

Tarea del tema 2: Estudio de distintos corpus

Busca información en la web sobre distintos corpus: Brown, Susanne y Penn Treebank. Algunos de ellos no son gratuitos, pero si es accesible su descripción, conjunto de etiquetas que utilizan, etc.

Debes entregar un documento, que contenga una sección para cada uno de los siguientes puntos:

- **Descripción de los corpus**
- **Comparativa** concisa (tabla o similar) de distintos aspectos que consideres relevantes:
 - Tipo de etiquetado: etiquetado léxico (*POS tagging*), sintáctico, etc.
 - Tamaño del corpus,
 - Tamaño del conjunto de etiquetas,
 - Temáticas incluidas,
 - Procedencia de los textos: periódicos, transcripciones de habla, etc.
- Breve **análisis** (5 líneas como máximo) de qué corpus, el de Susanne o el de Brown, es más apropiado en función de sus características para extraer información estadística significativa referente a cuales son las etiquetas y las parejas de etiquetas consecutivas que aparecen más frecuentemente en los textos.