

# Tratamiento del lenguaje natural

## ❑ Tema 5-II: Tratamiento del lenguaje natural

### ❑ Comprensión de LN

- ❑ Análisis morfo-léxico

- ❑ Análisis sintáctico

- ❑ **Análisis semántico**

- ❑ **Análisis pragmático**

### ❑ **Generación de LN**

- ❑ Métodos empíricos

- ❑ Recuperación de información

# Análisis semántico

- ❑ El árbol de análisis sintáctico de una frase es tan sólo el primer paso hacia la comprensión de la frase
- ❑ Hay que producir una **representación de su significado**
  - ❑ Comprensión del LN: proceso de correspondencia de una forma de entrada a otra representación de salida útil para una cierta tarea
  - ❑ No existe “El Lenguaje” en el que representar todos los significados
    - ❑ Cualquiera de los sistemas de representación vistos sería un candidato
    - ❑ Y aún nos faltaría definir el vocabulario que se utilizará (los predicados concretos, los marcos o lo que corresponda...)
  - ❑ Llamaremos a este lenguaje final, sea el que sea, **lenguaje objeto**
- ❑ Su elección depende de lo que haya que hacer con los significados una vez contruidos
  - ❑ Aplicación de LN: lenguaje objeto diseñado para procesar LN
    - ❑ Con primitivas como las de las DCs o CYC
  - ❑ LN como interfaz: lenguaje objeto = entrada a otro programa
    - ❑ Aún así, conviene usar representación intermedia BC

# Análisis semántico

- ❑ El objetivo de interpretación semántica que se plantea depende totalmente de la aplicación a desarrollar

Aplicación	Salida
Interfaz en LN a Base de Datos	Lenguaje de consulta BD
Interfaz Sistema Operativo	Órdenes del SO
Sistema de diálogo	Respuesta adecuada en LN
Analizador de noticias	Relleno de una plantilla
Sintetizador de textos	Texto en LN
Traductor Automático	Texto en LN
Sistema de comprensión de texto	Lenguaje de representación de conocimiento

# Análisis semántico

## ❑ Tipos de aplicaciones

- ❑ Más simples: salida producida directamente a partir de la entrada
- ❑ Más complejas: se usa representación intermedia del significado

## ❑ Análisis semántico

- ❑ Procesamiento léxico: buscar las palabras en un diccionario y desambiguar (en lo posible) teniendo en cuenta el contexto local
  - ❑ Las conexiones estructurales entre la estructura sintáctica y la semántica permiten seleccionar entre los distintos posibles resultados del análisis sintáctico: sólo aquellos cuyo significado tenga sentido
- ❑ Procesamiento a nivel de oración: representación semántica

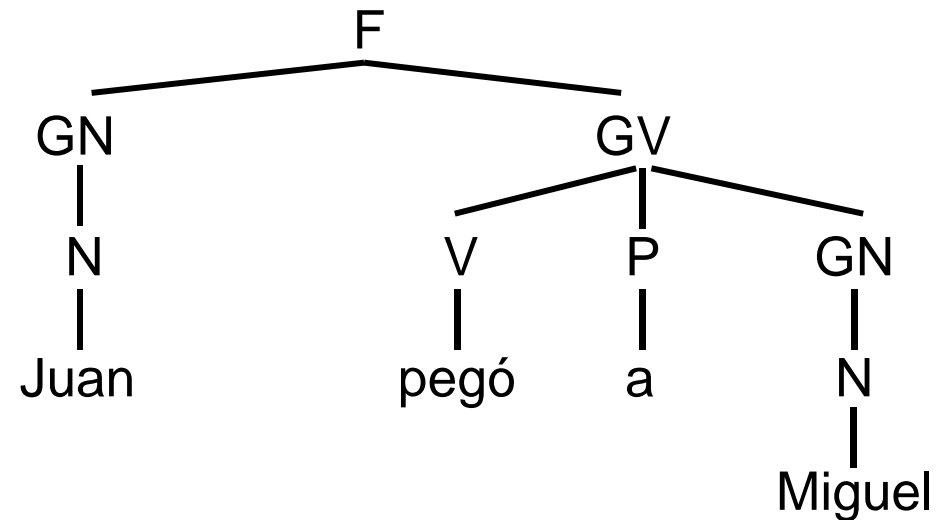
## ❑ Algunos de los métodos utilizados

- ❑ Encaje de patrones (*pattern-matching*)
- ❑ Gramáticas semánticas
- ❑ Gramáticas de casos
- ❑ Formas lógicas

# Análisis sintáctico vs. análisis semántico

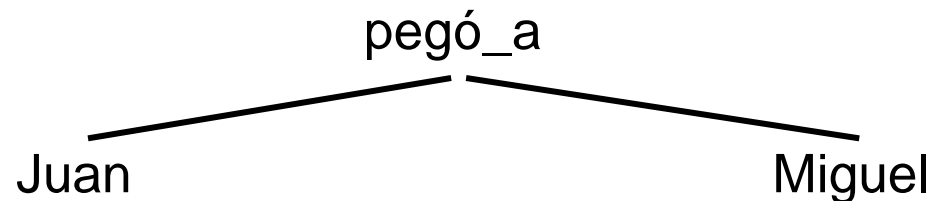
□ Ejemplo: Juan pegó a Miguel

□ Ejemplo de árbol de análisis sintáctico



□ Ejemplo de árbol de análisis semántico

□ El significado del verbo es la parte central de la semántica



# Encaje de patrones

- ❑ Es un mecanismo muy primitivo utilizado en sistemas como ELIZA (simulaba a un psicólogo)
- ❑ No se puede considerar una verdadera interpretación semántica, ya que sólo daba respuestas aparentemente adecuadas en función de los patrones que había reconocido
  - ❑ No hay comprensión

❑ Por ejemplo:

Patrón

Respuesta

X **always** Y

Can you think of a specific example?

