



Facultad de Ciencias Económicas y Empresariales

Tendencias de Investigación sobre la Transición Energética en España

Autor: Nicola Masotti Villanueva

Director: Jenny Alexandra Cifuentes Quintero

MADRID | Abril 2026

Resumen

El auge de la investigación sobre transición energética ha generado un volumen creciente de publicaciones académicas que dificulta identificar sus principales ejes temáticos y anticipar su evolución futura. Este trabajo examina la producción científica sobre transición energética en España mediante una metodología que combina minería de textos, modelado de tópicos, análisis de series temporales y técnicas bibliométricas. Para ello, se compiló un corpus de 1.262 artículos publicados entre 2000 y 2025 en Scopus. Tras un proceso de preprocesamiento textual, disminuyeron a 1040 y se aplicó el algoritmo LDA para identificar las estructuras temáticas latentes del corpus, con el número óptimo de tópicos determinado mediante el índice de coherencia Cv. Posteriormente, se ajustaron modelos ARIMA para proyectar la evolución temporal de cada tópico hasta 2030, previa verificación de la estacionariedad mediante el test de Dickey-Fuller Aumentado aplicado de forma iterativa. Finalmente, se realizó un análisis bibliométrico con VOSviewer centrado en redes de coocurrencia de palabras clave, acoplamiento bibliográfico, co-citación y coautoría internacional.

Los resultados revelan tres grandes líneas temáticas: el análisis de ciclo de vida, la huella de carbono y la descarbonización sectorial (39,2 % del corpus); la gobernanza energética europea, el análisis macroeconómico y la política climática (38,8 %); y el mercado eléctrico, la inversión renovable y los marcos regulatorios (22,0 %). Las proyecciones ARIMA apuntan a un crecimiento sostenido de la producción académica hasta 2030 para los Tópicos 0 y 2, cuyas series alcanzaron la estacionariedad mediante diferenciación. La proyección del Tópico 1 tiene carácter exploratorio, al no haber alcanzado su serie la estacionariedad requerida. Los mapas de VOSviewer confirman la coherencia de estos resultados y muestran que España ocupa el nodo central de la red de coautoría internacional, con una dimensión europea consolidada y una notable proyección iberoamericana.

Palabras clave: transición energética, LDA, ARIMA, España, descarbonización.

Abstract

The growth of research on energy transition has generated an increasing volume of academic publications, making it difficult to identify its main thematic axes and anticipate its future evolution. This study examines the scientific output on energy transition in Spain through a methodology combining text mining, topic modelling, time series analysis and bibliometric techniques. A corpus of 1,262 articles published between 2000 and 2025 in Scopus was compiled. Following a textual preprocessing stage, the corpus was reduced to 1040 and the LDA algorithm was applied to identify the latent thematic structures of the corpus, with the optimal number of topics determined through the Cv coherence index. Individual ARIMA models were then fitted to project the temporal evolution of each topic up to 2030, after verifying stationarity through an iterative application of the Augmented Dickey-Fuller test. Finally, a bibliometric analysis was conducted with VOSviewer, focusing on keyword co-occurrence networks, bibliographic coupling, co-citation and international co-authorship.

The results reveal three major thematic lines: life cycle assessment, carbon footprint and sectoral decarbonisation (39.2 % of the corpus); European energy governance, macroeconomic analysis and climate policy (38.8 %); and electricity markets, renewable investment and regulatory frameworks (22.0 %). ARIMA projections point to sustained growth in academic output up to 2030 for Topics 0 and 2, whose series achieved stationarity through differencing. The projection for Topic 1 is exploratory in nature, as its series did not reach the required stationarity. The VOSviewer maps confirm the coherence of these findings and show that Spain occupies the central node of the international co-authorship network, with a consolidated European dimension and a notable Ibero-American projection.

Keywords: energy transition, LDA, ARIMA, Spain, decarbonisation.

Índice general

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos del Trabajo de Grado	5
1.2.1. Objetivo General	5
1.2.2. Objetivos Específicos	5
1.3. Organización de la memoria	6
2. Técnicas bibliométricas para el análisis de las tendencias emergentes en la transición energética española: Revisión de la literatura	7
3. Metodología	13
3.1. Recopilación y exploración de los datos	14
3.2. Procesamiento y preparación del texto	15
3.3. Análisis exploratorio de N-gramas	17
3.4. Modelado de tópicos (LDA)	18
3.5. Predicción de tendencias (ARIMA)	20
3.6. Análisis bibliométrico con VOSviewer	21
4. Resultados del Análisis Temático, Bibliométrico y de las Proyecciones	23
4.1. Recopilación y exploración de los datos	23
4.2. Procesamiento y preparación del texto	24
4.3. Análisis exploratorio	26
4.4. Modelado de tópicos (LDA)	28
4.5. Análisis de tendencias temporales con ARIMA	33
4.6. Análisis bibliométrico con VOSviewer	35
5. Conclusiones y trabajo futuro	40
Bibliografía	45

Índice de figuras

1.1. Anomalía de la temperatura media global. Fuente de datos: Our World in Data (2024). Elaboración propia	2
1.2. Evolución de las emisiones globales de CO2. Fuente de datos: (Epdata, 2023). Elaboración propia	3
1.3. Generación moderna de electricidad renovable por fuente. Fuente de datos: Our World in Data (s.f.). Elaboración propia	3
3.1. Esquema de la metodología de trabajo. Elaboración propia	14
3.2. Esquema del proceso del modelo LDA. Elaboración propia	19
4.1. Evolución anual del número de publicaciones en el corpus (2000–2026). Elaboración propia	25
4.2. Nube de palabras más frecuentes en el corpus. Elaboración propia	27
4.3. Unigramas más relevantes del corpus. Elaboración propia	27
4.4. Bigramas más relevantes del corpus. Elaboración propia	28
4.5. Trigramas más relevantes del corpus. Elaboración propia	29
4.6. Evolución del índice de coherencia Cv en función del número de tópicos K. Elaboración propia	30
4.7. Mapa de distancia intertópica del modelo LDA con K=3 (pyLDAvis). Elaboración propia	30
4.8. Distribución de tópicos dominantes en el corpus (N=1.040). Elaboración propia	33
4.9. Evolución histórica y proyección del número de publicaciones por tópico (2000–2030). Elaboración propia.	35
4.10. Red de coocurrencia de palabras clave de autor (umbral mínimo: 5 ocurrencias). Elaboración propia con VOSviewer	36
4.11. Red de coocurrencia de la totalidad de palabras clave (umbral mínimo: 10 ocurrencias). Elaboración propia con VOSviewer	37
4.12. Red de acoplamiento bibliográfico por revistas (umbral mínimo: 5 documentos por revista). Elaboración propia con VOSviewer	38
4.13. Red de co-citación de referencias (umbral mínimo: 5 citas). Elaboración propia con VOSviewer	39

4.14. Red de coautoría por países (umbral mínimo: 5 documentos por país, 40 países representados). Elaboración propia con VOSviewer	39
---	----

Índice de tablas

2.1. Resumen de los estudios bibliométricos y analíticos sobre la transición energética.	12
4.1. Resumen de los tópicos identificados mediante LDA. Fuente: elaboración propia.	32
4.2. Resultados del test ADF iterativo por tópico. Fuente: elaboración propia. . .	34
4.3. Resumen de modelos ARIMA ajustados por tópico. Fuente: elaboración propia.	34

Acrónimos

<i>ACV</i>	Análisis de Ciclo de Vida
<i>ADF</i>	Augmented Dickey-Fuller (Test de Dickey-Fuller Aumentado)
<i>ARIMA</i>	AutoRegressive Integrated Moving Average
<i>BOW</i>	Bag of Words
<i>COP28</i>	Conferencia de las Partes 28
<i>CSV</i>	Comma-Separated Values
<i>IPCC</i>	Intergovernmental Panel on Climate Change
<i>LDA</i>	Latent Dirichlet Allocation
<i>LLM</i>	Large Language Model
<i>ML</i>	Machine Learning
<i>NLTK</i>	Natural Language Toolkit
<i>ODS</i>	Objetivos de Desarrollo Sostenible
<i>PLN</i>	Procesamiento del Lenguaje Natural
<i>PNIEC</i>	Plan Nacional Integrado de Energía y Clima
<i>PRISMA</i>	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
<i>RF</i>	Random Forest
<i>TF-IDF</i>	Term Frequency-Inverse Document Frequency

