada v_j del vocabulario, $V_j = \{c | v_j \text{ in } c\}$

Para cada categoría $t_l, T_l = \{c | t_l \in t(c)\}$

Para toda $v_j, t_l, P(v_j|t_l) = \frac{|V_j \cap T_l|}{\sum_j |V_j \cap T_l|}$

Para cada $t_l, P(t_l) = \frac{\sum_j |V_j \cap T_l|}{\sum_l \sum_j |V_j \cap T_l|}$

Desambiguación: Seleccionar s_k que maximice

$$\log P(t(s_k)) + \sum_{v_j \text{ in } c} \log P(v_j|t(s_k))$$

WSD

*Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros*

Análisis
semántico

Pragmática

WSD con tesauros (IV)

Resultados

- Muy util (90-100%) si los sentidos de una palabra están altamente determinados por el tema de interés.
 Por ejemplo, **bass** (instrumento musical, pescado)
- No es util (30%), en cambio, si un sentido de una palabra no es característico de ningún dominio:
 Por ejemplo, **interest** (curiosidad, *ventaja, financiero, acción de bolsa)

WSD

*Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros*

Análisis
semántico

Pragmática

Densidad Conceptual

Agirre y Rigau (1995) proponen utilizar redes léxico-semánticas para desambiguación de sentidos.

Estructuradas como grafos, donde

- Los nodos representan sentidos (conceptos)
- Los arcos representan relaciones semánticas:
 1. La relación de sinonimia es implícita entre las palabras que están en el mismo nodo.
 2. Generalización (hiperonimia) o especificación (hiponimia)
 3. Relación es-parte-de (holonimia) o tiene-parte (meronimia).
 4. Relación de antonimia.
 5. Relación causa-efecto
 6. etc...

WSD

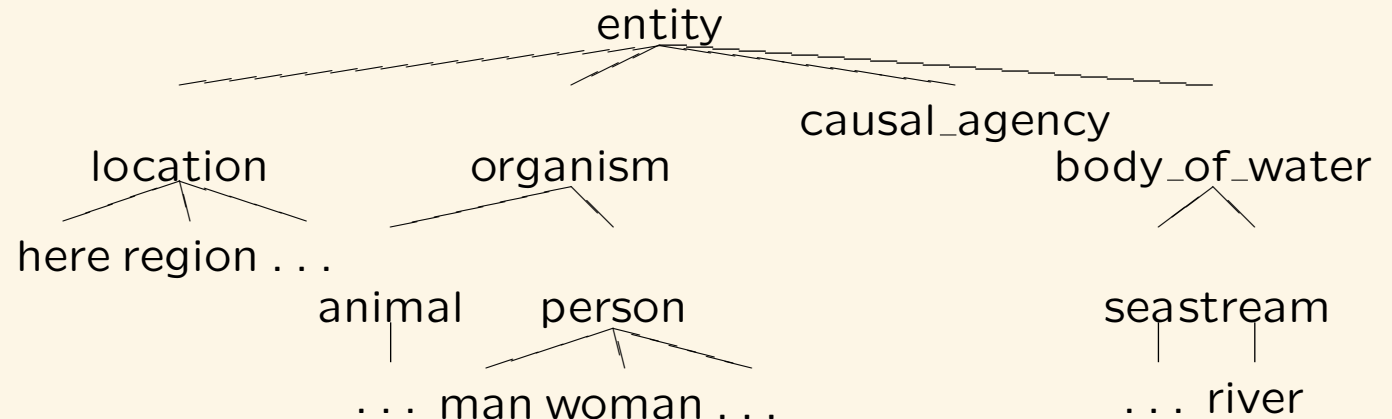
Introducción
 Bayesiana
 T. I.
 Diccionarios
 Tesoros
Dens. conc.
 Signaturas
 Combinación
 No supervis.
 Otros

Análisis
 semántico

Pragmática

Densidad Conceptual (II)

Ejemplo con la relación de hiperonimia



Si añadimos una relación de hiponimia entre *entidad causal* y *persona*, vemos que la estructura no es de árbol, sino de grafo.

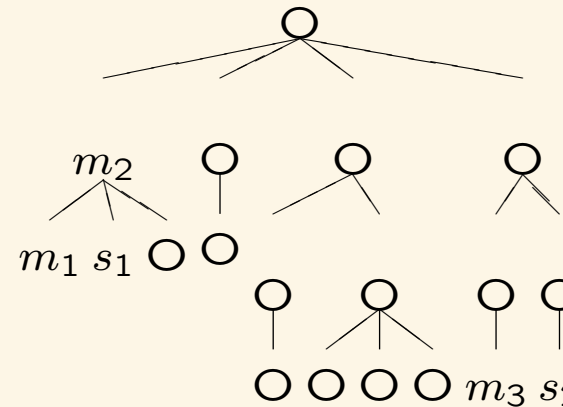
WSD

Introducción
 Bayesiana
 T. I.
 Diccionarios
 Tesoros
Dens. conc.
 Signaturas
 Combinación
 No supervis.
 Otros

Análisis
 semántico

Pragmática

Densidad Conceptual (III)



El cazador encontró plantas como la mimosa y el llantén

$w = mimosa$

s_1 , hipónimo de *planta* (m_2), al igual que *llantén* (m_1).

s_2 , hipónimo de *persona*, al igual que *cazador* (m_3).

