

## Tema 4

# Interpretación semántica

Ingeniería del Lenguaje Natural

Curso 2007-2008

<http://www.dlsi.ua.es/assignatures/iln/>

## Índice

1. Introducción: Importancia del nivel semántico en PLN.
2. Componentes de la semántica.
3. Modelos de representación del significado.
4. Proceso de interpretación semántica.

ILN

2

## ¿Dónde estamos?

- Nivel fonológico → sonido
- Nivel morfo-léxico → palabra
- Nivel sintáctico → sintagma
- **Nivel semántico** → significado\*
- Nivel pragmático → texto

Niveles de  
Representación  
Lingüística

ILN

3

## Valor de la semántica

- **Objetivo** principal de la mayoría de los sistemas de PLN:
  - llegar al significado de las expresiones lingüísticas.
- Ejemplo:
  - Sistema de Búsqueda de Respuestas**
  - ¿Quién ganó el Premio Nóbel de Economía en 1994?*
  - Interpretar la pregunta... (una persona....)
  - ...para localizar la respuesta.

ILN

4

## Valor de la semántica

---

- Es el aspecto lingüístico más **difícil** de procesar:
  - no existe una teoría general del significado que pueda ser implementada;
  - aproximaciones:
    - completar el análisis sintáctico con información semántica;
    - anotación de sentidos en dominios restringidos;
    - etc.

ILN

5

## Índice

---

1. Introducción: Importancia del nivel semántico en PLN
2. Componentes de la semántica
3. Modelos de representación del significado
4. Proceso de interpretación semántica

ILN

6

## Componentes de la semántica

---

### Significado oracional:

Formado por el significado de las unidades (palabras, sintagmas, cláusulas...) que componen la oración

### **PRINCIPIO DE COMPOSICIONALIDAD:**

- palabras con significado léxico,
  - nombres, verbos, adjetivos y adverbios;
- palabras con significado gramatical,
  - preposiciones, artículos, determinantes;
- relación sintáctica entre palabras.

ILN

7

## Componentes de la semántica

---

### **Principio de Composicionalidad (R. Montague):**

El significado de una expresión compleja es una función del significado de cada una de las unidades simples que la componen y de las reglas sintácticas mediante las cuales se combinan (de las relaciones sintácticas que se establecen entre ellas).

ILN

8

## Componentes de la semántica

### ❑ Rasgos (preferencias) de selección semántica.

- ❑ Requerimientos semánticos que los verbos imponen a sus argumentos.
- ❑ Ejemplos:
  - ❑ *Andar*: sujeto = [+animado]
  - ❑ *Razonar*: sujeto = [+humano]
  
  - ❑ *Escribir*: sujeto = [+humano]  
CC(con) = [+Instrumento]
  - ❑ *Ordeñar*: sujeto = [+humano]  
objeto = [-humano] [+animado]

ILN

9

## Componentes de la semántica

### ❑ Roles o papeles semánticos.

- ❑ Relaciones semánticas abstractas entre el verbo y sus argumentos.
- ❑ Basado en la estructura argumento-predicado
- ❑ Tipos:
  - ❑ **Agente**: entidad que causa, realiza y controla la acción;
  - ❑ **Tema**: entidad afectada por la acción;
  - ❑ **Paciente**: entidad animada afectada por la acción
  - ❑ **Beneficiario**;
  - ❑ **Locación**;
  - ❑ etc.

ILN

10

## Componentes de la semántica

### ❑ Ejemplos:

- “Pedro [**agente**] dijo al jefe [**receptor**] que no vendría [**tópico**] esta noche [**tiempo**]”
- “La piedra [**causa**] rompió la ventana [**tema**]”
- “la ventana [**tema**] se rompió por una piedra [**causa**]”
- “Tu hermano [**agente**] rompió la ventana [**tema**] con una piedra [**instrumento**]”
- “La ventana [**experimentante**] se cerró”

ILN

11

## Componentes de la semántica

### ❑ Actos de habla.

- ❑ Ámbito de la pragmática
- ❑ Operación mediante la cual un hablante emite una cadena fónica de acuerdo con unas reglas gramaticales y una intención.
- ❑ Tipos:
  - ❑ **Acto locutivo**: preferencia lingüística
  - ❑ **Acto ilocutivo**: significado del acto dentro de un contexto de interacción social. Intención(es) con la que se realiza el acto locutivo. Ej: “Buenos días” = saludar; “¿Cómo te llamas?” = investigar/preguntar; “Hace calor” = asertivo.
    - ❑ **Actos de habla indirectos**: cortesía, etc. (“Hace calor” “¿Tienes hora?”)
  - ❑ **Acto perlocutivo**: consecuencia en el receptor del acto ilocutivo.

