

# Extracción no supervisada de patrones sintácticos para la identificación de relaciones de oposición léxica en español

**Unsupervised Extraction of Syntactic Patterns for Identifying Lexical Opposition Relations in Spanish**

Alejandro Pimentel-Alarcón 

IIMAS, Universidad Nacional Autónoma de México

Gerardo Sierra  

Instituto de Ingeniería, Universidad Nacional Autónoma de México

Alexander Gelbukh  

CIC, Instituto Politécnico Nacional

Alec Sánchez-Montero  

Instituto de Ingeniería, Universidad Nacional Autónoma de México

## Resumen

Los recursos léxicos son difíciles, costosos y requieren mucho tiempo para su producción y mantenimiento. En este trabajo, se propone una metodología para la extracción automática de relaciones léxicas de oposición, en particular de antónimos y cohipónimos, mediante patrones sintácticos yuxtapuestos, con el fin de contribuir a la construcción y expansión de recursos léxicos. A partir de un extenso corpus textual en español, reglas específicas y herramientas de análisis textual desarrolladas en Python, se implementó una metodología basada en tres módulos: (1) extracción de patrones de repetición, (2) filtración de cohipónimos mediante simetría, y (3) binarización de múltiples pares candidatos a antónimos. El sistema mostró un nivel de precisión de hasta 0.68 en la clasificación de los cohipónimos, de acuerdo con una evaluación basada en anotaciones humanas (1000 pares extraídos del corpus BLESS y 3000 de WordNet, clasificados como “cohipónimos” y “no cohipónimos” con acuerdos entre anotadores de 82.93 % y 62.17 %, respectivamente). Los resultados demuestran que esta metodología es capaz de identificar relaciones léxicas de oposición de manera eficiente, sin depender de recursos lingüísticos externos, lo cual puede contribuir al enriquecimiento y mantenimiento automatizado de recursos léxicos.

## Palabras clave

extracción automática de antónimos; extracción automática de cohipónimos; lingüística computacional; relaciones léxicas; patrones sintácticos

## Abstract

Lexical resources are difficult, costly and time-consuming to produce and maintain. In this paper, we

propose a methodology for the automatic extraction of lexical opposition relations, in particular antonyms and co-hyponyms, by means of juxtaposed syntactic patterns, in order to contribute to the construction and expansion of lexical resources. Using an extensive Spanish text corpus, specific rules and textual analysis tools developed in Python, we implemented a methodology based on three modules: (1) extraction of repetition patterns, (2) filtering of co-hyponyms by symmetry, and (3) binarization of multiple candidate pairs of antonyms. The system showed a precision level of up to 0.68 in classifying cohyponyms, according to an evaluation based on human annotations (1000 pairs extracted from the BLESS corpus and 3000 from WordNet, classified as “cohyponyms” and “non-cohyponyms” with inter-annotator agreements of 82.93 % and 62.17 %, respectively). The results demonstrate that this methodology is able to identify oppositional lexical relations efficiently, without relying on external linguistic resources, which can contribute to the automated enrichment and maintenance of lexical resources.

## Keywords

automatic antonym extraction; automatic cohyponym extraction; computational linguistics; lexical relations; syntactic patterns

## 1. Introducción

La extracción de relaciones léxicas es una tarea esencial para la construcción y expansión de recursos léxicos, pues implica la identificación de conexiones sistemáticas entre palabras dentro de un conjunto de textos. Dentro de estas relaciones, la oposición mantiene un papel clave en la organización del léxico de un texto, aunque ha reci-



DOI: 10.21814/lm.17.2.475

This work is Licensed under a



Creative Commons Attribution 4.0 License

bido menos atención en contraste con otros tipos de relaciones léxicas o semánticas. Los enfoques más comunes para la extracción automática de las relaciones léxicas se basan en la búsqueda de patrones en textos no estructurados (Al-Yahya et al., 2016). Sin embargo, estos métodos suelen presentar limitaciones de precisión y cobertura, especialmente en la identificación de relaciones de oposición. Además, en los modelos distribucionales como los *word embeddings*, muchos tipos de similitud quedan agrupados dentro de una proximidad general en el espacio vectorial (Scheible et al., 2013), lo cual dificulta la distinción entre relaciones semánticas específicas, como las de oposición. En contraste, un enfoque lingüístico permite identificar con mayor precisión patrones que manifiestan contraposición semántica, sin depender de representaciones vectoriales abstractas, al operar sobre estructuras sintácticas explícitas.

El objetivo de este trabajo es desarrollar un enfoque no supervisado para la extracción de relaciones léxicas de oposición (antonimia y cohiponimia) en español, a partir de un corpus de texto plano no estructurado, con el fin de contribuir a la construcción y expansión de recursos léxicos en español. El enfoque propuesto se basa en la identificación de patrones sintácticos dentro de ventanas funcionales de contexto, sin recurrir a etiquetados manuales o a herramientas externas. Con base en la teoría de Mettinger et al. (1994), se implementa un algoritmo compuesto por tres módulos interdependientes: yuxtaposición (que genera pares candidatos a oposición), simetría (que filtra pares asimétricos) y binarización (que identifica antónimos a partir de cohipónimos).

La estructura del presente artículo es la siguiente: en la sección 2 se introduce el concepto de relaciones léxicas, con énfasis en la antonimia y la cohiponimia; en 3, se revisan los enfoques existentes en la extracción de relaciones léxicas dentro del procesamiento del lenguaje natural; la sección 4 describe las características del corpus utilizado en la investigación; a continuación, en 5, se presenta la técnica utilizada para identificar los patrones sintácticos; después, la sección 6 detalla el algoritmo propuesto para la extracción de pares de oposición; luego, en 7 se muestran los pares extraídos; y, finalmente, en 8, se resumen los principales hallazgos del trabajo, con posibles aplicaciones y futuras líneas de investigación.

## 2. Relaciones léxicas

En general, las palabras pueden clasificarse en dos grandes categorías: léxicas (aquellas que tienen un significado concreto y específico, como sustantivos, verbos, adjetivos y algunos adverbios) o funcionales (las que no tienen un significado propio, sino que funcionan para establecer relaciones sintácticas entre las unidades léxicas dentro de una oración) (Conde, 2005). Como mencionan Cruse (1986), el significado de las unidades léxicas se construye a partir de un número indefinido de relaciones con el contexto. Por lo tanto, se puede hablar de unidades léxicas cuando toman una relación semántica particular al ponerse en relación con otras unidades léxicas; el significado de cada unidad se revela mediante las relaciones que guarda con su contexto y con el sistema de la lengua.

