Procesamiento de lenguaje natural

¿Qué son los modelos como el de cadenas ocultas de Markov?
El modelo oculto de Markov es un modelo probabilístico usado en la predicción de procesos aleatorios, es decir, no tenemos ningún tipo de medición o conocimiento sobre cómo se pueden hacer esas predicciones. Se denomina oculto porque no sabemos los estados intermedios, sino que simplemente partimos de probabilidades generales a partir de observaciones y vemos cómo se relacionan todas esas observaciones entre sí.

A continuación, mostramos un ejemplo de cadenas ocultas de Markov, supongamos que el clima puede ser soleado o lluvioso, y la probabilidad de cambiar de un estado al otro está dada por la siguiente tabla:

Estado Actual	Probabilidad Soleado (Siguiente dia)	Lluvioso (Siguiente dia)
Soleado	0.8	0.2
Lluvioso	0.4	0.6

- Si hoy es soleado, hay un 80% de probabilidad de que mañana también sea soleado y un 20% de que sea lluvioso.
- Si hoy es lluvioso, hay un 40% de probabilidad de que mañana sea soleado y un 60% de que siga lluvioso.

Simulación:
Supongamos que hoy es soleado, y simulamos el clima durante 7 días:

Día 1: Soleado
Día 2: Soleado (probabilidad = 0.8)
Día 3: Lluvioso (probabilidad = 0.2)
Día 4: Lluvioso (probabilidad = 0.6)
Día 5: Soleado (probabilidad = 0.4)
Día 6: Soleado (probabilidad = 0.8)
Día 7: Soleado (probabilidad = 0.8)

(Cada transición depende únicamente del estado del día anterior)

¿Cómo se llaman los 3 corpus que podemos en español en spanishNLPModelCorpus?
Los corpus de *spanishNLPModelCorpus* se llaman AnCora, CoNLL-2002 y WikiNER.

- AnCora
  AnCora es un conjunto de corpus de textos anotados en español y catalán. Fue desarrollado por el grupo TALP de la Universidad Politécnica de Cataluña, este corpus es ideal para tareas de análisis sintáctico, semántico y de etiquetado morfosintáctico.

- CoNLL-2002
  Este corpus fue creado para la conferencia CoNLL-2002, centrada en el reconocimiento de entidades nombradas (NER) en español y neerlandés. Es una referencia clásica en tareas de NER.

- WikiNER
  WikiNER es un corpus multilingüe (incluido español) extraído de Wikipedia. Fue creado para tareas de reconocimiento de entidades nombradas y anotado automáticamente mediante herramientas de PLN.

¿Qué es la gramática y la sintaxis de una lengua? Pon ejemplos.
La gramática y la sintaxis son conceptos fundamentales en el estudio de cualquier lengua, ya que describen las reglas que permiten construir y comprender oraciones de manera adecuada.

- **La gramática:**
  Es el conjunto de reglas que rigen el uso y la estructura de una lengua. Incluye todos los aspectos que determinan cómo deben formarse palabras, frases y oraciones para que sean correctas y significativas.

  Ejemplo de regla gramatical:
  En español, los adjetivos suelen concordar en género y número con el sustantivo:
  "La casa bonita" (correcto).
  "La casa bonito" (incorrecto).

- **La sintaxis:**
  Es una parte de la gramática que se ocupa específicamente del orden y la relación entre las palabras en una oración para formar estructuras coherentes.
  Ejemplo del buen y mal uso sintáctico:
  Cambiar el orden puede alterar o romper el significado:

  "El gato persigue al ratón." (correcto y lógico).
  "El ratón persigue al gato." (cambia el significado).
  "Gato el ratón al persigue." (incorrecto sintácticamente).

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

- **Spacy (con su modelo en español)**
  Es una biblioteca de procesamiento de lenguaje natural que incluye herramientas para etiquetado gramatical, análisis sintáctico y reconocimiento de entidades nombradas.

- **Stanford CoreNLP**
  Es un sistema avanzado de PLN que incluye análisis sintáctico, etiquetado gramatical y más. Aunque es más conocido para inglés, soporta español mediante modelos adicionales.

- **Freeling**
  Es una herramienta de análisis lingüístico diseñada específicamente para lenguas romances como el español.

Explica la diferencia entre morfología, semántica y pragmática.

- **Morfología:**
  La morfología estudia la estructura de las palabras y cómo se forman mediante morfemas (raíces, prefijos, sufijos).
  Ejemplo:
  La palabra corriendo →
  "Corr-" (raíz, acción de correr).
  "-iendo" (morfema verbal que indica gerundio).

- **Semántica:**
  Analiza el significado literal de palabras y oraciones en un contexto neutral.
  Ejemplo:
  El gato persigue al ratón.
  En semántica, analizamos el significado de las palabras:
  Gato = animal felino.
  Perseguir = correr detrás de algo.

- **Pragmática:**
  Estudia el significado en contexto y el uso del lenguaje en situaciones reales, considerando intención y contexto.
  Ejemplo:
  Hace frío aquí. (Literalmente, describe la temperatura, pero pragmáticamente, podría ser una petición de cerrar la ventana).