X **you** Y **me**

Why do you think I Y you?

I **remember** X

Why do you remember X just now?

My {family member} **is** Y

Who else in your family is Y?

**Yes, he is**

You seem very positive

# Ejemplo de funcionamiento de Eliza

Patrón

X **always** Y

X **you** Y **me**

Respuesta

Can you think of a specific example?

Why do you think I **Y** you?

Paciente: People **always** stare at me

Eliza: Can you think of a specific example?

Paciente: For instance **you** keep staring at **me**

Eliza: Why do you think I **keep staring at you?**

❑ En inglés funciona mejor por el tipo de morfología del idioma

# Gramáticas semánticas

- ❑ Incorporan conocimiento sintáctico, semántico y pragmático en un conjunto de reglas que constituye una gramática
  - ❑ El resultado del análisis es una descripción semántica de la frase
- ❑ Las categorías son semánticas en vez de sintácticas
  - ❑ La posición de la palabra en la frase se usa para dar significado
  - ❑ Suele haber una acción semántica asociada a cada regla
- ❑ Simplifican muchos aspectos (concordancias, ...)
- ❑ Se utilizan en interfaces a programas como bases de datos, sistemas expertos, sistemas operativos, etc.
- ❑ Por ejemplo, el sistema LIFER para acceso a una base de datos sobre barcos
  - ❑ En la pregunta hay que identificar por qué atributo me están preguntando y por qué barco concreto



# Sistema LIFER

- Gramática semántica (*sin acciones semánticas asociadas*):

```
pregunta --> comienzo, [the], atributo, [of], barco.  
comienzo --> [what-is]; [tell-me]; [can-you-tell-me].  
atributo --> [length]; [beam]; [class].  
barco --> [the], nombrebarco; nombreclase, [class,  
ship].  
nombrebarco --> [kennedy]; [enterprise].  
nombreclase --> [kitty-hawk]; [lafayette].
```

- *atributo y barco es lo que interesa (no categorías sintácticas)*

- Ejemplos de preguntas admitidas:

What is the length of the kennedy

Tell me the class of the enterprise

What is the beam of lafayette class ship

# Sistema LIFER

- ❑ En Prolog:

```
?- pregunta([what, is, the, length, of, the,  
            kennedy], []).
```

- ❑ Como lo que queremos es construir una interfaz a la base de datos, no basta con analizar la frase sino que hay que devolver (o producir) el nombre del barco y la característica por la que se pregunta

- ❑ Supongamos que la base de datos también está implementada en Prolog

```
barco_atrib_val(kennedy, length, 1200).  
barco_atrib_val(kennedy, beam, 250).  
barco_atrib_val(kennedy, class, lafayette).  
barco_atrib_val(enterprise, length, 1100).  
...
```

# Sistema LIFER

- La gramática quedaría (*con acciones semánticas*):

```
pregunta(Nombre, Atributo) -->
    comienzo, [the],
    atributo(Atributo),
    [of], barco(Nombre).

comienzo --> [what,is]; [tell,me];
    [can,you,tell,me].

atributo(Atributo) -->
    [Atributo], {es_atributo(Atributo)}.

barco(Nombre) -->
    [the], nombre_barco(Nombre);
    nombre_clase(Nombre), [class,ship].

nombre_barco(Nombre) -->
    [Nombre], {es_nombre_barco(Nombre)}.

nombre_clase(Nombre) -->
    [Nombre], {es_nombre_clase(Nombre)}.
```

# Sistema LIFER

- Y el “diccionario” así:

```
es_atributo(Atributo):- barco_atrib_val(_, Atributo, _), !.  
es_nombre_barco(Nombre):-  
    barco_atrib_val(Nombre, _, _), !.  
es_nombre_clase(Nombre):-  
    barco_atrib_val(_, class, Nombre), !.
```

- Una vez analizada la pregunta hay que hacer la consulta a la B.D.  
`barco_atrib_val(Nombre, Atributo, Val)` y escribir `Val`

- Por ejemplo

```
procesa_pregunta(Xs) :-  
    pregunta(Nombre, Atributo, Xs, []),  
    barco_atrib_val(Nombre, Atributo, Val),  
    write('El valor del atributo es '), write(Val).
```

```
?- procesa_pregunta(['what-is-the-length-of-the-kennedy']).  
El valor del atributo es 1200
```

# Ventajas e inconvenientes

- ❑ Método eficiente para la interpretación del LN en dominios reducidos, eliminando mucha ambigüedad sintáctica
  - ❑ El resultado del análisis se puede usar inmediatamente
    - ❑ En el ejemplo del sistema LIFER, se han juntado todas las fases en una puesto que sólo interesa el resultado del análisis semántico
    - ❑ Y se ha hecho la “conexión con la B.D.” inmediatamente
- ❑ Carecen de abstracciones sintácticas, por lo que muchas construcciones similares se repiten en distintas categorías semánticas. Debido a ello, crecen demasiado rápidamente
  - ❑ Al poder necesitarse muchas reglas, el proceso de análisis puede resultar muy costoso
- ❑ Evolucionan hacia la integración de abstracciones sintácticas y semánticas en la misma gramática
  - ❑ Se evitan las ambigüedades de análisis estrictamente sintácticas teniendo en cuenta la semántica