Las relaciones léxicas (o relaciones semánticas) entre las palabras permiten organizar el léxico de una lengua y comprender el significado de cada unidad en su contexto. De acuerdo con Lyons (1977), el sentido de una unidad léxica resulta de las múltiples relaciones que esta mantiene con otras expresiones dentro de la misma lengua. Por ejemplo, la palabra “perro”, además de denotar una clase de seres del mundo real, se conecta de diversas maneras con términos como “animal”, “galgo”, “terrier”, o “mastín”. Esto implica que el sentido de las palabras es una construcción relacional que depende tanto de las conexiones entre palabras dentro de un mismo léxico (relaciones intraléxicas) como de las que se establecen con términos de otros campos semánticos (relaciones interléxicas).

Entre las relaciones léxicas más destacadas se encuentran la similitud, la sinonimia, la hiperonimia, la hiponimia, la antonimia y la homonimia. En ciencias cognitivas, se diferencia entre similitud, que ocurre entre palabras con significados semejantes (como “gato” y “felino”), y asociación, que conecta palabras con significados distintos pero relacionados (como “coche” y “gasolina”). La sinonimia, por otro lado, implica una relación más estrecha, que coincide con la relación de identidad en la que dos palabras pueden intercambiarse en ciertos contextos sin alterar el significado de la oración (Yule, 2020).

Mientras que la sinonimia establece una relación de equivalencia entre palabras, la hiperonimia y la hiponimia son relaciones jerárquicas. Un término es *hiperónimo* cuando su significado abarca el de otros términos más específicos, llamados *hipónimos*; por ejemplo, “animal” es el hiperónimo de los hipónimos “perro”, “gato” y “elefante”. Por su parte, la antonimia se da en-

tre categorías que sólo cuentan con dos elementos, o bien en una línea gradual de elementos con dos extremos bien definidos (por ejemplo, “blanco” y “negro”). Finalmente, la homonimia es una relación entre palabras que tienen la misma forma, aunque presentan significados diferentes, como “banco” en el contexto de una institución financiera o un objeto para sentarse.

## 2.1. Relaciones de oposición

Las relaciones de oposición pueden adoptar diversas formas, desde oposiciones graduales hasta relaciones de exclusión mutua. Dentro de esta subclase, la antonimia es crucial, pues establece pares de términos que se diferencian por rasgos semánticos específicos, aunque no toda oposición entre palabras responde a los mismos principios estructurales. Los cohipónimos son un caso particular de oposición más gradual, ya que agrupan términos mutuamente excluyentes que comparten un mismo hiperónimo.

Para estudiar esta diversidad, Mettinger et al. (1994) distingue dos tipos fundamentales de oposición: sistemática y no sistemática. Por un lado, la oposición sistemática consiste en una relación estable de términos opuestos, independiente del contexto y anclada en la estructura semántica del lenguaje (por ejemplo, pares como *amor/odio* o *grande/pequeño*). Esta marcada separación de los términos suele utilizarse para resaltar contrastes en distintos contextos lingüísticos, como en el “Poema 20” de Pablo Neruda:<sup>1</sup>

Es tan **corto** el *amor*,  
y es tan **largo** el *olvido*.

En esta frase, la oposición sistemática *corto-largo* intensifica el contraste entre las ideas mencionadas (*amor/olvido*). La oposición no sólo destaca la diferencia entre el par de ideas, también crea una tensión emocional que resalta la naturaleza extrema de ambos conceptos.

Por otro lado, en la oposición no sistemática, el contraste se manifiesta en contextos específicos donde las dos palabras aparecen juntas para resaltar sus diferencias. Aunque no existe una contraposición inherente en su significado, se percibe una dicotomía, como en el par *amor/dinero* en frases populares como:

Con **dinero** tienes a la pareja que quieras,  
sin **dinero** tienes a la pareja que te **ama**.

La percepción de limitación en la coocurrencia de los términos se evidencia en relaciones que no son binarias, como lo expresa Alberto Moravia al describir una oposición no sistemática diferente (*amor/matrimonio*):

El **amor** es un juego;  
el **matrimonio**, un negocio.

Los contextos lingüísticos determinan las relaciones de oposición, y la frecuencia con la que se repiten estas relaciones genera la percepción de contraste en pares como *cabeza/cola* o *teoría/práctica*. Aunque el interés principal de Mettinger et al. (1994) se centra en la oposición sistemática, sus hallazgos indican que la oposición no sistemática es más frecuente al analizar patrones en el uso del lenguaje. A pesar de esta observación, la clasificación de oposiciones en términos tan específicos rara vez se emplea en tareas de extracción automática de relaciones léxicas. En la mayoría de los estudios, el término “antónimos” se utiliza de manera general, sin considerar la distinción detallada entre oposiciones sistemáticas y no sistemáticas.

## 3. Trabajo previo

La hiperonomia y la sinonimia han sido las relaciones léxicas más estudiadas en la literatura, aunque también existen trabajos enfocados en la antonimia y en la cohiponimia. Por lo general, estos enfoques siguen la tendencia de utilizar patrones y conjuntos de *palabras semilla*, términos iniciales seleccionados manualmente que sirven para identificar relaciones léxicas. Sin embargo, estos métodos manuales pueden ser limitados por su falta de escalabilidad y flexibilidad.

Lobanova et al. (2010), en su trabajo de extracción automática de relaciones léxicas, utilizan pares de antónimos semilla para detectar patrones en un corpus de neerlandés y extraer más pares de antónimos. Este proceso revela que los patrones que generan antónimos como *pobre-rico* también producen cohipónimos percibidos como opuestos en contextos específicos (por ejemplo, *atlético-flaco*).

A su vez, Mettinger et al. (1994) describen estructuras con palabras de oposición, como *X y Z* o *ni X ni Y*, que han sido estudiadas por otros autores como Justeson & Katz (1991), Fellbaum (1995) y Jones (2003). Estas estructuras coinciden con patrones simétricos (véase subsección 3.1). Por su parte, Jones et al. (2007) utilizan estos patrones para medir la estandarización de antónimos en corpus generales, aunque no se han

<sup>1</sup>[https://www.cervantesvirtual.com/obra-visor/poema-20-0/html/ffc28ba4-9c97-4592-97a8-f338821299f1\\_2.html](https://www.cervantesvirtual.com/obra-visor/poema-20-0/html/ffc28ba4-9c97-4592-97a8-f338821299f1_2.html)

encontrado investigaciones que los empleen para la extracción automática de antónimos.

Roth & Upadhyay (2019) detectan antónimos entre palabras mediante un método mixto, con un enfoque particular en distinguir sinónimos y antónimos a través de patrones y marcadores discursivos. En sus resultados, demuestran que las personas usan marcadores específicos para expresar contraste entre palabras. Asimismo, Roth & Im Walde (2014) emplean marcadores de discurso en inglés, como los del cuadro 1, para identificar relaciones discursivas de manera efectiva. El método de Roth & Upadhyay (2019) se distingue por integrar un módulo de traducción que asocia traducciones con modelos entrenados en varios idiomas.

