



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Departamento de sistemas informáticos y
computación

Evaluación de etiquetadores morfosintácticos para el español

Trabajo Lingüística Computacional

Máster Universitario en Inteligencia Artificial, Reconocimiento de Formas
e Imagen Digital

Autores

Andrés Marín Galán
Francisco Javier Gil-Terrón Rodríguez

2021 - 2022

Tabla de contenidos

1. Introducción	4
2. Etiquetador hmm con juego de categorías completas y reducidas	5
2.1 Corpus original o completo.....	5
2.2 Corpus reducido.....	6
2.3 Conclusiones	7
3. Prestaciones respecto a cantidad de datos de aprendizaje (hmm)	8
4. Métodos de suavizado para palabras desconocidas.....	10
5. Evaluación del resto de etiquetadores.....	12
6. Evaluación del paquete Freeling	14
7. Preguntas propuestas.....	15

1. Introducción

En este trabajo trataremos de emplear diferentes etiquetadores morfosintácticos, a través del paquete NLTK (*Natural Language Toolkit*) para evaluar sus respectivas prestaciones realizando distintas pruebas sobre el corpus en español *cess-esp*. Así pues, mediante la variación de distintos parámetros de entrada, se estudiará cómo afecta dicha variación a los resultados obtenidos, en términos de la precisión o *accuracy*, de los etiquetadores.

Algunas de las pruebas que se van a realizar son la modificación del tamaño del conjunto de entrenamiento o la variación del tamaño de sufijo empleado en el método de suavizado para palabras desconocidas, además de comparar los resultados obtenidos por diferentes etiquetadores.

Para ello, se hará uso de la técnica de validación cruzada o *cross-validation* de talla diez, es decir, se dividirá el corpus en diez particiones y se emplearán nueve para entrenar el etiquetador y la restante se utilizará para testear, de manera que se harán diez ejecuciones distintas donde la partición sea cada una de las partes, y el conjunto de entrenamiento las nueve restantes.

2. Etiquetador hmm con juego de categorías completas y reducidas

Comenzaremos con la evaluación del etiquetador hmm tanto para el corpus que contiene las categorías originales, como para el corpus con las categorías reducidas que creamos en prácticas anteriores.

En esta primera evaluación, se empleará el etiquetador hmm basado en modelos de Markov mediante la técnica de *cross-validation* con diez particiones. Para ambos casos, tanto sobre el corpus original, como el reducido, se trabajará sobre ellos después de haberlos barajado.

Los resultados obtenidos, para cada una de las particiones han sido los siguientes:

2.1 Corpus original o completo

Partición	Precisión	Intervalo de confianza
0	0.893	[0.868 - 0.918]
1	0.894	[0.869 - 0.918]
2	0.891	[0.866 - 0.916]
3	0.895	[0.871 - 0.92]
4	0.898	[0.874 - 0.922]
5	0.894	[0.869 - 0.918]
6	0.892	[0.868 - 0.917]
7	0.89	[0.865 - 0.915]
8	0.895	[0.871 - 0.92]
9	0.897	[0.873 - 0.921]

