



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

# Etiquetadores morfosintàctics para el español

Santiago Millán Giner | Lingüística Computacional | MIARFID

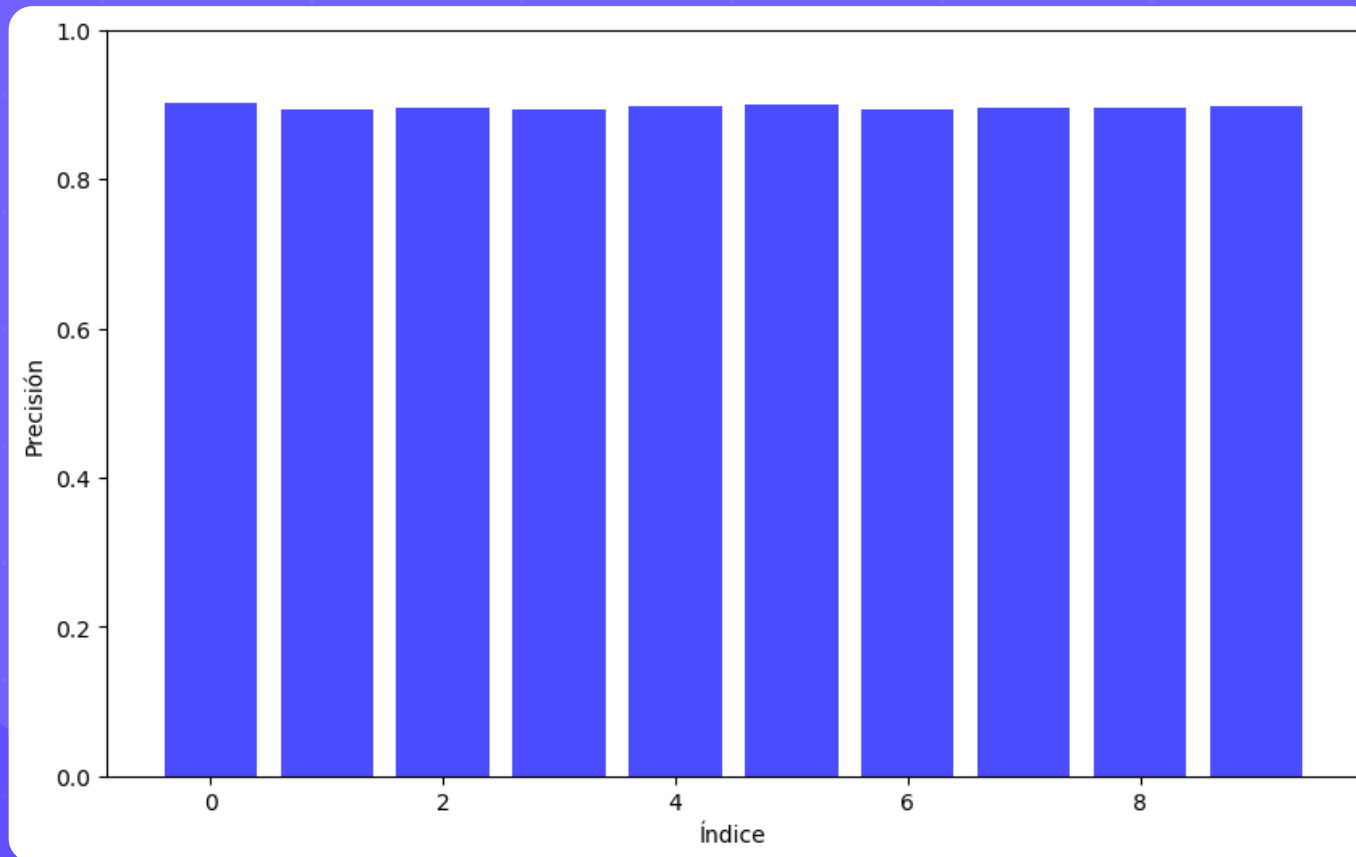
# 0.- Particionar el corpus y cálculo de la precisión