CONTRAST	but, although, rather ...
RESTATEMENT	indeed, specifically, ...
INSTANTIATION	(for) example, instance, ...

**Cuadro 1:** Ejemplos de marcadores discursivos utilizados en Roth & Im Walde (2014).

Este último método para la detección de relaciones logra un alto rendimiento, con un valor  $F_1$  de hasta 0.7 en contextos monolingües y mejoras de hasta 0.099 puntos cuando se incorporan las traducciones de los marcadores discursivos. No obstante, presenta dos desventajas considerables. En primer lugar, requiere estudios lingüísticos previos enfocados en patrones para corpus generales, lo cual puede excluir casos especializados. En segundo lugar, el algoritmo sólo detecta la relación cuando las palabras coocurren exactamente en el patrón esperado.

Aunque los cohipónimos han recibido menos atención que los antónimos, han sido abordados en investigaciones relevantes. Por ejemplo, Jana & Goyal (2018) presentan un método supervisado para detectar cohipónimos, basado en una representación especial de grafos (Riedl & Biemann, 2013), construidos a partir de n-gramas sintácticos con *Google Books* como corpus. Específicamente, para calcular la similitud entre dos palabras en un grafo, Jana & Goyal (2018) recurren a cinco métricas propias de grafos: similitud estructural, camino más corto, camino más corto ponderado, densidad de conexión en la intersección de vecindades, y densidad de conexión en la unión de vecindades. La principal hipótesis del trabajo sostiene que, cuando dos palabras comparten una relación de cohiponimia, su similitud distribucional será mayor en comparación con otras relaciones como la hiperonimia o la meronimia.

### 3.1. Patrones simétricos

Un concepto relevante para esta investigación es el de *patrones simétricos*. Aunque no están específicamente diseñados para encontrar antónimos o cohipónimos —y generalmente no se consideran un método explícito para detectar oposición—, existen casos cuyos resultados pueden generar ciertas oposiciones, al grado de que puede considerarse la herramienta más cercana a la metodología desarrollada en esta investigación. Se puede observar cierta oposición en los resultados que logra extraer, especialmente en los patrones especiales que se mencionarán más adelante.

Davidov & Rappoport (2006) desarrollaron un algoritmo no supervisado para la extracción automática de patrones simétricos. Este algoritmo consiste en extraer secuencias de 3 a 5 tokens, donde se encuentren 2 palabras intercambiables y de 1 a 3 palabras fijas. El corpus se recorre en busca de los patrones más frecuentes que coincidan con esta plantilla. Los patrones se aceptan si el orden de aparición de la palabra menos frecuente es al menos un 15% (u otro valor asignado) respecto al de la palabra más frecuente.

## 4. Corpus

Wikipedia es una enciclopedia en línea colaborativa mantenida por voluntarios, cuyos artículos están bajo la licencia de *Creative Commons*. Esta característica, junto con la riqueza y diversidad de su contenido, la convierte en una fuente ideal para el análisis textual en diversos estudios de investigación. Para los experimentos de este trabajo, se utilizó el *dump* —o la copia completa de todos los artículos y metadatos de la enciclopedia en un formato estructurado— de Wikipedia en español correspondiente a enero de 2018<sup>2</sup>, el cual contiene 3.4 GB de texto plano, lo que corresponde a más de 500 millones de palabras en total y a más de 10 millones de palabras únicas.

Como parte del preprocesamiento y la normalización, se empleó *WikiExtractor*<sup>3</sup> para eliminar *boilerplate*<sup>4</sup> y convertir todas las letras a minúsculas.

<sup>2</sup>[https://en.wikipedia.org/wiki/Wikipedia:Database\\_download](https://en.wikipedia.org/wiki/Wikipedia:Database_download)

<sup>3</sup>Esta herramienta de Giuseppe Attardi está disponible en: <https://github.com/attardi/wikiextractor>

<sup>4</sup>*Boilerplate* son partes de documentos que forman parte de la estandarización; en *Wikipedia*, es código dentro de los artículos que les da formato y estructura.

## 5. Ventanas funcionales

Las ventanas funcionales son una herramienta para identificar y analizar patrones lingüísticos dentro de un corpus. En el contexto de esta investigación, se entienden como rangos o secuencias de palabras en las que los elementos de un sintagma nominal interactúan con otras estructuras gramaticales, como verbos, adjetivos u otros sintagmas. Estos patrones son la base para observar cómo las palabras se agrupan y se comportan dentro de un contexto determinado y, en consecuencia, se pueda extraer más fácilmente las relaciones léxicas. Los experimentos de esta sección tienen como objetivo explorar las posibilidades de las palabras funcionales y sus contextos para añadir información implícita al texto.

En primer lugar, se seleccionaron tipos de palabras funcionales para los experimentos y se organizaron en una lista conforme a su categoría: preposiciones, artículos y conjunciones. A partir de esta lista, se extrajeron pares de palabras léxicas, o de contenido, relacionadas con un elemento funcional mediante la siguiente expresión:

$$C_1 \ F \ C_2 \quad (1)$$

Donde  $C_1$  es una palabra de contenido,  $F$  es un elemento funcional, y  $C_2$  es otra palabra de contenido. En el cuadro 2, se presentan los resultados iniciales más frecuentes de esta extracción. Como puede observarse en dicho cuadro, las palabras obtenidas conforman, en varias ocasiones, expresiones o locuciones como “debajo del umbral” o “dar a conocer”. Estas secuencias reflejan una interacción funcional entre las palabras de contenido ( $C_1$  y  $C_2$ ) y los términos funcionales ( $F$ ), que forman unidades con significado específico. Sin embargo, en la mayoría de los casos, las expresiones extraídas tienden a estar sesgadas por la naturaleza del corpus enciclopédico utilizado. Esto se evidencia en construcciones recurrentes como “especie de ...”, que, aunque son semánticamente válidas, no necesariamente representan un uso lingüístico general, sino que están muy asociadas al estilo informativo y descriptivo característico de los artículos encyclopédicos.