Tabla 1 - Etiquetado hmm con el corpus original

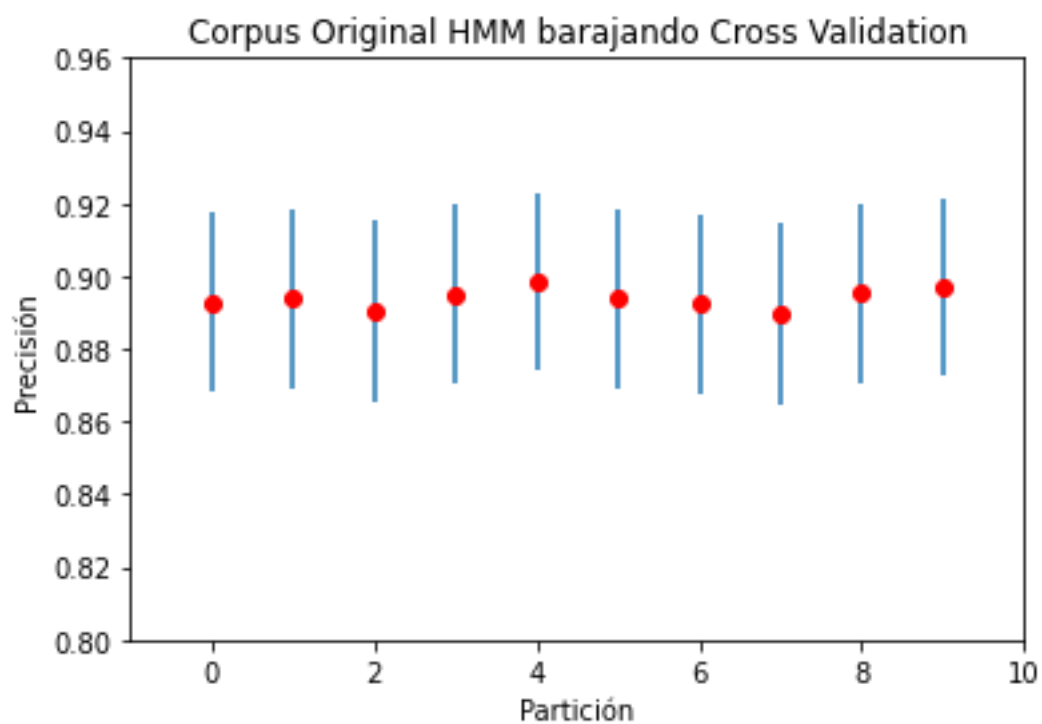


Figura 2 - Etiquetado hmm con el corpus original

2.2 Corpus reducido

Partición	Precisión	Intervalo de confianza
0	0.929	[0.909 - 0.95]
1	0.925	[0.904 - 0.946]
2	0.927	[0.906 - 0.947]
3	0.925	[0.904 - 0.946]
4	0.927	[0.906 - 0.947]
5	0.927	[0.906 - 0.948]
6	0.925	[0.904 - 0.946]
7	0.927	[0.906 - 0.948]
8	0.924	[0.903 - 0.945]
9	0.923	[0.902 - 0.945]

Tabla 2 - Etiquetado hmm con el corpus reducido

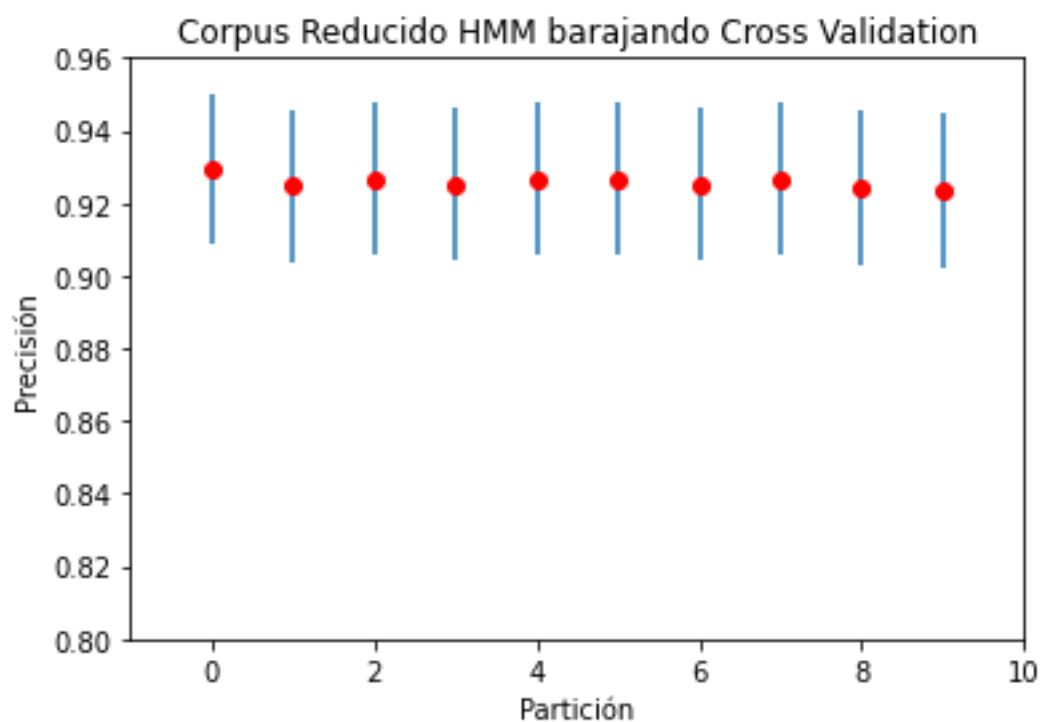


Figura 2 - Etiquetado hmm con el corpus reducido

2.3 Conclusiones

Como se puede observar, las prestaciones son mejores en el caso del corpus con categorías reducidas, puesto que la precisión obtenida es mayor para cada una de las diez particiones de la validación cruzada. Estos resultados siguen lo esperado, ya que el corpus reducido simplifica las categorías haciéndolas más cortas y unificando en grupos algunas de estas categorías, reduciendo su variabilidad, además de eliminar palabras con respecto al corpus completo.

