

Hacia una extracción semiautomática de metáforas conceptuales en un corpus de economía a partir del procesamiento de lenguaje natural.

Nieves Uribe, Stephany y Molina Mejia, Jorge Mauricio.

Cita:

Nieves Uribe, Stephany y Molina Mejia, Jorge Mauricio (2023). *Hacia una extracción semiautomática de metáforas conceptuales en un corpus de economía a partir del procesamiento de lenguaje natural*. *Estudios de Lingüística Aplicada*, 41 (76), 39-67.

Dirección estable: <https://www.aacademica.org/jorge.mauricio.molina.mejia/72>

ARK: <https://n2t.net/ark:/13683/pqc6/Dvk>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. *Acta Académica* fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

Hacia una extracción semiautomática de metáforas conceptuales en un corpus de economía a partir del procesamiento de lenguaje natural

Towards a semi-automatic extraction of conceptual metaphors in a corpus of economics from natural language processing

Stephany Nieves Uribe

Universidad de Antioquia, Facultad de Comunicaciones y Filología, Semillero de investigación: Corpus Ex Machina
stephany.nievesu@udea.edu.co

Jorge Mauricio Molina Mejía

Universidad de Antioquia, Facultad de Comunicaciones y Filología, Grupo de Estudios Sociolingüísticos, Semillero de investigación: Corpus Ex Machina
jorge.molina@udea.edu.co

Recepción: 14 de diciembre del 2021

Aceptación: 23 de agosto del 2022

doi: 10.22201/enallt.01852647p.2023.76.1027

Resumen

Este artículo plantea el fenómeno lingüístico de la metáfora conceptual aplicado al dominio de la economía desde la perspectiva de la lingüística computacional y de corpus a partir del análisis de *n*-gramas y colocaciones. Para tal fin se ha constituido un corpus de artículos del campo de la economía, el cual se encuentra dividido en dos subcorpus, ambos representados por revistas académicas en dicha área. El análisis textual del reconocimiento automático de metáforas se realizó mediante herramientas y programas informáticos que tienen como finalidad el procesamiento de lenguaje natural, entre estos, algunas librerías especializadas de Python y la herramienta LncsBox. La idea consiste en verificar una hipótesis inicial, según la cual en el discurso empleado por los economistas se tiende a utilizar, en gran medida, metáforas conceptuales que ayudan a comprender aquellos fenómenos ligados al tema de la economía, las cuales subyacen claramente dentro de las publicaciones académicas. Los resultados obtenidos ofrecen un acercamiento a la visualización de las metáforas conceptuales más utilizadas en el campo económico, validando la hipótesis inicial.

Palabras clave: análisis metafórico; ciencias económicas; lingüística cognitiva; lingüística computacional; expresiones metafóricas

Abstract

This article presents the linguistic phenomenon of conceptual metaphor applied to economics from the computational and corpus linguistics perspectives, approached through the analysis of *n*-grams and collocations. To this end, a corpus of articles on economics has been created, divided into two sub-corpora represented by academic journals in that area. The textual analysis of the automatic recognition of metaphors was carried out using computer tools and software for natural language processing, including some specialized Python libraries and the LncsBox tool. The aim is to explore the initial hypothesis, i.e., the discourse used by economists involves a tendency to use, to a large extent, conceptual metaphors that help to understand phenomena related to economics and that clearly underlie academic publications. Finally, the results from this research work offer an approach to visualizing the most commonly used conceptual metaphors in economics, which would validate the hypothesis mentioned above.

Keywords: metaphorical analysis; economic sciences; cognitive linguistics; computational linguistics; metaphorical expressions

1. Introducción

En términos generales, la metáfora conceptual refiere a una teoría aplicada habitualmente por la lingüística cognitiva, focalizada en el análisis de las categorías conceptuales utilizadas por el ser humano para organizar y dar sentido a sus experiencias vivenciales (Soriano, 2012). Esta variante de la metáfora permite comprender y experimentar pensamientos abstractos por medio de la asociación o vinculación a otros conceptos. De esta manera, se toman elementos por lo general adquiridos de modo experiencial con el fin de abordar o comparar una abstracción con una situación física, mediante un ejercicio de reflexión cognitivo que es empleado con diversos propósitos, como comprender mejor una idea, desenvolverse en el entorno social (Lakoff & Johnson, 2001) o simplemente como ornamento lingüístico (Alarcón, Díaz, Vergara, Vásquez & Torres, 2018).

Algunos trabajos que analizan el empleo de metáforas conceptuales en textos especializados en economía se han centrado en aspectos relacionados con los campos semánticos, como es el caso de Goegebeur (2016), quien examina un corpus de textos en español y holandés. También se encuentra la propuesta de Langer (2015), la cual efectúa una revisión de metáforas en textos de economía en inglés a partir de un mapeo de lexías ligadas a la construcción de dichas metáforas. En lo que respecta a la lengua castellana, es posible encontrar algunos trabajos como el de Rivera-León (2017), en el cual se estudia la metáfora conceptual en el campo de la enseñanza del español como lengua extranjera (ELE); o lo planteado por Loma-Osorio Fontecha (2011), relativo a la producción léxica metalingüística de las metáforas en temas económicos desde un

* Este artículo se encuentra enmarcado dentro del proyecto “Propuesta metodológica para el etiquetado de un corpus lingüístico con fines de identificación de metáforas conceptuales en español”, financiado por el Comité para el Desarrollo de la Investigación (Codi), Universidad de Antioquia (Colombia), código 2020-34124.

estudio comparativo español-inglés. Cabe resaltar que en ninguno de estos dos últimos trabajos se considera el estudio de corpus o el empleo de herramientas computacionales para la identificación de metáforas. Por otra parte, dichas investigaciones permiten evidenciar algunas metáforas exclusivas del ámbito de la economía, así como otras relacionadas con diferentes campos del conocimiento humano.

En el ejercicio de análisis que se llevará a cabo en el presente artículo se adopta un enfoque fundamentado en la lingüística de corpus, según una perspectiva computacional, sobre metáforas conceptuales aplicadas en el ámbito de los textos académicos en español, y que versan sobre temas de economía en general. El trabajo se centra en la hipótesis de que se trata de un fenómeno que se evidencia en tales discursos con el fin de dar a entender al lector conceptos abstractos y propios de esta área por medio de palabras más conocidas o concretas. Es así como el objetivo consiste en comprobar dicha hipótesis para identificar e interpretar el contexto de las metáforas más frecuentes en este tipo de publicaciones con métodos que involucran lenguaje de programación, como es el caso de Python y la herramienta LanksBox (Brezina, Weill-Tessier & McEnery, 2020). En síntesis, este procedimiento tiene por finalidad demostrar de qué manera dicho fenómeno cognitivo constituye una construcción que ayuda a reflejar y entender una visión de mundo. Para ello, se ha constituido un corpus de artículos académicos en español, extraído de dos revistas de economía, que pertenece a importantes instituciones de educación superior de Colombia.