A continuación, se llevó a cabo una búsqueda de construcciones comunes que permitieran el intercambio de una de sus partes, con el objetivo de identificar más palabras compatibles con la estructura. En otras palabras, se buscó un patrón que facilitara el reemplazo de componentes dentro de la construcción, lo que posibilitaría encontrar múltiples coincidencias relacionadas de manera similar a los patrones simétricos. Para ello,

$C_1$	$F$	$C_2$
especie	de	coleóptero
especie	de	peces
debajo	del	umbral
especie	de	arácnido
medalla	de	plata
continuación	se	brinda
especie	de	pez
selección	de	fútbol
dio	a	conocer
pez	de	mar
género	de	plantas
especie	de	anfibios

**Cuadro 2:** Pares de palabras más comunes alrededor de un elemento funcional

se continuó con la búsqueda de ventanas que incluyeran dos palabras de contenido, de la siguiente manera:

$$F_1 \ C_1 \ F_2 \ C_2 \ F_3 \quad (2)$$

Donde las  $F_n$  representan elementos funcionales, y las  $C_n$  representan palabras léxicas, o de contenido. Con esta estructura, se puede reemplazar una de las dos partes del par para encontrar palabras relacionadas e intentar formar campos semánticos alrededor de la palabra fija. En la columna izquierda del cuadro 3, se presentan las palabras asociadas a “coleóptero” (seleccionada a partir de los resultados iniciales anteriores), con la palabra “especie” como referencia en la ventana funcional: *una - de - de la*. De igual manera, en la columna derecha del mismo cuadro, se muestra el caso inverso: las palabras asociadas a “especie” (seleccionada también a partir de los resultados iniciales), si se toma como referencia la palabra “coleóptero”.

especie de $X$	$X$ de coleóptero
coleóptero	especie
peces	subespecie
pez	género
anfibios	
mantis	
roedor	
murciélagos	
ave	
gecos	
musaraña	
escarabajo	

**Cuadro 3:** Campos semánticos de coleóptero y especie

Como se refleja en estos cuadros, la distribución mediante palabras funcionales permite establecer asociaciones y relaciones, no sólo dentro de un campo semántico compartido, sino también a través del uso y el contexto en estructuras específicas de escritura. En este caso, inicialmente, se obtiene un grupo de especies y seres vivos, seguido por un modo de clasificación.

De manera similar, en el cuadro 4 se presentan las palabras asociadas a “fútbol” (tomada de los resultados iniciales) al considerar la palabra “selección” en la ventana funcional: *la - de - de*. Este proceso se repite para el caso inverso: se muestran las palabras asociadas a “selección” al considerar la palabra “fútbol”. Aunque estos ejemplos se enfocan en el campo del fútbol, cabe destacar que la elección de estos campos semánticos, como el de los seres vivos en el caso anterior, se basa en la disponibilidad de construcciones frecuentes en el corpus. No se utilizó un recurso semántico específico para extraer sólo información sobre seres vivos o fútbol. El estudio, por tanto, no se limita a un campo semántico específico, sino que analiza diversas áreas en función de las coincidencias y patrones emergentes en el corpus.

<b>selección de X</b>	<b>X de fútbol</b>
fútbol	selección
balonmano	asociación
baloncesto	federación
rugby	liga
voleibol	escuela
béisbol	confederación
polo	unión
críquet	eurocopa
costa	copa
sóftbol	sección
waterpolo	división
fútsal	academia

**Cuadro 4:** Campos semánticos de fútbol y selección

Los cuadros 4 y 3 ilustran cómo se agrupan las palabras relacionadas para formar campos semánticos. Cada patrón de ventanas funcionales genera dos campos semánticos: el primero, al permitir cambios en la segunda palabra (presentado en el subcuadro izquierdo de ambos cuadros de ejemplo); y el segundo, al permitir cambios en la primera palabra (mostrado en el cuadro derecho de los mismos cuadros).

## 5.1. Construcciones genitivas

El término “genitivo” hace referencia a una relación de posesión o pertenencia; en español esta función se expresa comúnmente mediante la preposición “de”. En las construcciones genitivas, esta palabra establece una relación entre los sustantivos e indica relaciones de propiedad, posesión, pertenencia, origen o materia entre ellos. Es una preposición de gran frecuencia y diversidad en español (Cruz Domínguez, 2011).

La mayoría de los patrones con ventanas funcionales siguen una estructura similar a los presentados en los cuadros anteriores, y la relación en torno a la preposición “de” se mantiene consistente con la variedad temática observada en los artículos de Wikipedia. En las ventanas funcionales, la preposición “de” es la más utilizada para relacionar palabras léxicas. Encontrar ejemplos con “de” en  $F_1$  sugiere divisiones inapropiadas en construcciones que deberían permanecer unidas, lo cual conduce a agrupaciones semánticas incorrectas. Esto lleva a considerar campos semánticos como “Del consejo de X de las” y “De todos los X en”. Ambos ejemplos comparten la presencia de la preposición “de” o la contracción “del” en  $F_1$ , lo que resulta en la segmentación errónea de estas frases en el marco de las ventanas funcionales. A continuación, se presentan los campos semánticos correspondientes para ilustrar cómo la posición del elemento funcional influye en la segmentación:

**Del consejo de X de las:** seguridad, rectores, redacción, administración, turismo.

**De todos los X en:** tiempos, cargos, santos, países, mamíferos, ciudadanos, tipos, tiempo, niños, niveles, puntos, derechos, existentes, trabajadores, siglos, programas, nacimientos, laboratorios, partidos, ejércitos.

Estos resultados menos favorables comparten la característica de que contienen “de” o “del” como el primer elemento funcional ( $F_1$ ). Este posicionamiento de los elementos funcionales “de” o “del” indica que las construcciones pueden estar siendo segmentadas de manera incorrecta, ya que estas preposiciones tienden a dividir frases que deberían mantenerse juntas. Para abordar este problema, se exploró un contexto más amplio para los patrones anteriores, lo que permitió encontrar ejemplos adicionales como los siguientes:

- La resolución del consejo de seguridad
- El líder de todos los tiempos

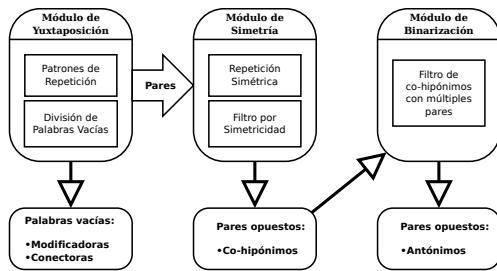
Al volver a aplicar la metodología, se lograron obtener campos semánticos más coherentes y precisos:

- resolución, reunión, presidencia, creación, decisión, aprobación, autoridad, presidenta, secretaría, sede
- líder, record, mejor, mundo, cine, ranking, cómic, segundo, metal

Estos resultados preliminares apoyan la hipótesis de que los patrones de palabras funcionales pueden ser utilizados para delimitar diversas unidades poliléxicas. Además, se destaca una tendencia a ubicar artículos en la posición  $F_1$ , lo cual constituye la base del experimento descrito en la sección 6.1, núcleo del mecanismo que explora un fenómeno poco investigado hasta ahora.

