

## Tokenización y lematización de un texto con Stanza

### Código

```
import stanza

texto = "Cómo funciona el procesamiento del lenguaje natural: técnicas y  
herramientas. El análisis sintáctico y semántico son dos técnicas  
principales que se utilizan con el procesamiento del lenguaje natural. La  
sintaxis es la disposición de las palabras en una oración para que tengan  
sentido gramatical. NLP utiliza la sintaxis para evaluar el significado  
de un idioma basado en reglas gramaticales. La semántica implica el uso y  
el significado de las palabras. NLP aplica algoritmos para comprender el  
significado y la estructura de las oraciones."

nlp = stanza.Pipeline('es')
doc = nlp(texto)

tokens = []

for sent in doc.sentences:
    for word in sent.words:
        tokens.append(word.lemma)

print(tokens)
```

### Salida

```
['cómo', 'funcionar', 'el', 'procesamiento', 'de', 'el', 'lenguaje', 'natural', ':', 'técnica', 'y', 'herramienta', '.', 'el',  
'análisis', 'sintáctico', 'y', 'semántico', 'ser', 'dos', 'técnica', 'principal', 'que', 'él', 'utilizar', 'con', 'el', 'proces  
amiento', 'de', 'el', 'lenguaje', 'natural', '.', 'el', 'sintaxis', 'ser', 'el', 'disposición', 'de', 'el', 'palabra', 'en', 'u  
no', 'oración', 'para', 'que', 'tener', 'sentido', 'gramatical', '.', 'NLP', 'utilizar', 'el', 'sintaxis', 'para', 'evaluar',  
'el', 'significado', 'de', 'uno', 'idioma', 'basado', 'en', 'regla', 'gramatical', '.', 'el', 'semántica', 'implicar', 'el', 'u  
so', 'y', 'el', 'significado', 'de', 'el', 'palabra', '.', 'NLP', 'aplicar', 'algoritmo', 'para', 'comprender', 'el', 'signific  
ado', 'y', 'el', 'estructura', 'de', 'el', 'oración', '.']
```