2. Marco teórico

2.1. Metáforas conceptuales

La metáfora conceptual ha sido definida por diversos autores a lo largo de la historia; Lakoff y Johnson (2001) la definen como una construcción mental que se manifiesta mediante el lenguaje, presente tanto en la vida diaria como en la construcción de la ex-

perencia. Las metáforas conceptuales constituyen un fenómeno cognitivo presente en la generalidad de las culturas, y que parte de dos elementos fundamentales o dominios experienciales (Kövecses, 2020): el *dominio fuente*, que es aquel conjunto de conocimientos derivado por lo general de la experiencia y de percepciones físicas y concretas de la vida, y el *dominio meta*, que se trata de aquel cúmulo de aprendizajes que, a diferencia del anterior, no nace de percepciones concretas, sino de análisis reflexivos y abstracciones (Soriano, 2012). Cristina Soriano afirma que a dicho fenómeno subyace un proceso en el que se entrecruzan ambos dominios, lo que permite crear relaciones entre ellos con fines adaptativos, explicativos o que pretenden la comprensión de los demás eventos del entorno.

Expresado en otros términos, la metáfora conceptual no se limita a una atribución lingüística propia de una cultura, sino que abarca un fenómeno abstracto que se da en toda la población y que se muestra de distintas maneras. Soriano (2012) reafirma esta postura al decir que: “son esquemas abstractos de pensamiento que se manifiestan de muchas formas, entre ellas el lenguaje. Estas expresiones lingüísticas pueden variar de una lengua a otra, aunque la metáfora conceptual sea la misma” (98), dejando claro, además, que es un esquema que va más allá del uso de la lengua y que se extiende a aspectos tales como los gestos, los comportamientos, el arte, o hasta los objetos y herramientas que cierto grupo construye en su cotidianidad. Asimismo, Lakoff y Johnson señalan que “las metáforas conceptuales se usan constantemente de manera automática, sin esfuerzo y sin darnos cuenta” (Soriano, 2012: 107), implicando que existen situaciones de tipo “inconsciente” en las que se manifiestan las metáforas conceptuales más básicas. No obstante, este suceso no siempre ocurre de esa manera, sobre todo si se considera el dominio meta y la reflexión sobre aprendizajes posteriores, lo que conduce a asociaciones conscientes.

En este sentido, es posible afirmar que la producción de metáforas también surge de un ejercicio completamente consciente, deliberado y complejo, dejando entrever cómo la interacción con

la experiencia es uno de los elementos más importantes en la construcción de categorías conceptuales. Para ello, conviene que los estímulos visuales, auditivos, sensitivos, olfativos, además de las situaciones vividas según la percepción de nuestro entorno, deriven en referentes constitutivos del *dominio fuente*, los cuales servirán en la construcción de las metáforas como puntos de partida para entender la realidad a través de acciones comparativas. De tal forma, el *dominio meta* puede estar entrelazado con conceptos más abstractos y complejos, no tan fácilmente comprensibles mediante la percepción, sino a través del uso racional de las herramientas cognitivas con las que cuenta un individuo.

Para entender mejor la teoría antes expuesta, citamos un ejemplo cotidiano propuesto por Lakoff y Johnson, la famosa metáfora *el tiempo es dinero*; en la cual se expresa el tiempo en términos de dinero, indicando que este bien material (en este caso, el dominio fuente) es apreciado, valioso e importante; por lo cual el tiempo (que corresponde al dominio meta) también es importante y está lleno de valor. Esta metáfora conceptual trae consigo diversas expresiones metafóricas tales como: *ahorrar tiempo*, *ganar tiempo*, *malgastar tiempo*, *prestar tiempo*, *invertir el tiempo en algo*, entre otras.

2.2. Metáforas y economía

El uso de metáforas propias del área de economía es bastante recurrente, ya que estas integran un recurso clave para desarrollar ideas abstractas, así como ajenas a la cotidianidad de las personas; es decir, para explicar conceptos especializados a través de un lenguaje ordinario y de ciertas asociaciones, con el objetivo de ser claros y contundentes. Múltiples autores han abordado el tema, como Henderson (1982), que explica tres tipos de metáforas: *metáfora imagen*, *metáfora genérica* y *metáfora específica*. En el caso particular del último tipo de metáfora, se advierte que contiene un valor heurístico y ayuda a nombrar los conceptos económicos, dado que “las metáforas específicas son las que provienen

de unidades terminológicas” (Gallego-Iniesta, 2022: 299); sus dominios desempeñan una función prominentemente referencial para denominar esa realidad del discurso económico.

Por otro lado, McCloskey (1983) asegura que, para formular ideas e hipótesis complejas, los economistas hacen uso de las metáforas conceptuales, lo que implica que sea un tipo de discurso “*heavily metaphorical*” (493). Por otro lado, varias investigaciones adelantadas por Boers (2000), Charteris-Black (2000), Charteris-Black y Ennis (2001), Charteris-Black y Mussolff (2003) y Skorczynska y Deignan (2006), estudian la metáfora convencional propuesta por Lakoff y Johnson dentro del área de economía, identificando ciertos campos semánticos frecuentes, como se muestra en el Cuadro 1.

CUADRO 1. Principales metáforas conceptuales en el campo de la economía según diversos estudios

| Charteris-Black (2000) | Boers (2000) | Charteris-Black y Ennis (2001) | Charteris-Black y Mussolff (2003) |
|------------------------|--------------|--------------------------------|-----------------------------------|
| Enfermedad | Guerra | Guerra | Guerra |
| Salud | Salud | Ser vivo | Salud |
| Ciclo de vida humano | Movimiento | Movimiento | Catástrofe natural |
| Familia | Camino | | |

Por su parte, Gotti (2008) habla sobre algunas ventajas que implica el uso de las metáforas conceptuales en el discurso económico. Dos de ellas son la concisión y la transparencia, refiriéndose la primera a la capacidad de desarrollar definiciones técnicas con mayor facilidad, siempre que el recurso metafórico remita a información que se encuentre en la mente del oyente de manera precedente; mientras que la segunda alude a la claridad que brindan las metáforas conceptuales al reflejar términos o conceptos complejos y abstractos.

Para finalizar esta sección, cabe señalar que, aparte de las teorías ya expuestas, existen estudios más recientes como los propuestos por María Muelas (2016, 2018, 2019a, 2019b), quien ha trabajado constantemente este fenómeno. Además de apoyarse en

bases teóricas, Muelas hace uso de diferentes herramientas y procedimientos actuales que han demostrado ser confiables y que siguen progresando debido a los avances informáticos. Esta línea de investigación ha planteado una clara relación entre el procesamiento de lenguaje natural y la lingüística para actuales y futuros estudios en este campo u otros relacionados.