ILN

12

# Índice

---

1. Introducción: Importancia del nivel semántico en PLN
2. Componentes de la semántica
3. Modelos de representación del significado
4. Proceso de interpretación semántica

ILN

13

# Modelos de representación

---

- ❑ ¿Cómo representar el significado?
  - ❑ **Medios** para representar el conocimiento:
    - **Lenguas naturales** → gran poder expresivo, herramienta de razonamiento, ambiguo.
    - **Lenguaje formal** → específico, representa conocimiento científico, no ambiguo.
  - ❑ **Problema** de la representación del conocimiento:  
Traducir lenguajes formales y naturales en un formalismo que pueda ser entendido por el ordenador, con los mecanismos adecuados de formalización y validación del razonamiento.

ILN

14

# Modelos de representación

---

- ❑ Características que debe cumplir el lenguaje formal para poder representar el significado:
  1. Ser **no ambiguo**: una única interpretación por cada oración.
  2. Forma **canónica**: la misma representación semántica para las diferentes formas sintácticas.
  3. Tener **reglas simples** de interpretación e inferencia: debe ser "**computable**", bien para obtener una respuesta, para realizar un proceso de traducción, etc.
  4. Tener una **estructura lógica determinada por la forma de la oración**: la estructura oracional debe guiar el proceso de construcción del significado.

ILN

15

# Modelos de representación

---

- ❑ Lógica de predicados
- ❑ Lógica modal
- ❑ Lambda-cálculo
- ❑ Redes semánticas
- ❑ *Frames*
- ❑ Lenguaje de la Forma Lógica.

ILN

16

## Lógica de predicados

---

- ❑ Símbolos básicos del lenguaje:
  - ❑ variables,
  - ❑ constantes (símbolos de función de aridad 0),
  - ❑ símbolos de predicado,
  - ❑ símbolos de función,
  - ❑ conectivas: no ( $\neg$ ), and ( $\wedge$ ), or ( $\vee$ ), implicación ( $\rightarrow$ ), equivalencia ( $\equiv$ ),
  - ❑ cuantificador existencial ( $\exists$ )
  - ❑ cuantificador universal ( $\forall$ ).

ILN

17

## Lógica de predicados

---

- ❑ **Sintaxis** de la lógica de predicados:
  - ❑ Un **término** es una variable o una forma de función.
  - ❑ Una **forma de función** es la aplicación de un símbolo de función sobre un número adecuado de términos.
    - ❑ Si  $f$  es un símbolo de función de aridad  $n$  y  $t_1, t_2, \dots, t_n$  son términos, entonces la forma correspondiente se expresa usualmente como  $f(t_1, \dots, t_n)$ .
    - ❑ Si  $n$  es cero la forma se escribe  $f$ .
  - ❑ Una **forma de predicado** es la aplicación de un símbolo de predicado sobre un número adecuado de términos.
    - ❑ Si  $P$  es un símbolo de predicado de aridad  $m$  y  $t_1, t_2, \dots, t_m$  son términos, entonces la forma correspondiente se representa usualmente como  $P(t_1, \dots, t_m)$ .
    - ❑ Si  $m$  es cero, la forma se escribe  $P$ .

ILN

18

## Lógica de predicados

---

- ❑ **Sintaxis** de la lógica de predicados:
  - ❑ Un **átomo** es una forma de predicado o una igualdad, eso es, una expresión ( $s=t$ ) donde  $s$  y  $t$  son términos.
  - ❑ El concepto de **fórmula** se define recursivamente como sigue:
    - ❑ Un átomo es una fórmula.
    - ❑ Si  $A$  es una fórmula, entonces  $\neg(A)$  es una fórmula.
    - ❑ Si  $A$  y  $B$  son fórmulas entonces  $(A \wedge B)$ ,  $(A \vee B)$ ,  $(A \rightarrow B)$  y  $(A \equiv B)$  son fórmulas.
    - ❑ Si  $A$  es una fórmula y  $x$  es una variable, entonces  $\exists x A$  y  $\forall x A$  son fórmulas.