WSD

Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros

Análisis
semántico

Pragmática

Densidad Conceptual (IV)

Cálculo de la Densidad Conceptual

Dada una palabra w , con sentidos $\{s_1, \dots, s_k, \dots, s_K\}$,
en un contexto $c = \{v_1, \dots, v_j, \dots, v_J\}$,

1. Marcar cada nodo del grafo que contiene $v_j \forall j$
2. Dividir el grafo de hiperonimia en subjerarquías J_k , de modo que cada una contenga exactamente un sentido s_k .
3. Para cada subjerarquía J_k ,
 - (a) Calcular su profundidad media h
 - (b) Calcular el número medio de hipónimos por nodo $nhyp$.
 - (c) Contar el número m_k de conceptos marcados en J_k .

$$(d) CD(s_k, m_k) = \frac{\sum_{i=0}^{m_k-1} nhyp^i}{descendientes(s_k)}$$

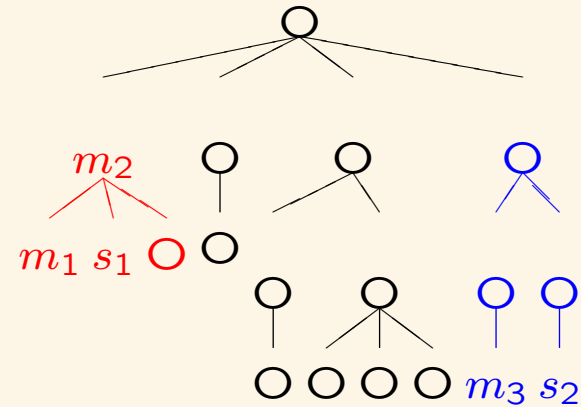
WSD

Introducción
 Bayesiana
 T. I.
 Diccionarios
 Tesoros
Dens. conc.
 Signaturas
 Combinación
 No supervis.
 Otros

Análisis
 semántico

Pragmática

Densidad Conceptual (V)



$$h_1 = 2$$

$$nhyp_1 = 3$$

$$m_1 = 2$$

$$CD_1 = \frac{4}{4} = 1$$

$$h_2 = 3$$

$$nhyp_2 = \frac{4}{3}$$

$$m_2 = 1$$

$$CD_2 = \frac{2.333}{5} = 0.467$$

WSD

*Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros*

Análisis
semántico

Pragmática

Densidad Conceptual (VI)

- Precisión del 53,9% en las palabras polisémicas de un texto.
- Fácilmente extensible a otras relaciones semánticas de la red léxico-semántica (meronimia, homonimia, etc.)

Desventajas:

- Altamente dependiente de la estructura de la red semántica.

WSD

*Introducción
Bayesiana*

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

Topic Signatures

Es una lista de las palabras v_j que ocurren en los contextos de una palabra w con un cierto sentido s_k , y pesos asignados según su importancia relativa para discriminar los sentidos.

Por ejemplo, consideremos los tres sentidos de la palabra *church*:

- Grupo de cristianos, grupo de personas que profesa una doctrina cristiana (e.g. Protestant Church).
- Edificio para el culto (especialmente cristiano).
- Servicio realizado en la iglesia.

WSD

*Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros*

Análisis
semántico

Pragmática

Topic Signatures (II)

I. Grupo de cristianos, grupo de personas que profesa una doctrina cristiana.

Palabra	Peso
African	59.51021957397461
Africa	55.09562301635742
Church	45.46275329589844
Protestant	26.547950744628906
Orthodox	19.718162536621094
religion	19.21417808532715
Catholic	18.881362915039062
leader	16.833858489990234
missionary	14.921016693115234
group	14.277392387390137
Christianity	13.294868469238281
mission	13.220355033874512
establish	11.054662704467773
recognize	10.822933197021484
culture	10.725310325622559

WSD

*Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros*

Análisis
semántico

Pragmática

Topic Signatures (III)

II. Edificio para el culto (especialmente cristiano).

Palabra	Peso
building	163.04629516601562
abbey	72.738037109375
St	56.681541442871094
Romanesque	51.661739349365234
inst	42.67708206176758
cathedral	42.16020202636719
parish	41.54650115966797
beautiful	31.534700393676758
old	30.53959083557129
Cathedral	28.35443687438965
basilica	27.967018127441406
reb	26.953947067260742
small	25.231870651245117
cross	24.15271759033203
Gothic	22.77438735961914
tower	22.77438735961914
portal	22.461620330810547

WSD

*Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros*

Análisis
semántico

Pragmática

III. Servicio realizado en la iglesia.

Palabra	Peso
service	301.1626892089844
music	141.82130432128906
marriage	87.49576568603516
Lutheran	69.7547378540039
Religious	43.224082946777344
attend	40.037288665771484
marry	35.330345153808594
divorce	29.0540828704834
appropriate	26.180206298828125
use	24.897682189941406
ceremony	23.636884689331055
school	19.30716896057129
Drumcree	18.18027687072754
instrumental	18.18027687072754
person	17.668134689331055
available	17.05433464050293
meet	16.69251251220703
student	16.400875091552734
lifetime	16.160245895385742
annual	16.160245895385742
attendance	16.03090476989746
Sunday	15.445367813110352

WSD

*Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros*

Análisis
semántico

Pragmática

Topic Signatures (V)

Para cada sentido s_k de la palabra w ,

1.
 - Localizar un corpus con etiquetación de los sentidos.
 - Realizar búsquedas en Internet poniendo palabras contextuales en las consultas.
2. Para cada sentido s_k de la palabra, recoger todas las palabras que aparecen en el contexto de w con sentido s_k
3. Calcular el peso de v_l como evidencia contextual de que estamos en el contexto de s_k .