3. Prestaciones respecto a cantidad de datos de aprendizaje (hmm)

En esta sección trataremos de estudiar la variación en las prestaciones del etiquetador hmm en función de la talla del conjunto de entrenamiento. Para ello se ha particionado el corpus reducido barajado en diez partes, las primeras nueve se utilizarán para entrenar de manera incremental y la última para testear. En otras palabras, primero se entrena únicamente con la primera parte y se irá aumentando el número de particiones con la que se realiza el entrenamiento hasta nueve particiones de manera progresiva. Obtenemos, como es lógico, una precisión ascendente conforme aumenta la cantidad de datos de entrenamiento, y un intervalo de confianza más acotado.

Nº de particiones	Precisión	Intervalo de confianza
1	0.845	[0.817 - 0.874]
2	0.875	[0.848 - 0.901]
3	0.891	[0.867 - 0.916]
4	0.903	[0.879 - 0.926]
5	0.909	[0.887 - 0.932]
6	0.916	[0.894 - 0.938]
7	0.922	[0.901 - 0.943]
8	0.927	[0.906 - 0.947]
9	0.929	[0.909 - 0.95]

Tabla 3 - Etiquetado hmm variando la cantidad de datos para entrenamiento

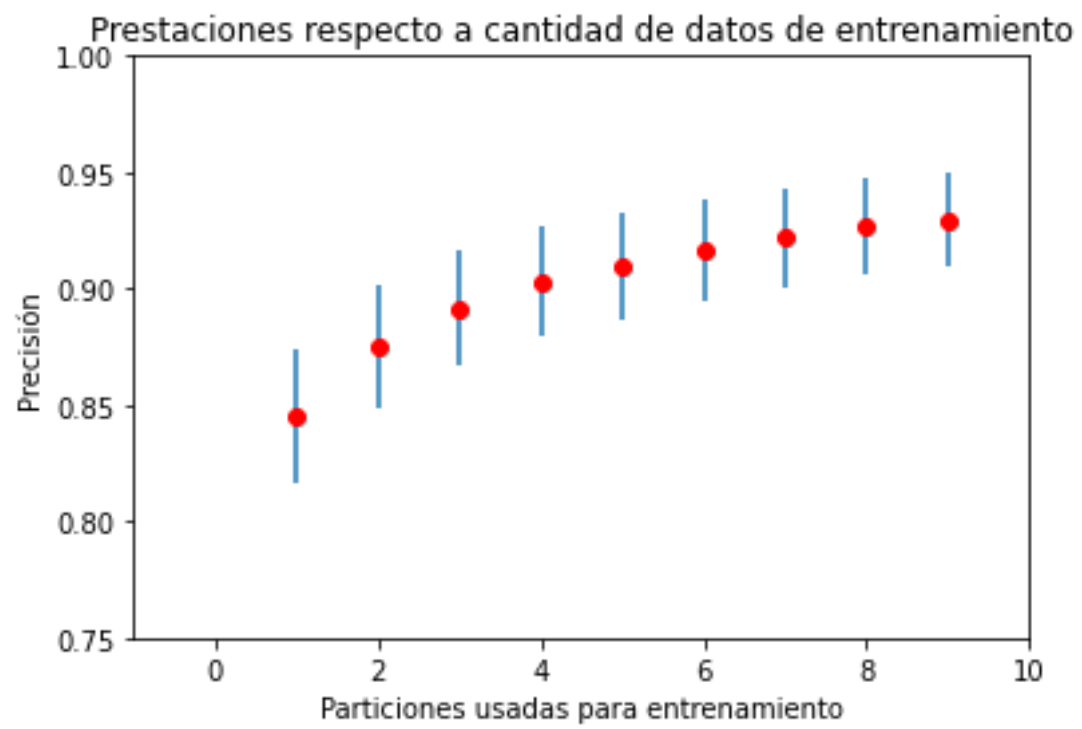


Figura 3 - Etiquetado hmm variando la cantidad de datos para entrenamiento

4. Métodos de suavizado para palabras desconocidas

En la actual sección se evaluará el etiquetador *tnt* en función del tamaño del sufijo para el método de suavizado *Affix Tagger* que se aplica a las palabras desconocidas, también haciendo uso del corpus reducido barajado. Para ello se ha medido la precisión para tamaños de sufijo de cero a seis donde se concluye que el mejor valor del tamaño del sufijo es tres por ser el que da la mejor precisión, y reduciendo la precisión al aumentar o disminuir el tamaño del sufijo con respecto a tres.

