



## *Licenciatura en Ingeniería en Sistemas Inteligentes*

Unidad de aprendizaje:  
**Procesamiento de Lenguaje Natural**

### ***Tema 1.3: Niveles de descripción del Lenguaje***

La unidad de aprendizaje (UA) de **Procesamiento de lenguaje Natural** tiene como área curricular **Herramientas para los sistemas** inteligentes y forma parte del núcleo sustantivo.

RECONOCIMIENTO DE PATRONES

INTRODUCCIÓN AL RECONOCIMIENTO DE PATRONES 3 0 4

ALGORITMOS DE AGRUPAMIENTO 2 1 5

REFORZAMIENTO DE PATRONES SINTÁCTICO-ESTRUCTURAL 4 1 9

ALGORITMOS DE CLASIFICACIÓN 2 1 5

REDES NEURONALES SUPERVISADAS 4 1 9

REDES NEURONALES NO SUPERVISADAS 4 1 9

DESCUBRIMIENTO DE CONOCIMIENTO A PARTIR DE DATOS

SISTEMAS BASADOS EN CONOCIMIENTO 3 2 8

MINERÍA DE DATOS I 2 2 6

MINERÍA DE DATOS II 2 2 6

ÁRBOLES DE DECISIÓN 3 2 8

INTERPRETACIÓN DE IMÁGENES 2 2 6

EVALUACIÓN E INTERPRETACIÓN DEL CONOCIMIENTO 1 2 4

HERRAMIENTAS PARA LOS SISTEMAS INTELIGENTES

LÓGICA MATEMÁTICA 2 1 5

INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL 2 1 5

INTRODUCCIÓN AL TRATAMIENTO DE IMÁGENES 4 1 9

PROCESAMIENTO DE LENGUAJE NATURAL 3 1 7

ROBÓTICA 3 1 7

ALGORITMOS GENÉRICOS 3 2 8

LÓGICA DIFUSA 4 1 9

TALLER CON WEKA 1 2 4

TALLER DE MINERÍA DE DATOS 1 2 4

REDES NEURONALES MODULARES 1 2 4