# Ventajas e inconvenientes

- ❑ Tras muchos experimentos de uso de gramáticas semánticas en una gran variedad de dominios, las conclusiones fundamentales parecen ser
  - ❑ Gran utilidad para producir interfaces restringidas para el LN muy rápidamente
  - ❑ Pero no ofrecen una solución global al problema de la comprensión del lenguaje natural
    - ❑ Imposibilidad de capturar generalizaciones lingüísticas importantes

# Gramáticas de casos

- ❑ Aproximación alternativa a la combinación de las interpretaciones sintáctica y semántica
  - ❑ Se usan reglas para describir las regularidades sintácticas
  - ❑ Las estructuras producidas por las reglas se corresponden con relaciones semánticas en lugar de sintácticas
- ❑ Se realiza un análisis de la frase para determinar los distintos casos gramaticales (papeles o roles) que juegan las palabras que la componen (agente, objeto, lugar...)
- ❑ El significado de una frase suele representarse con un marco cuya *cabeza* es el verbo principal y cuyos *slots* se corresponden a los distintos casos

**imprimió**

**agente: Susana**

**objeto: Archivo**

# Gramáticas de casos

- El objetivo es que distintas frases que significan lo mismo (distintos árboles sintácticos, misma semántica) deberían representarse con un mismo marco

- *“Susana imprimió el archivo”*

- *“El archivo fue impreso por Susana”*

**imprimió**

**agente: Susana**

**objeto: Archivo**

- No hay un acuerdo claro acerca del conjunto correcto de casos

- Algunos obvios:

- agente, instrumento, lugar, origen, destino, beneficiario, objeto, tiempo...

- Dirigido por las entradas léxicas asociadas con los verbos

- Análisis basado en expectativas (*las que da el verbo*)



# Gramáticas de casos

- Por ejemplo, la frase

*“En el río Juan pescó una trucha para Pedro con una caña”*

podría dar como resultado del análisis

**[PESCAR**

**[frame de casos**

**agente: JUAN**

**objeto: TRUCHA**

**instrumento: CAÑA**

**lugar: RÍO**

**beneficiario: PEDRO]**

**[forma verbal**

**tiempo: pasado**

**voz: activa]**

**]**

# Gramáticas de casos

- ❑ Se extraen el verbo principal y el sujeto
  - ❑ Se localiza el verbo en el diccionario: varias posibles entradas
- ❑ Se analiza el resto de la frase buscando componentes para rellenar las ranuras vacías de la estructura verbal
- ❑ En los marcos de casos se suelen representar explícitamente ciertas heurísticas (**marcadores**) sobre cómo localizar el caso en la frase junto con restricciones sobre sus posibles valores
  - ❑ **Marcadores posicionales**: sirven para establecer cuál es la posición habitual del caso dentro de la frase
    - ❑ Ejemplo: el objeto de la frase suele ser el complemento directo, el cual suele ser el primer nombre tras el verbo
  - ❑ **Marcadores léxicos**: palabras con las que suele comenzar el caso
    - ❑ Ejemplo: preposiciones que introducen el caso (lugar → “en”)
  - ❑ **Marcadores semánticos**: establecen restricciones sobre los valores que puede tomar un caso
    - ❑ Ejemplo: un actor es un ser animado

# Gramáticas de casos: ejemplo

- Para analizar frases como “*Copia los ficheros ejecutables de la biblioteca del sistema a mi directorio*” necesitaríamos algo como

[COPIAR

[objeto:

marcador posicional: objeto\_directo

marcador semántico: <estructura de información> ]

[origen:

marcador léxico: “de” | “desde”

marcador semántico: <almacén de información> |  
<dispositivo de entrada> ]

[destino:

marcador léxico: “a” | “en”

marcador semántico: <almacén de información> |  
<dispositivo de salida> ]

]

# Formas lógicas

**Objetivo:** Representación del significado de la oración, llamada **Forma Lógica (FL)**

**Definiciones:**

Al proceso de emparejar una oración con su FL se conoce como **interpretación semántica**;

Al proceso de emparejar una FL con el lenguaje de representación final del conocimiento se denomina **interpretación contextual**

**Ejemplo:** Las siguientes formas lógicas son similares bajo diferente contexto.

*Dime los ríos que nacen en Madrid y desembocan en Valencia*  
preg(X, río(X) & nacer(X,madrid) & desembocar(X,valencia))

*Dime los vuelos que salen de Madrid y llegan a Valencia*  
preg(X, vuelo(X) & salir(X,madrid) & llegar(X,valencia))

# Formas lógicas

## Modelo estructural:

- El principio de **composicionalidad**: supone que el significado de una expresión compleja está en función de los significados de sus partes y de las reglas sintácticas mediante las cuales se combinan
- *La unidad básica es el sentido de una palabra*, o en general de una unidad léxica
- Los diferentes sentidos de las palabras pueden organizarse en conjuntos de clases de objetos en los que se clasifican las palabras:

