

Introducción al Procesamiento de Lenguaje Natural Diciembre de 2007

Consideraciones generales

- La prueba es sin material escrito.
- Escriba nombre y C.I. en todas las hojas.
- Numere todas las hojas.
- En la primera hoja, indique el total de hojas.
- Comience cada ejercicio en una hoja nueva.
- Utilice las hojas de un solo lado.
- Entregue los ejercicios en orden.
- El total de puntos es 80.

Ejercicio 1 [16 puntos]

Para el texto que se da a continuación, responda las preguntas que se plantean.

.....

El **vendedor** era un muchacho correntino, bajo y de pelo cortado al rape, que usaba siempre botines amarillos. El **otro**, encargado de los libros, era un hombre hecho ya, muy flaco y de cara color paja. Creo que nunca lo vi reírse, mudo y contraído en su Mayor con estricta prolijidad de rayas y tinta colorada. Se llamaba Figueroa; era de Catamarca.

Ambos, comenzando por salir juntos, trabaron estrecha amistad y como ninguno tenía familia en Laboulaye, habían alquilado un caserón con sombríos corredores de bóveda, obra de un escribano que murió loco allá.

Los dos primeros años no tuvimos la menor queja de nuestros hombres. Poco después comenzaron, cada uno a su modo, a cambiar de modo de ser.

El vendedor —se llamaba Tomás Aquino— llegó cierta mañana a la barraca con una verbosidad exuberante. Hablaba y reía sin cesar, buscando constantemente no sé qué en los bolsillos. Así estuvo dos días. Al tercero cayó [con un fuerte ataque de gripe]; pero volvió después de almorzar, inesperadamente curado. Esa misma tarde, Figueroa tuvo que retirarse con desesperantes estornudos preliminares que lo habían invadido de golpe. Pero todo pasó en horas, a pesar de los síntomas dramáticos. Poco después se repitió lo mismo, y así, por un mes: la charla delirante de Aquino, los estornudos de Figueroa, y cada dos días un fulminante y frustrado ataque de gripe.

Esto era lo curioso. Les aconsejé que se hicieran examinar atentamente, pues no se podía seguir así. Por suerte todo pasó, regresando ambos a la antigua y tranquila normalidad, el vendedor entre las tablas, y Figueroa con su pluma gótica.

.....

Fragmento de *Las rayas* de Horacio Quiroga

- Decir cuál de los dos sintagmas nominales en negrita (*El vendedor* y *El otro*) es el antecedente del elemento anafórico *lo* (en negrita y subrayado). ¿En qué criterio se basó para identificarlo?
- Para la oración *Ambos, allá* (marcada en negrita), decir cuáles son los verbos conjugados e identificar las proposiciones que éstos determinan. Mostrar cómo se coordinan o subordinan las diferentes proposiciones, utilizando corchetes.
- Indicar cuál es el sujeto y cuál es el complemento indirecto de la oración (simplificada) *Les aconsejé ... atentamente*, marcada en el texto con subrayado. Justificar.
- Analizar el grupo de palabras *con un fuerte ataque de gripe*, marcado en el texto con corchetes:
 - ¿Qué tipo de grupo o sintagma es?
 - ¿Cuál es el elemento principal del sintagma, es decir, el que le da la categoría?
 - ¿Qué sintagmas más pequeños contiene? Dar el tipo y la palabra principal de cada sintagma identificado.

Ejercicio 2 [16 puntos]

Considere los **verbos** en el idioma ficticio Itio:

1. Los verbos en infinitivo tienen dos terminaciones: *tu* (como *itu*, que en español quiere decir *correr*) y *nu* (como *valnu* y *dinu*, que quieren respectivamente decir *saltar* y *meditar en medio del caos*).
2. Hay tres personas en Itio, y no se distingue plural de singular.
3. Para la conjugación en presente, a los verbos se les agregan los siguientes sufijos:

Terminación	Persona	Sufijo	Ejemplo
tu	1era	-fu	itfu
tu	2da	-fa	itfa
tu	3era	-fe	itfe
nu	1era	-nu	valnu
nu	2da	-no	valno
nu	3era	-ne	dine

Observar que al agregar el sufijo, los verbos terminados en *tu* pierden la *u*, y los terminados en *nu*, pierden toda la terminación.