Capítulo 1

Introducción

1.1. Motivación

En el contexto actual de emergencia climática, la transición energética se ha convertido en una necesidad estratégica para garantizar la sostenibilidad ambiental y avanzar hacia un modelo de desarrollo que preserve el equilibrio ecológico. Este proceso implica abandonar progresivamente los combustibles fósiles en favor de fuentes renovables, pero trasciende lo puramente tecnológico, ya que incorpora dimensiones económicas, sociales y políticas de enorme alcance. Tal como se observa en la Figura 1.1, la anomalía de la temperatura media global muestra una tendencia creciente sostenida desde mediados del siglo XIX, con un incremento especialmente acusado desde la década de 1980. Esta evolución confirma que el calentamiento global ya no constituye una proyección hipotética, sino una realidad medible, dado que la temperatura media mundial ha aumentado en torno a 1,5 °C respecto a los niveles preindustriales (IPCC, 2023). Este escenario nos sitúa en un umbral crítico que compromete la estabilidad de los ecosistemas y la seguridad humana. Si las emisiones continúan al ritmo actual, los escenarios científicos señalan que el calentamiento podría superar los 2 °C antes de 2050, lo que intensificaría fenómenos extremos como olas de calor, sequías prolongadas y ascenso del nivel del mar, reforzando la urgencia de acelerar la transición hacia sistemas energéticos más sostenibles.

La evolución de las emisiones globales de CO_2 refuerza este diagnóstico. Como evidencia la Figura 1.2, las emisiones han experimentado un incremento exponencial desde mediados del siglo XX, superando ampliamente los 35.000 millones de toneladas anuales en la actualidad. Este crecimiento acelerado está estrechamente vinculado al uso intensivo de combustibles fósiles en los sectores energético, industrial y del transporte, y constituye el principal impulsor del calentamiento global. Aunque el Acuerdo de París (2015) marcó un hito en la coordinación internacional para limitar el aumento de la temperatura global, la implementación de sus compromisos ha sido más lenta de lo previsto (UNFCCC, 2015). En la COP28 (2023) se reiteró la necesidad de reducir de manera urgente las emisiones y acelerar

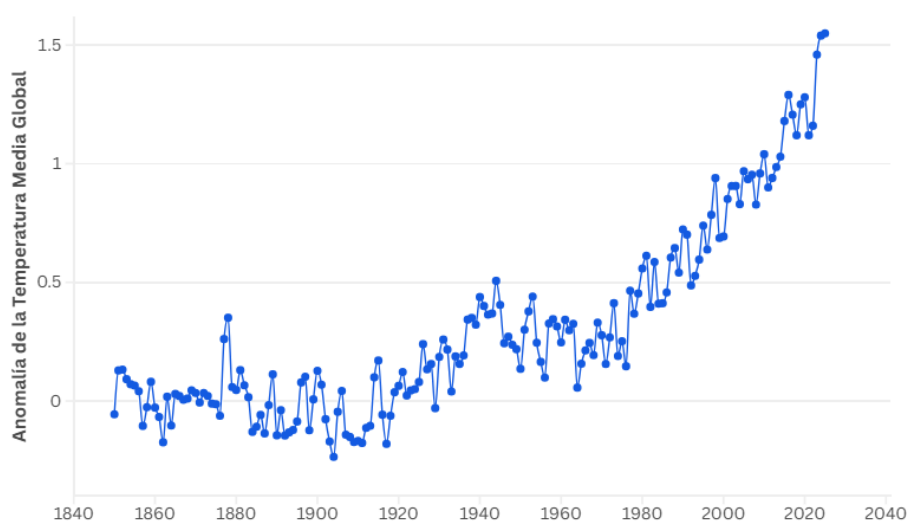


Figura 1.1: Anomalía de la temperatura media global. Fuente de datos: Our World in Data (2024). Elaboración propia

la transición energética, pero los desafíos económicos, geopolíticos y sociales continúan dificultando la adopción de medidas más ambiciosas (UNFCC, 2023). A pesar de esta situación, los avances en generación eléctrica renovable evidencian que existen alternativas tecnológicamente viables para reducir la dependencia del carbono. La Figura 1.3 muestra un aumento sostenido de las fuentes renovables modernas, con un crecimiento particularmente destacado de la energía eólica y solar en las dos últimas décadas (Ember, 2024). Mientras que la hidroelectricidad ha mantenido un desarrollo más estable, la rápida expansión de tecnologías como la solar fotovoltaica y la eólica demuestra la capacidad del sistema energético global para transformar sus patrones de generación. Sin embargo, aun con este crecimiento, la velocidad de despliegue sigue siendo insuficiente para compensar el ritmo de aumento de las emisiones, lo que destaca la urgencia de acelerar la adopción de energías limpias y fortalecer las políticas públicas orientadas a la descarbonización.

En este panorama de creciente presión climática a escala mundial, España se posiciona como un país firmemente comprometido con la transición hacia un modelo energético más sostenible, aunque enfrenta desafíos estructurales de considerable envergadura. El Plan Nacional Integrado de Energía y Clima (PNIEC 2023–2030) establece metas ambiciosas que incluyen alcanzar un 48 % de energías renovables sobre el consumo final de energía, mejorar en un 43 % la eficiencia energética y garantizar que aproximadamente el 81 % de la generación eléctrica provenga de fuentes renovables para el año 2030 (Ministerio para la Transición Ecológica y el Reto Demográfico, 2023). Estas metas reflejan la voluntad de situar al país en una senda de descarbonización acelerada. Sin embargo, la reducción de emisiones no ha avanzado al mismo ritmo que la expansión de las renovables, lo que evidencia que la transición energética requiere no solo tecnologías limpias, sino también transformaciones

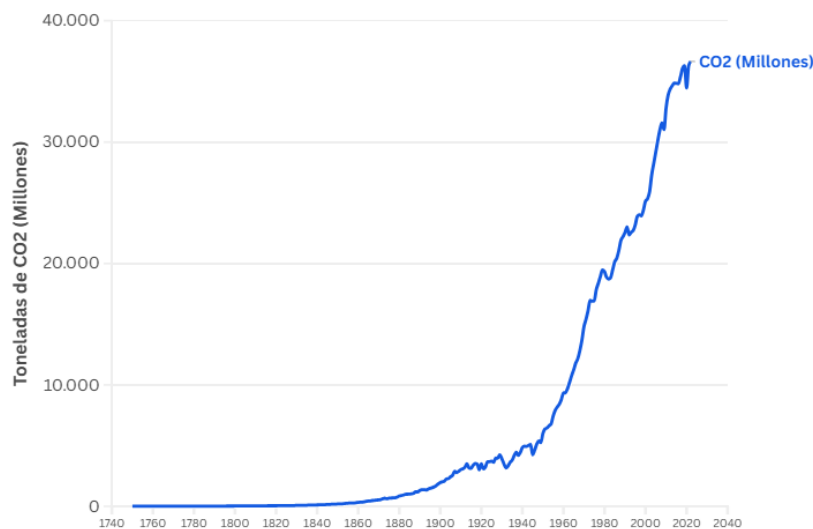


Figura 1.2: Evolución de las emisiones globales de CO2. Fuente de datos: (Epdata, 2023). Elaboración propia

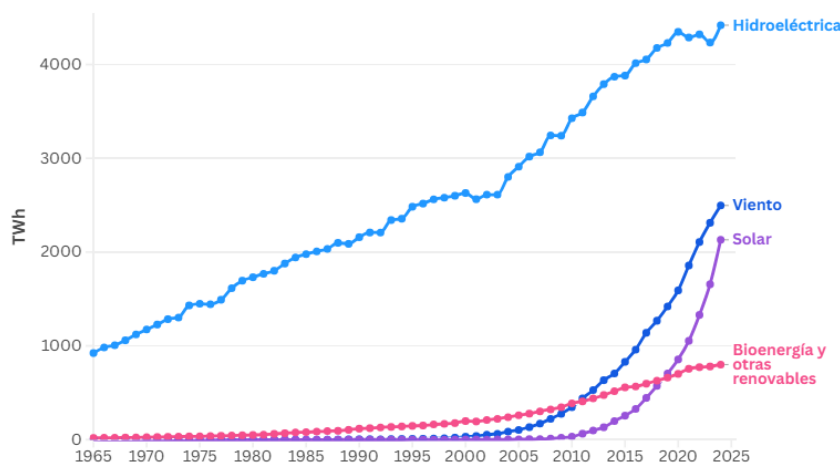


Figura 1.3: Generación moderna de electricidad renovable por fuente. Fuente de datos: Our World in Data (s.f.). Elaboración propia

profundas en las infraestructuras, la regulación y los patrones de consumo. A ello se suma la elevada dependencia energética del exterior, que continúa siendo uno de los principales retos estructurales. Según el Banco de España, más del 70 % de la energía consumida en el país sigue siendo importada, con una especial vulnerabilidad en el caso del gas natural (Balteanu y Viani, 2023). Esta situación expone a la economía española a fluctuaciones en los precios internacionales y a tensiones en los mercados globales, lo que evidencia la necesidad de impulsar aún más la diversificación de fuentes, fortalecer la producción interna de energías renovables y avanzar hacia una mayor autonomía energética.