RAZONAMIENTO BASADO EN CASOS 1 2 4

GRAFICACIÓN 1 2 4

PROGRAMACIÓN CON MATLAB 1 2 4

AGENTES INTELIGENTES 1 2 4

MÁQUINAS DE SOPORTE VECTORIAL 1 2 4

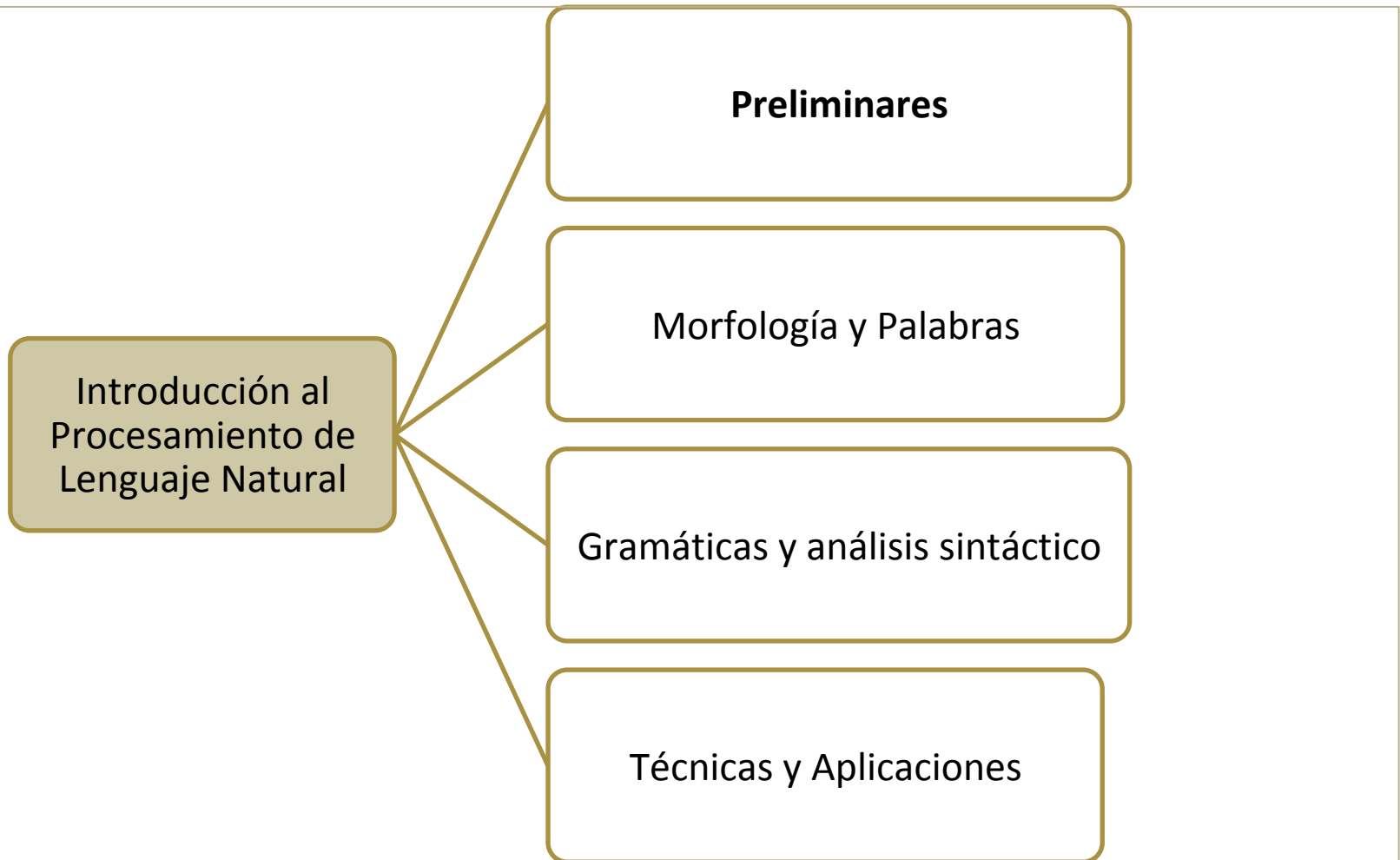
# Objetivo de la Unidad de Aprendizaje

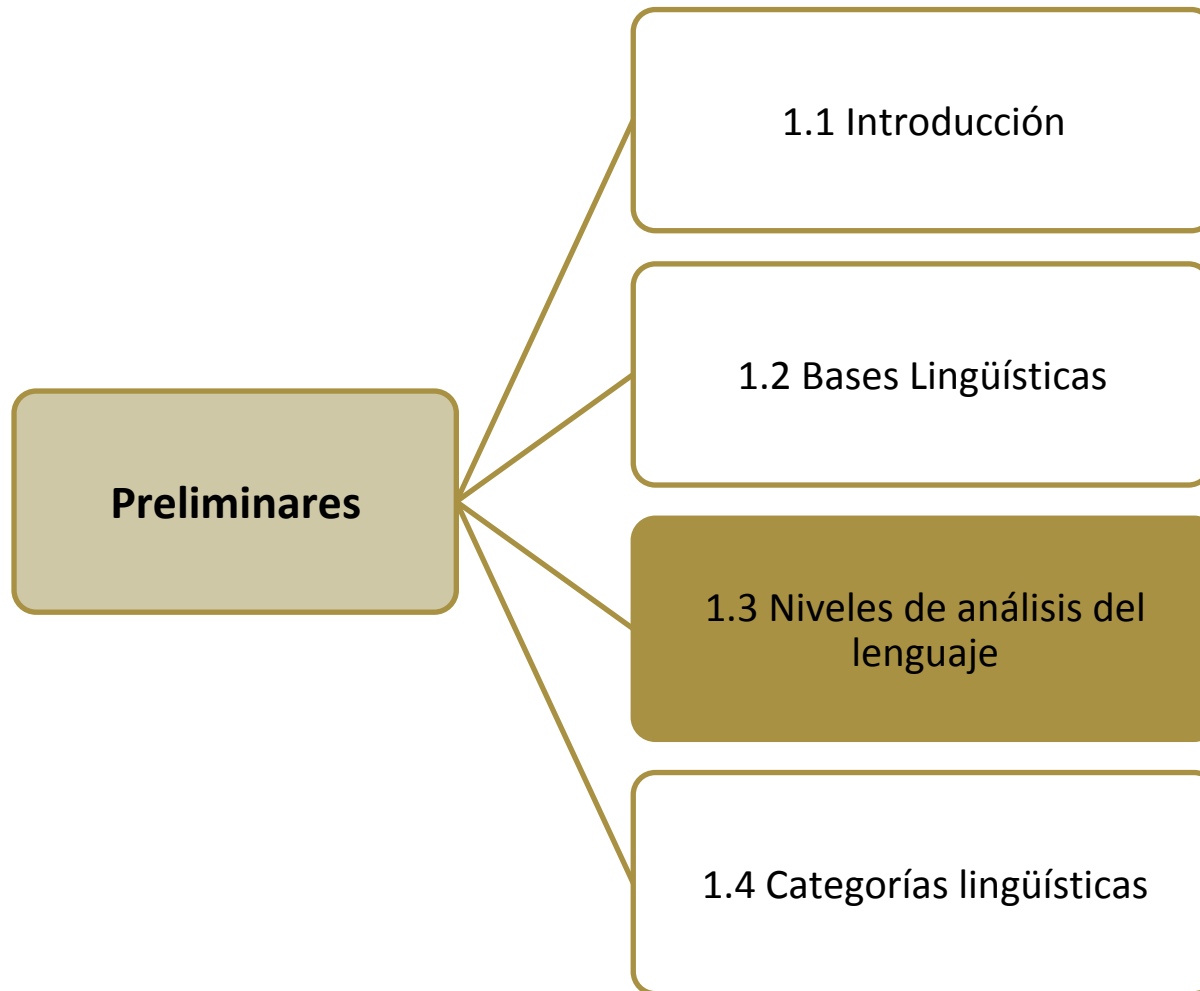
Proporcionar a los alumnos de conocimiento sobre las técnicas que se emplean para el desarrollo de sistemas concentrándose en los aspectos que conciernen a los métodos estadísticos de tratamiento de texto y destacando las siguientes tareas de clasificación, recuperación de documentos, la resolución de la ambigüedad léxica y el etiquetado sintáctico.

El presente material tiene como objetivo cubrir el tema: Niveles de análisis del Lenguaje

El alumno será capaz de conocer todo lo concerniente a los Niveles de Análisis del Lenguaje para comprender el Procesamiento de Lenguaje Natural.

# Estructura de la Unidad de Aprendizaje





# Comprender el Lenguaje Natural

La comprensión exige:

- Extraer el significado individual de las palabras
- Comprender las relaciones entre las palabras
- Referir el significado literal al contexto de actuación del sistema

Todo esto se alcanza a través de un **análisis de los componentes del lenguaje a diferentes niveles.**

# Niveles de descripción del lenguaje

---

Nivel fonético