3. Metodología propuesta para el análisis de metáforas en economía

3.1. Constitución del corpus

Para la recopilación de los textos, se requirió que pertenecieran al campo de las ciencias económicas. Esto debido a que su práctica discursiva se desarrolla en el ámbito académico, en el cual se establece un diálogo con personas pertenecientes a dicha área, como economistas, estudiantes, investigadores, docentes, entre otros. Se emplearon dos revistas especializadas: *Cuadernos de Economía* (a partir de ahora CDE), de la Universidad Nacional de Colombia, que tiene como objetivo “publicar en el ámbito académico nacional y latinoamericano, principalmente, los avances intelectuales en teoría, metodología y aplicaciones económicas” (CDE, 2021), y *Revista de Economía Institucional* (RDEI), de la Universidad Externado de Colombia, que se trata de “una publicación plural, abierta a la discusión y difusión de trabajos teóricos e investigaciones en el área de la economía institucional” (RDEI, 2021).

Con el fin de alcanzar el objetivo propuesto por este estudio, se recopilaron únicamente los artículos en español de los números 79 y 81 del volumen 39 de CDE, así como el contenido total del volumen 22 de la RDEI, que incluye los números 42 y 43. El corpus total está constituido por 42 artículos, 21 para cada revista a estudiar. Todos ellos fueron publicados durante el mismo año, lo cual permitió contar con una base sólida, equilibrada y actualizada.

Dado que en esta investigación se quiso explorar una nueva perspectiva, y puesto que las herramientas emergentes operan de

manera más rápida y compleja los datos que involucran el procesamiento de lenguaje natural, se determinó trabajar con Python. Este lenguaje se utilizará como un instrumento para la detección de expresiones metafóricas a partir de *scripts*, algoritmos y estructuras de datos que permiten realizar limpiezas, búsquedas, análisis, clasificaciones, etc., de manera mucho más rápida y exhaustiva que otros lenguajes de programación.

Las bibliotecas de Python posibilitan recolectar información desde internet, lo que facilitó consultar, analizar y compilar el corpus electrónico directamente de la página web de SciELO (Scientific Electronic Library Online). Esta biblioteca virtual propicia la publicación y difusión electrónica de revistas científicas bajo un esquema de acceso abierto, mediante el “sistema Creative Commons (CC) que promueve la reutilización y distribución de los artículos indexados en sus colecciones” (SciELO, 2021). Igualmente, por medio de Python se logra descargar de modo eficiente los artículos en un archivo de texto plano (.txt) para su análisis en paralelo con LanCSBox (Brezina *et al.*, 2020). Esta herramienta posibilita la comparación de concordancias y la visualización del contexto en que se da cada una de ellas, además de que constituye un recurso alternativo para la validación de las metáforas conceptuales y, como paso adicional, para la observación de su distribución en las publicaciones.

3.2. *Procesamiento del corpus*

Para la recopilación de los textos electrónicos se realiza un procedimiento usualmente llamado *web scraping*,¹ esto es, la extracción de información de sitios web. Como mencionamos previamente, en este caso la recolección del contenido se realizó desde la pá-

¹ El procedimiento y las indicaciones para efectuar el *web scraping* han sido desarrollados a partir de las herramientas propuestas por Python Software Foundation (PyPI, 2020) y pueden descargarse libremente en el sitio: <https://pypi.org/project/beautifulsoup4/>

gina de SciELO (science.org), que contiene las revistas de acceso abierto que integran el objeto de estudio. La recolección se facilitó debido a que los datos de todos los textos se encuentran codificados homogéneamente en lenguaje HTML. Lo antes señalado ofrece la posibilidad de que los textos sean almacenados y analizados de forma directa en Python, además de que puedan descargarse en un texto plano.

En la Figura 1 se muestra la función que permite realizar el web *scraping*, y a partir de la cual solo se obtienen partes estratégicas del artículo, tales como: título, resumen y cuerpo. Esto con el fin de generar un menor “ruido” y una mayor rapidez en el procesamiento, puesto que en estos artículos suele hallarse otro tipo de contenido que no es necesario para el análisis textual, como pueden ser las referencias bibliográficas, el resumen en otros idiomas, etcétera.

```

1 def extract_1(q):
2     parrafos = list()
3     for p in q:
4         parrafos.append(p.text)
5     return parrafos
6
7 def get_texto_url(webs):
8     a = list()
9     for url in webs:
10        try:
11            web = requests.get(url)
12            soup = BeautifulSoup(web.text, 'html.parser')
13            titulo = soup.find_all('p', class_="title")
14            resumen = [soup.select_one('.trans-abstract')]
15            articulo = soup.find_all('div', 'body')
16
17            a1 = extract_1(titulo)
18            a2 = extract_1(resumen)
19            a3 = extract_1(articulo)
20
21            a.append(a1+a2+a3)
22        except:
23            print("Error en la url {}, response {}".format(url,webs))
24
25        # print("Se han añadido un total de {} caracteres a la variable texto".format(Long))
26    return a

```

FIGURA 1. Función de web *scraping* para la extracción de artículos

Una vez recopilados los 42 artículos y exportados en formato texto plano (*.txt), se llevaron a LanCSBox para ser examinados de

manera detallada y analizar elementos importantes relativos a la cantidad de *tokens* (palabras), *types* (formas únicas) y lemas presentes en cada subcorpus; tal información puede corroborarse en el Cuadro 2.² En este se aprecia que el porcentaje de *tokens* en ambas revistas es bastante homogéneo, lo que posibilita obtener resultados más amplios y equilibrados.

CUADRO 2. Resumen de *tokens*, *types* y *lemas* a partir de LancsBox 5.1.2.

| Corpus | Cantidad de <i>tokens</i> | Porcentaje de <i>tokens</i> respecto al total | Cantidad de <i>types</i> | Cantidad de <i>lemas</i> |
|--------|---------------------------|---|--------------------------|--------------------------|
| CDE | 153 844 | 48.2% | 13 373 | 10 310 |
| RDEI | 165 261 | 51.8% | 16 048 | 12 380 |
| Total | 319 105 | 100% | 23 132 | 17 770 |

La gran cantidad de *tokens* encontrados hace necesario un procedimiento que facilite la selección de los posibles candidatos de metáforas conceptuales. Siguiendo esta línea, en cualquier modelo de estudio de procesamiento de lenguaje natural resulta indispensable delimitar la proporción de datos mediante la eliminación de elementos poco relevantes para la investigación; esto se realiza a partir de un tratamiento que se efectúa sobre todos los documentos para que queden, en lo posible, lo más depurados y homogéneos, lo que favorece su análisis posterior. Este tratamiento comienza con una limpieza para excluir los signos de puntuación y otras expresiones que generen algún tipo de “ruido”, llamadas *stopwords*,³ que corresponden a palabras tales como artículos, preposiciones, pronombres, entre otros. Este conjunto de vocablos

² Dicho análisis se elaboró de manera paralela con Python, el cual arrojó datos bastante similares y, en algunos casos, iguales.