## **Ontologías**

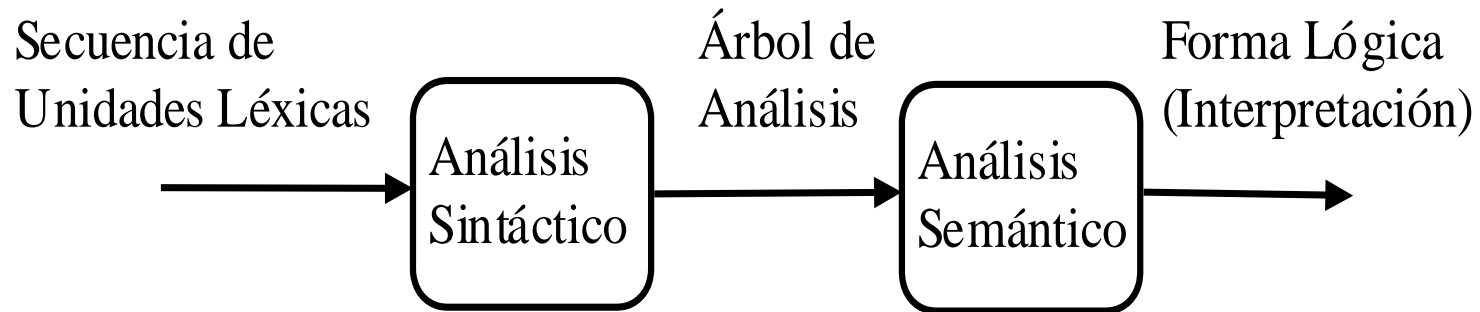
- Existen dos aproximaciones principales al proceso de interpretación semántica:

*interpretación guiada por la sintaxis*

*interpretación guiada por la semántica*

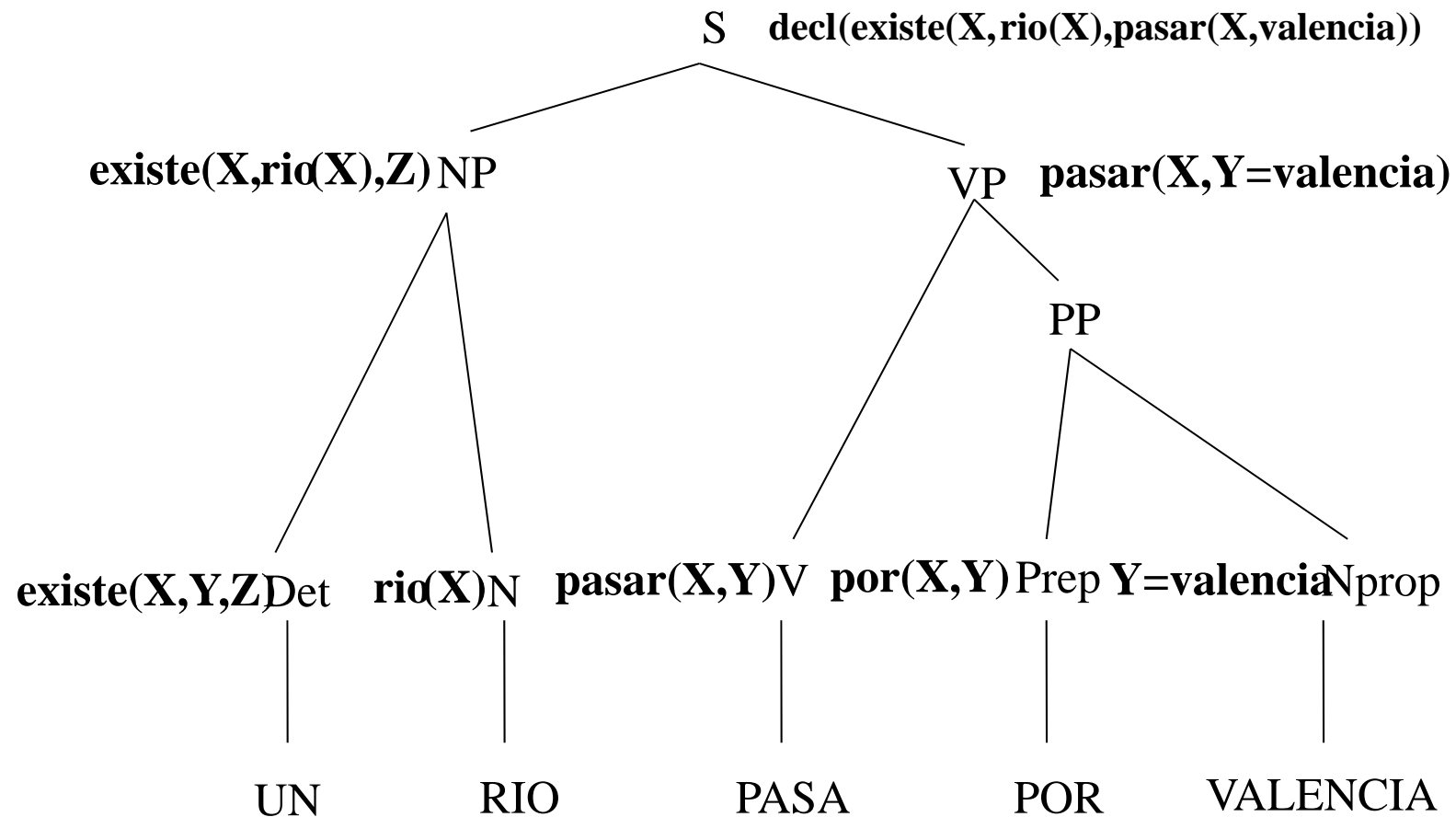
# Formas lógicas

Interpretación guiada por la sintaxis: propiedad de *composicionalidad*



# Formas lógicas

Ejemplo: Interpretación guiada por la sintaxis: “Un río pasa por Valencia”



# Formas lógicas

Ejemplo: Interpretación guiada por la semántica: "Un río pasa por Valencia"

**Fundamento:** si lo que finalmente se desea obtener es el significado de una oración, quizás lo más eficiente sea que el método de análisis se guíe por la semántica y que utilice restricciones sintácticas limitadas

**Representación:** *marcos* (frames) que se corresponden con acciones y tipos de objetos y *slots* para guardar instancias de acciones, objetos u otros marcos.