---

Nivel fonológico

---

Nivel léxico

*Léxicos computacionales*

---

Nivel morfológico

*Analizadores morfológicos*

---

Nivel sintáctico

*Analizadores sintácticos*

---

Nivel semántico

*Redes semánticas (WordNet) Corpus*

---

Nivel pragmático

---

Nivel ilocutivo

---

# Nivel de análisis: Fonética

Estudia la acústica (cómo se articulan) y fisiología (cómo funcionan los órganos –lengua, paladar, labios, nariz, etc.) al producir los sonidos

➤ Los sonidos son materiales y audibles

saco, zapato, cepa

grafías diferentes → mismo fonema /s/



# Nivel de análisis: **Fonológico**

- Es el tratamiento de los sonidos para detectar unidades de expresión (palabras).
- Parte de la lingüística que estudia los fonemas o descripciones teóricas de los sonidos vocálicos y consonánticos que forman una lengua.
- Se aplica en el **procesamiento del lenguaje oral**.



# Nivel de análisis: Fonológico

Investiga qué diferencias fónicas están relacionadas con diferencias de significado, qué relaciones establece un fonema con los demás dentro de un paradigma

**c**ala /k/

**p**ala /p/

**m**ala /m/

**s**ala /s/

**b**ala /b/

**-a**la (el fonema omitido determina la diferencia)

Un fonema es la representación abstracta de un sonido que pertenece a la lengua, al sistema.

