

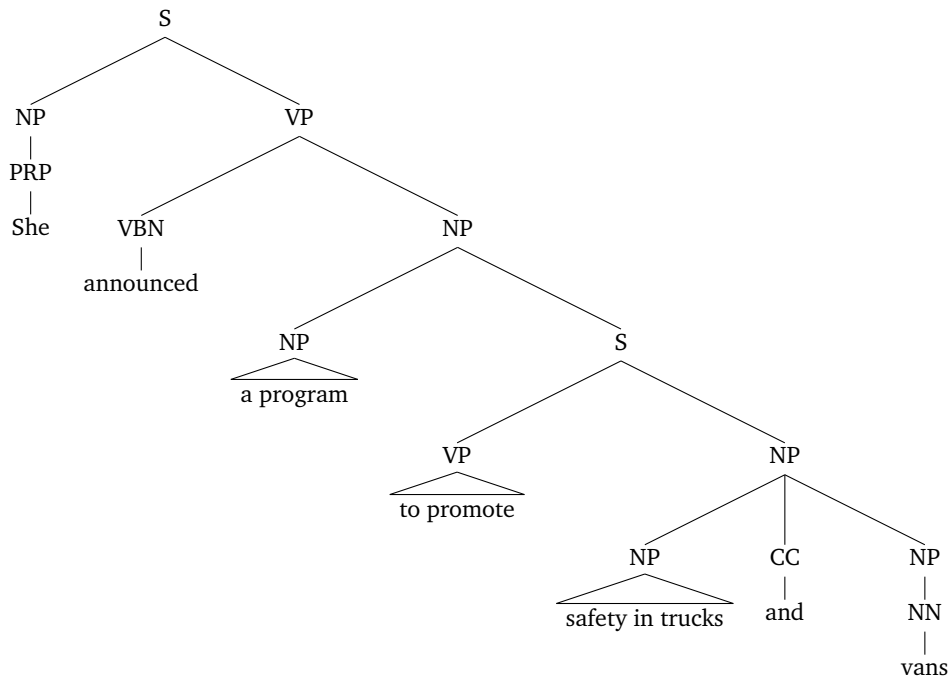
Processament del Llenguatge Humà (GIA-PLH)

Exercicis sobre anàlisi de constituents

CFGs

Ejercicio 1.

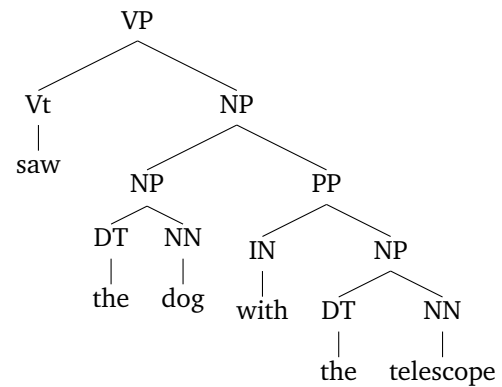
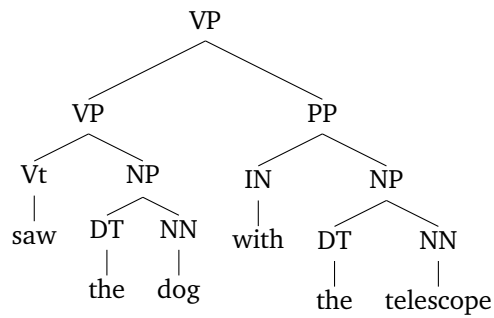
Considera la oración "She announced a program to promote safety in trucks and vans" y el siguiente análisis sintáctico correspondiente a una de las posibles interpretaciones (el programa promueve la seguridad en camiones y también promueve furgonetas):



1. Dibuja los árboles de constituyentes de, al menos, otras tres interpretaciones de esta oración
2. Dibuja los árboles de, al menos, dos interpretaciones para cada una de las siguientes oraciones:
 - *The post office will hold out discounts and service concessions as incentives*
 - *They are hunting lions and tigers*
 - *Monty flies like mosquitoes*

Ejercicio 2.

Supón la oración "saw the dog with the telescope", su árbol de constituyentes correcto (izquierda) y uno producido por un analizador sintáctico (derecha): and we are given the gold parse tree (left) and the predicted parse tree (right):



¿Cuales son la precisión y la cobertura del árbol producido por el analizador?

Ejercicio 3.