ILN

19

## Lógica de predicados

---

- ❑ **Semántica** de la lógica de predicados:
  - ❑ **Teoría de Modelos**: determinación de la semántica de las fórmulas lógicas.
  - ❑ Una **interpretación** es el contexto en el cual una fórmula toma el valor cierto o falso.
  - ❑ Si la interpretación de una fórmula toma un valor cierto se dice que es **modelo** de la fórmula.
  - ❑ Cada modelo de una fórmula denota una semántica posible para dicha fórmula.

ILN

20

# Lógica de predicados

## ❑ Ejemplos de representación:

Todos los ríos pasan por Valencia:  $(\forall x) R(x) \rightarrow P(x, \text{valencia})$

El Turia es un río:  $R(\text{turia})$

El Turia pasa por Valencia:  $P(\text{turia}, \text{valencia})$

Un río grande pasa por Valencia:  $(\exists x) R(x) \wedge G(x) \wedge P(x, \text{valencia})$

donde:

- ✓ x es una variable,
- ✓ R y G son predicados unarios (sustantivos y adjetivos),
- ✓ P es un predicado binario (verbos),
- ✓ Turia y Valencia son términos (objetos del dominio, nombres propios),
- ✓  $\exists, \forall$  son los cuantificadores (determinantes).

ILN

21

# Lógica de predicados

## ❑ Ventajas de la lógica de predicados:

- ❑ Mecanismos de inferencia poderosos y simples, como la resolución, que permiten el razonamiento acerca del conocimiento de forma fácil.
- ❑ En el ejemplo anterior, a partir de las fórmulas 1 y 2, puede inferirse la 3, mediante instanciación universal y *modus ponens*.

## ❑ Desventajas:

- ❑ Insuficiente para representar muchas sutilezas de las lenguas naturales:
  - ❑ resolución de problemas de ambigüedad,
  - ❑ cuantificadores de las lenguas naturales,
  - ❑ representación de los adverbios, aspectos temporales y modales, etc.

ILN

22

# Lógica Modal

- ❑ Ampliación de la lógica clásica.
- ❑ Expresan conocimiento relativo al tiempo, la posibilidad o la creencia.
- ❑ Extiende la sintaxis de la lógica de primer orden con dos operadores modales:

- ❑ el operador **modal universal**  $\Delta$ , y
- ❑ el operador **modal existencial**  $\nabla$ .

$$\Delta F \equiv \neg \nabla \neg F$$

•Interpretación de $\Delta$	•Interpretación de $\nabla$	•Lógica
<ul style="list-style-type: none"> <li>• necesario</li> <li>• siempre</li> <li>• deber</li> </ul>	<ul style="list-style-type: none"> <li>• posible</li> <li>• alguna vez</li> <li>• poder</li> </ul>	<ul style="list-style-type: none"> <li>• Alética</li> <li>• Temporal</li> <li>• Deontica</li> </ul>

ILN

23

# Lambda-cálculo

- ❑ Ampliación de la lógica clásica.
- ❑ Se expande la lógica de predicados con el lambda operador ( $\lambda$ -operador).
- ❑ Si F es un forma lógica, entonces  $\lambda x F$  representará un conjunto especificado por F con respecto a la variable x (generalización o abstracción).
  - ❑  $\text{querer}(\text{juan}, \text{maria}) \rightarrow \text{querer}(X, \text{maria}) \rightarrow \lambda x [\text{querer}(x, \text{maria})]$   
Representa el conjunto de entidades (x) que quieren a maría, es decir, la propiedad "querer a maría".
  - ❑  $\lambda y [\text{enviar}(\text{juan}, \text{maria}, y)]$   
representa el conjunto de objetos (y) que juan manda a maria.
  - ❑  $\lambda x, y [\text{enviar}(x, \text{maria}, y)]$   
representa el conjunto de pares (x,y) que hacen cierto el predicado *enviar*.