WSD

Introducción
 Bayesiana
 T. I.
 Diccionarios
 Tesoros
 Dens. conc.
Signaturas
 Combinación
 No supervis.
 Otros

Análisis
 semántico

Pragmática

Topic Signatures (VI)

Agirre et al., (2000) utiliza la función de peso χ^2 :
 $\{ (v_1, freq_{k,1}), \dots, (v_L, freq_{k,n}) \}$

$$m_{k,l} = \frac{\sum_k freq_{k,l} \cdot \sum_l freq_{k,l}}{\sum_{k,l} freq_{k,l}} \quad (3)$$

The weight for word v_l as contextual cue in favour of the sense s_k is

$$w_{k,l} = \begin{cases} \frac{(freq_{k,l} - m_{k,l})^2}{m_{k,l}}, & \text{if } freq_{k,l} > m_{k,l} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Es posible utilizar otras funciones de peso.

WSD

Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros

Análisis
semántico

Pragmática

Topic Signatures (VII)

Desambiguación:

Dada la palabra w , con sentidos $s_1, \dots, s_k, \dots, s_K$, en el contexto $c = \{v_1, \dots, v_l, \dots, v_L\}$,

1. Para cada sentido s_k ,
 - (a) Coger su *topic signature*

$$\mathcal{T}_l = \{(t_{k1}, \text{weight}_k(t_{k1}), \dots, (t_{kn}, \text{weight}_k(t_{kn}))\}$$

- (b) $\text{score}(s_k) = \sum_{v_l \text{ in } c} \text{weight}_k(v_l)$.

2. Escoger $\text{argmax}_{s_k} \text{score}(s_k)$

WSD

Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros

Análisis
semántico

Pragmática

Combinación de varios sistemas

Wilks y Stevenson (2001) describen un algoritmo que combina información de varias fuentes:

- Parte-del-lenguaje de la palabra.
- Solapamiento con las definiciones de diccionarios.
- Preferencias seleccionales (con métricas de Teoría de la Información).
- Categorías del tesoro Roget's.
- Colocaciones.

WSD

*Introducción
Bayesiana*

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis
semántico

Pragmática

Combinación de varios sistemas (II)

La información de todas estas fuentes se utiliza para caracterizar cada aparición de w en un contexto, y se le proporciona a un algoritmo de aprendizaje automático.

Wilke y Stevenson (2001) utilizan un algoritmo de aprendizaje basado en memoria, TiMBL.

La clasificación se realiza con el sistema ya entrenado.

WSD

Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros

Análisis
semántico

Pragmática

Combinación de varios sistemas (III)

Cada uno de los métodos por separado tenían una precisión de entre 50% y 80%.

El sistema completo alcanza el 90.37% de precisión sobre todas las palabras (ambiguas y no ambiguas), con una polisemia media de 14.62 sentidos por cada palabra.

La evaluación realizada (incluye desambiguación de PoS tag, y usa categorías del Roget's) es más sencilla que usando otros diccionarios con definiciones más finas.

WSD

Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros

Análisis
semántico

Pragmática

Métodos no supervisados

Los métodos no supervisados son los que no necesitan datos de entrenamiento.

Se utilizan para agrupar diversas apariciones de una palabra w en función de los contextos en que aparece.

No pueden etiquetar palabras con los sentidos.

Posibles métodos:

- Algoritmo EM.
- Clustering, utilizando características contextuales.

WSD

Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros

Análisis
semántico

Pragmática

Métodos no supervisados (II)

Ejemplo

Tenemos varias contextos de *banco*:

1. Cambié este cheque de 20.000 pesetas por dinero en efectivo en el banco.
2. Ayer estuvimos todos los amigos sentados en el banco hasta las siete.
3. Tengo la hipoteca de mi casa en este banco. Son 100.000 pesetas al mes.
4. Este banco tiene las hipotecas más baratas.
5. Los bancos de madera son más cómodos para sentarse que estos nuevos de cemento que ha puesto el ayuntamiento.

