

## Introducción al Procesamiento de Lenguaje Natural

Junio 2023

### Consideraciones generales

- i) La prueba es sin material escrito.
- ii) Escriba nombre y C.I. en todas las hojas.
- iii) Numere todas las hojas.
- iv) En la primera hoja, indique el total de hojas.
- v) Comience cada ejercicio en una hoja nueva.
- vi) Utilice las hojas de un solo lado.
- vii) Entregue los ejercicios en orden.
- viii) El total de puntos es 40.

### Ejercicio 1 [10 puntos]

Para cada afirmación diga si es Verdadera o Falsa. Justifique.

- i) En cualquier oración, el grupo que cumple la función sintáctica de sujeto respecto al verbo principal tiene siempre el rol semántico agente respecto a este mismo verbo.
- ii) La relación de hiperonimia es simétrica, es decir, si  $p_1$  es hiperónimo de  $p_2$ , entonces  $p_2$  es hiperónimo de  $p_1$ .
- iii) El algoritmo de Lesk para desambiguación semántica consiste en elegir el significado cuya glosa comparte más palabras con el contexto de la palabra a desambiguar.
- iv) La morfología derivativa aplicada a los verbos nos permite generar todas sus conjugaciones.
- v) Word embeddings es un método de clasificación supervisada que permite predecir, por ejemplo, si un tweet tiene sentimiento positivo o negativo.
- vi) La Precisión de la clase C mide la cantidad de elementos bien clasificados de esta clase sobre el total de elementos clasificados como clase C.
- vii) Un modelo de lenguaje permite obtener la probabilidad de que una secuencia de palabras ocurra en cierto lenguaje.
- viii) Las redes neuronales solo pueden utilizarse para clasificación secuencial, tomando N palabras de entrada y devolviendo N elementos de salida.
- ix) Clustering es un método de aprendizaje no supervisado.
- x) LDA es un modelo de recuperación de información.

### Ejercicio 2 [10 puntos]

Dadas las oraciones:

*María camina lento.*  
*Pedro camina lento.*

*María es lento.*  
*Pedro es lento.*

- a) Diga en qué ejemplos la palabra *lento* tiene categoría adjetivo y en qué ejemplos tiene categoría adverbio. Justifique.
- b) Escriba una gramática libre de contexto que genere las oraciones dadas.
- c) Aplique el algoritmo CKY para reconocer la siguiente oración: *María camina lento*.
- d) ¿La gramática propuesta permite descartar los ejemplos agramaticales siguientes?
  - \* *Pedro es lento*.
  - \* *María es lento*.

Si no lo hace, comente alguna alternativa para resolver este problema.

### Ejercicio 3 [10 puntos]

a) Suponga que tiene un corpus compuesto por las siguientes oraciones:

<S> Juan come arroz y María come pan </S>

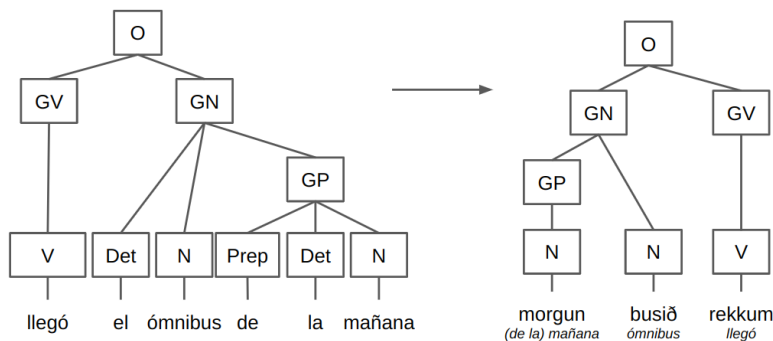
<S> María comerá arroz </S>

Calcule la probabilidad de las siguientes oraciones mediante un modelo de lenguaje de bigramas, asumiendo que la probabilidad del bigrama de los tokens "A B" se puede estimar como  $(\#AB + 1)/(\#A + |V|)$ . Indique cuál de las dos oraciones es más probable según este modelo.

<S> María come pan </S>

<S> María comerá pan </S>

b) Suponga que se quiere construir un sistema de traducción automática del español al feroés basado en transferencia sintáctica. Escriba las reglas de transferencia para dicho sistema, basándose en la siguiente traducción de ejemplo:



### Ejercicio 4 [10 puntos]

- Utilizando el algoritmo de cálculo de la distancia de Levensthein entre dos palabras, calcule la distancia entre CAMINO y CASIN indicando luego el camino utilizado para calcularla.
- En un sistema de recuperación de información basado en el modelo vectorial se pretende asignar pesos a un documento A en base a 4 términos: banco, plaza, fuente, economía

Se pretende seguir un modelo tf-idf para la asignación de pesos. Se pide:

- Defina en qué consiste el modelo tf-idf.
- Se tiene para cada término en el documento A la siguiente frecuencia de aparición de cada uno:  
banco (4) – plaza (5) – fuente (2) - economía (3)  
y que la frecuencia inversa de cada uno es:  
banco (2) – plaza (3) – fuente (3) - economía (4).  
Calcule el vector de pesos del documento A.