# Nivel de análisis: Fonológico

- No hay correspondencia biunívoca entre fonema y grafía (letras)
  - /b/ → {b, v}
  - /y/ → {ll, y (inicial de sílaba)}
  - /k/ → {k, qu, c (ca,co,cu)}
  - /s/ → {s, z, c (ce,ci)}
- En el español de México hay
  - 22 fonemas
  - 31 letras incluida la h
  - h (no representa sonido) tiene valor ortográfico

- Sistemas de análisis y síntesis del habla
- Reconocimiento de voz
- Generación de voz
- Estos desarrollos se dirigen hacia el interés de diseñar programas para el dictado y reconocimiento automático.
- No presentan aún desarrollos significativos puesto que no se logra modelar flexiones y variaciones en entonaciones particulares en las lenguas

# Nivel de análisis: **Léxico**

El concepto de léxico encierra varios significados, todos ligados al mundo de lingüística. Léxico es el vocabulario de un idioma o de una región, el diccionario de una lengua o el caudal de modismos y voces de un autor.

- Distingue entre palabras ortográficas y palabras gramaticales.
- Obtiene información léxica de diccionarios, ontologías..



# Nivel de análisis: **Léxico**

Ejemplo:

- “Ese no es el léxico apropiado para un niña de diez años”
- “Un profesional de las ciencias de la comunicación debe manejar un léxico acorde a su responsabilidad social”
- “El léxico de los adolescentes actuales parece limitarse a unas pocas palabras”

# Ambigüedad léxica

➤ “Pinchó la rueda de delante”

“rueda” puede ser nombre o verbo (*part-of-speech tagging - POS-tagging*)

➤ “Vio el banco”

“banco” puede ser el mueble para sentarse, la entidad financiera o un grupo de peces (*word sense disambiguation - WSD*)



# Léxicos computacionales

- **Diccionarios de uso común en soporte electrónico**
- **Volcado automático a una estructura de base de datos**
- **Vocabulario controlado**
- **Granularidad uniforme**
- **Aportan diferentes tipos de información léxica**

# Nivel de análisis: Morfológico

La morfología estudia la estructura de las palabras y su relación con las categorías del lenguaje.

- El objetivo del análisis morfológico automático es llevar a cabo una clasificación morfológica de las palabras.
- Por ejemplo, el análisis de la palabra *gatos* resulta en gato+Sustantivo+Genero+Plural,

