

Etiquetado Automático de Marcadores Discursivos

Dennis Daniel González Durán¹, Marta Lourdes Baguer Díaz-Romañach², Lydia Amalia Castro Odio³

¹Facultad de Matemática y Computación, Universidad de La Habana

²Facultad de Matemática y Computación, Universidad de La Habana

³Facultad de Artes y Letras, Universidad de La Habana

Resumen

Los marcadores discursivos son unidades lingüísticas invariables (formadas por estructuras diversas) que cumplen funciones fuera del marco estrictamente oracional relacionadas con el contexto de la comunicación, y su interpretación (Martín Zorraquino y Portolés Lázaro (1999)). Esencialmente orientan las inferencias que puedan derivarse de los enunciados y contribuyen a la coherencia, la cohesión y la eficacia del discurso para que este sea comprendido satisfactoriamente. Los marcadores forman un grupo heterogéneo que desde la lingüística se ha intentado sistematizar sobre todo a partir del papel que cumplen en el segmento discursivo en que se insertan. El análisis del comportamiento de estas unidades en muestras reales de uso, en diferentes contextos, resulta de mucha utilidad para los lingüistas por lo que se propone la implementación de un etiquetador automático de estas unidades para permitir su localización y estudio en grandes volúmenes de textos (específicamente, la herramienta está pensada para los corpus compilados por el proyecto Corespuc, marco en que se inserta el presente estudio).

Gran cantidad de los trabajos similares encontrados en la literatura se centran en clasificar los marcadores discursivos basándose en sus funcionalidades en los textos. Entre estas investigaciones destaca la de Robledo y Nazar (2018), que propone una clasificación para estas unidades usando técnicas de agrupamiento. No son muchos los trabajos relacionados directamente con el etiquetado de marcadores discursivos. Entre los existentes es de interés el de García Toro et al. (2022), que implementa un etiquetador de estas unidades utilizando modelos de aprendizaje automático, aunque no proporciona una clasificación para los marcadores discursivos encontrados.

Se propone una metodología basada en el trabajo de García Toro et al. (2022), extendiendo su etiquetador a una herramienta con la funcionalidad de clasificar las partículas discursivas etiquetadas.

Palabras clave: Marcadores discursivos, etiquetado automático, aprendizaje automático, clasificación.

Referencias

- García Toro, A., Porta Zamorano, J., & Moreno-Sandoval, A. (2022). A Discourse Marker Tagger for Spanish using Transformers. *Procesamiento del Lenguaje Natural*, 68, 123-132. <https://doi.org/10.26342/2022-68-09>
- Martín Zorraquino, M. A., & Portolés Lázaro, J. (1999). Los marcadores del discurso. En *Gramática descriptiva de la lengua española*. Editorial Espasa-Calpe.
- Robledo, H., & Nazar, R. (2018). Clasificación automatizada de marcadores discursivos. *Procesamiento del Lenguaje Natural*, 61, 109-116. <https://doi.org/10.26439/2018-61-12>