**Barajar las frases** > Dividir en 10 **para validación cruzada**

Los **intervalos de confianza** que se calculan son del **95%**

# 1.1- Hidden Markov Model

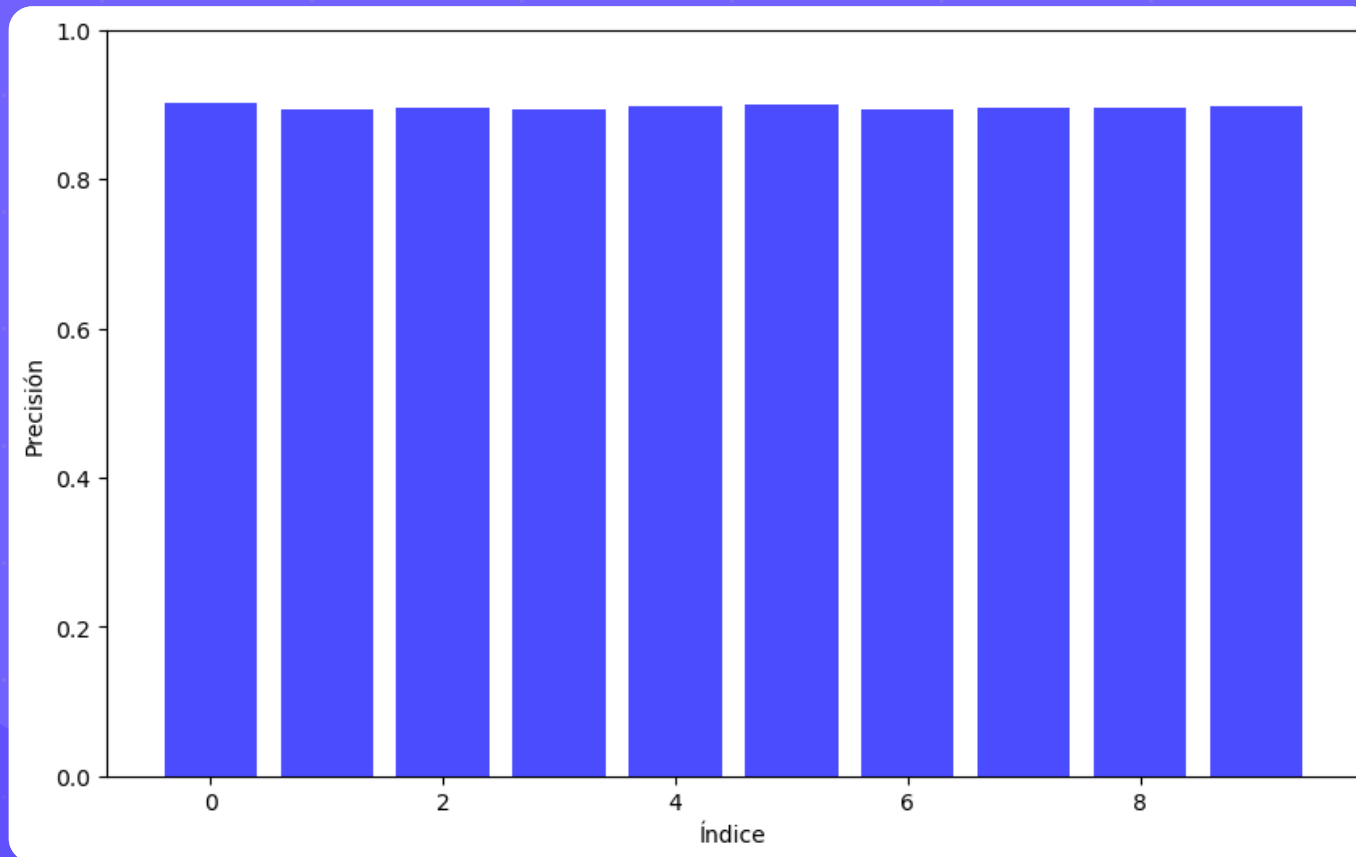
**89.65**  $\pm 0.21\%$  = [0.8944, 0.8986]



Precisión	
0	0.902726
1	0.894018
2	0.895029
3	0.894077
4	0.897296
5	0.900041
6	0.893084
7	0.895966
8	0.895916
9	0.897070

