

## Introducción al Procesamiento de Lenguaje Natural

Noviembre 2021

### Consideraciones generales

- i) La prueba es sin material escrito.
- ii) La duración es de 3 horas
- iii) Escriba nombre y CI al inicio
- iv) El total de puntos es 70

### Ejercicio 1 [ 21 puntos]

a) Para cada afirmación diga si es Verdadera o Falsa. Justifique los casos de que su respuesta sea Falso.

- i) La arquitectura Encoder-Decoder se utiliza para la traducción automática.
- ii) WordNet es un algoritmo de desambiguación semántica de las palabras.
- iii) El Recall es una medida que puede ser usada para validar la similitud de dos documentos.
- iv) La lematización de una palabra **x** es el proceso que permite generar todas las posibles palabras derivadas de esa palabra **x**.
- v) El algoritmo de parsing CKY funciona como un validador de oraciones sintácticamente correctas.
- vi) En la frase *La casa de papel* la palabra **de** tiene categoría adverbio.
- vii) En la frase *Juan se sentó bajo el árbol* la palabra **bajo** tiene la categoría preposición.

### Ejercicio 2 [ 16 puntos]

a) ¿Qué es un constituyente en una gramática? Mencione al menos dos constituyentes de la gramática del español. ¿Qué nombre reciben los constituyentes en una gramática de dependencias?

b) Considere una gramática con el siguiente conjunto de reglas de producción:

O → GN GV | GV

GV → V GN | V | V GN GP

GN → Det Nom | Nom | GN GP

GP → Prep GN

Nom → verano | primavera | caramelo | luna

Det → el | la

Prep → de | desde

V → miro | como

Sea la oración: "*Miro la luna de verano*"

- i) Realice el análisis sintáctico aplicando el algoritmo de Earley
- ii) Realice el análisis utilizando la notación bracket

### Ejercicio 3 [ 10 puntos]

a) Sea la siguiente gramática con anotaciones semánticas:

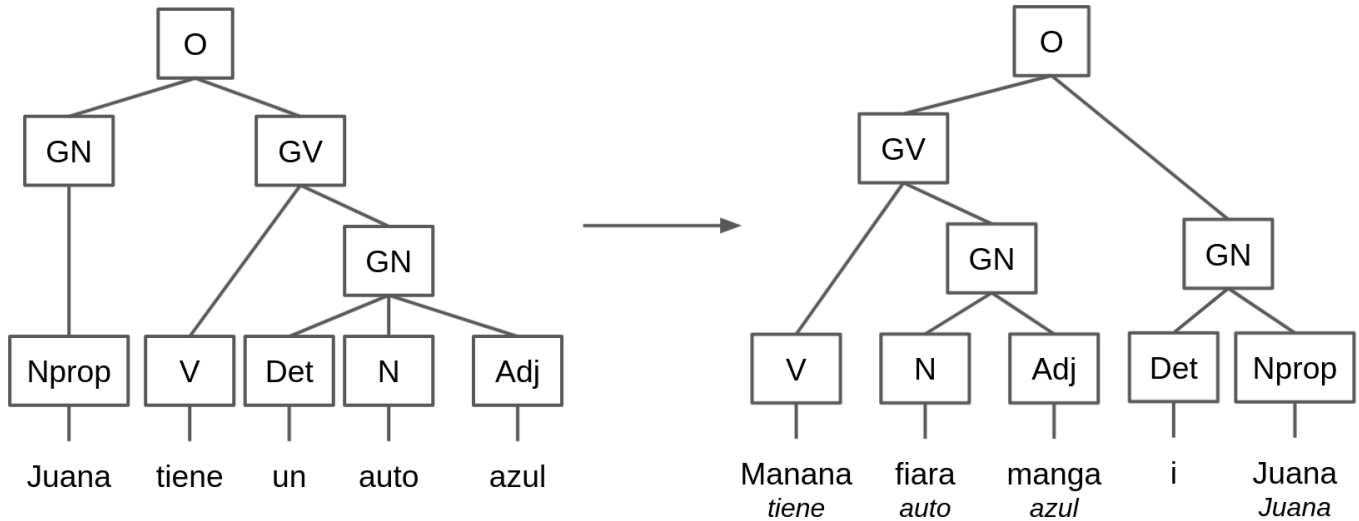
$o \rightarrow gn\ gv$	$o.sem = gn.sem(gv.sem)$
$gn \rightarrow det\ nom$	$gn.sem = det.sem(nom.sem)$
$gn \rightarrow npropio$	$gn.sem = npropio.sem$
$gv \rightarrow v$	$gv.sem = v.sem$
$nom \rightarrow n$	$nom.sem = n.sem$
$nom \rightarrow n\ adj$	$nom.sem = adj.sem(n.sem)$
$det \rightarrow un$	$det.sem = \lambda P. \lambda Q. \exists x P(x) \wedge Q(x)$
$det \rightarrow todo$	$det.sem = \lambda P. \lambda Q. \forall x P(x) \rightarrow Q(x)$
$n \rightarrow libro$	$n.sem = \lambda x. libro(x)$
$n \rightarrow pájaro$	$n.sem = \lambda x. pájaro(x)$
$npropio \rightarrow Pedro$	$npropio.sem = \lambda P. P(pedro)$
$v \rightarrow habla$	$v.sem = \lambda x. habla(x)$
$adj \rightarrow verde$	$adj.sem = \lambda P. \lambda x. verde(x) \wedge P(x)$

Utilizando las reglas anteriores realice una derivación para calcular la representación semántica de la oración:

*Todo pájaro verde habla*

**Ejercicio 4 [ 8 puntos]**

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



- b) Utilizando el algoritmo de programación dinámica, calcule la distancia de Levenshtein entre las palabras ARBOL y MARMOL. Una vez calculada, identifique el camino para obtenerla.

**Ejercicio 5 [ 15 puntos]**

- a) ¿En qué consiste un esquema de clasificación supervisada? Explique la diferencia entre modelos de clasificación probabilistas generativos y discriminativos, y nombre un método como ejemplo para cada enfoque.
- b) Defina brevemente qué entiende por análisis de sentimiento. Mencione al menos dos tareas de PLN que se utilizan en esa tarea.
- c) ¿Sobre qué objetivo o tarea está definido skip-gram con negative sampling? ¿Cómo se forman los ejemplos negativos en skip-gram con negative sampling?