Considera la siguiente CFG:

$S \rightarrow NP VP$	$DT \rightarrow the$	$NN \rightarrow park$
$NP \rightarrow DT NN$	$NN \rightarrow man$	$VB \rightarrow saw$
$NP \rightarrow NP PP$	$NN \rightarrow dog$	$IN \rightarrow with$
$PP \rightarrow IN NP$	$NN \rightarrow cat$	$IN \rightarrow under$
$VP \rightarrow VB NP$		

1. ¿Cuántos árboles genera la gramática para la oración "the man saw the dog in the park"?
2. ¿Cuántos para la oración "textitthe man saw the dog in the park with the cat"?

Ejercicio 4.

Considera la siguiente CFG:

$S \rightarrow NP VP$	$DT \rightarrow the$	$NNS \rightarrow cats$
$NP \rightarrow DT NN$	$NN \rightarrow man$	$NNS \rightarrow parks$
$NP \rightarrow DT NNS$	$NN \rightarrow dog$	$VB \rightarrow see$
$NP \rightarrow NP PP$	$NN \rightarrow cat$	$VB \rightarrow sees$
$PP \rightarrow IN NP$	$NN \rightarrow park$	$IN \rightarrow in$
$VP \rightarrow VB NP$	$NNS \rightarrow dogs$	$IN \rightarrow with$
$VP \rightarrow VP PP$		

Esta gramática sobregenera oraciones incorrectas para el inglés, como por ejemplo:

the dog see the cat
the dog in the park see the cat
the dog in the park see the cat in the park
the dogs sees the cat
the dogs in the park sees the cat
the dogs in the park sees the cat in the park

1. Modifica la gramática para que todas las oraciones generadas respeten la concordancia de 3a persona entre el sujeto y el verbo

PCFGs

Ejercicio 5.

Usando la siguiente PCFG en CNF:

$S \rightarrow NP VP$	1.0	$P \rightarrow with$	1.0
$NP \rightarrow NP PP$	0.4	$V \rightarrow saw$	1.0
$PP \rightarrow P NP$	1.0	$NP \rightarrow astronomers$	0.1
$VP \rightarrow V NP$	0.7	$NP \rightarrow ears$	0.18
$VP \rightarrow VP PP$	0.3	$NP \rightarrow saw$	0.04
		$NP \rightarrow stars$	0.18
		$NP \rightarrow telescopes$	0.1

sobre la oración *astronomers saw stars with ears*

- ¿Cuántos análisis correctos se obtienen?
- Escríbelos junto con sus probabilidades.

Ejercicio 6.

Dada la siguiente PCFG:

$S \rightarrow NP VP$	1.0	$N \rightarrow time$	0.4
$NP \rightarrow N N$	0.25	$N \rightarrow flies$	0.2
$NP \rightarrow D N$	0.4	$N \rightarrow arrow$	0.4
$NP \rightarrow N$	0.35	$D \rightarrow an$	1.0
$VP \rightarrow V NP$	0.6	$ADV \rightarrow like$	1.0
$VP \rightarrow V ADV NP$	0.4	$V \rightarrow flies$	0.5
		$V \rightarrow like$	0.5

y la oración *time flies like an arrow*

1. Escribe 2 árboles de constituyentes que esa PCFG genere para la oración dada.
2. Computa la probabilidad de cada uno de esos árboles.
3. Convierte la gramática a CNF y simula el comportamiento del algoritmo PCKY sobre la oración. Proporciona el chart final con toda la información involucrada en el proceso.

Ejercicio 7.

Supongamos que los siguientes árboles están dentro del treebank de entrenamiento. Cada árbol se observó dentro del treebank las veces indicadas bajo ellos.

S	S	S	S	S
$\begin{array}{c} \wedge \\ A \quad A \\ \quad \\ a \quad a \end{array}$	$\begin{array}{c} \wedge \\ B \quad B \\ \quad \\ a \quad a \end{array}$	$\begin{array}{c} \wedge \\ A \quad A \\ \quad \\ f \quad g \end{array}$	$\begin{array}{c} \wedge \\ A \quad A \\ \quad \\ f \quad a \end{array}$	$\begin{array}{c} \wedge \\ A \quad A \\ \quad \\ g \quad f \end{array}$
75	10	325	8	428

1. ¿Qué PCFG se consigue a partir de ellos usando máxima verosimilitud (MLE)?
2. Dada la gramática obtenida:
 - ¿Cuál es el análisis mas probable para la secuencia "a a"?
 - ¿Es un resultado razonable? Justifica tu respuesta.