A estos desafíos se añaden barreras adicionales que condicionan la velocidad y la eficacia de la transición energética. La dependencia del gas natural mantiene un peso significativo en

sectores relevantes, mientras que las limitaciones regulatorias dificultan la creación de un entorno estable que incentive inversiones sostenidas en energías renovables y en tecnologías de descarbonización. Aunque el PNIEC constituye un marco estratégico relevante, persisten carencias en materia de incentivos, coherencia normativa y planificación de largo plazo. Estas dificultades coinciden con lo señalado por la literatura internacional, que muestra que el éxito de la transición energética depende tanto de los avances tecnológicos como de la existencia de políticas públicas consistentes, estables y orientadas a la descarbonización de sectores como el transporte y la industria (Geels, Sovacool, Schwanen, y Sorrell, 2017). En esta misma línea, (Kern y Rogge, 2016) destacan la necesidad de fortalecer las capacidades institucionales y de alinear las políticas energéticas con objetivos de sostenibilidad y competitividad, aspectos particularmente relevantes para España en su camino hacia un sistema energético más resiliente, autónomo y bajo en carbono.

En este escenario, el conocimiento científico adquiere una gran importancia para orientar y acelerar la transición energética. La investigación académica constituye una fuente fundamental para comprender la evolución de los sistemas energéticos y para analizar las alternativas que permiten reducir las emisiones y promover modelos más sostenibles. Además de documentar avances tecnológicos o evaluar la efectividad de las políticas públicas, la literatura especializada proporciona marcos conceptuales que ayudan a interpretar los cambios en curso y a anticipar necesidades futuras. Esta dimensión analítica y prospectiva convierte a la producción científica en un recurso estratégico para los distintos actores implicados en la descarbonización, desde administraciones públicas hasta empresas y organismos reguladores.

La evidencia empírica muestra que las tendencias de investigación actúan como indicadores tempranos de los cambios que probablemente marcarán la agenda energética en el corto y medio plazo (Alvial-Palavicino, 2015). En este contexto, el análisis bibliométrico y la minería de texto han emergido como herramientas especialmente valiosas para identificar patrones, detectar temas emergentes y comprender la evolución de las prioridades científicas en el ámbito energético. Diversos estudios, como los de (Diaconescu, Marinas, Marinoiu, Popescu, y Diaconescu, 2024) y (Harichandan et al., 2022), muestran que estos métodos permiten mapear la dinámica del conocimiento y anticipar las áreas relevantes de investigación incluso antes de que las políticas públicas o las inversiones privadas se alineen con dichos desarrollos. Su capacidad para sintetizar grandes volúmenes de información y revelar las estructuras subyacentes de la producción científica ofrece una base importante para interpretar el rumbo de la investigación energética y para identificar las tecnologías, enfoques y marcos conceptuales que previsiblemente adquirirán mayor relevancia en el futuro.

En esta línea de investigación, que destaca la utilidad del análisis bibliométrico y de la minería de textos para comprender y anticipar las transformaciones del ámbito energético, se enmarca el presente Trabajo de Fin de Grado. El estudio plantea examinar la evolución de la investigación sobre transición energética en España mediante la construcción y análisis de un

corpus de publicaciones científicas recientes, con el objetivo de identificar los principales ejes temáticos, detectar tendencias emergentes y proyectar la evolución del volumen de estudios en los próximos años. Este enfoque ofrecerá una visión estructurada y basada en evidencia sobre las dinámicas de la producción científica nacional, aportando elementos de utilidad para la planificación estratégica, la orientación de políticas públicas y el fortalecimiento de la agenda investigadora en materia de sostenibilidad energética.

1.2. Objetivos del Trabajo de Grado

1.2.1. Objetivo General

El presente Trabajo de Fin de Grado en Business Analytics tiene como objetivo principal identificar y examinar las tendencias emergentes en la investigación sobre transición energética en España, analizando su evolución temporal a través de técnicas de minería de textos y modelado de tópicos aplicadas a publicaciones académicas recientes. Asimismo, se plantea elaborar un pronóstico del volumen de publicaciones con el propósito de anticipar las áreas de investigación que adquirirán mayor relevancia.

1.2.2. Objetivos Específicos

- Analizar el impacto y la evolución de la investigación sobre la transición energética en España, destacando los principales ejes temáticos y los cambios en los focos de interés académico a lo largo del tiempo, con el fin de comprender su desarrollo.
- Examinar los enfoques metodológicos existentes en la literatura sobre análisis textual y predicción de series temporales, con el propósito de establecer una base conceptual que sustente la aplicación de técnicas adecuadas en el estudio de las publicaciones académicas.
- Identificar y analizar las principales tendencias de investigación en torno a la transición energética en España, mediante la aplicación de técnicas de minería de textos y modelado de tópicos a un corpus de publicaciones académicas recientes.
- Desarrollar un marco de predicción del volumen de publicaciones sobre transición energética en España que permita anticipar tendencias y proyectar las áreas de mayor relevancia en la agenda investigadora.

1.3. Organización de la memoria

La memoria de este trabajo se organiza en cinco capítulos principales. El capítulo 1 presenta la introducción, en la que se expone la motivación del estudio, se definen los objetivos planteados y se describe la estructura general del documento. El capítulo 2 desarrolla el estado del arte, ofreciendo una revisión de los trabajos más relevantes en el ámbito de la transición energética y de los enfoques metodológicos empleados en investigaciones afines. En el capítulo 3 se detalla la metodología empleada, que comprende la recopilación y exploración de los datos, el preprocesamiento y preparación del texto, el modelado de tópicos y la predicción del volumen de publicaciones. El capítulo 4 describe los resultados más significativos derivados del análisis realizado. Finalmente, el capítulo 5 recoge las principales conclusiones del estudio y plantea posibles líneas de investigación futura.

Capítulo 2

Técnicas bibliométricas para el análisis de las tendencias emergentes en la transición energética española: Revisión de la literatura

En las últimas décadas, la transición energética se ha consolidado como uno de los ámbitos de investigación más relevantes y dinámicos en el marco de las ciencias ambientales, económicas y tecnológicas, debido a su papel relevante en la respuesta al cambio climático y a los retos asociados a la sostenibilidad. La necesidad de transformar los sistemas energéticos tradicionales hacia modelos más eficientes y basados en fuentes renovables ha situado a la reducción de emisiones de gases de efecto invernadero en el centro del debate científico y social. En este contexto, la investigación académica ha experimentado un notable crecimiento, reflejado en una producción científica cada vez más amplia sobre energías renovables, eficiencia energética y procesos de descarbonización. Son ejemplo de ellos los estudios de (Alizadeh, Moghaddam, Amjady, Siano, y Sheikh-El-Eslami, 2016), (ur Rehman, Sanjari, Elavarasan, y Jamal, 2026) y (Morte, Ofélia de Queiroz, Morgado, y de Medeiros, 2023). Esta literatura no solo documenta los avances tecnológicos y regulatorios alcanzados, sino que también aporta marcos analíticos para comprender las implicaciones económicas, sociales y territoriales de la transición energética. De este modo, las investigaciones científicas se convierten en una fuente relevante de conocimiento para interpretar la evolución del sector energético y para anticipar las tendencias que orientarán la innovación y las estrategias de sostenibilidad en los próximos años.