## 6. Metodología

Los experimentos iniciales proporcionaron detalles importantes sobre cómo utilizar las palabras funcionales para extraer palabras relacionadas, conforme al objetivo de este trabajo. Además, se comprobó que las palabras funcionales son eficaces para formar estructuras en las que se agrupan palabras de un mismo campo semántico, especialmente para denotar contraposición, contraste o aclaración de características compatibles. Este comportamiento apoya la hipótesis de que ciertas configuraciones gramaticales actúan como marcas fiables de relaciones semánticas específicas. La capacidad de rastrear directamente los patrones en el texto facilita también la trazabilidad y evaluación de los errores del sistema, algo que suele estar ausente en enfoques de “caja negra”. La figura 1 ilustra el proceso completo para la extracción de pares de oposición, compuesto por tres módulos consecutivos que refinan los pares de palabras de manera progresiva.



**Figura 1:** Método de extracción de pares de oposición. El proceso consta de tres módulos, cada uno con resultados secuenciales.

El primer módulo, de yuxtaposición, constituye la base para la construcción de un patrón inicial, o protopatrón, con posiciones para palabras funcionales y léxicas, pues genera los pares candidatos a oposición. El segundo módulo, de simetría, aplica un enfoque simétrico para refinar los pares, eliminar asimetrías y mejorar la calidad de la oposición. Este proceso es suficiente para generar listas de palabras con un grado de oposición comparable al de la cohíponimia. Finalmente, el tercer módulo, de binarización, diseñado para la extracción de antónimos, filtra los pares obtenidos en los pasos anteriores. Este paso culmina en la identificación de antónimos a partir de un conjunto de cohíponimos.

### 6.1. Módulo de repetición o yuxtaposición

Cuando se aborda la repetición, no debe interpretarse como instancias de figuras retóricas que utilizan la repetición para establecer un tema contextual. En este caso, la repetición se refiere a la que comúnmente se encuentra en la estructura de la escritura natural, similar a la yuxtaposición de ideas. Dubremetz (2017) señala que el lenguaje está lleno de repeticiones de palabras, muchas de las cuales pasan desapercibidas, ya que suelen ser “palabras vacías”. La repetición relevante para esta investigación ocurre cuando las estructuras se repiten de manera adyacente. En gramática, la yuxtaposición se utiliza para unir dos oraciones simples y formar una oración compuesta (Españaña & Madrid, 2001). Este fenómeno se refleja en los patrones propuestos, los cuales pueden incluir elementos conectores (como conjunciones o preposiciones), pero siempre de forma repetida. Por lo tanto, la estructura completa (palabra funcional - palabra léxica) se mantiene sin ningún conector adicional. Por ello, se pueden denominar *patrones yuxtapuestos*.

El método propuesto adapta parcialmente la técnica de Schwartz et al. (2015), con base en el uso de las palabras más frecuentes como palabras funcionales. Sin embargo, el enfoque que se basa en las primeras 1000 palabras más frecuentes, independientemente del idioma, presenta inconvenientes, ya que muchas de ellas no son funcionales. Las ventanas funcionales muestran dos sustantivos unidos por una palabra funcional. Se observa que el primer sustantivo está determinado por un artículo definido. Al unir sustantivos con artículos, se advierte que, similar al genitivo, los artículos pueden repetirse.

Esto llevó a buscar un patrón de repetición de artículos junto a sustantivos, a través de una expresión regular tipo *Perl* en Python.<sup>5</sup>

```
''(\w+) \w+ (\w+) \1 ''
```

Esta expresión regular captura dos aspectos clave. Se añadieron restricciones para evitar innecesarias repeticiones múltiples de una misma palabra. Además, se diseñó el Algoritmo 1, que genera listas de palabras funcionales según su posición y confiabilidad. Se añadieron verificaciones adicionales para evitar repeticiones y manejar casos atípicos. La condición de paro detecta palabras funcionales típicas de  $F_1$  en  $F_2$ , lo cual minimiza la aparición de patrones erróneos.

## 6.2. Módulo de simetría

Este módulo aplica un filtro a los pares extraídos del módulo anterior mediante la técnica de simetría. Un patrón simétrico permite que un par de palabras,  $X$  e  $Y$ , se intercambien dentro de una estructura, como se ejemplifica en construcciones como “ $X$  y  $Y$ ” o “tanto  $X$  como  $Y$ ”. La simetría se ajusta particularmente a un conjunto reducido de palabras, como las conjunciones. El módulo de simetría extrae diversas palabras funcionales conectoras y muestra grados variables de simetría en las estructuras sintácticas. Esta variabilidad puede ayudar a identificar relaciones más útiles en la búsqueda de oposición. Las construcciones genitivas, debido a la amplia variedad de relaciones que pueden expresar, no son confiables como identificadores de oposición. Para evaluar la simetría asociada a una palabra, se calcula una puntuación ( $S$ ) como la razón entre la frecuencia de la palabra en el contexto simétrico  $f_s$  y la frecuencia en el contexto total  $f_t$ , multiplicada por 100:

$$S = \frac{f_s}{f_t} \times 100$$

Este factor es ideal para filtrar palabras funcionales y excluir aquellas construcciones que representen diversas relaciones, especialmente las genitivas. El método de filtrado propuesto descarta los pares que provengan de la palabra fun-

<sup>5</sup>Se utilizó Python como lenguaje de programación, y se recurrió a expresiones regulares de estilo *Perl* por su capacidad para manejar de manera efectiva estructuras complejas. Dado el enfoque lingüístico del estudio, las expresiones regulares proporcionan una manera directa de capturar y manipular estructuras específicas sin recurrir a herramientas de análisis más complejas. Esta estrategia permite mantener un alto nivel de control interpretativo sobre los patrones sintácticos; al mismo tiempo, un modelo basado en reglas proporciona una solución transparente, reproducible y fácilmente adaptable a otros contextos lingüísticos o tipologías textuales.

ctional con la mayor disminución en la relación entre frecuencia y simetría.

## 6.3. Módulo de binarización

La cohiponimia y la antonimia están estrechamente relacionadas, ya que los antónimos son un subconjunto de los cohipónimos. La principal diferencia radica en la naturaleza binaria de los antónimos, que se dan entre categorías opuestas o en una escala gradual con dos extremos bien definidos (como “caliente” y “frío”). Suponemos que, si dos palabras son antónimas, también son cohipónimos binarios. A partir de esta premisa, se propone un método para extraer antónimos de cohipónimos mediante un algoritmo simple: del conjunto de pares de cohipónimos, se extraen aquellos que incluyen una palabra sin ningún otro par, lo cual indica que dicha palabra tiene una única contraparte y aumenta la probabilidad de que sean antónimos.

## 7. Extracción de pares

En esta sección se describe el proceso y se presentan ejemplos de los resultados de la metodología.