ILN

24

## Redes semánticas

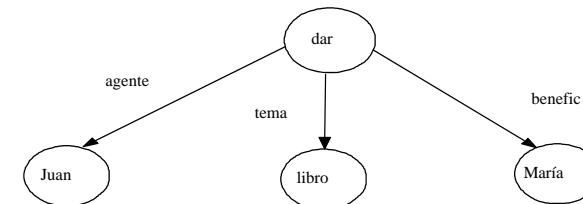
- ❑ Una red consiste en nodos etiquetados y arcos etiquetados.
  - ❑ Las **constantes** son representadas por nodos.
  - ❑ Los **operadores** o **funciones unarias** se representan mediante arcos etiquetados con el nombre de la función.
  - ❑ Cada **predicado** se representa mediante un nodo.
    - ❑ Los **argumentos** del predicado son nodos que se conectan a éste mediante arcos etiquetados.

ILN

25

## Redes semánticas

- ❑ La oración “Juan da un libro a María”, que podría expresarse mediante el predicado lógico  $\text{dar}(\text{juan}, \text{libro}, \text{maría})$ ,  
Para representarlo mediante red semántica:  
Nodos → predicado y argumentos  
Arcos → relaciones entre ellos: roles semánticos.



ILN

26

## Redes semánticas

- ❑ Las redes semánticas tienen el poder de representación del cálculo de predicados.
- ❑ Permiten representar bases de conocimiento mediante relaciones jerárquicas entre nodos.
  - ❑ Cada nodo representa clases de entidades, instancias o valores de atributos.
  - ❑ Los arcos representan las relaciones entre los nodos: de herencia, de instanciación o atributos.

ILN

27

## Frames

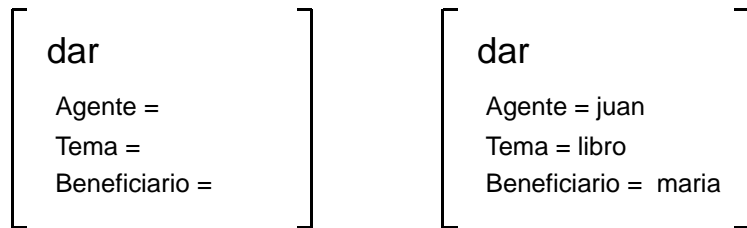
- ❑ **Frame** = un conjunto de atributos, llamados *slots*, ranuras o huecos.
- ❑ Cada *slot* tiene valores asociados y posibles restricciones sobre esos valores.
- ❑ Cada *frame* describe:
  - ❑ alguna entidad del dominio
  - o,
  - ❑ algún objeto (instancia) particular.
- ❑ **FrameNet**

ILN

28

## Frames

- Los *frames* son equivalentes a las redes semánticas.



ILN

29

## Lenguaje de la Forma Lógica

- Un lenguaje de la forma lógica es un lenguaje de representación del significado oracional basado en el cálculo de predicados de primer orden (CPPO).
- Permite la interpretación de:
  - las palabras\*,
  - las conectivas,
  - los cuantificadores\*,
  - plurales\*,
  - aspectos modales y temporales,
  - papeles o roles temáticos.
  - actos del habla.

ILN

30

## Lenguaje de la Forma Lógica

- Interpretación de las **palabras**:
- Unidad básica de **representación del significado**:
  - el **sentido** de las palabras que es el significado de una palabra en un contexto determinado.
- Cada palabra contribuye a la construcción del significado con:
  - un **término** si es un objeto del dominio,
  - un **predicado** si describe una relación entre términos o una propiedad de estos,
  - un **cuantificador**.

ILN

31

## Lenguaje de la Forma Lógica

- Nombres propios**: Término o constante.
- Nombres comunes**: Predicados unarios.
- Verbos**: Predicados  $n$ -arios, donde  $n$  depende del número de términos o complementos del verbo.
- Artículos, adjetivos indefinidos**: cuantificadores

ILN

32



# Lenguaje de la Forma Lógica

## □ Ejemplo: “Juan planta un árbol”

- CPPO:  $(\exists x) A(x) \wedge P(\text{juan}, x)$
- LFL : existe (X, arbol(X), plantar(juan,X)),  
donde:
  - **juan** es un término introducido por el nombre propio “Juan”.
  - **arbol(X)** es un predicado unario introducido por el nombre común “árbol”.
  - **plantar(X,Y)** es un predicado binario introducido por el verbo “plantar”.
  - **existe(X,B,F)**, es el cuantificador existencial introducido por el artículo “un”.