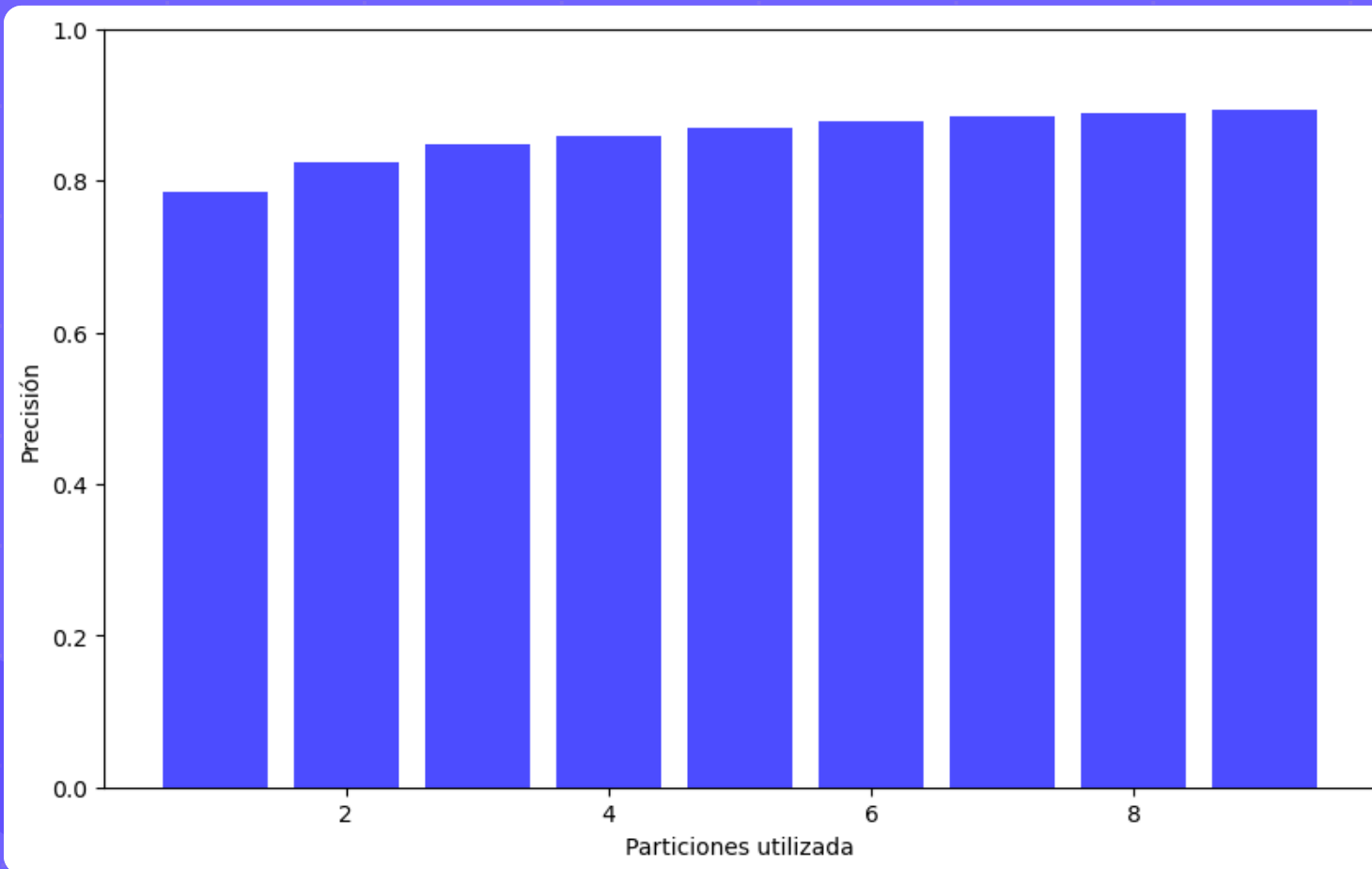
# 1.2- Hidden Markov Model reducido

**92.60**  $\pm 0.10\%$  = [0.9250, 0.9269]



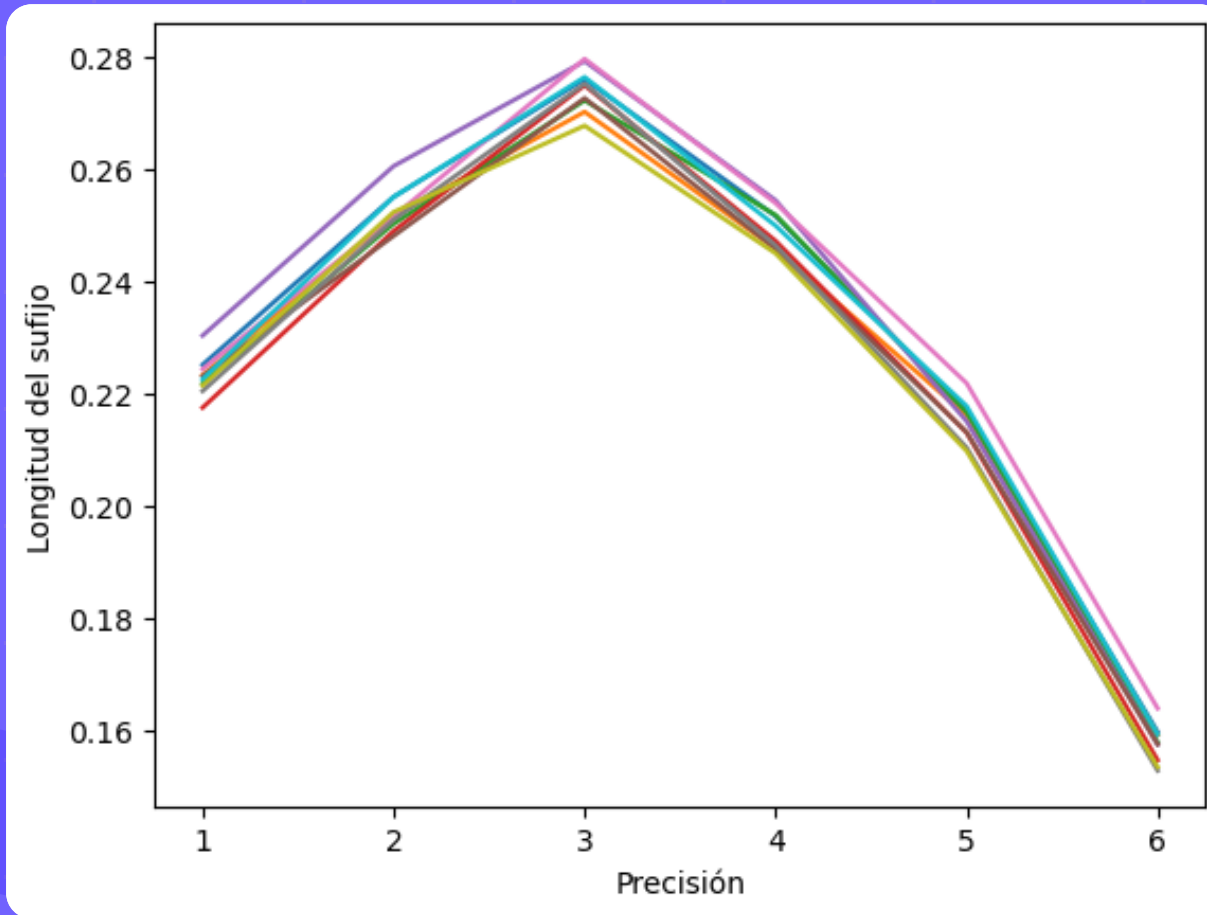
Precisión	
0	0.926726
1	0.926661
2	0.925778
3	0.923956
4	0.924122
5	0.926262
6	0.925230
7	0.925887
8	0.926190
9	0.928783

## 2.- HMM aumentando el tamaño del entrenamiento



Precisión	
1	0.785229
2	0.824803
3	0.846997
4	0.859672
5	0.868496
6	0.878175
7	0.885395
8	0.889887
9	0.894059