³ De acuerdo con la comunidad de desarrollo de soluciones para el procesamiento del lenguaje natural GeeksforGeeks (2020), las *stopwords*, también llamadas palabras vacías, son aquellas palabras sin significado relevante dentro del contexto de cada idioma. Estas son excluidas del corpus, permitiendo el posterior análisis de las palabras de contenido de los diferentes textos mediante el procesamiento de lenguaje natural.

Las etapas antes mencionadas consisten en: a) normalización, se convierten todas las palabras en mayúsculas o minúsculas; b) lematización, una palabra flexionada o derivada se cambia por su lema; c) *stemming*, se relacionan los afijos de las palabras para obtener la raíz y, por último, d) tokenización, se divide el texto en oraciones y a su vez en palabras. Para que este proceso completo sea óptimo, cómodo y fácil de ejecutar en la siguiente fase (la creación de los *n*-gramas), se procede a la elaboración de una función o algoritmo de procesamiento del lenguaje correspondiente al conjunto de las etapas.

La función quedó determinada como Ptex (Procesamiento de textos), a la cual se le dan simples instrucciones para que ejecute la construcción de unigramas, bigramas o trigramas, según lo que se desee obtener en la salida. Es importante señalar que un *n*-grama es un conjunto de *n* elementos ($n = 1, 2, 3, \dots, n$) que se encuentran ligados de manera consecutiva en un texto. Estos modelos son bastante utilizados en aplicaciones de análisis textual. Para su construcción, en la Figura 4 mostramos varios parámetros llamados con nombres seleccionados arbitrariamente, algunos de ellos se activan con *True* y se desactivan con *False* o *None*, y otros pueden activarse a partir de palabras o de números según sea el caso.

- “lema”: lematiza el texto.
- “regex”: aplica alguna expresión regular según corresponda y según se requiera.
- “check”: corrige la ortografía si es el caso y elimina el idioma deseado.
- “stop_w”: activa el conjunto de palabras que generan ruido en el análisis.
- “ListSW”: agrega alguna palabra que no esté dentro del conjunto de *stopwords* predefinido.
- “removeSW”: elimina una palabra del conjunto de *stopwords* para que sea analizada.
- “min_lett”: número mínimo de letras de una palabra para que pueda ser analizada.

- “max_lett”: número máximo de letras de una palabra para que pueda ser analizada.
- “min_freq”: número mínimo de frecuencia en el unigrama.
- “min2_freq”: número mínimo de frecuencia en el bigrama.
- “min3_freq”: número mínimo de frecuencia en el trigram.
- “n_gram”: crea el unigrama, bigrama, trigram, donde $n = 1, 2$ o 3 , correspondientemente.
- “ran = [0 - n]”: crea un número de salidas o barras (conjunto de palabras) para el gráfico, donde $n = 1, 2 \dots 50$.

```

1 txt = PText(Texto, lemma=None, regex = r'[A-Za-z]+', check=None,
2   stop_w=True, listSW= ['tabla', 'ntabla'], removeSW= ['tiempo'],
3   min_lett=4, max_lett=None, min_freq=None,
4   ngram=1, min2_count=2, min3_count=2)
5
6 txt.PltFreqWord(n_gram=1, ran = [0,20])

```

FIGURA 4. Parámetros para la creación de unigramas, bigramas y trigramas

4. Análisis de resultados

El reconocimiento de los dominios cognitivos de la metáfora no siempre es tarea fácil para el hablante; esto se debe a que el desarrollo de categorización en algunos casos puede ser, según la lingüística cognitiva, un tanto subjetivo, ya que suele basarse en experiencias culturales de cada individuo, en las cuales las proyecciones abstractas se interpretan mediante una proyección concreta. El tratamiento aplicado en el presente estudio para la identificación de metáforas obedeció un criterio de mayor frecuencia en los subcorpus, siguiendo una graficación en orden descendente, de entre 20 y 40 barras, según se trate de unigramas, bigramas o trigramas. De esta manera, al ejecutar el análisis para cada revista de forma independiente, se seleccionaron únicamente las expresiones que tenían una mayor frecuencia en el grupo de n -gramas, ya que estas representaban la variación del conjunto de posibles metáforas conceptuales por medio de diagramas de barras, en los cuales el eje “X” corresponde a la categoría comparada y el eje “Y” a la frecuencia de cada forma.

Antes de la depuración de los textos se contaba con un total de 153 844 *tokens* para el caso de CDE y 165 261 para el de RDEI. No obstante, en este punto aún se incluían las *stopwords*, que serían eliminadas posteriormente. Después de la limpieza de datos se obtuvo un total de 6 013 palabras únicas para CDI y 7 320 para RDEI, seleccionadas a partir de una frecuencia mayor a dos, esto quiere decir que cada palabra debía estar presente por lo menos dos veces en cada subcorpus. La suma total de los vocablos para CDE correspondió a 71 049 y 75 895 para RDEI; estos valores pertenecen a los *tokens* con los que se trabajará para el primer *n*-grama correspondiente.

Los *n*-gramas se realizaron para los léxicos en su estado original, es decir, las palabras tal cual aparecen en el texto. No se efectuó ningún proceso de normalización, lematización o *stemming* de las palabras, dado que la modificación de estas no tuvo mucha incidencia en los agrupamientos y se consideró pertinente estudiar los posibles dominios de las metáforas conceptuales sin la transformación de los vocablos. En las Figuras 5 y 6, aparecen los primeros 20 términos lexicales con mayor frecuencia y su valor en orden descendente.⁴

Conforme a los unigramas presentados, se observa que hay palabras presentes en ambos subcorpus, como es el caso de *economía*, *trabajo*, *nivel*, *mercado*, entre otras; estos vocablos podrían expresar alguna metáfora conceptual. Sin embargo, como no se tiene una contextualización más definida, no se puede afirmar con exactitud que corresponden a expresiones metafóricas ligadas a estas palabras, ya que se necesita de una construcción más compleja para entender cabalmente su sentido.

Una vez obtenida la lista de frecuencias se procedió a elaborar un análisis estadístico que permitiera observar los porcentajes de los datos y verificar si la muestra era representativa. Para que

⁴ Los *n*-gramas que se muestran son resultado de todo el proceso descrito anteriormente. No han sido filtrados de manera manual; este es un procedimiento automático que se logra mediante la función ya mencionada.

