

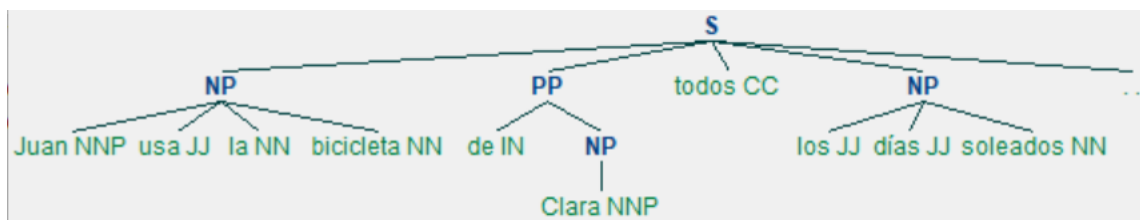
TEMA 2 – HOJA DE EJERCICIOS III

Se proponen diferentes ejercicios relacionados con análisis sintáctico.

1. Dada la frase “Juan usa la bicicleta de Clara todos los días soleados.”, se pide realizar un análisis sintáctico parcial de forma que se obtengan los *chunks* que se muestran a continuación. Será necesario, por tanto, definir una gramática con las reglas que puedan identificar dichos *chunks*.

Analizando los *chunks* de las figuras, se puede ver que hay errores. Por ejemplo, “usa” lo ha etiquetado como adjetivo (JJ) cuando es un verbo. Esto ocurre porque, a pesar de que se puede especificar el idioma español a la hora de tokenizar con NLTK, no lo permite cuando se hace el PoS, por eso comete errores en el part-of-speech de textos en español. Aún así, hay que definir una gramática para que se obtenga la salida de las figuras.

```
(S
  (NP Juan/NNP usa/JJ la/NN bicicleta/NN)
  (PP de/IN (NP Clara/NNP))
  todos/CC
  (NP los/JJ días/JJ soleados/NN)
  ./.)
```



Una vez resuelto, se pide traducir la frase en español a inglés e identificar *chunks* con la misma gramática. ¿Cuál es el resultado? Debería de identificar menos *chunks*.

2. A partir del contenido del fichero de texto “Cycling.txt”, se quiere hacer *chunking* para detectar diferentes sintagmas. Utilizar la clase `RegexParser` de NLTK para crear gramáticas que contengan los patrones que se quieren identificar. Se quieren identificar los siguientes *chunks* (ponerlo todo en la misma gramática):
 - a. *Chunks* formados por dos nombres.
 - b. *Chunks* formados por un determinante y un nombre.

- c. *Chunks* formados por una preposición, puede que un determinante y un nombre.
 - d. *Chunks* formados por un verbo, un adjetivo y un nombre.
-
- 3. Utilizando el chunker de NLTK que reconoce Entidades Nombradas (Named Entity, NE) (https://www.nltk.org/api/nltk.chunk.ne_chunk.html), se pide reconocer las NE presentes en el texto del archivo "Cycling.txt".
 - 4. Mismo ejercicio que el anterior, pero para el contenido del archivo "Health_IA.txt", que es texto en español. Probar a hacerlo con la librería NLTK. ¿El resultado es correcto?
 - 5. Mismo ejercicio que el anterior, pero utilizando spaCy.