En este escenario, diversos autores a nivel internacional han propuesto estrategias metodológicas orientadas a analizar de manera sistemática, automatizada y fiable la creciente producción científica en torno a la transición energética. Entre los estudios más relevantes destaca el trabajo de (Harichandan et al., 2022), que constituye una de las aproximaciones

más detalladas en este ámbito. Los autores analizaron cuatro décadas de publicaciones académicas relacionadas con sostenibilidad y energía mediante la aplicación de la metodología PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), combinada con herramientas bibliométricas como *VOSviewer* y *Bibliometrix*. El estudio examinó un total de 2.169 publicaciones y permitió identificar las principales tendencias asociadas a la transición desde sistemas energéticos basados en combustibles fósiles hacia fuentes renovables. Los resultados evidencian una creciente relevancia de temáticas como la gobernanza energética, la justicia ambiental y la participación comunitaria, junto con el papel central de la innovación tecnológica y de las políticas públicas en el proceso de descarbonización. Asimismo, el análisis señala áreas con elevado potencial de desarrollo futuro, entre ellas los marcos regulatorios, las infraestructuras de almacenamiento y transferencia de energía, y la integración de energías renovables en ecosistemas locales. De manera complementaria, el estudio identifica los términos más frecuentes en la literatura, tales como “transición energética”, “política energética” y “energías renovables”, lo que refuerza la centralidad de estos conceptos en la agenda investigadora contemporánea.

Junto a los análisis de carácter general, la literatura reciente incorpora estudios bibliométricos centrados en tecnologías renovables específicas, que permiten profundizar en dinámicas de investigación más concretas dentro del proceso de transición energética. En este sentido, el trabajo de (Obaideen et al., 2023) constituye una referencia destacada al analizar de manera sistemática la investigación sobre energía solar, una de las tecnologías relevantes en los escenarios globales de descarbonización. A partir de la base de datos *Scopus*, los autores examinan 276.989 publicaciones, de las cuales 126.513 se vinculan explícitamente con los Objetivos de Desarrollo Sostenible (ODS). Mediante el uso de indicadores bibliométricos, análisis de palabras clave y herramientas de visualización, el estudio identifica vacíos de investigación y propone la necesidad de alinear de forma más explícita la agenda científica en energía solar con el conjunto de los ODS. Sus resultados muestran que la literatura ha evolucionado desde enfoques predominantemente técnicos hacia perspectivas más amplias que incorporan aspectos relacionados con la sostenibilidad de materiales, la economía circular y las políticas energéticas. Una lógica metodológica similar se observa en el estudio de (Aytac, 2025), que desarrolla un análisis bibliométrico de la investigación en energía eólica a partir de 25.842 publicaciones indexadas en *Web of Science*. Este trabajo identifica a los países e instituciones más relevantes, como China, EE. UU. e India, que lideran la producción de investigaciones en este campo. Además, destaca los principales clústeres temáticos, siendo los más destacados los relacionados con energía y combustibles sostenibles, tecnología científica y sostenible y ingeniería eléctrica. También se observa una creciente orientación hacia las dimensiones socioeconómicas de la energía eólica, especialmente en lo que respecta a la aceptación social de los proyectos y la planificación territorial.

En una línea complementaria, el trabajo de (Kut y Pietrucha-Urbanik, 2023) profundiza en el análisis bibliométrico de la investigación en energías renovables desde una perspectiva

comparada a nivel nacional. Los autores examinan tres décadas de publicaciones académicas recopiladas de *Web of Science*, centrando su estudio en dos países europeos, Polonia y Alemania. Los resultados muestran que Alemania presenta una trayectoria investigadora más consolidada y diversificada, caracterizada por una mayor amplitud temática y una adopción temprana de tecnologías renovables. En contraste, la producción científica de Polonia refleja un desarrollo más condicionado por su histórica dependencia del carbón y por los retos asociados a la seguridad energética. A pesar de compartir un marco geográfico y político común dentro de la Unión Europea, el estudio destaca que las prioridades de investigación, el ritmo de innovación y la orientación de las políticas energéticas difieren de manera significativa entre ambos países. Estos resultados muestran que la transición energética presenta dinámicas nacionales específicas que difícilmente pueden captarse mediante análisis globales, lo que a su vez destaca la relevancia de estudios focalizados en contextos nacionales para comprender en profundidad la evolución de la investigación en este ámbito.

Desde una perspectiva metodológica, la investigación reciente también resalta el papel creciente de las técnicas de *machine learning* (ML) en el análisis de sistemas energéticos y en el estudio de las energías renovables. En este marco, el trabajo de (Ajibade, Bekun, Adedoyin, Gyamfi, y Adediran, 2023) realiza una revisión bibliométrica de la producción científica sobre la aplicación de ML en energías renovables durante el período 2012–2021. A partir de los registros extraídos de la base de datos *Scopus*, los autores identifican un incremento sostenido en el uso de algoritmos predictivos, lo que refleja su progresiva consolidación como herramientas habituales en este campo. El estudio muestra que modelos como las redes neuronales artificiales, las máquinas de soporte vectorial, *Random Forest* y diversos enfoques híbridos se emplean de forma recurrente para la predicción de la generación renovable, la optimización de sistemas energéticos y la planificación de la demanda. Además del análisis cuantitativo de las publicaciones, el trabajo ofrece una clasificación de los algoritmos más utilizados y de sus principales aplicaciones, destacando que el ML se ha convertido en un componente de gran relevancia del avance científico y metodológico asociado a la transición energética.

En esta línea de trabajos que combinan enfoques bibliométricos avanzados con técnicas de análisis estructural, el estudio de (Diaconescu et al., 2024) ofrece una de las panorámicas más amplias sobre la investigación en transición energética a escala global. El autor analiza 11.456 artículos científicos recopilados de *Web of Science* con el objetivo de identificar las principales dinámicas, actores y patrones temáticos que configuran el campo. Para ello, emplea algoritmos de *clustering* y análisis de co-ocurrencia apoyados en herramientas como *VOSviewer*, lo que permite detectar comunidades temáticas, visualizar la proximidad conceptual entre términos clave y examinar la evolución temporal de los focos de investigación. Los resultados muestran que la producción científica se organiza en torno a clústeres consolidados, entre los que destacan las líneas de investigación vinculadas a la energía solar, seguidas por la energía eólica, los estudios sobre mitigación del cambio climático y los análisis de los

impactos ambientales y económicos de la transición energética. Asimismo, el trabajo destaca un interés creciente por cuestiones relacionadas con la gobernanza, la seguridad energética, la planificación estratégica y la integración de las energías renovables en los sistemas eléctricos, evidenciando el carácter multidisciplinar y progresivamente complejo del campo.

Más recientemente, la literatura ha comenzado a incorporar enfoques que combinan el análisis estructural de la producción científica con técnicas de carácter predictivo, ampliando el alcance de los estudios bibliométricos tradicionales. En esta línea, el trabajo de (Zou, Guo, Li, y Wu, 2024) integra métodos de minería de textos y modelos de predicción temporal para analizar la orientación futura de la investigación en política energética. A partir de un corpus de 13.967 resúmenes de publicaciones relacionadas con la transición energética en China, los autores aplican el modelo LDA (*Latent Dirichlet Allocation*) para identificar 26 tópicos principales presentes en la literatura. Posteriormente, estos resultados se combinan con modelos ARIMA (*AutoRegressive Integrated Moving Average*) con el fin de predecir la evolución temporal de cada tópico en un horizonte de 36 meses. Este enfoque permite anticipar las líneas de investigación que presentan un mayor potencial de crecimiento, ofreciendo una visión prospectiva de las prioridades científicas y de la innovación tecnológica en el ámbito energético.

Con el fin de sintetizar de manera estructurada los principales enfoques revisados, la Tabla 2.1 presenta un resumen comparativo de los estudios bibliométricos y analíticos más relevantes en el ámbito de la transición energética. En particular, se recogen los trabajos seleccionados, el tamaño y origen de los conjuntos de datos, los métodos de adquisición empleados, los objetivos de cada contribución, los algoritmos aplicados y los resultados más destacados. Este resumen permite identificar patrones metodológicos comunes, así como diferencias en el alcance empírico y en el foco temático de los análisis, aportando una base clara para justificar el enfoque adoptado en el presente estudio.