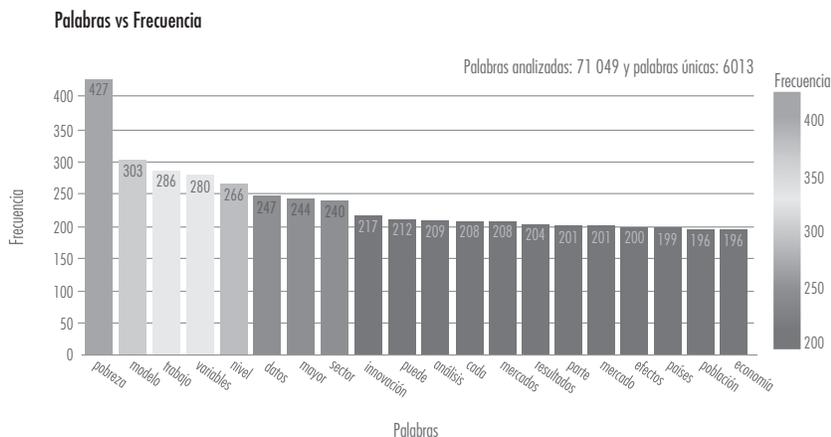


FIGURA 5. Unigrama CDE

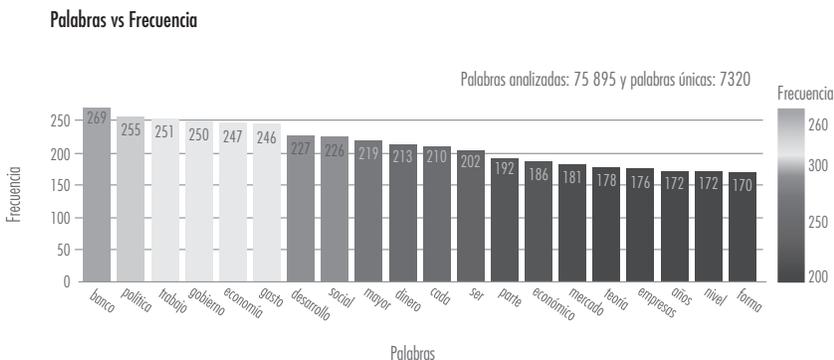


FIGURA 6. Unigrama RDEI

el estudio reflejara con mayor precisión al corpus se consideró la cantidad total de texto seleccionado. El Cuadro 3 muestra que en el corpus de CDE se hallaron 205 palabras diferentes con una frecuencia mayor a 60, de las cuales el promedio de aparición correspondió a 110 veces; su mediana fue de 92, y el vocablo que más se repitió fue *pobreza* (427 ocurrencias). En el caso de la revista RDEI se obtuvieron resultados similares, se hallaron 200 unidades

léxicas distintas que aparecían más de 60 veces; la frecuencia promedio fue de 107.2 y la mediana de 93; la palabra *banco* fue la más utilizada (269 ocurrencias). De acuerdo con estos resultados estadísticos se comprueba que ambas revistas presentan un subcorpus representativo y que muestran gran similitud, lo cual indica que el corpus es homogéneo y se puede hacer un correcto análisis.

CUADRO 3. Descriptiva del corpus elaborado a partir de Python

| | CDE | | RDEI |
|-------|------------|-------|------------|
| freq | | freq | |
| count | 205.000000 | count | 200.000000 |
| mean | 110.580488 | mean | 107.290000 |
| std | 54.907316 | std | 46269865 |
| min | 61.000000 | min | 61.000000 |
| 25% | 72.000000 | 25% | 72.000000 |
| 50% | 92.000000 | 50% | 93.000000 |
| 75% | 134.000000 | 75% | 127.250000 |
| max | 427.000000 | max | 269.000000 |

Con el fin de obtener las posibles metáforas conceptuales se analizó los resultados tanto de los bigramas como de los trigramas. Es importante aclarar que, al igual que en la realización de los unigramas, para estas deducciones se excluyeron los resultados que tienen solamente una ocurrencia y se graficaron las categorías con un mayor nivel de frecuencia.

Para la realización de los bigramas correspondientes a CDE y RDEI, se obtuvieron las primeras 40 parejas de palabras con su respectivo índice de ocurrencia, como se muestra en las Figuras 8 y 9. En el caso de CDE se analizó un total de 22 886 expresiones, de las cuales 3269 son únicas, mientras que en el de RDEI se examinaron 17 977 unidades léxicas, 3187 únicas. En ambos gráficos se visualiza la formación de algunos bigramas con conceptos muy

generales, no obstante, también se observa la presencia de posibles candidatos a metáforas conceptuales con frecuencia alta.

Palabras vs Frecuencia

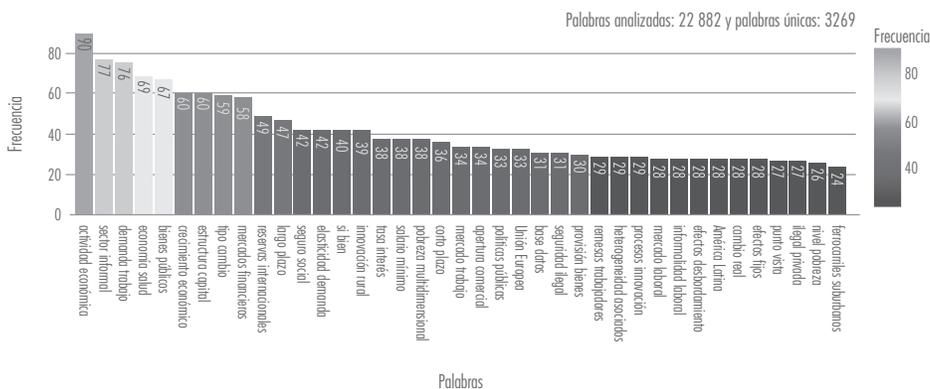


FIGURA 7. Bigramas para cde

En el subcorpus de CDE (Figura 7), se presentan posibles metáforas como *crecimiento económico*, cuyo dominio pertenece al de la economía como un ser vivo o parte del ciclo de vida humano; también se observa la pareja *estabilidad financiera*, que se encuentra en el ámbito bien sea de movimiento o situación, tal como sucede con *actividad económica*, *nivel pobreza*⁵ (nivel de pobreza) y *precios activos*. Además, *corto plazo* y *largo plazo*, se podrían clasificar dentro del dominio de tiempo como un objeto físico medible. Igualmente, podría añadirse la ocurrencia de *efectos desbor-*

⁵ El bigrama *nivel pobreza* corresponde a nivel de pobreza, puesto que, como se mencionó previamente, las *stopwords* eliminan cierto tipo de palabras (artículos, preposiciones, pronombres, etc.), ya que no son significativas para el procesamiento del lenguaje natural. Esto permite que los *n*-gramas se centren en las palabras relevantes para el análisis, como sucede en este caso y en otros subsiguientes, de los que se ofrecerá su posible equivalente entre paréntesis.

damiento (efectos de desbordamiento), que tiene como referente al mercado evocado como un líquido o una catástrofe natural.

En cuanto a los artículos de RDEI (Figura 8), estos revelan potenciales metáforas como *pobreza multidimensional*, aludiendo a la pobreza como un fenómeno físico medible en varios espacios; por otro lado, está *nivel educativo* que, al igual que *nivel pobreza* (nivel de pobreza), se encuentra ubicada dentro del dominio de movimiento o situación; además se detecta *moneda soberana*, que se clasifica dentro del dominio de la economía como persona. Finalmente, también se incluyen las metáforas: *largo plazo* y *crecimiento económico*, declaradas y explicadas con anterioridad.

