

# ¿Capturan los Modelos del Lenguaje el Significado?

## Reflexiones desde la Lingüística

Lucía Pitarch Ballesteros<sup>1</sup>, Jorge Gracia del Río<sup>1</sup>, Jordi Bernad<sup>1</sup>

<sup>1</sup> Afiliación: Sistemas de Información Distribuidos (SID)  
Instituto de Investigación en Ingeniería de Aragón (I3A)  
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.  
Tel. +34-976762707, e-mail: [lpitarch@unizar.es](mailto:lpitarch@unizar.es)

### Resumen

Los Modelos de Lenguaje han revolucionado la lingüística automatizando diferentes procesos semánticos, sin embargo, tareas complejas como la metáfora presentan margen de mejora. Este trabajo reflexiona desde una perspectiva lingüística sobre posibles puntos de mejora de estos modelos mediante un enfoque híbrido que aúne la semántica distribucional con la cognitiva, multimodal y multilingüe.

### Introducción

Los modelos del lenguaje pre-entrenados (PLMs) basados en la arquitectura transformer [1] han revolucionado el campo del procesamiento del lenguaje natural (PLN), que es la rama de la inteligencia artificial (IA) que estudia la interacción entre los ordenadores y el lenguaje humano. La capacidad de estos modelos para generar texto coherente y contextualmente relevante ha llevado a muchos a preguntarse si, en su esencia, han logrado capturar el significado del lenguaje humano. Sin embargo, esta pregunta fundamental es más compleja de lo que parece, especialmente cuando se la aborda desde una perspectiva lingüística.

Según Escandell, el significado emerge de diversas fuentes interconectadas [2]. En primer lugar, existe un significado referencial, que surge del proceso de asociar palabras como "mesa" con el objeto real que representan en el mundo físico. Esta conexión directa entre la palabra y la realidad es una base fundamental de la comprensión. En segundo lugar, el significado se construye a través de acuerdos culturales y sociales, donde un grupo poblacional converge en denominar una serie de experiencias o estados emocionales como "felices" o "tristes". Estos significados son intrínsecamente intersubjetivos y dependen de un contexto compartido. Finalmente, Escandell señala los procesos introspectivos, en los que al pensar una palabra, generamos una imagen

mental o una serie de asociaciones cognitivas que contribuyen a su significado individual.

Paradójicamente, el funcionamiento de la mayoría de los modelos del lenguaje contemporáneos no se basa directamente en estas suposiciones sobre el significado. En cambio, su poder reside en lo que se conoce como semántica distribucional. Esta teoría, arraigada en las ideas del lingüista Ferdinand de Saussure sobre el lenguaje como un sistema de signos, postula que el significado se define por las palabras de las que suele estar rodeada [9]. En este marco, el significado de una palabra se construye a partir de sus relaciones de coocurrencia con otras palabras en grandes corpus de texto. Así, surge una duda crucial: ¿es suficiente para capturar el significado realizar operaciones probabilísticas de coocurrencia de palabras para inferir qué palabras suelen acompañarlas?

### La Evidencia de la Semántica Distribucional

En nuestro trabajo "No clues, good clues: Out of context Lexical Relation Classification" [5] mostramos cómo modelos del lenguaje tipo BERT [1] fine-tuneados con prompts muy sencillos, eran capaces de capturar una cantidad significativa de información semántica, por lo menos en lo referente a relaciones léxico-semánticas, con una correlación de Spearman muy cercana a los resultados humanos sobre 0.80 en la identificación de hiperónimos (muy cerca del acuerdo humano entre anotadores, el cual ronda el 0.83 y el 0.86) y con métricas F1 de entre 0.60 para sinónimos y por encima del 0.70 y de 0.80 para otras relaciones (hiperonimia, meronimia y antonimia). Estos resultados, basados en patrones de coocurrencia, sugieren que la semántica distribucional proporciona al modelo una base de conocimiento léxico-semántico considerable, aunque con margen de mejora para algunas relaciones como la sinonimia.