PASAR: AGENTE: río  
OBJETO: valencia



# Formas lógicas

## *El Lenguaje de la Forma Lógica (LFL):*

El LFL es el lenguaje de representación del significado de la oración

La unidad básica de representación del significado es el **sentido** de las palabras en el contexto de la oración y serán los átomos o constantes de la representación

**términos:** constantes que describen objetos en el mundo, incluyendo objetos abstractos tales como eventos o situaciones.

**Predicados:** constantes que describen las relaciones que se establecen entre los términos y propiedades de éstos.

# Formas lógicas

## Interpretación de las palabras

o en general de una proposición del lenguaje, se forma mediante un predicado seguido por un número apropiado de términos que son sus argumentos.

- Nombres propios (Juan, Luis,...): Términos
- Nombres comunes (perro, árbol,...): Predicados unarios
- Verbos (correr, querer,...): Predicados  $n$ -arios, donde  $n$  depende de los términos del verbo, etc.

## Ejemplo

"*Juan planta un árbol*" , según el LFL sería:

**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 *planta*,
- **existe(X,B,F)** es el cuantificador existencial introducido por el artículo *un*.

# Formas lógicas

## Interpretación de las conectivas

**Operadores lógicos:** disyunción ( $\vee$ ), conjunción ( $\wedge$ , &), implicación ( $\supset$ ,  $\rightarrow$ ) o negación ( $\neg$ ).

Una proposición negada se representa: *negación proposición*

Ejemplo: *Juan no planta un árbol*: **no** ( $\text{existe}(X, \text{arbol}(X), \text{plantar}(\text{juan}, X))$ )

La combinación de dos o más proposiciones mediante una conectiva se representa usualmente:

*conectiva proposición1 proposición2* o bien,  
*proposición1 conectiva proposición2*

Ejemplo:

*Juan planta un árbol y Juan riega un árbol*  
 $\text{existe}(X, \text{arbol}(X), \text{plantar}(\text{juan}, X)) \ \# \ \text{existe}(X, \text{arbol}(X), \text{regar}(\text{juan}, X))$

# Formas lógicas

## Interpretación de los cuantificadores

En el Cálculo de Predicados de Primer Orden hay dos cuantificadores: para todo ( $\forall$ ) y existe ( $\exists$ ).

*Otros: muchos, pocos, algunos, varios, un, unos, etc., Son introducidos por los determinantes del nombre: artículos, adjetivos indefinidos, demostrativos, numerales, etc.*

**Cuantificadores generalizados** de la forma  $q(K, S1, S2)$ ,

- $q$  es el cuantificador ("el", "los", "uno", "muchos", "otro", etc.),
- $K$  es una variable introducida por la cuantificación
- $S1$  y  $S2$  son fórmulas donde aparece  $K$ .

Si por ejemplo, se está tratando el sujeto de una frase,  $S1$  se corresponde con la representación del sujeto y  $S2$  con la del predicado, donde  $S1$  y  $S2$  también pueden estar cuantificados.

**Ejemplo:** *Muchos hombres pasean* será  $\text{muchos}(X, \text{hombre}(X), \text{pasear}(X))$ .

# Formas lógicas

## Los cuantificadores

- **Cuantificador existencial restringido:** representado por  $\text{ex}(x, B, F)$  que equivale a la fórmula  $\exists x (B \wedge F)$  donde  $x$  es una variable libre en  $B$ , siendo  $B$  y  $F$  Formas Lógicas.
- **Cuantificador universal restringido:** representado por  $\text{todo}(x, B, F)$  que equivale a la fórmula  $\forall x (B \supset F)$  donde  $x$  es una variable libre en  $B$ , siendo  $B$  y  $F$  Formas Lógicas
- **Cuantificador universal negado restringido:** representado por  $\text{ningun}(x, B, F)$  que equivale a la fórmula  $\forall x (B \supset \neg F)$  donde  $x$  es una variable libre en  $B$ , siendo  $B$  y  $F$  Formas Lógicas

## Interpretación de los plurales

Los plurales en castellano introducen un nuevo tipo de ambigüedad, es decir un plural puede interpretarse de diferentes formas:

- plural **colectivo** la propiedad del predicado se aplica al conjunto de entidades afectadas.
  - *Juan y María se encontraron*  $\Rightarrow$  *encontrarse (Juan, María)*.
- plural **distributivo** la propiedad del predicado debe distribuirse entre todas las combinaciones posibles entre las entidades afectadas.
  - *Juan y María saben Latín y Griego*,
    - $\Rightarrow$   $\text{saber}(\text{Juan}, \text{Latín}) \wedge \text{saber}(\text{María}, \text{Latín}) \wedge \text{saber}(\text{Juan}, \text{Griego}) \wedge \text{saber}(\text{María}, \text{Griego})$ .
- plural **respectivo** que debe distribuirse apropiadamente.
  - *Juan y María saben Latín y Griego respectivamente*
    - $\Rightarrow (\text{Juan}, \text{Latín}) \wedge \text{saber}(\text{María}, \text{Griego})$ .

