

Introducción al Procesamiento de Lenguaje Natural

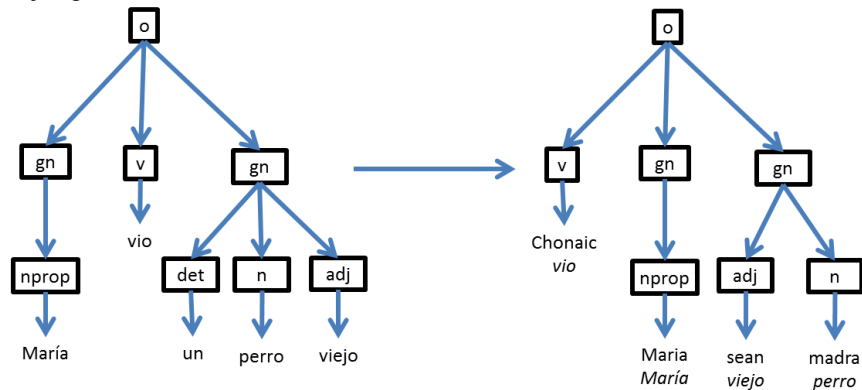
Diciembre de 2017

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 [6 puntos]

- i) Dibuje el Triángulo de Vauquois, indicando qué representan los lados derecho e izquierdo del triángulo y ubicando dentro del triángulo los siguientes métodos de traducción automática: transferencia sintáctica, traducción directa, traducción basada en interlingua.
- ii) Suponga que se quiere construir un sistema de traducción automática del español al irlandés basado en transferencia sintáctica. Infiera las reglas de transferencia que utilizaría dicho sistema, basándose en la siguiente traducción de ejemplo:



Ejercicio 2 [12 puntos]

Considere una gramática G con las siguientes reglas:

- O → GN GV | GN
- GN → Det GN | GP GN | Nom
- GV → V GN GP | V GN | V GP GP
- GP → Prep GN | Prep
- Nom → Juan | redonditos | ricota | pasta | música | letra
- Prep → de | con | en | a
- V → escucha | come | baila
- Det → el | la | los | un

- i) Aplique el algoritmo CKY para la entrada “Juan escucha a los redonditos de ricota” considerando las reglas de la gramática G. ¿Qué salida devuelve el algoritmo? Justifique.
- ii) Dibuje **al menos 2** árboles sintácticos posibles para la oración dada. Dé, en cada caso, una representación del análisis basada en la estructura “bracketed notation”.
- iii) ¿A qué tipo de método de análisis sintáctico pertenece el algoritmo CKY? Describalo brevemente.

Ejercicio 3 [10 puntos]

- i) Describa brevemente (no más de diez líneas) **dos** de los métodos vistos en el curso para identificación de hiponimia.
- ii) Considere la siguiente gramática con anotaciones semánticas:

$o \rightarrow gn\ gv$	$o.sem = gn.sem(gv.sem)$
$gn \rightarrow nprop$	$gn.sem = nprop.sem$
$gn \rightarrow det\ nom$	$gn.sem = det.sem(nom.sem)$
$nom \rightarrow n$	$nom.sem = n.sem$
$nom \rightarrow n\ adj$	$nom.sem = adj.sem(n.sem)$
$gv \rightarrow v$	$gv.sem = v.sem$
$gv \rightarrow v\ gn$	$gv.sem = v.sem(gn.sem)$
$nprop \rightarrow Pedro$	$nprop.sem = \lambda P . P(pedro)$
$neg \rightarrow no$	$neg.sem = \lambda P . \lambda x . \neg P(x)$
$n \rightarrow perro$	$n.sem = \lambda x . perro(x)$
$n \rightarrow hombre$	$n.sem = \lambda x . hombre(x)$
$v \rightarrow caza$	$v.sem = \lambda P . \lambda y . P(\lambda x . caza(y, x))$
$v \rightarrow vuela$	$v.sem = \lambda x . vuela(x)$
$adj \rightarrow negro$	$adj.sem = \lambda P . \lambda x . P(x) \wedge negro(x)$
$det \rightarrow todo$	$det.sem = \lambda P . \lambda Q . \forall P(x) \rightarrow Q(x)$
$det \rightarrow ningún$	$det.sem = \lambda P . \lambda Q . \forall P(x) \rightarrow \neg Q(x)$
$det \rightarrow un$	$det.sem = \lambda P . \lambda Q . \exists x P(x) \wedge Q(x)$

Dibuje el árbol sintáctico y derive la expresión lógica asociada a la oración: *Ningún hombre vuela.*

Ejercicio 4 [7 puntos]

- i) Calcule la distancia de Levenshtein entre "ORACION" y "CREATIVO". Realice un alineamiento entre las dos palabras, que indique la secuencia de operaciones realizadas. ¿Qué sucede si el costo de sustitución se cambia a 2 (manteniendo los costos de inserción y borrado)? ¿Y a 10? Justifique.
- ii) Demuestre que si $L(T)$ es la relación asociada a un transductor T , entonces los lenguajes asociados a cada una de las proyecciones de T son lenguajes regulares.

Ejercicio 5 [5 puntos]

Sea el siguiente texto:

El presidente de Cutcsa Juan Salgado valoró los últimos cambios anunciados por la Intendencia de Montevideo en el Corredor Garzón y auguró que la comuna no cometa los mismos errores con las modificaciones previstas para la Avenida 18 de Julio.

En declaraciones recogidas por el periodista Leo Sarro para Radio Montecarlo, Salgado valoró que "al fin se están tomando las medidas que se tenían que haber tomado desde el inicio".

Salgado se refería a la decisión de la comuna de eliminar el corredor en el centro de Colón, algo que para el empresario "llega tarde pero llega".

En un proceso de extracción de información sobre el texto anterior, dé:

- i) Al menos 3 entidades con nombre que pertenezcan a distintas categorías.
- ii) Al menos 2 relaciones binarias indicando sus argumentos.