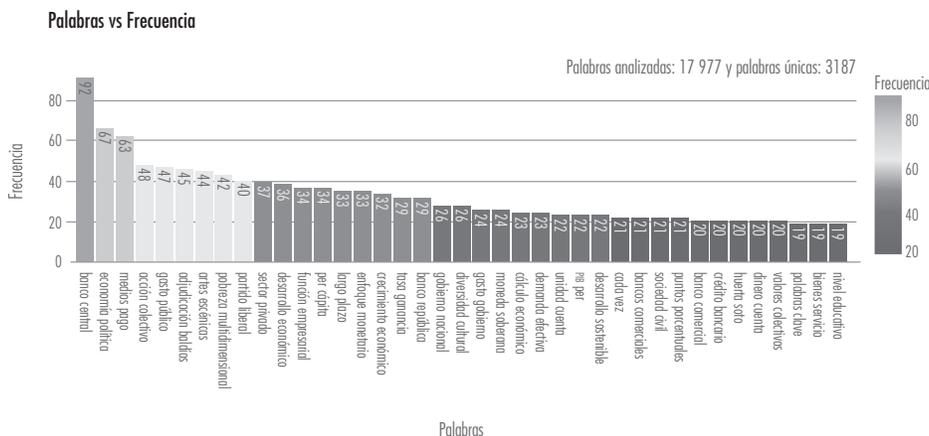


FIGURA 8. Bigramas para RDEI

Según la construcción de los trigramas, se advierte que las palabras analizadas disminuyen para una y otra revista, contando con 7073 y 3550 expresiones para CDE y RDEI, respectivamente. En consecuencia, se reducen los vocablos únicos: 1896 para CDE y 1463 para RDEI; debido a esto, existe una menor cantidad de agrupamientos y por lo tanto las frecuencias son menores. Asimismo, existen conjuntos que, como en el caso de los bigramas, pueden ser

una persona con cualidades y, por último, la expresión *medir pobreza multidimensional* (medir la pobreza multidimensional), en la cual la pobreza conlleva un elemento físico medible, tal como en el análisis de su bigrama, pero de manera mucho más clara, ya que en este caso se cuenta con un mayor contexto.

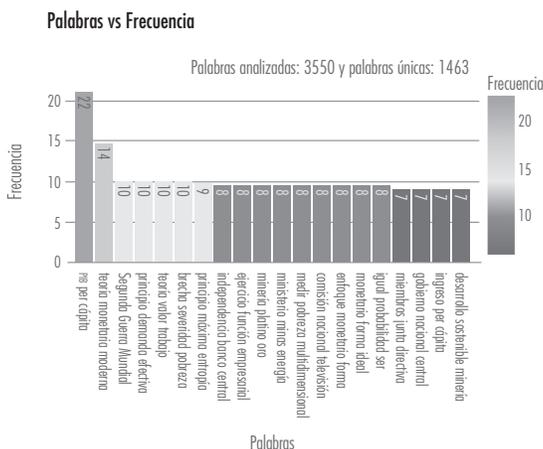


FIGURA 10. Trigramas RDEI

De acuerdo con las palabras analizadas de los n -gramas desarrollados previamente, se observa que efectivamente ambas revistas incurren en la utilización de metáforas conceptuales. En el Cuadro 4 se identifica a grandes rasgos que CDE, a pesar de tener inicialmente un menor número de vocablos que RDEI, tiene mayor incidencia en la utilización de metáforas conceptuales. En las dos revistas se sugiere que entre más grande sea el n -grama (en este caso el trigramas), disminuye el número de las palabras únicas y por ende las variantes a analizar. Sin embargo, la proporción entre palabras analizadas y únicas tiende a aumentar —como se especifica en la variable del porcentaje— debido a que es más difícil encontrar un conjunto de dos o más lexías que se repita constantemente. En este sentido, se identifica que, en la ejecución de los bigramas y trigramas, tanto las palabras analizadas como las palabras únicas dismi-

nuyen más rápidamente para RDEI, por lo que se advierte que estas publicaciones no frecuentan las mismas expresiones, sino que, por el contrario, utilizan más diversidad en el lenguaje.

CUADRO 4. Porcentaje de acuerdo con el *n*-grama

| <i>n</i> -gramas | Palabras analizadas | | Palabras únicas | | Porcentaje | |
|------------------|---------------------|--------|-----------------|-------|------------|-------|
| | CDE | RDEI | CDE | RDEI | CDE | RDEI |
| Unigramas | 71 049 | 75 895 | 6 013 | 7 320 | 8.5% | 9.6% |
| Bigramas | 22 886 | 17 977 | 3 269 | 3 187 | 14.3% | 17.7% |
| Trigramas | 7 073 | 3 550 | 1 896 | 1 463 | 26.8% | 41.2% |

Conforme a lo anterior y a lo indicado en los gráficos de barras, es posible identificar que las expresiones metafóricas aparecen con mayor facilidad en los bigramas y los trigramas, es decir, que las palabras candidatas a ser metáforas conceptuales predominan en lexías complejas, mientras que los resultados de los unigramas no ofrecen información significativa, ya que se necesita de un contexto más específico para declarar que efectivamente se trata de una construcción metafórica.

Es necesario aclarar que esta metodología de procesamiento de lenguaje natural con Python debe ser supervisada, puesto que se trata de un lenguaje de programación que ayuda a la comprensión de una gran cantidad de datos. Debido a esto, se requiere examinar con detenimiento sus resultados mediante un análisis semi-automático, empleando algunas librerías de Python y otras herramientas especializadas en el análisis de corpus textuales.

Así, se recurrió a un determinado módulo de la librería de Python Scikit Learn (2022) para realizar un análisis TF-IDF,⁶ el cual asoció cada palabra (o expresión metafórica) con un número

⁶ TF-IDF es una medida estadística cuyo valor corresponde al peso de dos productos. La ecuación utilizada en Python se documenta en https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

que representaba qué tan relevante era según la totalidad de la revista estudiada; este número está comprendido entre 0 y 1, donde el más cercano a 1 es más significativo. Igualmente, se recurrió a LncsBox (Brezina *et al.*, 2020), para evaluar si una determinada expresión metafórica estaba repartida de manera homogénea en todos los documentos de cada uno de los subcorpus (se ofrece un ejemplo en la Figura 11), y bajo qué contexto se encontraban las metáforas más recurrentes (Figura 12).