De manera general, la literatura revisada se apoya principalmente en bases de datos bibliográficas de amplia cobertura internacional, destacando *Web of Science* y *Scopus* como fuentes predominantes para la recopilación de publicaciones. El volumen de los análisis varía significativamente según el objetivo y el nivel de granularidad de cada estudio, desde revisiones con unos pocos miles de registros, como el análisis de (Harichandan et al., 2022) con 2.169 publicaciones, hasta estudios de gran escala, como el de (Obaideen et al., 2023), basado en 276.989 publicaciones. En términos metodológicos, los estudios revisados combinan técnicas bibliométricas clásicas y avanzadas, análisis de redes y co-ocurrencia, algoritmos de *clustering* temático y herramientas de visualización como *VOSviewer* y *Bibliometrix*. Además, se observa una progresiva incorporación de técnicas de aprendizaje automático para tareas de clasificación, predicción y optimización en sistemas energéticos (Ajibade et al., 2023). Dentro de este conjunto de enfoques, el modelado de tópicos mediante LDA adquiere especial relevancia, al permitir identificar de forma automatizada estructuras temáticas latentes a partir de grandes corpus textuales. En particular, el trabajo de (Zou et al., 2024) ejemplifica

el potencial de integrar LDA con modelos de series temporales, en este caso ARIMA, para analizar la dinámica temporal de los tópicos y anticipar su evolución en horizontes de corto plazo. Este enfoque resulta especialmente pertinente en investigaciones orientadas a detectar tendencias emergentes y a caracterizar la agenda científica de manera prospectiva.

Finalmente, la síntesis comparada de la literatura permite identificar ejes temáticos recurrentes que estructuran la investigación internacional en transición energética. En primer lugar, se observa un núcleo tecnológico dominado por el despliegue y la integración de energías renovables, especialmente energía solar y eólica, así como cuestiones asociadas al almacenamiento y a la infraestructura energética. En segundo lugar, emerge un eje transversal de descarbonización y mitigación del cambio climático, que conecta la producción científica con objetivos ambientales globales. En tercer lugar, adquieren peso creciente las dimensiones de gobernanza, seguridad energética y planificación estratégica, junto con perspectivas socioeconómicas vinculadas a la aceptación social, la justicia ambiental y la participación comunitaria. Estos ejes muestran el carácter multidisciplinar del campo y sugieren que la transición energética se estudia cada vez más como un fenómeno sistémico, donde la innovación tecnológica se articula con marcos regulatorios, dinámicas institucionales y efectos territoriales.

A partir de este marco de referencia, se identifica la oportunidad de profundizar en el análisis de la transición energética desde una perspectiva integrada que combine el estudio temático de la producción científica con una dimensión temporal y prospectiva. Si bien la literatura existente ha aportado contribuciones relevantes a nivel global, y tecnológico, persiste la necesidad de aplicar estos enfoques de forma específica al contexto español, atendiendo a sus particularidades regulatorias, energéticas y socioeconómicas. En este sentido, el presente Trabajo de Fin de Grado se propone analizar la evolución de la investigación sobre transición energética en España mediante la construcción de un corpus de publicaciones académicas y la aplicación de técnicas de minería de textos y modelado de tópicos, con especial énfasis en LDA, así como métodos de análisis temporal orientados a anticipar tendencias futuras. Este enfoque permitirá no solo identificar los principales ejes temáticos que han caracterizado la agenda investigadora nacional, sino también proyectar su evolución en el corto plazo, contribuyendo así a una comprensión más informada de las dinámicas científicas que acompañan el proceso de transición energética en España.

Tabla 2.1: Resumen de los estudios bibliométricos y analíticos sobre la transición energética.

Referencia	Tamaño del Dataset	Método de Adquisición	Objetivo	Algoritmo	Resultados
(Harichandan et al., 2022)	2.169 publicaciones	<i>Web of Science</i> y <i>Scopus</i> (1981–2021). Filtrado por palabras clave relacionadas con transición energética y sostenibilidad.	Mapear 40 años de investigación global sobre transición energética e identificar tendencias temáticas, redes de colaboración y evolución conceptual.	Bibliometría clásica: <i>VOSviewer</i> , <i>Bibliometrix</i> , co-ocurrencia, análisis de redes.	Detecta transición desde estudios tecnológicos hacia gobernanza, justicia ambiental e innovación. Señala vacíos en almacenamiento energético y políticas de descarbonización.
(Obaideen et al., 2023)	276.989 publicaciones (126.513 vinculadas a ODS)	<i>Scopus</i> ; filtrado por términos relacionados con “ <i>solar energy</i> ” y clasificación respecto a los ODS.	Analizar la investigación global sobre energía solar y su alineación con los ODS.	Bibliometría; análisis de palabras clave; paneles de visualización; co-ocurrencia.	La energía solar se consolida como tecnología central para la descarbonización. Identifica vacíos entre energía solar y ciertos ODS. Aumenta interés por economía circular y sostenibilidad de materiales.
(Aytac, 2025)	25.842 publicaciones	<i>Web of Science</i> , filtrado por “ <i>wind energy</i> ” y categorías científicas asociadas.	Examinar la evolución científica de la energía eólica e identificar clústeres temáticos, países líderes y líneas de investigación emergentes.	Bibliometría avanzada; clustering; análisis longitudinal; visualización con <i>VOSviewer</i> .	Crecimiento acelerado desde 2018. Sector dominado por China y EE.UU. Expansión hacia estudios socioeconómicos: aceptación social, planificación territorial e integración en red.
(Kut y Pietrucha-Urbanik, 2023)	≈ 10.000 publicaciones (Polonia y Alemania)	<i>Web of Science</i> ; filtrado por país, afiliación institucional y temática.	Comparar la evolución de la investigación en renovables entre Polonia y Alemania para detectar diferencias nacionales y prioridades científicas.	<i>Co-word analysis</i> ; co-citación; análisis de redes de coautoría.	Alemania presenta trayectoria más diversificada e innovadora. Polonia condicionada por dependencia del carbón. Evidencia dinámicas científicas diferenciadas a nivel nacional.
(Ajibade et al., 2023)	≈ 8.000 publicaciones	<i>Scopus</i> (2012–2021). Filtrado temático por “ <i>machine learning</i> ” y “ <i>renewable energy</i> ”.	Analizar la evolución de la investigación en aprendizaje automático aplicado a energías renovables e identificar los algoritmos más utilizados y sus principales aplicaciones.	<i>Random Forest</i> , métodos híbridos; bibliometría; clasificación de algoritmos.	Aumento significativo en el uso de ML para predicción renovable, optimización energética y análisis de demanda. Identifica áreas emergentes y algoritmos predominantes.
Diaconescu et al. (2024)	11.456 publicaciones	Extracción desde <i>Web of Science</i> con filtros temáticos sobre energías renovables.	Identificar clústeres temáticos y analizar la evolución de los focos de investigación en energías renovables.	<i>Clustering</i> temático; co-ocurrencia; visualización con <i>VOSviewer</i> .	Identifica clústeres consolidados (solar, eólica, mitigación climática). Muestra evolución hacia gobernanza, planificación estratégica y seguridad energética.
(Zou et al., 2024)	13.976 resúmenes	<i>Web of Science</i> ; extracción automática de resúmenes; filtrado temático sobre políticas energéticas.	Detectar temas principales en investigación sobre transición energética y predecir su evolución temporal.	LDA (<i>topic modeling</i>) y ARIMA (<i>forecasting</i>).	Identifica 26 tópicos y predice un crecimiento sostenido en la mayoría. Demuestra la utilidad de combinar minería de texto y predicción temporal para anticipar tendencias científicas.

Capítulo 3

Metodología

En este capítulo se expone la metodología empleada para el desarrollo del presente Trabajo de Fin de Grado, cuyo objetivo es analizar la evolución temática y la dinámica temporal de la investigación académica sobre la transición energética en España. El procedimiento seguido se articula en seis etapas principales, representadas en la Figura 3.1, que integran técnicas de minería de texto, modelado de tópicos, predicción de series temporales y análisis bibliométrico.

