

## Introducción al Procesamiento de Lenguaje Natural

### Diciembre de 2010

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 [ 14 puntos]

1. Represente en lógica las siguientes oraciones:
  - a) Juan y Pedro corren.
  - b) Juan canta y baila.
  
2. Se propone la siguiente gramática (parcial) para dar cuenta de las expresiones:

REGLA	TIPO	SEMANTICA
$o \rightarrow gn\ gv$	t	$o.sem = gn.sem(gv.sem)$
$gn \rightarrow npropio$	$\langle\langle e,t \rangle, t \rangle$	$gn.sem = npropio.sem$
$gv \rightarrow v$	$\langle e,t \rangle$	$gv.sem = v.sem$
$cgn \rightarrow conj\ gn$	?	?
$gn \rightarrow gn\ cgn$	?	?
$cgv \rightarrow conj\ gv$	?	?
$gv \rightarrow gv\ cgv$	?	?
$npropio \rightarrow Juan$	$\langle\langle e,t \rangle, t \rangle$	$\lambda P.P(j)$
$npropio \rightarrow Pedro$	$\langle\langle e,t \rangle, t \rangle$	$\lambda P.P(p)$
$v \rightarrow corren$	$\langle e,t \rangle$	$\lambda x.corre(x)$
$v \rightarrow canta$	$\langle e,t \rangle$	$\lambda x.canta(x)$
$v \rightarrow baila$	$\langle e,t \rangle$	$\lambda x.baila(x)$

- i. Agregue regla(s) preléxica(s) para la conjunción “y”, indicando su tipo y valor semántico. Notar que todas las reglas son binarias, la conjunción se aplica primero al argumento que le sigue en el texto.
- ii. Complete la gramática propuesta en las partes donde aparece “?”.

- iii. Realice la derivación completa para los ejemplos del punto 1).

**Ejercicio 2 [ 14 puntos]**

1. Describa brevemente las ventajas y desventajas de aplicar estrategias puramente *top-down* o *bottom-up* para el análisis sintáctico.
2. Considere la siguiente gramática (muy simplificada).

O           →  GV  
GV          →  V GN | V GN GP  
GN          →  N | N GP  
GP          →  P GN  
N           →  tomo | agua | manzana | manzanas | Pepe | María  
V           →  tomo | toma | tomamos | toman  
P           →  de | con | para

Aplique el algoritmo Early para la entrada *Tomo agua de manzana*.

3. Analice las posibles ambigüedades sintácticas de la oración dada y explique si es posible detectarlas mediante el algoritmo aplicado en 2).
4. Si se agrega a la gramática de 2) la regla:

O           →  GN GV

¿Qué oraciones agramaticales serían reconocidas? Justifique.

Ejercicio 3 [ 12 puntos]

1. Describa brevemente el modelo del canal ruidoso y muestre una aplicación.
2. Se desea aplicar un modelo HMM para resolver la asignación de categorías léxicas a un texto. Del análisis del corpus de entrenamiento, se estiman las siguiente probabilidades:

Probabilidades de transición  $P(\text{tag}_j | \text{tag}_i)$

	<b>j:</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
<b>i</b>		</s>	PREP	DET	V	NN
0	<s>	0	0,2	0,05	0,1	0,1
1	PREP	0	0	0,5	0,1	0,1
2	DET	0	0	0	0	0,1
3	V	0,3	0,1	0,1	0,2	0,2
4	NN	0,2	0,2	0,1	0,1	0,3

Probabilidades de emisión diferentes de 0:

$$P(\text{sobre} | \text{PREP}) = 0,07$$

$$P(\text{sobre} | \text{V}) = 0,0012$$

$$P(\text{sobre} | \text{N}) = 0,00062$$

$$P(\text{el} | \text{DET}) = 0,4$$

Aplique el algoritmo de Viterbi para obtener la secuencia de tags más probable para la expresión "sobre el sobre"

3. ¿Cuál es la diferencia entre el algoritmo Viterbi y el algoritmo Forward para HMMs?