Tamaño sufijo	Precisión	Intervalo de confianza
0	0.899	[0.875 - 0.923]
1	0.929	[0.909 - 0.95]
2	0.939	[0.92 - 0.958]
3	0.948	[0.931 - 0.966]
4	0.946	[0.928 - 0.964]
5	0.937	[0.917 - 0.956]
6	0.925	[0.904 - 0.946]

*Tabla 4 - Etiquetado *tnt* variando el tamaño de sufijo para suavizado*

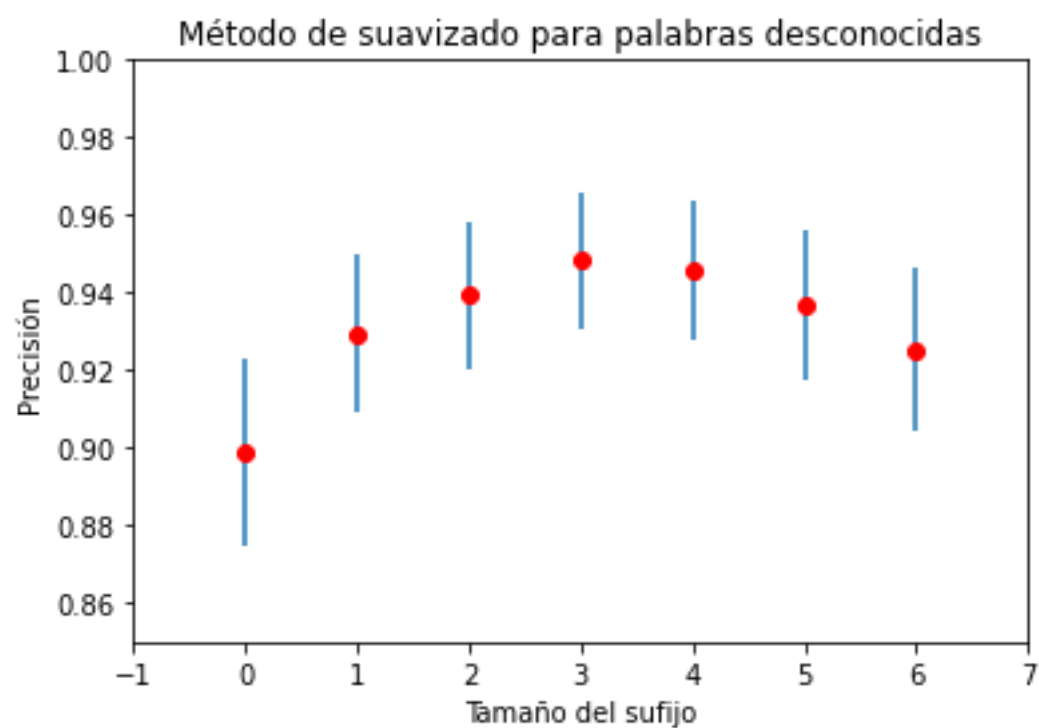


Figura 4 - Etiquetado tnt variando el tamaño de sufijo para suavizado

5. Evaluación del resto de etiquetadores

A continuación, los resultados obtenidos anteriormente por los etiquetadores hmm y tnt serán contrastados con otros etiquetadores morfosintácticos. Estos serán el etiquetador Brill con etiquetado inicial *Unigram Tagger* y hmm, etiquetador CRF y etiquetador Perceptrón, utilizando, una vez más, el corpus reducido barajado.

En este caso no se ha empleado la validación cruzada para estas pruebas y se tomará la evaluación de una única partición.

Etiquetador	Precisión	Intervalo de confianza
Brill-Unigram Tagger	0.897	[0.873 - 0.922]
Brill-Hmm Tagger	0.926	[0.905 - 0.947]
CRF	0.958	[0.942 - 0.974]
Perceptrón	0.968	[0.955 - 0.982]