1. **Recopilación y exploración de los datos:** En una primera fase se llevó a cabo la recopilación de artículos científicos procedentes de la base de datos Scopus, empleando una estrategia de búsqueda basada en combinaciones de términos clave relacionados con la transición energética en España. Se aplicaron filtros por tipo de documento, idioma y periodo temporal. Los resultados obtenidos fueron estructurados para su procesamiento posterior, incluyendo datos como el título, resumen, autores, filiación institucional, palabras clave y año de publicación.
2. **Procesamiento y preparación del texto:** Una vez definido el corpus, se aplicaron técnicas de limpieza y normalización sobre los resúmenes de los artículos seleccionados. Este proceso incluyó la eliminación de registros duplicados y de ruido editorial introducido por las editoriales, normalización del texto (conversión a minúsculas, eliminación de signos de puntuación y caracteres especiales), tokenización, eliminación de palabras vacías (*stopwords*) y lematización. Estas tareas permitieron transformar el texto en una representación estructurada adecuada para su análisis automático.
3. **Análisis exploratorio de N-gramas:** Como etapa preliminar al modelado de tópicos, se realizó un análisis exploratorio de unigramas, bigramas y trigramas con el fin de identificar combinaciones de términos y conceptos dominantes en el corpus. Este análisis aportó una primera aproximación descriptiva que facilitó la interpretación posterior de los tópicos obtenidos.

4. **Modelado de tópicos (LDA):** La fase central del estudio consistió en la aplicación del algoritmo *Latent Dirichlet Allocation* (LDA) sobre el corpus obtenido. Esta técnica de modelado probabilístico no supervisado permite identificar estructuras temáticas latentes a partir de grandes volúmenes de texto, extrayendo tópicos representativos. El número óptimo de tópicos se determinó mediante métricas de coherencia semántica, buscando maximizar la interpretabilidad y consistencia de los resultados.
5. **Análisis temporal y predicción del volumen de publicaciones:** Finalmente, se construyó una serie temporal a partir del número anual de publicaciones científicas sobre transición energética en España. Sobre esta serie se aplicó un modelo autorregresivo integrado de medias móviles (ARIMA), con el fin de analizar la evolución histórica de la producción científica y realizar una predicción a corto plazo de su comportamiento futuro.
6. **Análisis bibliométrico con VOSviewer:** Como complemento al análisis temático, se llevó a cabo un estudio bibliométrico mediante la herramienta VOSviewer. Este análisis incluyó mapas de co-ocurrencia de palabras clave, acoplamiento bibliográfico entre artículos y redes de coautoría por país y de cocitación. El objetivo fue identificar los núcleos temáticos del campo, las comunidades de conocimiento más consolidadas y los patrones de colaboración científica internacional, con especial atención al papel de las instituciones españolas en las redes globales de investigación.

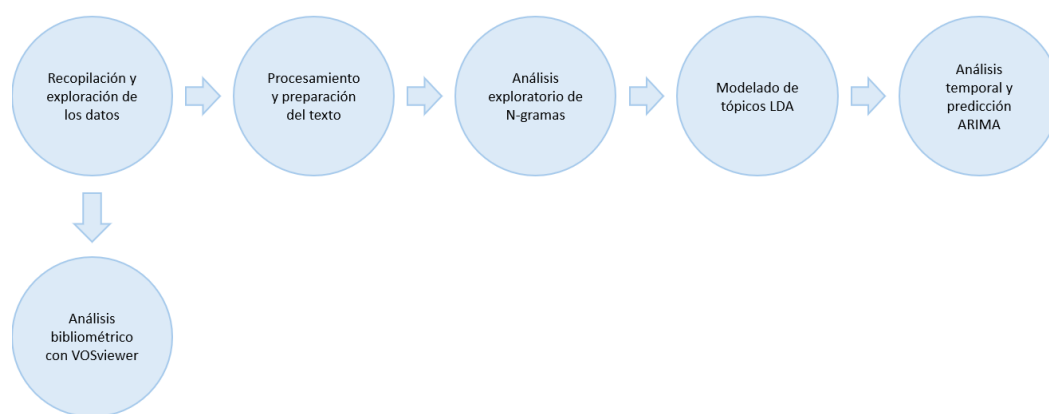


Figura 3.1: Esquema de la metodología de trabajo. Elaboración propia

3.1. Recopilación y exploración de los datos

En esta primera etapa se empleó la base de datos científica Scopus, seleccionada por su amplia cobertura internacional y su elevado nivel de calidad editorial en publicaciones revisadas por pares. El objetivo fue construir un corpus bibliográfico representativo que permitiera

abordar el análisis temático y temporal de la literatura reciente sobre transición energética en España. Para ello, se diseñó una consulta de búsqueda estructurada a partir de términos clave relacionados con la transición energética y el contexto geográfico español. La búsqueda se construyó combinando conceptos centrales como energías renovables, descarbonización, eficiencia energética o cambio climático, con referencias explícitas a España como ámbito de estudio, tanto en el título como en el resumen o las palabras clave de los artículos. Esta combinación se definió con el fin de maximizar la relevancia temática y garantizar que los documentos recopilados abordaran efectivamente el contexto energético español.

Con el fin de asegurar la calidad y homogeneidad del corpus, se aplicaron filtros adicionales directamente en Scopus. La búsqueda se limitó a artículos científicos (*ar*) y revisiones (*re*), excluyéndose otros tipos de documentos no revisados. Asimismo, se consideraron exclusivamente publicaciones en inglés, dada su predominancia como lengua de difusión científica internacional. Finalmente, se estableció un intervalo temporal que incluyó únicamente publicaciones posteriores a 1999, con el propósito de capturar la evolución más reciente del debate académico en torno a la transición energética. Una vez obtenidos los resultados, los registros bibliográficos fueron exportados en formato CSV compatible con las herramientas de análisis utilizadas en las siguientes fases del estudio. Con el objetivo de adaptar el corpus a los distintos tipos de análisis, se realizaron dos exportaciones diferenciadas.

La primera exportación, destinada al análisis bibliométrico mediante VOSviewer, incluyó información completa de citación y referencias. En concreto, se exportaron los campos correspondientes a autores, título del documento, año de publicación, fuente, número de citas, DOI, afiliaciones institucionales, resumen, palabras clave del autor, palabras clave indexadas y referencias citadas. Estos metadatos permiten realizar análisis de coautoría, coocurrencia de términos, redes de colaboración institucional y mapas de citación. En paralelo, se preparó un segundo conjunto de datos orientado específicamente al modelado de tópicos y al análisis temporal. Esta exportación incluyó únicamente el año de publicación y el resumen de cada artículo, con el fin de construir un conjunto de datos simplificado para el procesamiento textual. A partir de este archivo, se generó un conjunto de datos con dos columnas (fecha y resumen) que fue utilizado en las fases posteriores de preprocesamiento, aplicación del modelo LDA y construcción de la serie temporal del volumen anual de publicaciones.

3.2. Procesamiento y preparación del texto

A partir del conjunto de datos preparado para el modelado de tópicos, se llevó a cabo un proceso de preprocesamiento textual con el objetivo de reducir la complejidad del lenguaje natural y mejorar la calidad de la información utilizada en las fases posteriores del análisis. Este procedimiento se aplicó exclusivamente sobre los resúmenes de los artículos científicos recopilados en la base de datos Scopus, dado que estos resumen el contenido principal de cada

publicación y permiten capturar los conceptos más importantes presentes en la literatura.

En una primera etapa, se llevó a cabo la eliminación del ruido editorial introducido por las editoriales en los resúmenes exportados desde Scopus. Este fenómeno, frecuente en exportaciones grandes de bases de datos bibliográficas, consiste en la inclusión al final de los resúmenes de textos de copyright pertenecientes a editoriales como Elsevier o MDPI, entre otras. A continuación, se realizó una normalización básica del texto. Todos los caracteres fueron convertidos a minúsculas con el fin de evitar diferencias entre palabras causadas por el uso de mayúsculas. Posteriormente, se eliminaron signos de puntuación, números, caracteres especiales y espacios redundantes que no aportan contenido semántico relevante para el análisis. Este proceso permitió obtener una representación textual más uniforme y adecuada para su procesamiento automático. Asimismo, se procedió a la estandarización de determinados términos que podían aparecer con diferentes variantes dentro del corpus. Por ejemplo, se unificaron expresiones como decarbonization y decarbonisation, o combinaciones como renewable energies y renewable energy, con el objetivo de evitar la fragmentación de conceptos que en realidad representan la misma idea. Este paso contribuyó a mejorar la coherencia semántica del corpus antes de aplicar las técnicas de análisis textual.