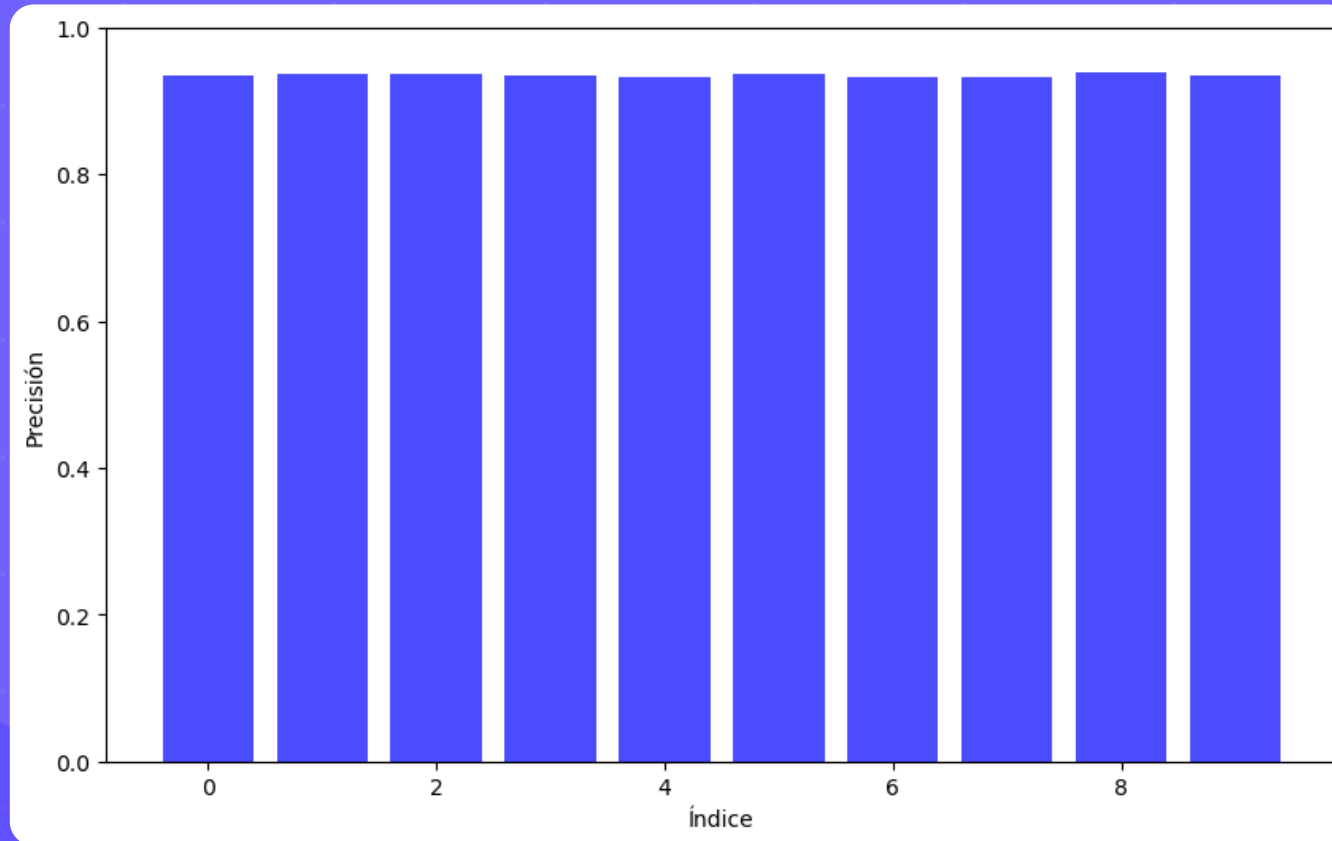
## 3.1 .- Affix Tagger



suffix_len	Intervalo
1	<b>22.30</b> $\pm$ 0.24% = [0.2206, 0.2254]
2	<b>25.24</b> $\pm$ 0.26% = [0.2498, 0.2550]
3	<b>27.44</b> $\pm$ 0.27% = [0.2717, 0.2771]
4	<b>27.44</b> $\pm$ 0.27% = [0.2717, 0.2771]
5	<b>21.50</b> $\pm$ 0.26% = [0.2125, 0.2176]
6	<b>15.76</b> $\pm$ 0.24% = [0.1552, 0.1599]