Tabla 5 – Evaluación de otros etiquetadores

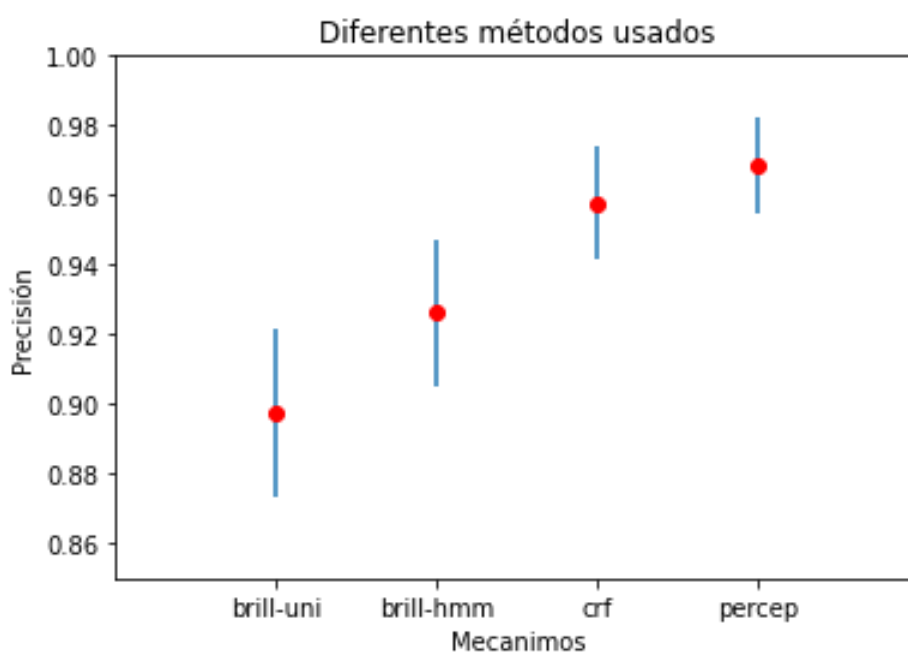


Figura 5 – Evaluación de otros etiquetadores

Como se puede ver, los peores resultados los obtenemos con el etiquetador Brill usando etiquetado inicial Unigram Tagger, con unos valores similares a los que se obtuvieron con hmm usando entre tres y cuatro particiones para entrenar, o a los obtenidos con tnt con un sufijo de tamaño para el método de suavizado aplicado a palabras desconocidas.

Estas prestaciones las mejora Brill si usamos como etiquetador inicial hmm, con valores similares a hmm con casi todo el conjunto de entrenamiento (unas ocho particiones) o a tnt con un tamaño de sufijo de uno (no tendremos en cuenta la parte decreciente tras el máximo con un tamaño de tres).

El siguiente etiquetador es CRF, cuyas prestaciones superan tanto el máximo obtenido mediante hmm empleando el conjunto entero de entrenamiento (nueve particiones) como a tnt usando tres como tamaño de sufijo para el suavizado, el mejor de los valores para esto.

Finalmente, el etiquetador Perceptrón, consigue mejorar los resultados incluso de CRF, obteniendo por tanto las mejores prestaciones de entre todos los etiquetadores usados hasta este momento.

6. Evaluación del paquete Freeling

Para finalizar, se hará uso de la herramienta Freeling sobre el texto Alicia.txt proporcionado, para realizar un etiquetado morfosintáctico. Tras su descarga e instalación, a partir del fichero **Alicia.txt** se ha generado uno nuevo con el análisis realizado por Freeling mediante el comando:

```
analyze -f es.cfg --outlv tagged <Alicia.txt >output.tag
```

Posteriormente se ha eliminado el segundo y cuarto elemento de cada línea y pasado a **output.txt** (que será adjuntado junto con la memoria), de manera que obtenemos un archivo de texto con cada palabra y su categoría etiquetada por Freeling en el formato palabra / categoría.

En cuanto a la herramienta, tras su uso podemos hacer un breve análisis de los siguientes aspectos:

- Problemas de instalación: No nos ha funcionado todas las versiones ni en todos los sistemas operativos por lo que hemos tenido bastantes problemas en el proceso.
- Facilidad de uso: no ofrece una buena facilidad de uso siendo los comandos toscos y poco intuitivos.
- Documentación: posee poca documentación, pero la poca existente es entendible y funciona.
- Funcionalidad: parece hacer un buen trabajo a la hora de asignar categoría a las palabras.

7. Preguntas propuestas

¿Alguien ha probado a contrastar las prestaciones obtenidas por la librería propuesta (NLTK) con otras librerías enfocadas al tratamiento de lenguaje natural, como pueden ser CoreNLP o spaCY?

Nosotros hemos optado por llevar a cabo el contraste con los etiquetadores propuestos en la guía del trabajo, estos son: etiquetador Brill con etiquetado inicial Unigram Tagger y hmm, CRF y Perceptrón.

¿A la hora de evaluar los distintos etiquetadores, se ha tenido en consideración el tiempo de computo que ha requerido cada uno?

En nuestro caso, pese a no haberlo citado en la memoria, hemos podido observar que la ejecución del etiquetador Perceptrón era considerablemente más rápida que CRF y el etiquetador Brill con etiquetado inicial de hmm, siendo solo superado en tiempo de ejecución por el etiquetador Brill con etiquetado inicial Unigram Tagger, aunque este obtenía las peores prestaciones en términos de precisión.