### 7.1. Módulo de yuxtaposición

Para comenzar, se aplica la expresión regular para obtener pares de palabras, como los del cuadro 5. Al analizarlos, se observa que predominan las palabras funcionales, como se había anticipado. Los resultados de los experimentos anteriores indican que en la posición conectora ( $F_2$ ) aparecen palabras como *de*, *y*, *en*, *a*, *con*, *por*, mientras que en la posición  $F_1$  se encuentran artículos como *la*, *el*, *los*, *las*. La preposición genitiva *de* aparece con mayor frecuencia en  $F_2$ . Aunque existen casos inesperados (“ciudad de”, “provincia de”, etc.), los filtros subsecuentes aplicados a las listas evitan que afecten al algoritmo.

Al aplicar el algoritmo 1, se obtuvieron los siguientes resultados:

**L<sub>F1</sub>:** la, el, los, las, del, su, un, al, se, the, san, más, una, lo, sus, dos, le.

**L<sub>F2</sub>:** de, y, en, a, por, es, que, con, como, para, desde, era, o, hacia, hasta, fue, contra, durante, entre, sobre, son.

Estos resultados indican que las palabras funcionales se agrupan según su posición en listas que desempeñan funciones específicas. De esta

<b>F<sub>2</sub></b>	<b>F<sub>1</sub></b>
de	la
y	la
y	de
ciudad	de
y	el
en	la
de	los
provincia	de
universidad	de
era	de
en	el
a	la

**Cuadro 5:** Pares de repetición extraídos directamente mediante expresión regular

forma, los elementos de  $L_{F1}$  pueden actuar como artículos (o palabras similares), mientras que los de  $L_{F2}$  pueden funcionar como preposiciones y conjunciones. Además de las listas de palabras funcionales, se obtienen pares de palabras léxicas. El cuadro 6 muestra el tipo de pares obtenidos durante este proceso.

<b>C<sub>1</sub></b>	<b>C<sub>2</sub></b>
guerra	independencia
final	copa
presidencia	república
bandera	ceremonia
humanidad	unesco
imagen	virgen
copa	uefa
mayoría	población
capital	provincia
letra	canción
norte	sur
agricultura	ganadería
organización	federación
clasificación	copa
casa	cultura
copa	liga
santos	últimos
población	ciudad
pedro	pablo
universidad	república

**Cuadro 6:** Pares de palabras léxicas extraídas de los patrones de repetición.

En el cuadro 7 se presentan los pares obtenidos para “fútbol”, igualmente seleccionada arbitrariamente como se hizo en el experimento de ventanas funcionales. En la columna izquierda (cuadro 7), se muestran los pares donde “fútbol” está en  $C_1$ , cuando se extrae  $C_2$ . En la columna de-

recha, se presentan los pares donde “fútbol” está en  $C_2$ , cuando se extraen las palabras en  $C_1$ .

<b>C<sub>1</sub> = “fútbol”</b>	<b>C<sub>2</sub> = “fútbol”</b>
deporte	baloncesto
baloncesto	rugby
club	club
rugby	béisbol
país	profesionalismo
atletismo	interés
equipo	país
béisbol	tenis
tenis	atletismo
ciclismo	equipo
hockey	ciclismo
voleibol	campeonato

**Cuadro 7:** Pares de “fútbol” resultado de los patrones de repetición.

Al comparar estos resultados con los obtenidos en las ventanas funcionales para la misma palabra “fútbol”, se observa una similitud, a pesar de tratarse de un contexto diferente. En las ventanas funcionales, los campos semánticos se obtenían a partir de una palabra pivote, utilizada como punto de partida para extraer información contextual o semántica, lo cual implicaba un cambio de contexto. En este nuevo experimento, no hay cambio de contexto, ya que las palabras aparecen en el mismo entorno que la palabra de interés. La ventaja de los pares de repetición radica en que no necesitan una palabra pivote para extraer campos semánticos, sino que basta con la palabra de interés para identificar las relaciones y conexiones semánticas dentro del contexto. Esto conlleva una extracción más directa de información sin necesidad de depender de una palabra como referencia.

## 7.2. Módulo de simetría

Un aspecto relevante en el cuadro 7 es la presencia de elementos repetidos en ambas columnas. Como era previsible, las palabras repetidas indican una relación más fuerte con la palabra de interés y, además, muestran una tendencia a coincidir en cohiponimia. Por ejemplo, la palabra “deporte”, a pesar de su relación de hiperonimia con “fútbol”, desaparece en la segunda columna, mientras que otros deportes forman parte de la intersección. Esta discrepancia podría explicarse por el hecho de que la hiperonimia no establece una relación simétrica, a diferencia de otras relaciones como la cohiponimia.

También se encuentran palabras con relaciones asimétricas (“país”, “club”, “equipo”) que permanecen en la intersección. En este punto, se aplicó el módulo de simetría para extraer los pares que ofrecen información esencial sobre la oposición. El cuadro 8 presenta los resultados de aplicar simetría en el caso de “fútbol”, mientras que el cuadro 9 muestra una muestra aleatoria de 14 elementos.

boxeo
ciclismo
béisbol
hockey
tenis
club
básquet
atletismo
básquetbol
equipo
balonmano
rugby

**Cuadro 8:** Intersección simétrica de pares de repetición para fútbol.

música	televisión
biblioteca	oficina
ciudad	posguerra
cuadrado	triángulo
cola	garganta
inglés	italiano
escultura	música
salvación	redención
gato	ratón
plomo	cinc
descenso	ascenso
justicia	caridad
guitarrista	bajista
occidente	norte

**Cuadro 9:** Intersección simétrica de pares de repetición para una muestra aleatoria.

En los resultados, la aplicación de la simetría destaca una preferencia por pares de palabras que pertenecen a la misma categoría semántica o que comparten un hiperónimo. Esto sugiere una tendencia hacia la obtención de cohipónimos vinculados a la oposición semántica y relevantes para las ontologías. Los ejemplos extraídos revelan una mejora en la calidad de la relación, lo cual está en línea con los objetivos de la investigación.

En esta fase, se aplicó el filtro de simetría para detectar el mayor cambio de rango entre las palabras funcionales, con el objetivo de identificar automáticamente la funcionalidad genitiva.

Aunque la funcionalidad genitiva establece relaciones asimétricas, su uso diverso en contextos (por ejemplo, “la mayoría de la acción” o “la acción de la mayoría”) puede generar coincidencias en patrones de simetría. No obstante, los pares simétricos resultantes no mantienen las mismas propiedades que aquellos con otros tipos de simetría. Por lo tanto, todos los pares de palabras extraídos por esta funcionalidad son descartados de los candidatos a oposición.