# Formas lógicas

## Roles temáticos

**Rol temático** o *casos*: relaciones semánticas entre el verbo y sus argumentos

En el ejemplo *Isabel cruzó el océano con el velero*, *Isabel*, *océano* y *velero*, no juegan el mismo papel (rol) semántico en cada una de las oraciones. *Isabel* es el **actor**, el *océano* es el objeto (**Rol de Tema**) y el *velero* es el instrumento (**Rol de Instrumento**):

existe(X, tema(océano(X)), past(cruzar(agente(isabel), X)))

AGENTE: realiza la acción (generalmente el sujeto)

TEMA: para el verbo transitivo X será la respuesta a la pregunta ¿Qué es X?. Para verbos intransitivos el TEMA es el sujeto que no será AGENTE.

INSTR: material o herramienta usada para realizar el evento (velero).

CO-AGENTES: existen ejemplos donde se describen situaciones donde dos personas realizan una acción juntas *María viaja a Egipto con Isabel*.

# Formas lógicas

## Tipos de oraciones

- Oraciones declarativas o aserciones: DECL  
*El Turia pasa por Valencia:* **decl**(pasar(turia,valencia))
- Oraciones interrogativas de cierto-falso: SI\_NO  
*¿El Turia pasa por Valencia?:* **sino**(pasar(turia,valencia))
- Oraciones interrogativas de tipo cantidad: CANT  
*¿Cuántos ríos pasan por Valencia?:* **cant**(X, rio(X) & pasar(X,valencia))
- Oraciones interrogativas de tipo general: PREG  
*¿Qué ríos pasan por Valencia?:* **preg**(X, rio(X) & pasar(X,valencia))
- Oraciones imperativas u órdenes: IMP  
*Dime los ríos que pasan por Valencia:* **imp**(X, rio(X) & pasar(X,valencia))



# Formas lógicas

## *Interpretación semántica y composicional*

### Objetivo:

Traducir las oraciones del Lenguaje Natural al Lenguaje de la Forma Lógica elegido.

El proceso se denomina **interpretación semántica**. Consiste en obtener la forma lógica asociada a la oración a partir del árbol de análisis.

### Método: *proceso composicional*

El significado de cada constituyente de la oración se deriva solamente a partir del significado de sus subconstituyentes. A partir de la interpretación de los constituyentes básicos, las unidades léxicas, y de manera incremental se construye la interpretación de constituyentes más complejos.

La composición semántica viene a suponer que la semántica de cualquier frase es una composición de las semánticas de sus subfrases.

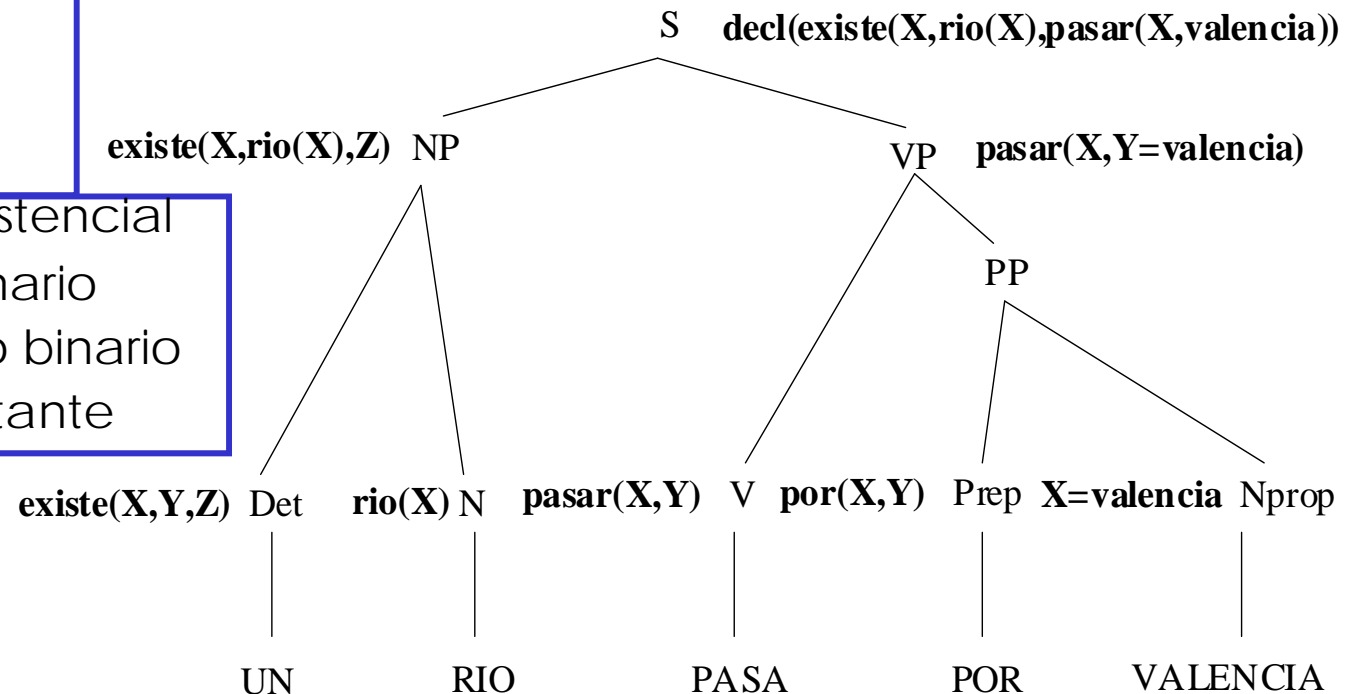
# Formas lógicas

## Procedimiento para el proceso composicional

- A cada nodo del árbol de análisis se le asigna una FL, que corresponde a la interpretación de la secuencia de palabras derivadas a partir de ese nodo.
- La FL de ese nodo se calcula a partir de las FL asociadas a los nodos hijos.
- La FL asociada al nodo raíz será la interpretación de la oración.