4. Al conjugar un verbo en presente, puede (opcionalmente) agregarse un sufijo que denota modalidad: *natnat*, para denotar *con violencia*, *uz*, para denotar *rápidamente*, y *top*, para denotar *sonriendo misteriosamente*. Por ejemplo, *valnunatnat* quiere decir *salto violentamente*.
5. Por cuestiones ortográficas, en Itio no están permitidas las vocales dobles. Cuando aparezcan al combinar morfemas, se convertirán en una sola.

- a) Describa brevemente los tres componentes principales de un analizador morfológico. Describa cada componente para un hipotético analizador morfológico para los verbos en Itio.
- b) Construya una expresión regular utilizando el álgebra de Xerox, que relacione un verbo en Itio conjugado en presente con su análisis morfológico. Utilice las siguientes marcas léxicas para la salida: *+Verb* (marca verbal), *+Primera* (primera persona), *+Segunda* (segunda persona), *+Tercera* (tercera persona), *+Violencia*, *+Rapido*, *+SonriendoMisteriosamente*. Suponga que las únicas raíces verbales son las que se presentaron en los ejemplos (*itu*, *valnu*, *dinu*), pero diseñe la solución de modo que pueda ampliarse a otras raíces fácilmente.

Ejemplos:

dinuz	dinu+Verb+Primera+Rapido
itfatop	itu+Verb+Segunda+SonriendoMisteriosamente
valno	valnu+Verb+Segunda

Ejercicio 3 [8 puntos]

Suponga que utiliza un Modelo de Markov de Estados Ocultos (HMM) para hacer POS tagging.

- a) Los HMMs son un caso particular de *inferencia bayesiana*. Aplique la fórmula de Bayes para mostrar cómo estima este método la probabilidad de una secuencia de tags, dada una secuencia de palabras.
- b) ¿Qué hipótesis principales asumen los HMMs para simplificar el cálculo de la fórmula obtenida en el punto anterior?

Ejercicio 4 [18 puntos]

- a) Describa brevemente las ventajas y desventajas de aplicar estrategias puramente *top-down* o *bottom-up* para el análisis sintáctico.
b) Considere la siguiente gramática en Forma Normal de Chomsky.

O	→	GV GN GN GV
GV	→	V GN
GN	→	Det N Det GNP
GNP	→	N GP
GP	→	Prep GN
N	→	sala masa hombre
Det	→	el la los
V	→	sala trae
Prep	→	con de

Aplice el algoritmo CYK para la entrada *El hombre sala la masa de la sala*.

Ejercicio 5 [16 puntos]

o	→	gn gv	[o.sem = gv.sem (gn.sem)]
gn	→	det nominal	[gn.sem = < det.sem x nominal.sem(x) >]
gn	→	npropio	[gn.sem = npropio.sem]
nominal	→	n	[nominal.sem = $\lambda x \text{ isa}(x, n.sem)$]
gv	→	v	[gv.sem = v.sem]
gv	→	v gn	[gv.sem = v.sem (gn.sem)]
v	→	lee	[v.sem = $\lambda x \lambda y \exists e \text{ leer}(e) \wedge \text{lee}(e, y) \wedge \text{leído}(e, x)$]
det	→	un	[det.sem = \exists]
n	→	libro	[n.sem = libro]
v	→	entrega	[v.sem = $\lambda x \lambda y \lambda z \exists e \text{ entregar}(e) \wedge \text{entrega}(e, z) \wedge \text{entregado}(e, y) \wedge \text{receptor}(e, x)$]

- a) Complete la gramática anterior con las reglas sintácticas y sus correspondientes componentes semánticas necesarias para reconocer :
i. Juan lee un libro.
ii. Juan entrega un libro a Pedro.
iii. Juan recibe un libro de Pedro.

La tercera oración es ambigua. Verifique que su gramática represente esta ambigüedad (genera más de un análisis).

- b) Utilizando la gramática completada realice la derivación de una representación semántica para la oración :
Juan entrega un libro a Pedro.

Ejercicio 6 [6 puntos]

- a) Explique brevemente en qué se diferencian los llamados Sistemas de Recuperación de Información de los Sistemas de Información tradicionales (basados en bases de datos).
b) Mencione cuáles son las características principales del Modelo Vectorial, que lo diferencian sustancialmente de su antecesor, el Modelo Booleano.