

Introducción al Procesamiento de Lenguaje Natural Diciembre de 2017

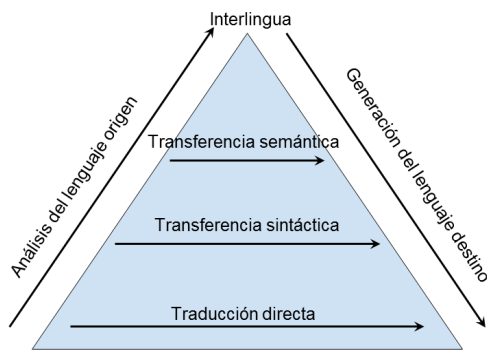
Consideraciones generales

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

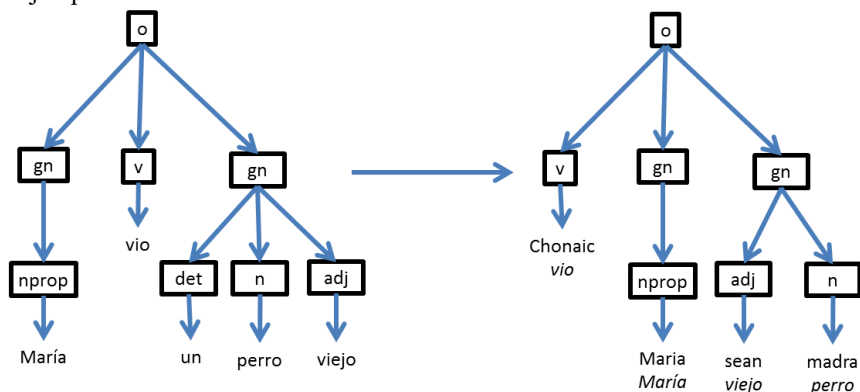
Ejercicio 1 [6 puntos]

- 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.

Solución:



- 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:



Solución:

o → gn₁ v gn₂ / o → v gn₁ gn₂
 gn → nprop / gn → nprop
 gn → det n adj / gn → adj n

Ejercicio 2 [12 puntos]

Considere una gramática G con las siguientes reglas:

- O → GN GV | GN
- GN → Det Nom | 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.

Solución:

Primero se verifica que la gramática esté en FNC (Forma Normal de Chomsky). Como no está, el primer paso es normalizarla.

Quedan entonces las siguientes reglas de producción:

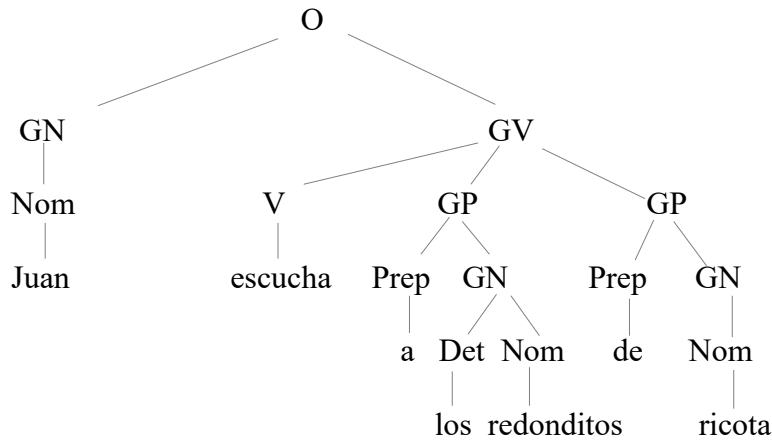
- O → GN GV | Det Nom | GP GN | Juan | redonditos | ricota | pasta | música | letra
- GN → Det Nom | GP GN | Juan | redonditos | ricota | pasta | música | letra
- GV → V X₁ | V GN | V X₂
- X₁ → GN GP
- X₂ → GP GP
- GP → Prep GN | de | con | en | a
- Nom → Juan | redonditos | ricota | pasta | música | letra
- Prep → de | con | en | a
- V → escucha | come | baila
- Det → el | la | los | un

Ahora, sobre esta gramática se aplica el algoritmo CKY, resultando el siguiente cuadro:

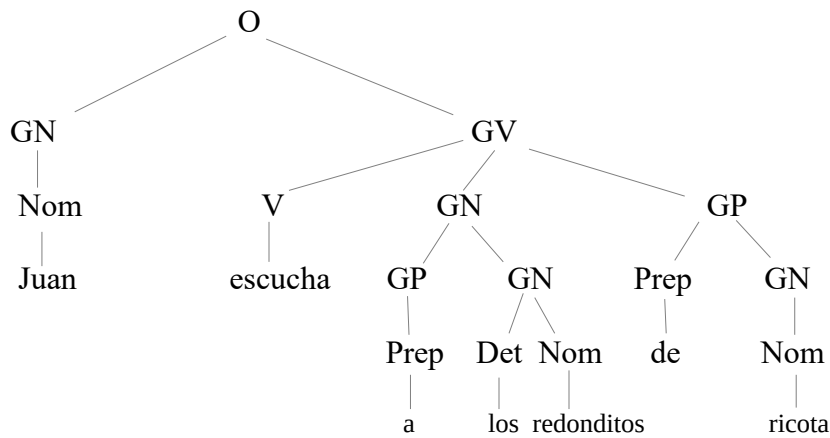
	Juan (1)	escucha (2)	a (3)	los (4)	redonditos (5)	de (6)	ricota (7)
0	O GN Nom				O	O	O
1		V			GV	GV	GV
2			Prep GP		O GP GN	X2	O GN X2
3				Det	GN O	X1	X1
4					O GN Nom	X1	X1
5						Prep GP	GP GN
6							O GN Nom

El algoritmo devuelve TRUE, indicando que la secuencia de entrada puede derivarse a partir de la gramática, al estar la variable O en la celda V₀₇

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”.