En el caso de tareas más complejas como la identificación e interpretación de metáforas, que

requieren un conocimiento profundo del mundo real y de los acuerdos socioculturales entre los hablantes, los modelos actuales muestran un margen considerable de mejora. Por ejemplo, en el conjunto de datos para la identificación de metáforas más utilizado como benchmark: VUA [4], los modelos actuales no han conseguido superar el 0.73 de F1[3]. Por otro lado, al preguntar a ChatGPT por metáforas en la frase: “[El cáncer] es un bache más que te pone la vida [...], pero que una vez te lo ha puesto solo tienes que luchar y seguir adelante.” no identifica la connotación violenta de *luchar* ni la personificación de *vida*, sólo reconoce la palabra *bache*.

## Exploración de enfoques híbridos

Continuando la línea de investigadores que abogan por la integración de información no distribucional, consideramos que para aportar una mayor granularidad a la información semántica en los modelos de lenguaje, se requiere añadir datos que trasciendan la mera coocurrencia.

Para ello, hemos trabajado en explorar la teoría de los mapas semánticos [10]. A través de comparar cómo se lexicaliza un mismo concepto en diferentes lenguas, podemos inyectar a los modelos de lenguaje información sobre cómo se organiza el significado a nivel sociocultural en las lenguas [6].

Otra de las vertientes que nos interesa es experimentar con la inyección de información cognitiva. En esa dirección, y en línea con Rai et al, (2016) [8], estamos trabajando en investigar la relación entre estas medidas psicolingüísticas y su impacto en la identificación de metáforas.

Finalmente, los trabajos multimodales también resultan claves para el procesamiento del lenguaje [7] y, consecuentemente, para una comprensión más holística del significado. Esta, sin embargo, es una línea de investigación que esperamos explorar en el futuro, reconociendo su potencial para cerrar la brecha entre el significado distribucional y el conocimiento encarnado en la experiencia sensorial.

## Conclusiones

La pregunta de si los modelos del lenguaje capturan el significado es compleja y multifacética. Si bien la semántica distribucional ha demostrado ser extraordinariamente eficaz para aprender relaciones léxico-semánticas y patrones contextuales, sus limitaciones se hacen patentes en tareas que exigen un conocimiento más profundo del mundo real y fenómenos lingüísticos complejos como la metáfora.

La investigación actual apunta hacia la necesidad de enfoques híbridos que complementen la información distribucional con información cognitiva, social y multimodal.

## REFERENCIAS

- [1]. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding (No. arXiv:1810.04805).
- [2]. Escandell Vidal, M. V. (2004). Fundamentos de semántica composicional (1. ed). Ed. Ariel.
- [3]. Ge, M., Mao, R., & Cambria, E. (2023). A survey on computational metaphor processing techniques: From identification, interpretation, generation to application. *Artificial Intelligence Review*, 56(S2), 1829-1895.
- [4]. Leong, C. W. (Ben), Beigman Klebanov, B., Hamill, C., Stemle, E., Ubale, R., & Chen, X. (2020). A report on the 2020 vua and toefl metaphor detection shared task. *Proceedings of the Second Workshop on Figurative Language Processing*, 18-29.
- [5]. Pitarch, L., Bernad, J., Dranca, L., Bobed Lisbona, C., & Gracia, J. (2023). No clues good clues: Out of context Lexical Relation Classification. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 5607-5625.
- [6]. Pitarch, L., Bobed Lisbona, C., Abián, D., Gracia, J., & Bernad, J. (2024). Building muscle, a dataset for multilingual semantic classification of links between entities. En N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 2580-2594). ELRA and ICCL.
- [7]. Qi, F., Lv, C., Liu, Z., Meng, X., Sun, M., & Zheng, H.-T. (2022). Sememe prediction for babelnet synsets using multilingual and multimodal information. *Findings of the Association for Computational Linguistics: ACL 2022*, 158-168.
- [8]. Rai, S., Chakraverty, S., & Tayal, D. K. (2016). Supervised metaphor detection using conditional random fields. *Proceedings of the Fourth Workshop on Metaphor in NLP*, 18-27.
- [9]. Saussure, F. de, & Saussure, F. de. (1980). *Curso de lingüística general* (C. Bally, Ed.). Akal ed.
- [10]. Vanhove, M. (Ed.). (2008). *From polysemy to semantic change: Towards a typology of lexical semantic associations*. John Benjamins Pub.