Al agruparlos, juntando las que tienen más palabras en común: $\{1, 3, 4\}, \{2, 5\}$

WSD

*Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros*

Análisis
semántico

Pragmática

Métodos no supervisados (III)

Aprendizaje con Expectation-Maximisation

Inicialización:

1. Inicializar el modelo. Por ejemplo, con clasificación Bayesiana, el modelo será $P(s_k)$ y $P(v_j|s_k)$.
2. Computar la probabilidad de que ocurra el corpus que tenemos dado ese modelo:

$$I(C|\mu) = \log \prod_{i=1}^I \prod_{k=1}^K P(c_i|s_k)P(s_k)$$

WSD

Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros

Análisis
semántico

Pragmática

Métodos no supervisados (IV)

Aprendizaje con Expectation-Maximisation

Expectación: Estimar la probabilidad de que s_k haya generado c_i :

$$h_{ik} = \frac{P(c_i|s_k)}{\sum_{k=1}^K P(c_i|s_k)}$$

Maximización: Reestimar los parámetros del modelo:

$$P(v_j|s_k) = \frac{\sum_{\{c_i:v_j \in c_i\}} h_{jk}}{Z_j}$$

$$P(s_k) = \frac{\sum_{i=1}^I h_{ik}}{\sum_{k=1}^K \sum_{i=1}^I h_{ik}}$$

WSD

*Introducción
Bayesiana
T. I.
Diccionarios
Tesauros
Dens. conc.
Signaturas
Combinación
No supervis.
Otros*

Análisis
semántico

Pragmática

Métodos no supervisados (V)

- El mejor número de sentidos en que se dividirán las apariciones de la palabra se puede obtener empíricamente parando en el momento en que la probabilidad del modelo $I(C|\mu)$ comience a aumentar lentamente conforme aumentamos k .

WSD

Introducción

Bayesiana

T. I.

Diccionarios

Tesauros

Dens. conc.

Signaturas

Combinación

No supervis.

Otros

Análisis

semántico

Pragmática

Otros métodos

Aunque los métodos que han dado mejores resultados son los que se basan en estimadores estadísticos, se han intentado otros métodos.

- Con reglas hechas a mano.
- Utilizando las preferencias seleccionales (obtenidas de diccionarios, como LDOCE).

Para algunos casos, puede ser necesario utilizar una combinación de métodos basados en aprendizaje y estos otros.

Índice

- **Desambiguación de sentidos**
 - **Análisis semántico**
 - **Análisis pragmático**
- Introducción
 - Clasificador Bayes
 - Teoría de la Información
 - WSD con diccionarios
 - WSD con tesauros
 - WSD con densidad conceptual
 - WSD con topic signatures
 - Combinaciones
 - WSD no supervisada
 - Otros métodos
 - **Introducción**
 - **Cálculo lambda**
 - **HPSG**
 - **Otros formalismos**
 - **Introducción**

WSD

**Análisis
semántico**

Introducción

Cálculo- λ

HPSG

Otros

Pragmática

Introducción (I)

La **Semántica** es el estudio del significado lingüístico.

Semántica computacional:

Modelar la manera en que se puede recomponer sistemáticamente el significado de las frases a partir del significado de sus componentes sintácticos. ■

Hipótesis:

La semántica está relacionada con la sintaxis.

Problemas:

- Definir un formalismo para representación de significado.
- Establecer una interfaz entre sintaxis y semántica.

WSD

**Análisis
semántico**

Introducción
Cálculo- λ
HPSG
Otros

Pragmática

Cálculo lambda (I) Composicionalidad

La hipótesis de composicionalidad significa que **el significado del todo puede derivarse del significado de cada una de las partes.**

Conjetura de Frege:

- En el siglo 19, Frege conjeturó que la composición semántica *siempre* consiste en la saturación de un componente de significado insaturado.
- Frege formalizó los significados insaturados como funciones.

WSD

**Análisis
semántico**

Introducción

Cálculo- λ

HPSG

Otros

Pragmática

Cálculo lambda (II)

- Representamos el mundo con un conjunto D de individuos o constantes, y un conjunto de relaciones que se dan entre ellos.

Por ejemplo,

pedro

maria

ana

hombre(pedro)

mujer(maria)

mujer(ana)

corre(pedro)

corre(ana)

ama(pedro, maria)

WSD

**Análisis
semántico**

Introducción
Cálculo- λ
HPSG
Otros

Pragmática

Cálculo lambda (III)

- Cada relación n -aria R puede considerarse un conjunto de tuplas de D^n , o una función

$$f : D^n \rightarrow \{true, false\}$$

$$f(a_1, \dots, a_n) = true \Leftrightarrow (a_1, \dots, a_n) \in R$$

- Usaremos B para denotar el conjunto $\{true, false\}$.
- La semántica de los nombres propios sería las constantes a las que representan en el lenguaje:
 - [[Pedro]] = *pedro*
 - [[María]] = *maria*
 - [[Ana]] = *ana*
 - [[Anita]] = *ana*
- La semántica de los verbos sería una relación o función: [[corre]] = *corre* : $D \rightarrow B$ tal que *corre*(x) = *true* si x corre.

WSD

**Análisis
semántico**

Introducción
Cálculo- λ
HPSG
Otros

Pragmática

Cálculo lambda (IV)