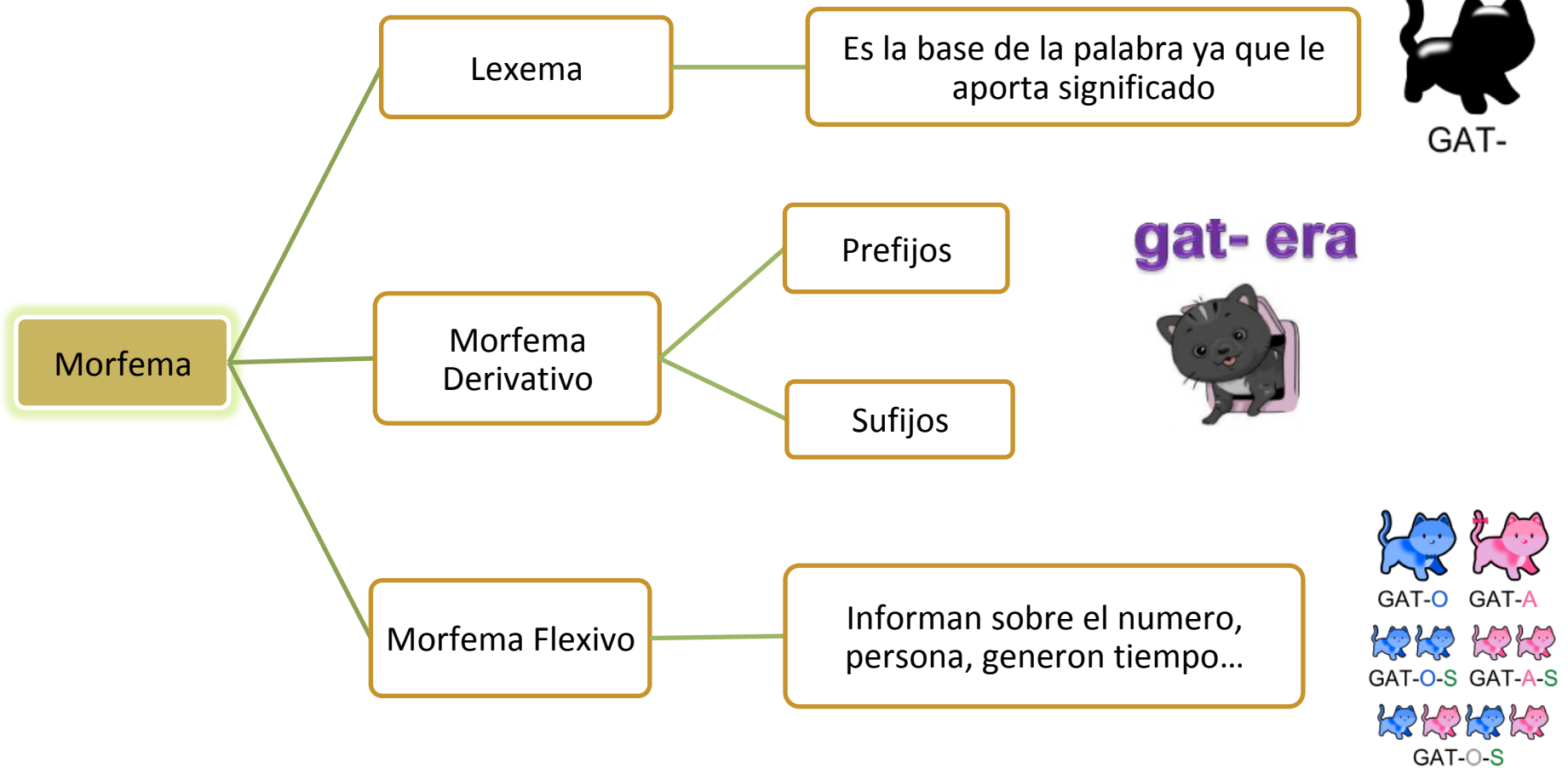
que nos indica que se trata de un sustantivo plural con género masculino y que su forma normalizada (lema) es gato.

## Nivele de análisis: **Morfológico**

Versión simple: utilización de *formarios* (listas de formas con información morfológica y los lexemas correspondientes)

- Morfemas = lexemas (o raíz) o gramemas

# Analizadores morfológicos:



# Analizadores morfológicos:

- Análisis y síntesis automático morfológico (sufijos, prefijos, raíces y flexiones)
- Morfotáctica: reglas de combinación de morfemas
- Variaciones fonológicas: cambios al combinar morfemas
- Hace falta la construcción de diccionarios robustos
- Aunque ha habido grandes desarrollos, se requiere una estandarización efectiva del lenguaje para la marcación

## Nivel de análisis: **Sintáctico**

- Incluye todo lo relacionado con el significado de las palabras.
- Se ocupa también de la clasificación de las palabras según su origen,
- Su uso, sus posibles transformaciones...



# Nivel de análisis: Sintáctico

- Reconoce, extrae y representa sintácticamente válidas (o inválidas):
- Es especialmente difícil automatizar en español donde no hay un orden fijo de las palabras que componen una oración

Ejemplo:

- ¿Estuviste en el parque el sábado?
- ¿El sábado estuviste en la parque?
- ¿Tu estabas en el parque el sábado?