### 7.3. Módulo de binarización

Tras la aplicación de la simetría, los pares de palabras mejoran cualitativamente su capacidad para diferenciar entre elementos de una misma categoría, con énfasis en cohipónimia y, en algunos casos, en la identificación de antonimia. Para completar la metodología y extraer palabras con un alto grado de oposición —y, por ende, más propensas a ser antónimos—, se aplicó el módulo de binarización. El cuadro 10 muestra una selección aleatoria de 20 pares obtenidos bajo esta restricción.

audición	visión
aimara	quechua
salarios	precios
calor	friño
compradores	vendedores
consciente	inconsciente
terroristas	rehenes
docentes	alumnos
oveja	cabra
suicidio	asesinato
tejidos	órganos
diablo	ángel
semitono	tono
emoción	razón
objetos	personajes
racionalismo	empirismo
muertos	vivos
melodías	letras
ventana	puerta
aritmética	geometría

**Cuadro 10:** Muestra aleatoria de pares binarios

Estos resultados muestran la generación de pares con una oposición significativa, muchos de los cuales son comúnmente usados como antónimos prototípicos (por ejemplo, *calor/friño*, *emoción/razón*). Este constituye el último paso de nuestra metodología para la extracción de palabras de oposición, lo que, de acuerdo con el objetivo delineado, permite obtener dos conjuntos de palabras: antónimos y cohipónimos, extraídos

de manera no supervisada a partir de un texto plano. La ausencia de entrenamiento supervisado refuerza la aplicabilidad del enfoque en escenarios con escasos recursos, a la vez que preserva la posibilidad de ajustar los criterios lingüísticos subyacentes, una característica cada vez más valorada frente al carácter opaco de los modelos extensos de lenguaje.

#### 7.4. Evaluación de la extracción

La evaluación de los pares extraídos automáticamente como cohipónimos se realizó mediante etiquetado manual, pues no se ha considerado conveniente llevar a cabo una evaluación mediante la comparación con recursos léxicos existentes. Para la evaluación se siguió el enfoque de Weeds et al. (2014). Se recurrió a dos etiquetadores externos y se utilizaron dos corpora: BLESS y WordNet, pues han sido cuidadosamente balanceados de tal manera que la mitad de los elementos son cohipónimos y los elementos que no lo son mantienen otro tipo de relación de similitud.

Los etiquetadores hicieron una clasificación binaria de los pares en “cohipónimos” y “no cohipónimos”. Para el caso del corpus que se construyó a partir de los datos de BLESS, se etiquetaron 1000 pares de palabras, con un porcentaje de acuerdo entre anotadores del 82.93 % y un coeficiente  $K$  de Cohen de 0.6295; para el corpus construido a partir de WordNet, se etiquetaron 3000 pares, con un porcentaje de acuerdo de 62.17 % y un coeficiente  $K$  de Cohen de 0.2674. En términos de precisión, BLESS mostró una media de 0.68, mientras que WordNet alcanzó el 0.66 en función del *gold standard* de los etiquetadores. Estos resultados, resumidos en el cuadro 11, indican que el sistema propuesto extrae de manera precisa pares de cohipónimos, con un acuerdo considerablemente mejor que otros enfoques como los patrones simétricos.

Corpus	Precisión
BLESS	0.68
WORDNET	0.66

Cuadro 11: Evaluación de precisión de pares de cohipónimos extraídos de los corpora BLESS y WORDNET.

Además, se evaluó la extracción de los pares de antónimos como una clasificación entre oposición sistemática y oposición no sistemática, a partir de los cohipónimos. Los candidatos a antónimos fueron clasificados con el apoyo de dos equipos de etiquetadores humanos. Para evitar el sesgo,

un equipo se dedicó exclusivamente a la clasificación de la oposición sistemática, mientras que el otro se encargó de la clasificación de la oposición no sistemática.

El equipo de oposición sistemática —una categoría más restrictiva de antónimos—, alcanzó un acuerdo moderado del 75.47 % y un coeficiente  $K$  de Cohen de 0.4234, con un 30.61 % de pares clasificados como antónimos puros. Por su parte, el equipo de oposición no sistemática —una clase menos restrictiva de antónimos—, obtuvo un acuerdo del 72.30 % y un coeficiente  $K$  de Cohen de 0.4232, con un 39.73 % de los pares reconocidos como antónimos ocasionales. Los resultados de la clasificación de los pares se muestran en la 12. La diferencia de percepción entre oposición sistemática y no-sistemática es pequeña, lo cual sugiere que el filtro por binarización funciona mejor para la primera, aunque aún es necesario mejorar los filtros para obtener antónimos más precisos.

Oposición	Pares clasificados
Sistémica	30.61 % Antónimos puros
No sistemática	39.73 % Antónimos ocasionales

Cuadro 12: Evaluación de pares de antónimos por categoría: oposición sistemática y no sistemática.

## 8. Conclusiones

En este trabajo se ha propuesto una metodología para la extracción automática de relaciones léxicas basadas en oposición, mediante la identificación de patrones sintácticos yuxtapuestos en textos. A lo largo de la investigación, se ha demostrado que esta metodología es efectiva para identificar pares de palabras con relaciones de oposición, como antónimos y cohipónimos, a partir de un corpus de texto no estructurado. Los resultados obtenidos indican que la binarización y el refinamiento de pares de palabras pueden mejorar la calidad de los pares extraídos automáticamente, con un alto grado de precisión para identificar antónimos y cohipónimos.

Sin embargo, se reconoce que el filtrado de los pares extraídos debe ser más preciso para asegurar que los antónimos identificados sean de naturaleza estricta y, al mismo tiempo, se minimice la inclusión de falsos positivos. La estrategia de binarización, aunque útil, aún presenta limitaciones, especialmente cuando se trata de identificar relaciones de oposición más sutiles o contextuales que no se ajustan a las categorías estrictas de antónimos puros. Además, el proceso de refinamiento de los pares podría beneficiarse de una

mayor consideración del contexto semántico más amplio en el que se encuentran las palabras, lo que permitiría detectar mejor las relaciones de oposición menos evidentes.

La evaluación del sistema con etiquetadores humanos ha ayudado a comprobar la validez de los pares extraídos. Este proceso ha proporcionado un indicador numérico para la precisión de los pares clasificados automáticamente como antónimos y cohipónimos. Además, la intervención de los etiquetadores constituye una oportuna evidencia para refinar y ajustar los parámetros del sistema de extracción automática en el futuro. Para futuras iteraciones, sería útil incorporar más métricas cuantitativas de rendimiento, con el fin de evaluar con mayor precisión la efectividad de los patrones extraídos y los ajustes realizados al sistema.