Posteriormente se llevó a cabo la tokenización del texto, proceso mediante el cual cada resumen fue dividido en unidades léxicas básicas o tokens. Esta transformación permite tratar cada palabra como una unidad independiente para su análisis estadístico. Tras la tokenización, se aplicó un proceso de eliminación de palabras vacías (stopwords), es decir, términos muy frecuentes que carecen de contenido semántico significativo, como artículos, preposiciones o conectores. Para esto se utilizó un diccionario base de stopwords del inglés, ampliado con tres categorías adicionales de términos: en primer lugar, residuos editoriales no eliminados en la fase anterior; en segundo lugar, términos propios del discurso académico genérico, como study, result, analysis o propose, cuya presencia sistemática en prácticamente todos los resúmenes introduce ruido al análisis temático; y en tercer lugar, nombres de países y regiones, dado que al tratarse de un análisis centrado en España estos términos aparecen de forma transversal en todos los tópicos sin aportar diferenciación temática. Asimismo, se excluyeron términos energéticos excesivamente genéricos como energy que por su presencia en el corpus tampoco contribuyen a distinguir unos tópicos de otros.

Una vez depurado el conjunto de tokens, se aplicó un proceso de lematización, mediante el cual cada palabra se redujo a su forma base o lema. Esta transformación permite agrupar variantes morfológicas de un mismo término bajo una única representación léxica, facilitando la consolidación semántica del corpus. Este procedimiento permite mejorar la calidad de las distribuciones de palabras utilizadas posteriormente en el modelado de tópicos. Finalmente, se eliminaron posibles registros duplicados y aquellos documentos cuyo contenido textual resultaba insuficiente tras el proceso de limpieza. Como resultado de este preprocesamiento se obtuvo un corpus estructurado compuesto por listas de términos normalizados para cada documento, adecuado para su utilización en el análisis exploratorio de n-gramas y en la

posterior aplicación del modelo de tópicos LDA.

3.3. Análisis exploratorio de N-gramas

Una vez completado el preprocesamiento del corpus, se procedió a realizar un análisis exploratorio basado en N-gramas, permitiendo descomponer los textos en secuencias de n palabras consecutivas. El propósito de esta fase es obtener una fotografía descriptiva del vocabulario dominante en la literatura sobre transición energética española, que sirva de referencia interpretativa para las etapas posteriores. Se trabajó con tres niveles de granularidad: unigramas, bigramas y trigramas. Mientras los primeros permiten detectar los términos individuales más relevantes del corpus, los bigramas y trigramas capturan expresiones técnicas compuestas que difícilmente podrían identificarse analizando palabras de forma aislada, como *life cycle*, *domestic hot water* o *internal rate return*.

La relevancia de cada secuencia se cuantificó mediante la métrica TF-IDF (*Term Frequency – Inverse Document Frequency*), que pondera la importancia de un término no únicamente en función de su frecuencia absoluta, sino también de su capacidad para discriminar entre documentos dentro del corpus (Das, Alphonse, et al., 2023). De este modo, términos que aparecen en prácticamente todos los resúmenes reciben una puntuación baja, al no ser informativos sobre ningún documento en particular, mientras que aquellos presentes en un subconjunto reducido de artículos obtienen puntuaciones más altas, al resultar más característicos de un contenido específico. Esta propiedad convierte al TF-IDF en una herramienta especialmente adecuada para corpus técnicos y especializados como el presente. Su cálculo se articula en dos componentes:

- **Frecuencia de término (TF)**: mide la proporción con la que un término t aparece en un documento d :

$$TF(t, d) = \frac{f_{t,d}}{n_d} \quad (3.1)$$

donde $f_{t,d}$ es la frecuencia del término t en el documento d , y n_d es el número total de términos en dicho documento.

- **Frecuencia inversa de documento (IDF)**: cuantifica la rareza de un término dentro del conjunto de documentos. A menor frecuencia de aparición global, mayor será su valor IDF. Se define como:

$$IDF(t) = \log \left(\frac{N + 1}{df(t)} \right) \quad (3.2)$$

donde N representa el número total de documentos del corpus, y $df(t)$ el número de

documentos que contienen el término t .

De esta manera, el valor final para cada término se obtiene multiplicando ambos componentes:

$$TF-IDF(t) = TF(t, d) \times IDF(t) \quad (3.3)$$

Esta métrica permite identificar aquellos N-gramas que no solo aparecen con frecuencia dentro de un documento específico, sino que además son poco comunes en el resto del corpus, lo que los convierte en elementos distintivos de su contexto. De este modo, el análisis TF-IDF ofrece una representación más informativa que el simple recuento de frecuencias, contribuyendo a una comprensión preliminar de la literatura académica asociada a la transición energética española. Los resultados obtenidos de este análisis exploratorio sirven como insumo preliminar para orientar la posterior etapa de modelado de tópicos, al facilitar la identificación de términos relevantes, coocurrencias significativas y posibles agrupamientos semánticos dentro del corpus textual.

3.4. Modelado de tópicos (LDA)

El modelado de tópicos es una técnica no supervisada de análisis de texto que permite extraer de forma automática estructuras temáticas latentes a partir de grandes colecciones documentales. Su principio fundamental consiste en identificar grupos de palabras que co-ocurren con frecuencia en los mismos documentos, infiriendo a partir de dichos patrones los temas subyacentes que organizan el contenido del corpus (Campbell, Hindle, y Stroulia, 2015). Su aplicación resulta especialmente adecuada en contextos como el presente, donde el volumen y la heterogeneidad de los documentos hacen inviable una revisión manual. Tal y como se recoge en la revisión de la literatura presentada en el capítulo 2, el modelo LDA ha sido adoptado como método principal de modelado de tópicos en estudios previos de características similares, como el de (Zou et al., 2024), que lo aplica sobre abstracts de publicaciones académicas sobre política energética china. Su consolidación en la literatura responde a su capacidad interpretativa, a su adaptabilidad a corpus de distintos tamaños y a la disponibilidad de implementaciones computacionales contrastadas.

LDA parte del supuesto de que cada documento está compuesto por una combinación de tópicos, y que cada tópico se define mediante una distribución probabilística sobre el vocabulario del corpus. La Figura 3.2 ilustra el proceso generativo subyacente al modelo. La distribución de tópicos por documento viene determinada por el hiperparámetro de Dirichlet α . Valores bajos de este parámetro favorecen que cada documento quede dominado por un número reducido de tópicos, lo que facilita interpretaciones más especializadas, mientras que valores altos producen distribuciones más uniformes entre tópicos, incrementando la diversidad temática por documento. Por su parte, la distribución de palabras dentro de cada tópico

ϕ_k está regulada por el hiperparámetro β . Un valor bajo concentra el peso en un vocabulario reducido y distintivo, mientras que un valor alto da lugar a tópicos léxicamente más difusos y de interpretación más compleja (Jelodar et al., 2019).

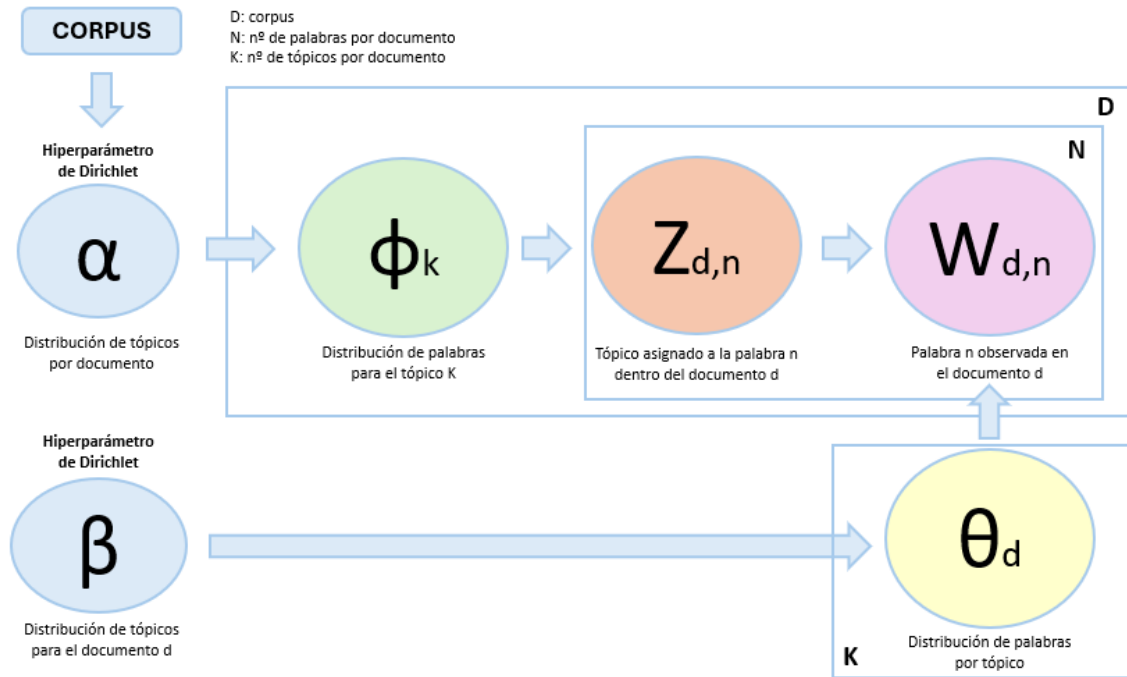


Figura 3.2: Esquema del proceso del modelo LDA. Elaboración propia

Durante el proceso generativo, la asignación de cada palabra n en un documento d sigue dos pasos sucesivos:

1. Se selecciona un tópico $Z_{d,n}$ a partir de la distribución de tópicos del documento d .
2. A partir del tópico seleccionado, se genera la palabra observada $W_{d,n}$ según la distribución de palabras asociada a dicho tópico.