*Interpretación de  
"un río pasa por  
Valencia"*

UN: cuantificador existencial  
RIO: un predicado unario  
PASAR: un predicado binario  
VALENCIA: una constante

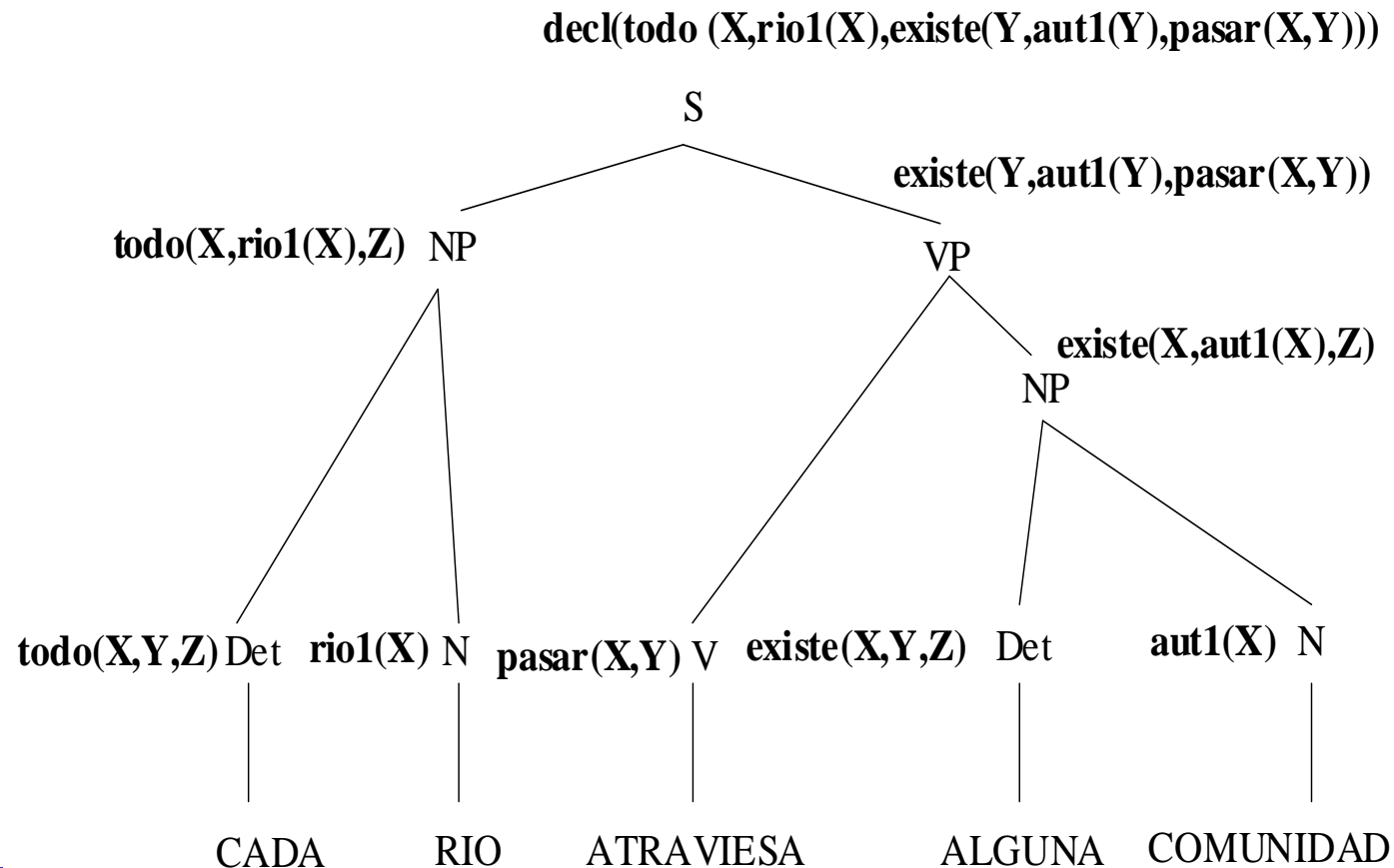


# Formas lógicas

## Problemática: Ambito de la cuantificación

Arbol de "Cada río atraviesa alguna comunidad" :

1)  $\text{decl}(\text{todo}(X, \text{rio1}(X), \text{existe}(Y, \text{aut1}(Y), \text{pasar}(X, Y))))$



# Formas lógicas

## Verificación de la consistencia semántica

Una oración puede ser correcta sintácticamente, pero no ser válida semánticamente dentro de un dominio.

- *El río Turia pasa por Valencia* sería válida en un dominio geográfico
- *El mar Mediterráneo pasa por Valencia* es correcta sintácticamente pero no lo es semánticamente.

Hay que observar la manera en que los objetos del universo se clasifican dentro de grupos por sus propiedades o rasgos semánticos. Un objeto consistiría de:

- un *tipo*: conjunto de rasgos denotando su pertenencia a un universo.
- un *referente* identificador del concepto genérico o individual.

