

Lingüística Computacional Ejercicio 11
---

1. (100 points) Utilizar el Corpus Brown como modelo para extraer los parámetros de un HMM: probabilidades de transición y probabilidades de emisión.
2. (50 points) Usar una implementación propia del algoritmo de Viterbi para obtener la secuencia más probable de categorías para la secuencia:

*She paid attention to them in ways that still feel connected to love. But this is only part of what makes the exhibition timely.* (Del *The Washington Post* del 26 de marzo de 2021)

### Observaciones importantes

El etiquetado del Corpus Brown es muy complejo, como hecho manualmente por expertos lingüistas. Vean en la web el repertorio de etiquetas y las combinaciones. Tomen las decisiones de simplificación que crean razonables **como lingüistas**. El resultado tiene que tener sentido **desde el punto de vista de la Lingüística no de la Informática**.

Remitan (i) la implementación que han escrito del algoritmo de Viterbi; (i) la justificación lingüística de las decisiones que han tomado sobre las etiquetas; (iii) la implementación de una función a la que pasamos el texto en 'raw' del Corpus Brown y devuelve como diccionarios las probabilidades de transición y las de emisión; (iv) la secuencia de categorías asignada al texto de *The Washington Post*