ILN

33

# Lenguaje de la Forma Lógica

## □ Interpretación de los cuantificadores

### □ Cuantificadores generalizados: *Three-Branched Quantification* (Abramson 1989):

- $q(K, S1, S2)$ , donde
  - $q$  es el cuantificador (“el”, “los”, “uno”, “muchos”, “otro”, ...),
  - $K$  es una variable introducida por la cuantificación,
  - $S1$  representa la restricción de la proposición (sujeto) y
  - $S2$  el cuerpo de la proposición (predicado).
- Por ejemplo: “Muchos hombres pasean” se interpreta
  - $\text{muchos}(X, \text{hombre}(X), \text{pasear}(X))$ .

### □ En el LFL se define un nuevo predicado, **num**, que actúa a modo de cuantificador:

Ejemplo: “Tres ríos pasan por Valencia”:  
 $\text{num}(X, \text{río}(X) \& \text{pasar}(X, \text{valencia}), 3)$

ILN

34

# Lenguaje de la Forma Lógica

## □ Interpretación de los plurales

- En castellano los plurales se introducen por determinantes  
“Los perros ladran”, tendría la forma lógica:  
 $\text{los}(X, \text{perro}(X), \text{ladrar}(X))$
- Los plurales introducen un **nuevo tipo de ambigüedad**:
  - Un **plural colectivo** en el que la propiedad del predicado debe aplicarse al conjunto entero de entidades afectadas.  
“Juan y María se encontraron”:  $\text{encontrarse}(\text{Juan}, \text{María})$ .
  - Un **plural distributivo** en el que la propiedad debe distribuirse entre todas las combinaciones posibles, para cada una de las entidades afectadas.  
“Juan y María saben Latín y Griego”:  $\text{saber}(\text{Juan}, \text{Latín}) \wedge \text{saber}(\text{María}, \text{Latín}) \wedge \text{saber}(\text{Juan}, \text{Griego}) \wedge \text{saber}(\text{María}, \text{Griego})$ .
  - Un **plural respectivo** que debe distribuirse apropiadamente.  
“Juan y María saben Latín y Griego respectivamente”:  $\text{saber}(\text{Juan}, \text{Latín}) \wedge \text{saber}(\text{María}, \text{Griego})$ .
- Esta ambigüedad se resuelve usando el **conocimiento del contexto y del mundo**.

ILN

35

# Índice

1. Introducción: Importancia del nivel semántico en PLN
2. Componentes de la semántica
3. Modelos de representación del significado
4. Proceso de interpretación semántica

ILN

36

## Proceso de interpretación semántica

### □ Interpretación **Semántica**:

- Proceso de emparejar una oración con su forma lógica.

### □ Principales **aproximaciones**:

#### □ Interpretación **guiada por la sintaxis**:

- la información semántica completa el análisis sintáctico;

#### □ Interpretación **guiada por la semántica**:

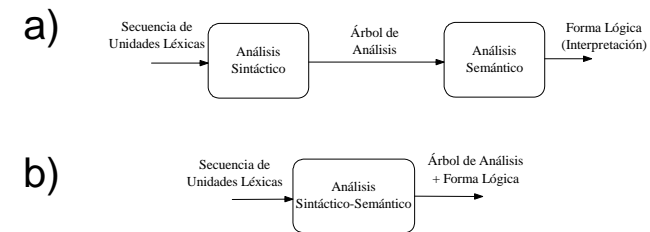
- la información sintáctica sólo aparece como restricciones limitadas.

ILN

37

## ... guiada por la sintaxis

### □ Proceso de interpretación semántica



ILN

38

## ... guiada por la sintaxis

- Obtiene la forma lógica asociada a la oración durante el análisis sintáctico (o a partir del árbol de análisis) mediante un proceso composicional.

### □ Principio de Composicionalidad (R. Montague)

- El significado de cada constituyente oracional debe derivarse a partir de:
  - el significado de sus constituyentes, y,
  - de la regla sintáctica aplicada.