Ejemplos:

rio(NOMRIO, CAUDAL, LONGITUD, NACE, DESEMBOCA)  
com\_aut(NOMBRE, EXTENSION, CAPITAL)  
ciudad(NOMBRE, HABITANTES, COMUNIDAD)

Cada vez que se reconoce un elemento sintáctico se debe verificar que su tipo semántico asociado es consistente con las restricciones

$s(\text{Sem}) \rightarrow np(X, \text{Scope}, \text{Sem}, \text{Rasgo}), vp(X, \text{Scope}, \text{Restric})$   
{**verificar(Rasgo, Restric)**}

# Análisis pragmático

- ❑ Análisis de los textos de forma global
  - ❑ La mayoría de los sistemas reales no lo incluyen
- ❑ Estudiar los fenómenos que trascienden el análisis frase a frase
  - ❑ Tener en cuenta el contexto
- ❑ Clases de fenómenos del lenguaje
  - ❑ **Anáfora o referencia anafórica:** uso de una palabra que hace referencia o reemplaza a otras palabras (que han aparecido antes, ya sea en la misma frase o en otras)
    - ❑ En particular, hace referencia al uso de pronombres
    - ❑ Por ejemplo, en muchas frases el sujeto hace referencia a un nombre que aparece en otra frase
      - ❑ *Juan está buscando a su perro. Él piensa que lo encontrará.*

# Análisis pragmático

## ❑ Tratamiento de las **anáforas**

### ❑ Técnicas basadas en mantener el **foco de atención**

- ❑ Los pronombres sucesivos se sustituyen por el foco de atención más cercano que concuerde sintáctica y semánticamente
- ❑ Se mantiene la lista de entidades del discurso que se han referenciado en las últimas frases

### ❑ Hace falta mucho conocimiento

- ❑ *Puse el disco en el armario. Después lo cogí.*
- ❑ *Puse el disco en el armario. Después lo cerré.*
- ❑ *Sustituir “lo” por un nombre de la frase anterior: ¿”disco” o “armario”?*

### ❑ Algunas referencias no son directas sino a través de *es-un* o *es-parte-de*

- ❑ *Compré un gato. El animal no me dejaba dormir.*
- ❑ *Compré un coche. Las ruedas estaban gastadas.*

# Análisis pragmático

- ❑ Más clases de fenómenos del lenguaje
  - ❑ **Elipsis:** omitir partes de la frase que tienen que completarse a partir del contexto para determinar el significado
    - ❑ Por ejemplo, sujeto omitido
    - ❑ *Juan perdió su perro. Paco también.*
      - ❑ Referencia a acciones y no únicamente a cosas o personas...
  - ❑ Construcciones extra-gramaticales
    - ❑ No respetan la gramática aunque podemos darle significado
  - ❑ Construcciones metalingüísticas
    - ❑ Referencias a frases anteriores
    - ❑ *“Aunque dije que iba a llover realmente no lo parece”*
  - ❑ Intenciones de los interlocutores
    - ❑ Difícil de tratar
    - ❑ La misma frase dicha por dos personas distintas puede significar cosas distintas según la intención del que habla

# Análisis pragmático

- ❑ Se necesita un conocimiento del dominio muy profundo para entender las relaciones entre distintas frases. Los programas que lo consiguen
  - ❑ o bien cuentan con unas bases de conocimiento muy grandes,
  - ❑ o bien el dominio de discurso está tan restringido que basta con una base de conocimiento más limitada
- ❑ El análisis pragmático sólo es realista si restringimos mucho el dominio
  - ❑ Sistemas de consulta a bases de datos: catálogos de pisos, ordenadores...
  - ❑ El vocabulario está predefinido y las frases tienen patrones claros
- ❑ La forma en la que se organiza el conocimiento es crítica para que los sistemas tengan éxito en esta fase



# Generación de lenguaje natural

- ❑ Aplicaciones en las que el lenguaje “enlatado” no es suficiente
  - ❑ Sistemas de diálogo
  - ❑ Sistemas expertos de soporte de decisiones → generación de explicaciones
  - ❑ Tutores inteligentes
  - ❑ Sistemas de traducción automática
- ❑ Dos componentes
  - ❑ Módulo estratégico
    - ❑ Cubre la generación profunda del lenguaje
    - ❑ Se ocupa de qué decir
    - ❑ Depende totalmente de la aplicación
  - ❑ Módulo táctico
    - ❑ Generación superficial del lenguaje
    - ❑ Se ocupa de cómo decirlo

# Generación de lenguaje natural

- ❑ Para la **generación estratégica** se tienen en cuenta
  - ❑ Objetivos
    - ❑ Qué información se pretende transmitir al usuario
  - ❑ Planes para la consecución de estos objetivos
    - ❑ Cómo organizar y fraccionar la información que vamos a transmitir
  - ❑ Modelos de usuario
    - ❑ La información que vamos a presentar depende del usuario
    - ❑ Permiten tener en cuenta
      - ❑ Lo que sabe el usuario
      - ❑ Para qué va a usar la información el usuario
    - ❑ Estáticos o dinámicos
      - ❑ Los modelos de usuario estáticos no cambian
      - ❑ Los modelos dinámicos son aquéllos en los que el conocimiento del usuario va cambiando.
        - ❑ Se aplican a sistemas de aprendizaje (por ejemplo, a tutores inteligentes)

# Generación de lenguaje natural

- ❑ La **generación táctica** se compone de dos tareas
  - ❑ Selección léxica
    - ❑ Elegir de un diccionario las palabras que se van a usar para transmitir la información
  - ❑ Selección sintáctica
    - ❑ Elegir qué tipo de frases se van a usar
    - ❑ Basada en el uso de DCGs