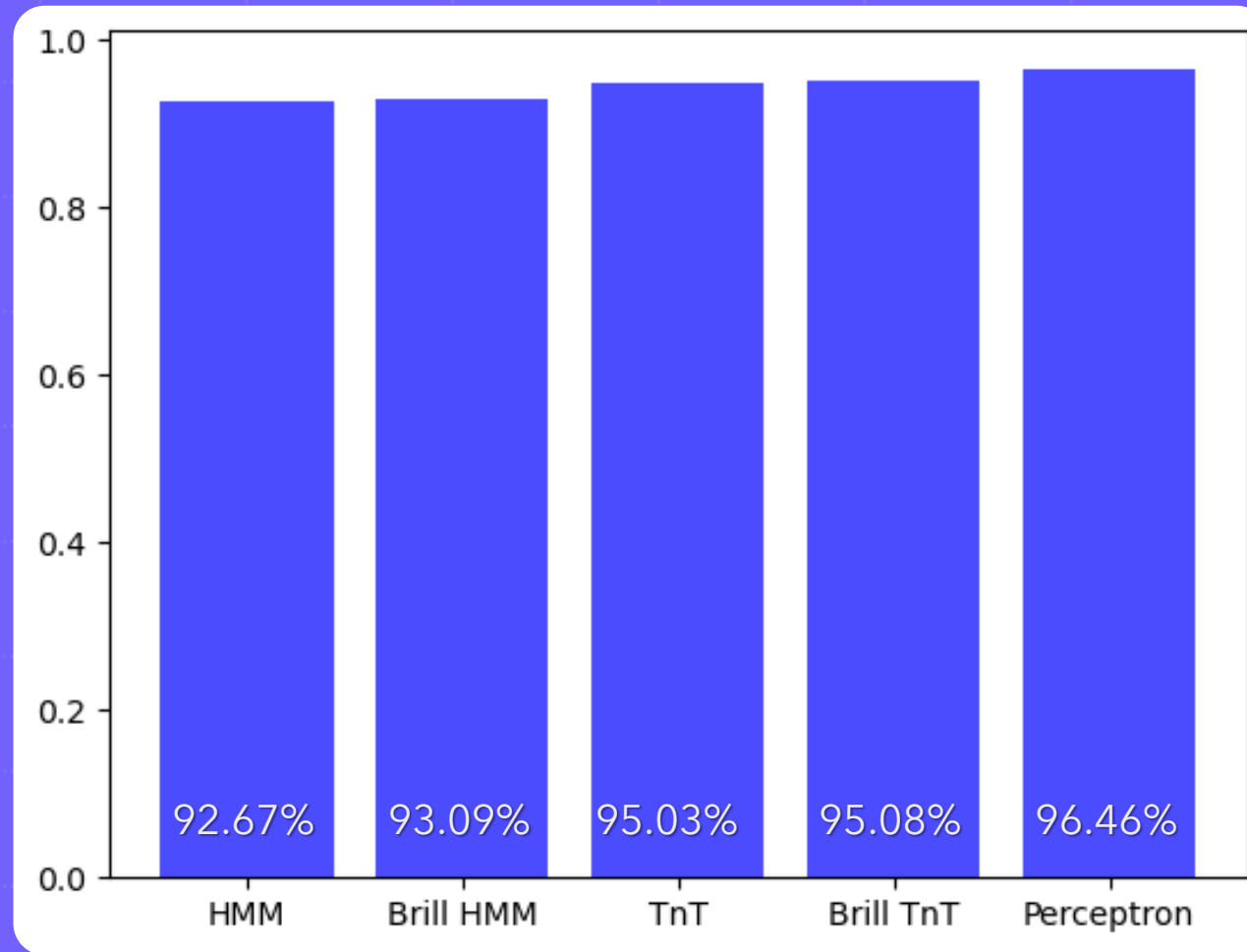
## 3.2 .- TnT usando Affix como suavizado

**93.46**  $\pm 0.15\%$  = [0.9330, 0.9361]



Precisión	
0	0.933620
1	0.937220
2	0.935816
3	0.933905
4	0.932098
5	0.936228
6	0.932554
7	0.932242
8	0.938109
9	0.933778

## 4.- Brill vs Perceptron





# 5.- Spacy

```
!pip install -U spacy  
!python -m spacy download es_core_news_sm
```

```
import spacy  
nlp = spacy.load("es_core_news_sm")  
doc = nlp(text)
```

## 5.- Spacy

Al procesar el texto, se devuelve un array con todos los tokens.

**.text** : Texto original del token  
**.pos\_** : Etiquetado morfosintáctico  
**.label\_** : Entidad

```
Alicia/PROPN/PERSON  
estaba/AUX/O  
soñando/VERB/O  
con/ADP/O  
un/DET/O  
conejo/NOUN/ANIMAL  
blanco/ADJ/O
```

## 5.- Spacy

A/ADP  
través/NOUN  
de/ADP  
la/DET  
tarde/NOUN  
color/NOUN  
de/ADP  
oro/NOUN

/SPACE

el/DET  
agua/NOUN  
nos/PRON  
lleva/VERB

...

La primera/MISC  
Conseguido al fin el silencio/MISC  
Ya/MISC  
País de las Maravillas/LOC  
Ponlo/LOC

## 5.- Spacy

```
!python -m spacy download es_core_news_lg
```

Mas/MISC

La primera/MISC

La segunda/MISC

Conseguido al fin el silencio/MISC

País de las Maravillas/LOC

Alicia, para ti este cuento infantil/MISC

Ponlo/LOC



**FIN**