Composicionalidad en semántica de Frege

- Si la regla sintáctica α que se aplica es de la forma

$$X \rightarrow \beta$$

entonces su significado es

$$[[\alpha]] = [[\beta]]$$

- Si la regla sintáctica α que se aplica es de la forma

$$S \rightarrow \beta, \gamma$$

entonces su significado es

$$[[\alpha]] = [[\gamma]]([[\beta]])$$

WSD

**Análisis
semántico**

Introducción

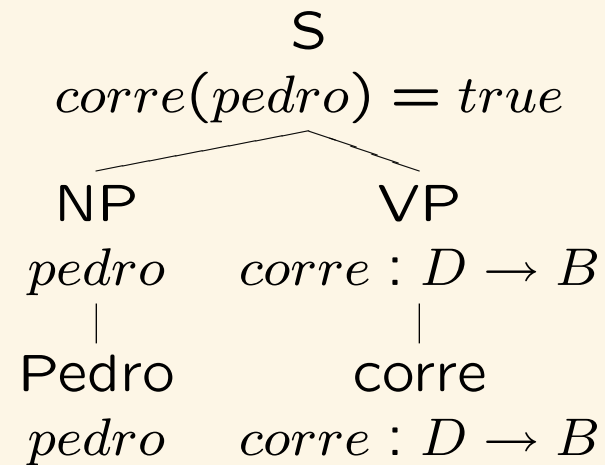
Cálculo- λ

HPSG

Otros

Pragmática

Cálculo lambda (V) Ejemplo



WSD

**Análisis
semántico**

Introducción
Cálculo- λ
HPSG
Otros

Pragmática

Cálculo lambda (VI)

Ejemplo

La notación

$$\lambda x : x \in D. corre(x)$$

representa la función tal que, para todo $x \in D$, toma el valor de $corre(x)$.

Si el dominio es conocido, se puede abreviar a $\lambda x. corre(x)$

Puede utilizarse para representar funciones con varios argumentos: $\lambda x. \lambda y. ama(y, x)$

Y para representar funciones parcialmente saturadas: $\lambda y. ama(y, ana)$

WSD

**Análisis
semántico**

Introducción
Cálculo- λ
HPSG
Otros

Pragmática

Cálculo lambda (VII)

Ejemplo

La semántica de los verbos, en notación, lambda, nos queda:

$$[[\text{corre}]] = \lambda x.\text{corre}(x)$$

$$[[\text{ama}]] = \lambda x.\lambda y.\text{ama}(y, x)$$

Podemos redefinir las reglas de composición:

- $\alpha \rightarrow \beta$, entonces $[[\alpha]] = [[\beta]]$
- $\alpha \rightarrow \beta, \gamma$, entonces $[[\alpha]] = [[\gamma]]([[\beta]])$
si el tipo de β coincide con el dominio de γ .
- $\alpha \rightarrow \beta, \gamma$, entonces $[[\alpha]] = [[\beta]]([[\gamma]])$
si el tipo de γ coincide con el dominio de β .

WSD

**Análisis
semántico**

Introducción

Cálculo- λ

HPSG

Otros

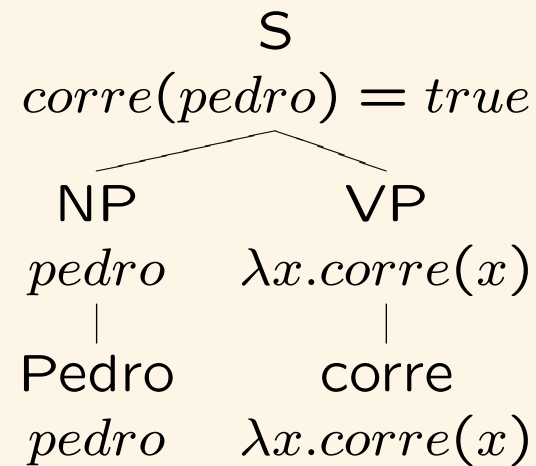
Pragmática

Cálculo lambda (VIII)

Ejemplo

$[[\text{Pedro}]] == \text{pedro}$

$[[\text{corre}]] == \lambda x.\text{corre}(x)$



Cálculo lambda (IX)

Ejemplo

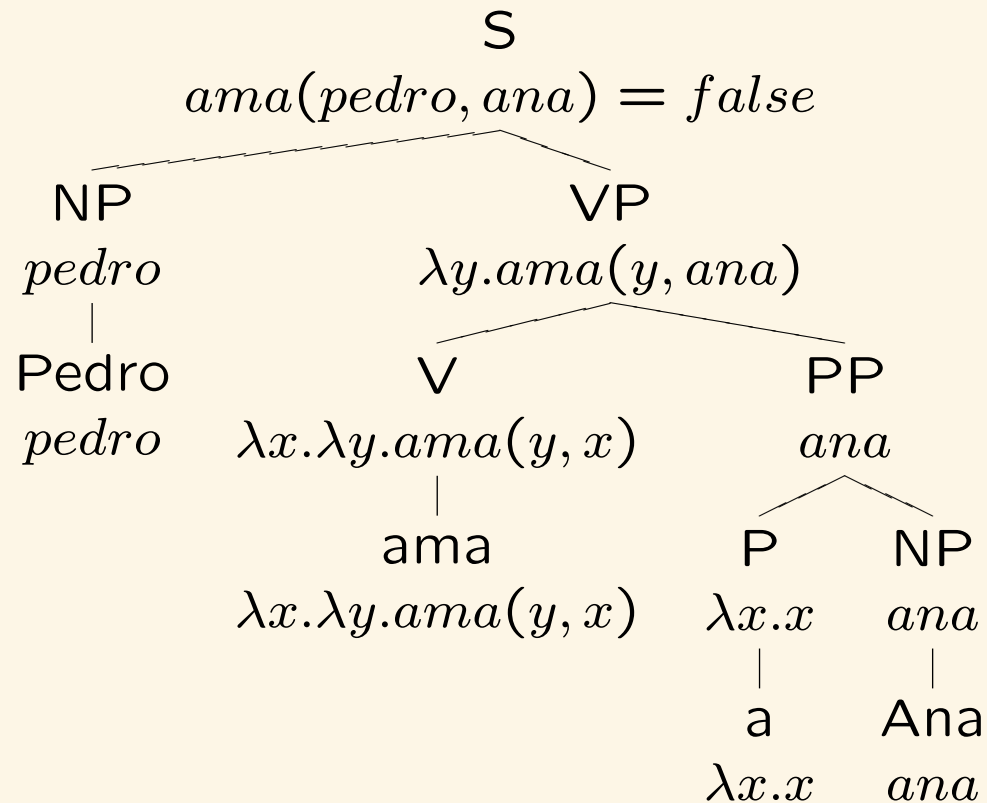
$[[\text{Pedro}]] == \text{pedro}$
 $[[\text{Ana}]] = \text{ana}$
 $[[\text{ama}]] = \lambda x.\lambda y.\text{ama}(y, x)$
 $[[\text{a}]] = \lambda x.x$