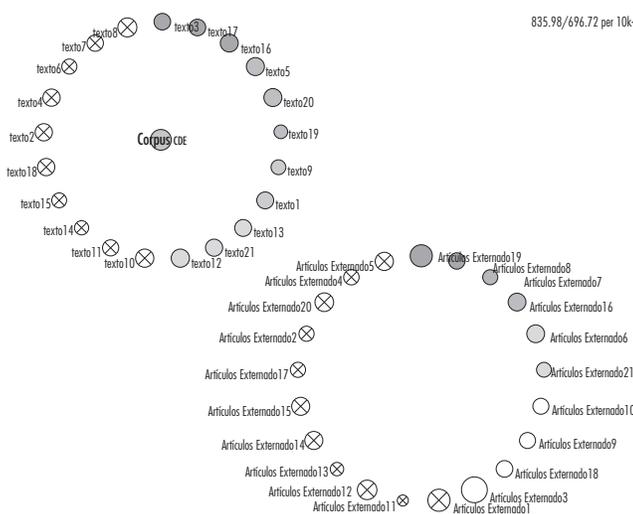


FIGURA 11. Dispersión de la expresión *crecimiento económico* en los corpus CDE y RDEI

Según el ejemplo mostrado en la Figura 11, el bigrama correspondiente a la expresión metafórica *crecimiento económico* aparece en 11 de los 21 textos del corpus CDE de manera no regular, además, puede observarse que tiene una importancia bastante grande: su TF-IDF corresponde al 0.7055. Este mismo bigrama apareció en 10 de los 21 documentos del corpus RDEI, en las mismas condiciones que en el caso CDE, y con una relevancia un poco mayor, 0.7057. Al realizarse búsquedas similares se obtuvieron los resultados mostrados en el Cuadro 5.

CUADRO 5. Análisis de expresiones metafóricas comparadas según la cantidad de documentos en que aparecen

| Expresión metafórica | Subcorpus | n-grama | Número de documentos | Análisis TF-IDF |
|--|-----------|----------|----------------------|-----------------|
| crecimiento económico | CDE | bigrama | 11 | 0.705520 |
| largo plazo | CDE | bigrama | 9 | 0.702562 |
| actividad económica | CDE | bigrama | 8 | 0.699221 |
| corto plazo | CDE | bigrama | 8 | 0.666605 |
| nivel [de] pobreza | CDE | bigrama | 3 | 0.688763 |
| efectos [de] desbordamiento | CDE | bigrama | 1 | 0.564494 |
| países [en] vías [de] desarrollo | CDE | trigrama | 5 | 0.372105 |
| elasticidad [de la] demanda [de] trabajo | CDE | trigrama | 1 | 0.666605 |
| estructura [de] capital [de] corto plazo | CDE | trigrama | 1 | 0.467359 |
| delitos [contra el] patrimonio económico | CDE | trigrama | 1 | 0.496299 |
| largo plazo | RDEI | bigrama | 13 | 0.705976 |
| crecimiento económico | RDEI | bigrama | 10 | 0.705714 |
| nivel educativo | RDEI | bigrama | 4 | 0.587469 |
| pobreza multidimensional | RDEI | bigrama | 2 | 0.635795 |
| nivel [de] pobreza | RDEI | bigrama | 1 | 0.707057 |
| moneda soberana | RDEI | bigrama | 1 | 0.593045 |
| independencia [del] banco central | RDEI | trigrama | 4 | 0.503216 |
| medir [la] pobreza multidimensional | RDEI | trigrama | 1 | 0.456894 |

Finalmente, se identificó en qué contexto aparecen las metáforas conceptuales más comunes o más frecuentes también mediante la herramienta LancsBox (Brezina *et al.*, 2020), para comprobar que efectivamente se trataba de una expresión metafórica. En la Figura 12 puede apreciarse el análisis de dos candidatos a expresiones metafóricas y sus respectivos contextos textuales; se seleccionó *crecimiento económico* del corpus CDE y *pobreza multidimensional* del RDEI. Gracias a LancsBox es posible analizar las palabras en un contexto oracional y observar el contexto derecho e izquierdo de ambas expresiones.

The screenshot displays the LancsBox interface. At the top, there's a search bar with 'crecimiento económico' entered. Below it, a list of search results is shown, including document titles and frequencies. The main part of the image shows a detailed view of a specific context, where a text snippet is displayed with a highlighted phrase. To the right of this snippet, a conceptual mapping is shown, indicating the source and target domains of the metaphor.

FIGURA 12. Análisis mediante LancsBox de los contextos oracionales de *crecimiento económico* y *pobreza multidimensional*

5. Conclusiones

De acuerdo con los resultados obtenidos, se confirma la hipótesis planteada sobre el uso de la metáfora conceptual en los discursos académicos de índole económico. Dichas construcciones son fenómenos abundantes en este campo, puesto que la metáfora brinda una cierta transparencia que permite entender conceptos abstractos y específicos, lo que posibilita efectuar una particular categorización al dar sentido a diversas situaciones. Es por esto que la relación asociativa que da lugar a las expresiones metafóricas suele ser muy semejante, ya que los individuos tienden a desarrollar una cognición más o menos alineada, la cual permite entender el mundo de cierta manera, sin que se convierta en un factor meramente aleatorio.

El estudio exhibe gran uniformidad en el uso de las expresiones metafóricas más comunes en este discurso académico. En algunos casos se encontraron las mismas metáforas en ambas revistas y a lo largo de sus diferentes publicaciones, las cuales fue-

ron escritas de manera paralela e independiente en el mismo año. Por ejemplo, *largo plazo*, expresión que aparece distribuida en 22 publicaciones en ambas revistas, con un total de 80 apariciones, o *crecimiento económico*, que se encuentra en 21 artículos, con una ocurrencia de 92 veces. En cuanto al análisis de la relevancia de estas dos expresiones (TF-IDF), en promedio es de 0.704943, lo cual significa que son bastante importantes en los textos analizados. De esta manera se comprueba que entre más ocurrencias tenga una expresión, mayor será su relevancia, estableciéndose una relación directamente proporcional.

Aunque ambas revistas puedan tener diferencias a nivel de escritura y tipos de artículos publicados, es interesante anotar que las metáforas predominantes son aquellas relacionadas con la pobreza, lo que refiere a la crisis económica que se desató en Colombia debido a la pandemia. Las metáforas asociadas a la pobreza no son expresiones conceptuales nuevas, por el contrario, son estructuras recurrentes en este tipo de discursos económicos. Este resultado posibilita examinar el uso general de dichas expresiones y comprobar la eficacia de las metáforas conceptuales al establecer dominios de interpretación comunes y uniformes dentro de un espacio; en este caso, el ámbito económico colombiano y los temas de actualidad.

En cuanto a la metodología, vale la pena resaltar la utilidad de las herramientas empleadas, ya que procesan grandes volúmenes de datos en poco tiempo, lo cual resulta muy beneficioso para el estudio de corpus y para cumplir a cabalidad el objetivo principal del trabajo. Lo anterior demuestra la clara relación entre la lingüística y el procesamiento de lenguaje natural en la realización de diversos estudios.