[Juan]_{GN} [escucha [a [los redonditos]_{GN}]_{GP} [de [ricota]_{GN}]_{GP}]_{GV}



[Juan]_{GN} [escucha [[a]_{GP}[los redonditos]_{GN}]_{GN} [de [ricota]_{GN}]_{GP}]_{GV}

iii) ¿A qué tipo de método de análisis sintáctico pertenece el algoritmo CKY? Describalo brevemente.

Pertenece a la familia de los métodos bottom-up. Los métodos bottom-up construyen el árbol de análisis sintáctico desde las hojas tratando de llegar a la raíz (símbolo inicial de la gramática). Pueden llegar a construir árboles que nunca lleguen al símbolo inicial, pero siempre va a “sugerir” árboles que sean consistentes con la secuencia de entrada.

Ejercicio 3 [10 puntos]

i) Describa brevemente (no más de diez líneas) **dos** de los métodos vistos en el curso para identificación de hiponimia.
 Solución:

- *Patrones léxico sintácticos:*

Si encontramos un patrón en el texto como “NP0 tales como NP1, {NP2, ... y|o NPn}, podemos inferir que todos los NP_i son hipónimos de NP0. En esta técnica se utiliza una colección de patrones de este estilo definidos a mano para buscar pares hiperónimo-hipónimo en un corpus. El éxito de esta técnica dependerá de la exhaustividad con la que se construyan los patrones, lo cual es muy difícil de realizar debido a que el lenguaje natural es muy creativo.

- *Supervisión distante:*

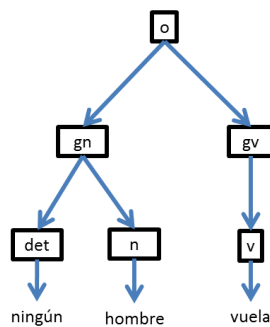
Para evitar tener que escribir los patrones a mano, esta técnica propone puede partir de una colección de pares hiperónimo-hipónimo ya conocida, buscar en un gran corpus todas sus ocurrencias y recopilar los contextos en donde aparecen. Utilizando los contextos se infieren los patrones de búsqueda o se construye un clasificador que sepa identificar pres hiperónimo-hipónimo en oraciones nuevas.

ii) Considere la siguiente gramática con anotaciones semánticas:

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

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

Solución:



$o.sem =$
 [sustitución según reglas]
 $= gn.sem(gv.sem) =$
 [sustitución según reglas]
 $= (det.sem(nom.sem))(v.sem) =$

[sustitución según reglas]
 $= ((\lambda P . \lambda Q . \forall x P(x) \rightarrow \neg Q(x))(n.sem))(\lambda x . vuela(x)) =$
 [sustitución según reglas]
 $= ((\lambda P . \lambda Q . \forall x P(x) \rightarrow \neg Q(x))(\lambda x . hombre(x)))(\lambda x . vuela(x)) =$
 [cambio de variable]
 $= ((\lambda P . \lambda Q . \forall x P(x) \rightarrow \neg Q(x))(\lambda y . hombre(y)))(\lambda x . vuela(x)) =$
 [aplicación funcional]
 $= (\lambda Q . \forall x (\lambda y . hombre(y))(x) \rightarrow \neg Q(x))(\lambda x . vuela(x)) =$
 [aplicación funcional]
 $= (\lambda Q . \forall x hombre(x) \rightarrow \neg Q(x))(\lambda x . vuela(x)) =$
 [cambio de variable]
 $= (\lambda Q . \forall x hombre(x) \rightarrow \neg Q(x))(\lambda z . vuela(z)) =$
 [aplicación funcional]
 $= \forall x hombre(x) \rightarrow \neg (\lambda z . vuela(z))(x) =$
 [aplicación funcional]
 $= \forall x hombre(x) \rightarrow \neg vuela(x)$

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.

Solución:

La distancia de Levenshtein implica costo 1 para las tres operaciones (inserción, borrado y sustitución). Por lo tanto, la distancia mínima es 5. Una posible alineación es la siguiente:

```
O R _ A C I O N
s  i  s  s  s
C R E A T I V O
```

Si el costo pasa a ser 2, entonces la distancia mínima pasa a ser 7, y una posible alineación sería:

```
_ O R _ A _ C I _ O N
i b  i  i b  i  b
C _ R E A T _ I V O _
```

Si se cambia el costo de sustitución a 10, el resultado no cambia (notar que ya hay un camino de mínima distancia que no incluye sustituciones).

- 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.

Solución:

La proyección de un transductor es el autómata finito resultante de tomar los estados y transiciones del transductor, etiquetando cada transición con el primer (o segundo) componente de las transiciones originales. Se obtiene un autómata finito que reconoce al lenguaje de entrada (o salida) del transductor, que será, por tanto, regular.

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 distintas categorías
- ii) al menos 2 relaciones binarias indicando sus argumentos

Solución:

- i) *Cutcsa; Intendencia de Montevideo; Radio Montecarlo --> Organización*
Juan Salgado; Lao Sarro; --> Persona
Corredor Garzón; Avenida 18 de Julio; Colón --> Lugar
- ii) *valorar(el presidente de Cutcsa Juan Salgado, últimos cambios anunciados por la Intendencia de Montevideo)*
eliminar_en(el corredor, el centro de Colón)