WSD

Análisis semántico

Introducción
Cálculo- λ
 HPSG
 Otros

Pragmática



WSD

**Análisis
semántico**

Introducción
Cálculo- λ
HPSG
Otros

Pragmática

Cálculo lambda (X)

El cálculo lambda (Montague, 1974),

- Es totalmente composicional: el significado de la frase se computa enteramente a partir de las interpretaciones de sus constituyentes.
- Existen herramientas comunes para trabajar con teorías de lógica de predicados de primer orden.

WSD

**Análisis
semántico**

Introducción

Cálculo- λ

HPSG

Otros

Pragmática

Cálculo lambda (XI)

- Aunque necesita usar lógicas de predicados de segundo orden para representar a los cuantificadores,

$$[[\text{nadie}]] = \lambda P.\forall x.\text{not}P(x)$$

la interpretación final de cada proposición viene expresada en lógica de primer orden.

- No resuelve la ambigüedad semántica:

Todo hombre ama a una mujer

$$\forall y(\text{hombre}(y) \Rightarrow \exists x(\text{mujer}(x) \wedge \text{ama}(y, x)))$$

$$\exists x(\text{mujer}(x) \wedge \forall y(\text{hombre}(y) \Rightarrow \text{ama}(y, x)))$$

WSD

**Análisis
 semántico**

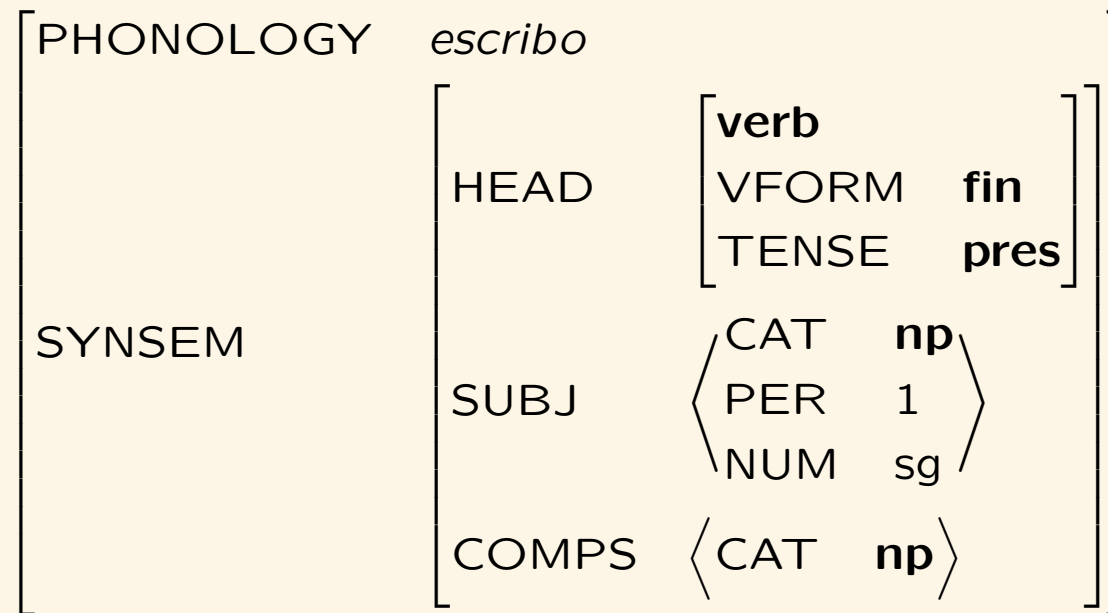
Introducción
 Cálculo- λ
HPSG
 Otros

Pragmática

Semántica en HPSG (I)

Como se vio en el tema 3, las HPSG

- Utilizan estructuras de rasgos para representar las palabras.
- En estas estructuras, se pueden introducir restricciones sobre los argumentos (preferencia seleccional, etc.)