Igualmente, la realización de los *n*-gramas evidencia que las metáforas conceptuales predominan en construcciones complejas; no se pueden recabar expresiones metafóricas en los unigramas, ya que se requiere de otras unidades léxicas para que pueda evidenciarse el proceso analógico y reflejar adecuadamente las formas de entender el mundo involucradas. Adicionalmente, la extracción

realizada con Python de las posibles metáforas generadas a partir de los n -gramas, al tratarse de un método semi-automático, no asegura que los resultados estén excluidos de ruido u otros problemas a resolver, ya que la interpretación de la metáfora conceptual es compleja e involucra relaciones semánticas y conceptuales de alto poder significativo, por lo cual se deben atender estas cuestiones de forma “manual”, con la ayuda de herramientas como LancsBox u otras similares.

Finalmente, en todo proceso de identificación de metáforas conceptuales es imprescindible el cotejo de los resultados obtenidos, esto con el fin de comprobar que los resultados se extienden de manera homogénea con respecto a la muestra estudiada. Por lo tanto, es de suma importancia que se pueda contar no solamente con programas computacionales como Python, sino también con otras herramientas, como las señaladas en esta investigación, para depurar de manera ideal los resultados y no correr el riesgo de caer en generalizaciones.

6. Referencias

- Alarcón, Paola; Díaz, Claudio; Vergara, Jorge; Vásquez, Victor, & Torres, Claudio (2018). Análisis de metáforas conceptuales sobre la imagen social del profesorado en estudiantes de pedagogía. *Onomázein*, 40, 1–27. doi: 10.7764/onomazein.40.01
- Brezina, Vaclav; Weill-Tessier, Pierre, & McEnery, Antony (2020). #LancsBox v. 5.x. [software]. <http://corpora.lancs.ac.uk/lancsbox>
- Boers, Frank (2000). Enhancing metaphoric awareness in specialised reading. *English for Specific Purposes*, 19(2), 137–147.
- Charteris-Black, Jonathan (2000). Metaphor and vocabulary teaching in ESP economics. *English for Specific Purposes*, 19(2), 149–165.
- Charteris-Black, Jonathan, & Ennis, Timothy (2001). A comparative study of metaphor in Spanish and English financial reporting. *English for Specific Purposes*, 20(3), 249–266.
- Charteris-Black, Jonathan, & Musolff, Andreas (2003). ‘Battered hero’ or ‘innocent victim’? A comparative study of metaphors for euro trading in British

- and German financial reporting. *English for Specific Purposes*, 22(2), 153–176.
- Gallego-Iniesta, Carmen Mateo (2022). *Metáfora, terminología y traducción: informes institucionales sobre la crisis económica en inglés, español y alemán*. Berlín: Frank & Timme.
- GeeksforgGeeks (2020). Removing stopwords with NLTK in Python. <https://www.geeksforggeeks.org/removing-stop-words-nltk-python/>
- Goergebeur, Lieselot (2016). *Lenguaje metafórico en la prensa económica y política sobre el 'brexit': estudio comparativo español-neerlandés* (Tesis de maestría inédita). Universidad de Gante, Gante.
- Gotti, Maurizio (2008). Defining the notion of 'specialized discourse'. En *Investigating specialized discourse* (pp. 17–22). Berna: Peter Lang.
- Henderson, Willie (1982). Metaphor in economics. *Economics*, 18, 147–1.
- Kövecses, Zoltán (2020). *Extended conceptual metaphor theory*. Cambridge: Cambridge University Press.
- Lakoff, George, & Johnson, Mark (2001). *Metáforas de la vida cotidiana*. Madrid: Cátedra.
- Langer, Tomáš (2015). Metaphors in economics: Conceptual mapping possibilities in the lectures of economics. *Procedia Economics and Finance*, 25, 308–317.
- Loma-Osorio Fontecha, Marciana (2011). La metáfora como recurso de producción léxica en la metalengua de la economía en español e inglés. *Lorenzo Hervás*, 20(extraordinario), 311–325.
- McCloskey, Donald (1983). The rhetoric of economics. *Journal of Economic Literature*, 21(2), 481–517.
- Morato, Jorge; Palacios, Vicente; Urbano, Julián; Sánchez-Cuadrado, Sonia, & Marrero, Mónica (2020). Módulo VIII. Técnicas de Procesamiento de Lenguaje Natural (PLN). OpenCourseWare.
- Muelas, María (2016). A cross-linguistic study of conceptual metaphors in financial discourse. En Jesús Romero-Trillo (Ed.), *Yearbook of corpus linguistics and pragmatics* (pp. 107–126). Cham: Springer.
- Muelas, María (2018). *Metaphor, economy and persuasion: A socio-cognitive approach to metaphor in times of elections* (Tesis de doctorado inédita). Universidad Autónoma de Madrid, Madrid.

- Muelas, María (2019a). Ideología, metáfora y persuasión en tiempos de elecciones: un estudio de corpus de noticias económicas en prensa británica y española. *Estudios Ingleses de la Universidad Complutense*, 27, 223–247.
- Muelas, María (2019b). Visual metaphors in economic discourse. An analysis of the interaction of conventional and novel visual metaphors in *The Economist*. *Current Approaches to Metaphor Analysis in Discourse*, 39, 347–366.
- PyPI (2020). Beautifulsoup4. <https://pypi.org/project/beautifulsoup4/>
- Rivera-León, Lorena (2017). Al rescate de la economía o cómo hablar con metáforas de asuntos serios. *Foro de Profesores de E/LE*, 13, 233–246.
- Scientific Electronic Library Online (2021). Declaración de acceso abierto. <https://scielo.org/es/sobre-el-scielo/declaracion-de-acceso-abierto/>
- Scikit Learn (2022). Sklearn.feature_extraction.text.TfidfVectorizer. https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html
- Soriano, Cristina (2012). La metáfora conceptual. En Iraide Ibarretxe-Antuñano & Javier Valenzuela (Coords.), *Lingüística cognitiva* (pp. 98–121). Madrid: Anthropos.
- Skorczynska, Hanna, & Deignan, Alice (2006). Readership and purpose in the choice of economics metaphors. *Metaphor and Symbol*, 21(2), 87–104.

7. Corpus de estudio

- Universidad Externado de Colombia (2020). *Revista de Economía Institucional*, 22(42), 3–323. http://www.scielo.org.co/scielo.php?script=sci_isuetoc&pid=0124-599620200001&lng=en&nrm=iso
- Universidad Externado de Colombia (2020). *Revista de Economía Institucional*, 22(43), 3–319. http://www.scielo.org.co/scielo.php?script=sci_isuetoc&pid=0124-599620200002&lng=en&nrm=iso
- Universidad Nacional de Colombia, Facultad de Ciencias Económicas (2020). *Cuadernos de Economía*, 39(79), 1–395. doi: 10.15446/cuad.econ
- Universidad Nacional de Colombia, Facultad de Ciencias Económicas (2020). *Cuadernos de Economía*, 39(81), 595–1033. doi: 10.15446/cuad.econ

