

## 3.2 Procesamiento de lenguaje natural

### 1. ¿Qué son los modelos como el de cadenas ocultas de Markov?

Los modelos de cadenas ocultas de Markov (Hidden Markov Models, HMM) son modelos estadísticos que se utilizan para representar sistemas cuya evolución está gobernada por un proceso probabilístico. Estos modelos consisten en:

- **Estados ocultos:** Los estados reales del sistema que no son directamente observables.
- **Observaciones:** Datos visibles o evidencias que dependen probabilísticamente de los estados ocultos.
- **Transiciones:** Probabilidades de pasar de un estado oculto a otro.
- **Emisiones:** Probabilidades de observar un dato dado un estado oculto.

Un ejemplo clásico de uso es el reconocimiento del habla, donde las palabras pronunciadas (estados ocultos) generan señales acústicas (observaciones) que se modelan probabilísticamente.

### 2. ¿Cómo se llaman los 3 corpus que podemos encontrar en español para realizar test en spanishNLPMoelCorpus?

Los corpus en español utilizados comúnmente en spanishNLPMoelCorpus para pruebas son el CORPES XXI, que abarca textos actuales en diversos dialectos del español. El CESS-ECE, diseñado para estudios sintácticos. Y, el Ancora un corpus anotado para el análisis sintáctico y semántico en español.

Estos corpus son utilizados para entrenar, evaluar y ajustar modelos lingüísticos en tareas de procesamiento del lenguaje natural (NLP).

### 3. ¿Qué es la gramática y la sintaxis de una lengua? Pon ejemplos.

- **La gramática** es el conjunto de reglas que definen cómo se forman palabras, frases y oraciones en una lengua. La gramática incluye la morfología (formación de palabras) y la sintaxis (estructura de oraciones).
  - Formación del plural: "gato" → "gatos" (regla gramatical para formar el plural añadiendo "s").

- Conjugación verbal: "yo hablo", "tú hablas" (reglas gramaticales para el presente de indicativo).
- **La sintaxis** es el componente de la gramática que se ocupa de cómo las palabras se combinan para formar oraciones coherentes y con significado.  
Ejemplo:
  - Oración correcta: "El perro corre rápidamente."
  - Oración incorrecta: "Corre rápidamente perro el." (violación de la sintaxis en español).

#### 4. ¿Qué herramienta web podría encontrar que realizase un etiquetado y unificación de una oración en español?

Una herramienta web que puede realizar etiquetado y unificación de una oración en español es Freeling. Es una plataforma de análisis lingüístico que ofrece:

Por un lado, etiquetado morfosintáctico que asigna categorías gramaticales a las palabras, luego la desambiguación léxica identifica el sentido correcto de las palabras; por último el análisis sintáctico representa la estructura jerárquica de una oración.

#### 5. Explica la diferencia entre morfología, semántica y pragmática.

- **La morfología** estudia la estructura interna de las palabras y cómo se forman. Se ocupa de morfemas (la unidad mínima con significado).
  - Morfología léxica: "niño" (raíz) + "-s" (sufijo plural) → "niños".
  - Morfología verbal: "correr" → "corrió" (cambio de raíz y sufijo para indicar tiempo pasado).
- **La semántica**: Se ocupa del significado literal de las palabras y oraciones.
  - Oración: "El cielo está azul." → Significado literal: El color del cielo es azul.
  - No analiza el contexto ni la intención.
- **La pragmática**: Estudia cómo el contexto afecta al significado y la interpretación de las palabras o frases.
  - "¿Puedes pasarme la sal?" → Literalmente es una pregunta, pero pragmáticamente es una solicitud.