Semántica en HPSG (II)

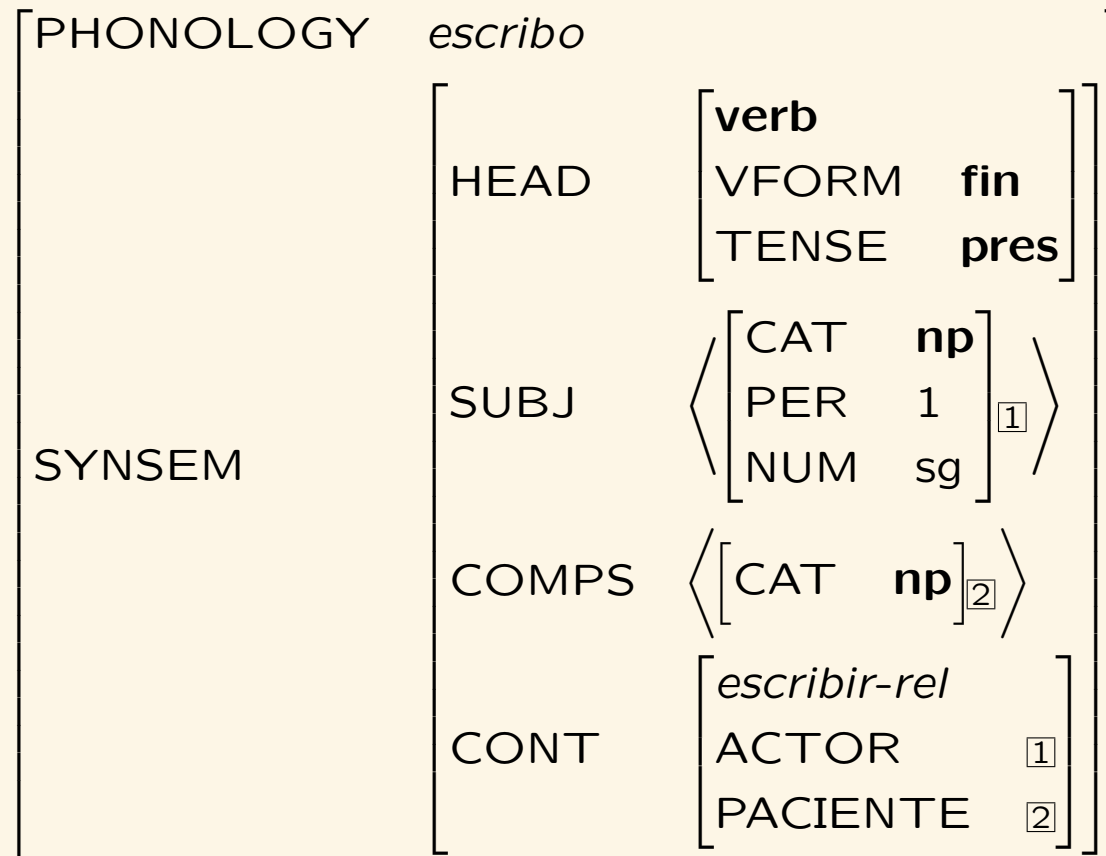
Es posible construir relaciones semánticas entre los distintos constituyentes directamente en la estructura de rasgos, indicando correferencias:

WSD

Análisis semántico

Introducción
 Cálculo- λ
HPSG
 Otros

Pragmática



WSD

**Análisis
semántico**

Introducción
Cálculo- λ
HPSG
Otros

Pragmática

Semántica en HPSG (III)

Conclusiones

- De esta forma, conforme se analiza una oración, en el atributo CONT se van unificando los significados de los constituyentes.
- Se puede extender para incluir tratamiento de cuantificadores.
- Como en el caso del cálculo lambda, se puede traducir el contenido del atributo CONT a alguna forma normal lógica, para razonar con ello.
- El interfaz entre sintaxis y semántica no es una traducción de la primera a la segunda, sino que las dos están codificadas en el mismo formalismo, completamente integradas.

WSD

**Análisis
semántico**

Introducción
Cálculo- λ
HPSG
Otros

Pragmática

Otros formalismos de semántica (I)

- En las teorías de semántica composicional, la relación entre el significado de una expresión y los significados de sus constituyentes es una función.
- Para tratar la ambigüedad de los cuantificadores, es necesario, en los formalismos anteriores, utilizar operadores no composicionales (p.ej., almacenamiento y reordenamiento de cuantificadores).
- Existen extensiones para generalizar esto con parámetros cuyos valores se puedan definir de maneras distintas, y que den lugar a interpretaciones diferentes.
- Otras extensiones tratan respuestas cortas en diálogos, contexto, etc.

Índice

- **Desambiguación de sentidos**
 - **Análisis semántico**
 - **Análisis pragmático**
- Introducción
 - Clasificador Bayes
 - Teoría de la Información
 - WSD con diccionarios
 - WSD con tesauros
 - WSD con densidad conceptual
 - WSD con topic signatures
 - Combinaciones
 - WSD no supervisada
 - Otros métodos
 - **Introducción**
 - **Cálculo lambda**
 - **HPSG**
 - **Otros formalismos**
 - **Introducción**

WSD

Análisis
semántico

Pragmática
Introducción

Pragmática (I)

La **Pragmática** es el estudio del significado en su contexto.

- Aunque poco estudiada hasta los años setenta, recientemente ha recibido más atención.
- Fuertemente impulsada por los sistemas de diálogo orales.
- Con aplicaciones al procesamiento y entendimiento de texto escrito.