ILN

39

## ... guiada por la sintaxis

### □ Proceso:

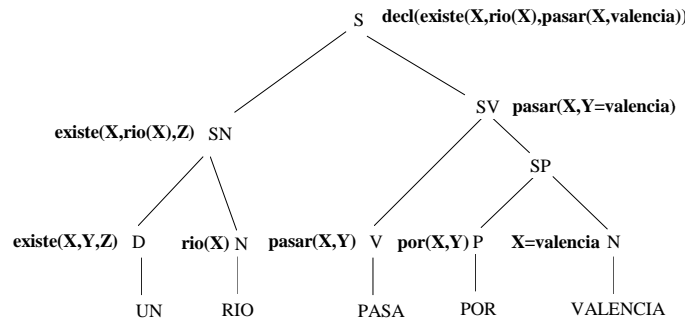
- Las hojas del árbol se etiquetan con la FL de las unidades léxicas.
- A cada nodo del árbol de análisis se le asigna una FL, que corresponde a la interpretación de la secuencia de constituyentes derivados a partir de ese nodo.
- La FL de cada nodo se computa a partir de las FL asociadas a los nodos hijos.
- La FL asociada al nodo raíz será la interpretación de la oración.

ILN

40

## ... guiada por la sintaxis

Interpretación de "Un río pasa por Valencia"

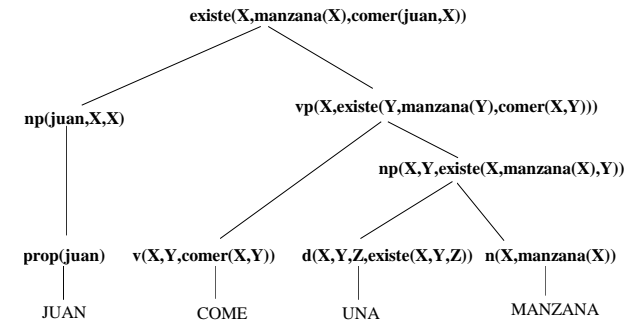


ILN

41

## ... guiada por la sintaxis

□ Ejemplo: *Juan come una manzana*



ILN

42

## ... guiada por la sintaxis

- Problemas para la teoría de la composicionalidad:
  - ámbito de los cuantificadores,
  - la interpretación de frases hechas,
  - resolución de ambigüedad de significado,
  - tratamiento de problemas lingüísticos:
    - extraposición sintáctica,
    - tratamiento de elipsis y coordinación,
    - tratamiento de la anáfora.

ILN

43

## ... guiada por la sintaxis

- Frases hechas, locuciones:
    - "Luis estiró la pata" = "Luis murió" = morir(Luis).
  - No se cumple la teoría de la composicionalidad.
  - Soluciones:
    - permitir significado semántico a la frase completa en lugar de construir el significado composicionalmente,
    - introducir nuevos sentidos a las palabras que aparecen en esas locuciones; en este caso que "estirar" significara también "morir" y que esperará un objeto de tipo "la pata".
  - Problema:
    - Ambigüedad: a veces se deben interpretar como tales locuciones o, a veces, con su sentido literal
- Es necesario el conocimiento del contexto.

ILN

44

## Proceso de interpretación semántica

---

### ❑ Interpretación **Semántica**:

- ❑ Proceso de emparejar una oración con su forma lógica.

### ❑ Principales **aproximaciones**:

#### ❑ interpretación **guiada por la sintaxis**:

- ❑ la información semántica completa el análisis sintáctico;

#### ❑ interpretación **guiada por la semántica**:

- ❑ la información sintáctica sólo aparece como restricciones limitadas.



## ... guiada por la semántica

---

- ❑ Método de análisis guiado por la semántica: utiliza restricciones sintácticas limitadas

### ❑ Uso de *frames*:

- ❑ conjunto de huecos que almacenan instancias de acciones, objetos u otros frames.

### ❑ Algoritmo de análisis:

- ❑ 1. Identificar la acción básica (verbo principal) descrita en la oración.
- ❑ 2. La acción determina un *frame*.
- ❑ 3. Relleno recursivo de huecos a partir del resto de elementos de la oración.

## ... guiada por la semántica

---

### ❑ Ejemplo:

"Un río pasa por Valencia"

- ❑ 1. *pasa* determina la acción principal → PASAR

### ❑ 2. selección del *frame*

PASAR: AGENTE <...>

OBJETO <...>

### ❑ 3. relleno de *slots*:

PASAR: AGENTE: río

OBJETO: valencia