Los resultados de una metodología de extracción automática no supervisada de relaciones de oposición, como la que se ha propuesto en este trabajo, son particularmente relevantes para el enriquecimiento de recursos léxicos, como ontologías y bases de datos de relaciones léxicas, puesto que implican nuevas posibilidades para automatizar su actualización y expansión. En consecuencia, los avances logrados en la identificación y clasificación de relaciones léxicas complejas, a través de la extracción de pares de antónimos y cohipónimos mediante patrones sintácticos yuxtapuestos, puede mejorar la precisión y la eficiencia de los sistemas de procesamiento del lenguaje natural. Además, al tratarse de una estrategia interpretable, reproducible y de bajo costo computacional, el enfoque propuesto representa una alternativa viable frente a soluciones basadas en modelos de aprendizaje profundo.

Para futuras líneas de investigación, se podrían desarrollar métodos más sofisticados para identificar y clasificar relaciones léxicas a través de un análisis semántico y sintáctico más detallado. Por ejemplo, se podría investigar estrategias de filtrado y refinamiento de pares que puedan incorporar más reglas heurísticas basadas en el contexto y la coocurrencia de términos en textos más variados. Además, se podrían proponer metodologías para integrar conocimientos de recursos léxicos ya existentes, como diccionarios y tesauros, para mejorar la precisión de las relaciones extraídas sin recurrir a modelos de aprendizaje automático.

## Agradecimientos

Este trabajo recibió el apoyo del proyecto UNAM-PAPIIT “Extracción automatizada de in-

formación para la catalogación del patrimonio documental” con clave IG400725, así como del proyecto SECIHTI “Grafos conceptuales para la construcción de diccionarios inversos en áreas de especialidad” con clave FC-2023-G-64. Agradecemos sinceramente a los revisores por sus valiosas observaciones, comentarios y sugerencias.

## Anexos

A continuación, se presenta el algoritmo utilizado para extraer los pares funcionales de oposición.

---

**Algoritmo 1** Extrae pares de palabras funcionales en dos grupos  $F_1$  y  $F_2$

---

**Require:** Lista ordenada por frecuencia de pares de repetición ( $L_R$ ).

**Ensure:** Dos listas ( $L_{F1}$  y  $L_{F2}$ ) de palabras funcionales.

```

1:  $T_{F2} \leftarrow L_R[0]$ 
2: for all  $R_1, R_2$  in  $L_R$  do
3:   if  $R_1 = T_{F2}$  then
4:     break
5:   end if
6:   if  $R_2$  in  $L_{F1}$  then
7:     continue
8:   end if
9:   Agregar  $R_1$  a  $L_{F1}$ 
10:  Agregar  $R_2$  a  $L_{F2}$ 
11: end for
12: return  $L_{F1}, L_{F2}$ 
```

---

## Referencias

Al-Yahya, Maha, Sawsan Al-Malak & Lulu AlHubayi. 2016. Ontological lexicon enrichment: The Badea system for semi-automated extraction of antonymy relations from Arabic language corpora. *Malaysian Journal of Computer Science* 29(1). 56–73.  
[doi 10.22452/mjcs.vol29no1.5](https://doi.org/10.22452/mjcs.vol29no1.5)

Conde, Xavier Frias. 2005. *An introduction to syntax according to generative grammar theories*. Romania Minor

Cruse, David Alan. 1986. *Lexical semantics*. Cambridge University Press

Cruz Domínguez, Isasema. 2011. *El sintagma nominal en la extracción de relaciones léxico-semánticas de contextos definitorios: el caso de la preposición “de”*: Universidad Nacional Autónoma de México. Tesis de licenciatura. ↗

- Davidov, Dmitry & Ari Rappoport. 2006. Efficient unsupervised discovery of word categories using symmetric patterns and high frequency words. En *21<sup>st</sup> International Conference on Computational Linguistics and the 44<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (COLING/ACL)*, 297–304. doi 10.3115/1220175.1220213
- Dubremetz, Marie. 2017. *Detecting rhetorical figures based on repetition of words: Chiasmus, epanaphora, epiphora*: Uppsala Universitet. Tesis Doctoral
- Española, Real Academia & España Madrid. 2001. *Diccionario de la lengua española*. Real Academia Española
- Fellbaum, Christiane. 1995. Co-occurrence and antonymy. *International Journal of Lexicography* 8(4). 281–303. doi 10.1093/ijl/8.4.281
- Jana, Abhik & Pawan Goyal. 2018. Network features based co-hyponymy detection. arXiv [cs.CL]. doi 10.48550/arXiv.1802.04609
- Jones, Steven. 2003. *Antonymy: A corpus-based perspective*. Routledge
- Jones, Steven, Carita Paradis, M Lynne Murphy & Caroline Willners. 2007. Googling for ‘opposites’: A web-based study of antonym canonicity. *Corpora* 2(2). 129–155. doi 10.3366/cor.2007.2.2.129
- Justeson, John S & Slava M Katz. 1991. Co-occurrences of antonymous adjectives and their contexts. *Computational linguistics* 17(1). 1–19. ↗
- Lobanova, Anna, Tom Van der Kleij & Jennifer Spenader. 2010. Defining antonymy: A corpus-based study of opposites by lexico-syntactic patterns. *International Journal of Lexicography* 23(1). 19–53. doi 10.1093/ijl/ecp039
- Lyons, John. 1977. *Semantics: Volume 2*, vol. 2. Cambridge University Press
- Mettinger, Arthur et al. 1994. *Aspects of semantic opposition in english*. Oxford University Press
- Riedl, Martin & Chris Biemann. 2013. Scaling to large data: An efficient and effective method to compute distributional thesauri. En *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 884–890. ↗
- Roth, Michael & Sabine Schulte Im Walde. 2014. Combining word patterns and discourse markers for paradigmatic relation classification. En *52<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics (COLING/ACL)*, 524–530. doi 10.3115/v1/P14-2086
- Roth, Michael & Shyam Upadhyay. 2019. Combining discourse markers and cross-lingual embeddings for synonym–antonym classification. En *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 3899–3905. doi 10.18653/v1/N19-1390
- Scheible, Silke, Sabine Schulte im Walde & Sylvia Springorum. 2013. Uncovering distributional differences between synonyms and antonyms in a word space model. En *6<sup>th</sup> International Joint Conference on Natural Language Processing (IJCNLP)*, 489–497. ↗
- Schwartz, Roy, Roi Reichart & Ari Rappoport. 2015. Symmetric pattern based word embeddings for improved word similarity prediction. En *19<sup>th</sup> Conference on Computational Natural Language Learning (CoNLL)*, 258–267. doi 10.18653/v1/K15-1026
- Weeds, Julie, Daoud Clarke, Jeremy Reffin, David Weir & Bill Keller. 2014. Learning to distinguish hypernyms and co-hyponyms. En *25<sup>th</sup> International Conference on Computational Linguistics (COLING)*, 2249–2259. ↗
- Yule, George. 2020. *The study of language*. Cambridge university press