WSD

Análisis
semántico

Pragmática
Introducción

Pragmática (II)

Resolución de anáfora

Una **anáfora** es el fenómeno lingüístico que consiste en referirse a una entidad mencionada previamente:

- **Anáfora** es la expresión que contiene la referencia.
- **Antecedente** es aquello a lo que se refiere.
- Cuando una anáfora se refiere a un antecedente, y los dos tienen el mismo referente en el mundo real, existe una relación de **correferencia**.

Ejemplo:

Miles Vorkosigan₁ dice que él₂ siempre estará agradecido al sargento Bothari.

WSD

Análisis
semántico

Pragmática
Introducción

Pragmática (III)

Resolución de anáfora

Puede darse el caso de que una expresión es tanto una anáfora de otra anterior, como un referente de una posterior. Se dice entonces que hay una **cadena de correferencia**.

He₁ who is in love with himself₂ has at least this advantage – he₃ won't encounter many rivals. – Georg Lichtenberg, "Aphorisms"

After Snow White₁ used a couple rolls of film taking pictures of the seven dwarfs, she₂ mailed the roll to be developed. Later she₃ was heard to sing, "Some day my₄ prints will come."

WSD

Análisis
semántico

Pragmática

Introducción

Pragmática (IV)

Resolución de anáfora

No todas las anáforas se refieren a sintagmas nominales. Algunos ejemplos son:

cláusulas Él llegó tarde₁. Eso₂ la enfadó mucho.

sintagmas verbales Alyson returned home three years ago, so it was inevitable that John would follow her, and yesteryear he did. ■

En la mayor parte de los casos, es de interés localizar los antecedentes de **pronombres** y **descripciones definidas**:

Pedro₁ era el hijo menor de un policía₂. El niño₁ le₂ cogió un día la pistola, pero afortunadamente estaba descargada.

WSD

Análisis
semántico

Pragmática

Introducción

Pragmática (V)

Resolución de anáfora

El procedimiento de resolución de anáforas en un texto incluye los siguientes pasos:

1. Identificar las posibles anáforas: pronombres y algunas descripciones definidas.
2. Para cada una de ellas, identificar posibles antecedentes.
3. Seleccionar el antecedente más probable, por ejemplo,
 - Aplicando restricciones semánticas (sexo, tipo de entidad, etc.)

Lupo₁ era el pastor alemán de Luis₂. El perro_? robó un hueso.

- Aplicando heurísticas sobre la distancia entre el antecedente y el pronombre.
- etc.

WSD

Análisis
semántico

Pragmática

Introducción

Pragmática (VI)

Resolución de anáfora

Aplicaciones:

- Extracción de información: localizar información de relevancia dentro de un texto.

En Colombia actúan el PLN, las FARC y el PLMF. Este último cometió el atentado de la semana pasada.

- Resúmenes automáticos: si se reduce la longitud del texto, es posible que se mantenga una anáfora y se elimine su antecedente.
- Resúmenes multidocumento: para identificar correferencias entre documentos distintos.

WSD

Análisis
semántico

Pragmática
Introducción

Pragmática (VII)

Resolución de anáfora

Aplicaciones (cont.):

- Sistemas de traducción automática: por ejemplo, para traducir a un lenguaje que marca el género de los pronombres de uno que no lo hace.
- Búsqueda de respuestas a preguntas: para identificar correferencias entre los pronombres de la pregunta y las posibles respuestas.

WSD

Análisis
semántico

Pragmática

Introducción

Pragmática (VIII)

Pragmática en sistemas de diálogo

Incluye:

- Utilización del contexto para desambiguar el significado semántico de las oraciones.
- Identificación de los tipos de enunciados:
 1. *Asertivos*: afirman un hecho.
 2. *Directivos*: ordenan una acción.
 3. *Comisivos*: Comprometen al hablante a la realización de una acción (prometer, jurar, ofrecer).
 4. *Expresivos*: expresan un estado psicológico (acción de gracias, felicitación, disculpas...)
 5. *Declaraciones*: oraciones que producen una correspondencia entre su contenido y la realidad (p.ej., dar nombre a un barco, una dimisión, etc.)

WSD

Análisis
semántico

Pragmática

Introducción

Pragmática (IX)

Pragmática en sistemas de diálogo

Principios de H: P. Grice:

Cantidad Dar la cant. de información adecuada.

Comió parte del pastel. ⇒ No se lo comió todo. ■

Calidad Dar una afirmación cierta.

Eso cuesta un ojo de la cara. ⇒ Eso cuesta mucho dinero. ■

Relevancia Dar una afirmación relevante en el contexto actual.

Hijo: Puedo ver la tele?
Padre: Es la hora del baño. ■

Forma Evitar la obscuridad y la ambigüedad. Ser breve y ordenado.

A. Estás listo?

B ...

A. Te pregunto si estás listo o no lo estás.

WSD

Análisis
semántico

Pragmática

Introducción

Pragmática (X) Conclusiones

- La pragmática implica tomar decisiones sobre las creencias del hablante, sus intenciones, y la relevancia de sus afirmaciones.
- Conforme aumente la versatilidad de los diálogos computacionales, será necesario incorporar las teorías lingüísticas y filosóficas sobre pragmática cada vez más.
- Son de esperar avances en la pragmática computacional en el futuro cercano.