# Nivel de análisis: **Semántico**

- Estudia los cambios de significación de palabras/grupos de palabras y el papel que desempeñan las categorías y sus combinaciones en el significado de la frase.
- Dependiendo de la categoría sintáctica se puede hacer referencia a diferentes tipos de la realidad.
- Construcción de redes semánticas donde se presenten conceptos y relaciones entre ellos

**Problema: Polisemia**

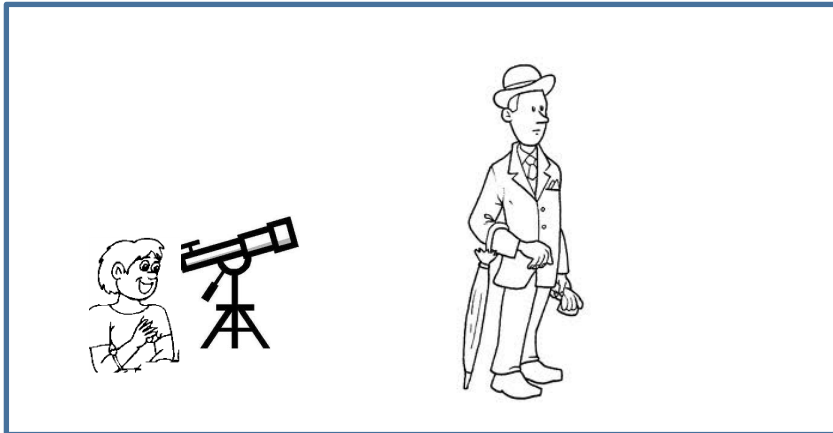
**Ejemplo: Gato**



# Nivel de análisis: **Semántico**

Otros problemas

*Vi a un hombre con un telescopio*



# Redes semánticas

- Una **red semántica** es una manera de representar el conocimiento basada en las **relaciones** que establecen los **conceptos** representados

En las redes lexicosemánticas se representan los **conceptos lexicalizados**.

- **EuroWordNet** es la red lexicosemántica más extendida, de manera que se ha constituido como un estándar de representación.

castellano, catalán e inglés

- Se utiliza para:

- El etiquetado automático de corpus
- Léxico computacional para obtener relaciones entre palabras
- Recuperación de información

...

- Encontrar sentidos y hacer separaciones entre ellos
  - Lexicografía: construcción de diccionarios generales y especializados
- Representación de conocimientos en general
  - Elaboración de terminologías especializadas
  - Ontologías

# Nivel de análisis: **Pragmático**

- Interpretación en un contexto determinado (incorpora referencias implícitas)
  - *“Le dio un libro”*
  - *“No les gustó”*
- Se encarga de como el contexto (situación) influye en la interpretación del significado
- Se preocupa como utilizamos el lenguaje para obtener diversos objetivos en el mundo que nos rodea. Utilizamos el lenguaje para informar, prometer, pedir o preguntar entre otras funciones.

# Nivel de análisis: Pragmático

El pragmatismo

Consiste en reducir

- Lo verdadero útil

Solo es verdadero  
aquello que

- Conduce al éxito
- Verificando los hechos

¿Me puedes pasar la sal?



- No es tal pregunta
- Es una forma cortés de pedir a alguien que nos haga un favor acercarnos la sal
- Hay elementos extralingüísticos



- Relaciones entre oración y el mundo externo.
- Como ya se tienen muchos problemas en la semántica, es difícil continuar el análisis en el nivel siguiente.

# Nivel de análisis: **Ilocutivo**

- Es la idea o el concepto de la frase, es decir, aquello que se dice.
- Es la intención o finalidad concreta del acto de habla.
- Tiene problemas de asignación de intenciones

*“Los platos están sucios”* (por tanto, ¡lávalos!)

# Conclusión

Para entender el procesamiento del lenguaje natural es necesario entender primero los **niveles de análisis del lenguaje**, que estudia todo lo referente al lenguaje humano, como el español, y requiere cierto grado de «entendimiento» de su contenido. Por otro lado esta ciencia desarrolla las herramientas que ayudan al lingüista en su trabajo cotidiano e incluso pueden llevar a descubrimientos lingüísticos nuevos. Ya que como bien sabemos el español (México) es difícil de procesar debido a los modismo que en nuestro país se ocupan.

# Referencias

- Domínguez Burgos, Andrés. Lingüística computacional: un esbozo. Boletín de lingüística, No. 18. pp. 104-119
- Mercado, Heiner (2008). Fundamentos de la LC [presentación de clase]
- Sidorov, Grigory (2001). Problemas actuales de lingüística computacional. Revista digital universitaria, Vol. 2, No. 1



**UAEM**