En términos operativos, el modelo trabaja sobre tres niveles jerárquicos: el corpus D como conjunto total de documentos, cada documento $d \in D$, compuesto por N palabras, y cada palabra $W_{d,n}$, generada por un tópico $Z_{d,n}$ vinculado a las distribuciones θ_d y ϕ_k . El proceso de entrenamiento invierte esta lógica generativa, a partir del texto observado, se infiere qué tópicos componen cada documento y qué palabras caracterizan cada tópico. Un aspecto relevante en la aplicación del modelo LDA es la elección del número óptimo de tópicos (K). Este parámetro no puede determinarse de forma teórica previa, sino que debe estimarse empíricamente. Un valor de K demasiado reducido agrupa bajo un mismo tópico contenidos conceptualmente distintos, mientras que uno excesivamente elevado genera fragmentaciones artificiales con tópicos redundantes o difícilmente interpretables. Para su determinación, se empleó la métrica de coherencia C_v , utilizada en la literatura por su robustez empírica y su

estrecha correlación con juicios humanos de interpretabilidad (Röder, Both, y Hinneburg, 2015).

Como complemento al análisis cuantitativo de coherencia, se empleó la herramienta pyLDavis para la validación visual del modelo. Esta herramienta proyecta los tópicos como círculos en un plano bidimensional, de modo que la distancia entre ellos refleja su grado de similitud semántica. Además, el tamaño de cada círculo representa la prevalencia relativa del tópico dentro del corpus, es decir, la proporción de documentos o palabras asociadas a ese tema en comparación con los demás. Una distribución espacialmente dispersa indica una segmentación temática bien diferenciada, mientras que solapamientos o proximidades excesivas entre círculos apuntan a posibles redundancias entre tópicos. Adicionalmente, pyLDavis ofrece un ranking de términos por tópico que pondera tanto su frecuencia como su exclusividad respecto al resto de tópicos, facilitando así la interpretación y etiquetado de cada tema identificado.

3.5. Predicción de tendencias (ARIMA)

Como complemento al análisis temático derivado del modelo LDA, se aplicó el modelo *Auto-Regressive Integrated Moving Average* (ARIMA) con el objetivo de predecir la evolución futura de la popularidad de los tópicos identificados. Este análisis permite cuantificar cómo ha evolucionado el interés académico por cada línea de investigación a lo largo del periodo 2000–2025, y el modelo ARIMA se aplica sobre dicha serie para proyectar su tendencia futura hasta el horizonte 2030. De este modo, se pueden distinguir los tópicos que experimentan una evolución ascendente, los que mantienen una presencia relativamente constante y los que evidencian una pérdida de relevancia en la literatura sobre transición energética en España.

El modelo ARIMA, desarrollado por (Box, Jenkins, Reinsel, y Ljung, 2015), se expresa en la forma $ARIMA(p, d, q)$, donde p es el orden del componente autorregresivo (AR), que relaciona el valor actual de la serie con sus p valores pasados, d es el orden de diferenciación necesario para hacer la serie estacionaria, y q es el orden del componente de media móvil (MA), que incorpora el efecto de los q errores de predicción anteriores (Fattah, Ezzine, Aman, El Moussami, y Lachhab, 2018).

1. **Contraste de estacionariedad:** Para que el modelo ARIMA pueda aplicarse correctamente, la serie temporal debe ser estacionaria, es decir, su media y varianza no deben variar a lo largo del tiempo. Para verificarlo, se aplicó el test de *Dickey-Fuller Aumentado* (ADF) (Said y Dickey, 1984), cuya hipótesis nula establece la presencia de no estacionariedad en la serie. Dado que las series de publicaciones científicas agregadas por año pueden presentar tendencia, el contraste se aplicó de forma secuencial: en primer lugar sobre la serie original y, cuando no era posible rechazar la hipótesis nula

($p > 0,05$), sobre su primera diferencia ($\Delta y_t = y_t - y_{t-1}$) y, de persistir la no estacionariedad, sobre su segunda diferencia ($\Delta^2 y_t$). Este procedimiento iterativo permite determinar el orden mínimo de diferenciación d necesario para alcanzar la estacionariedad, que se incorpora posteriormente como parámetro del modelo ARIMA. En el caso de que la serie no alcance la estacionariedad tras la segunda diferenciación, el modelo se ajusta igualmente con $d = 2$, si bien las proyecciones resultantes tienen un carácter exploratorio y deben interpretarse con cautela.

2. **Identificación de parámetros:** Una vez evaluada la estacionariedad de cada serie, los parámetros p y q se determinaron de forma automática mediante el algoritmo `auto_arima`, que evalúa sistemáticamente las combinaciones candidatas dentro de los rangos $p \in [1, 3]$ y $q \in [0, 2]$, seleccionando la combinación (p, d, q) que minimiza el criterio de información de Akaike (AIC). Este enfoque garantiza una selección objetiva y reproducible de los parámetros del modelo.
3. **Predicción:** Con el modelo ARIMA(p, d, q) óptimo ya validado, se proyectó la tendencia de popularidad de cada tópico hasta el horizonte temporal 2030, acompañada de intervalos de confianza al 95 % que permiten cuantificar la incertidumbre asociada a las proyecciones. Cabe señalar que, a diferencia de (Zou et al., 2024), que emplea series mensuales con aproximadamente 156 observaciones, la disponibilidad de datos anuales en Scopus limita cada serie a 26 observaciones (2000–2025), lo que restringe la complejidad de los modelos ajustables y amplía los intervalos de confianza de las proyecciones. Esta limitación es común en estudios bibliométricos basados en exportaciones estándar de Scopus y no invalida el análisis de tendencias a largo plazo.

3.6. Análisis bibliométrico con VOSviewer

Como complemento al análisis textual basado en LDA, se realizó un análisis bibliométrico de la literatura mediante el software VOSviewer, herramienta ampliamente utilizada en la investigación académica para la construcción y visualización de redes de conocimiento científico a partir de datos bibliográficos. A diferencia del modelado de tópicos, centrado en el análisis del contenido textual de los abstracts, el análisis bibliométrico permite evaluar la estructura relacional de la literatura, en particular las conexiones entre conceptos, autores y países, ofreciendo así una perspectiva complementaria sobre la organización del campo de estudio (Van Eck y Waltman, 2010). Los datos bibliográficos necesarios para construir las redes se exportaron directamente desde Scopus en formato CSV, incluyendo campos como palabras clave de autor, palabras clave de índice, afiliaciones institucionales y referencias citadas. Se realizaron cuatro tipos de análisis de redes, cada uno orientado a responder una pregunta distinta sobre la estructura de la literatura:

1. **Coocurrencia de palabras clave:** Este análisis permite identificar los términos que coocurren de forma recurrente en los mismos artículos, lo que facilita la delimitación de los clústeres temáticos presentes en la literatura, así como la visualización de la agrupación y articulación de las principales líneas de investigación sobre transición energética en España.
2. **Acoplamiento bibliográfico (*bibliographic coupling*):** Dos fuentes se consideran bibliográficamente acopladas cuando publican artículos con referencias compartidas, lo cual sugiere la existencia de afinidades en los problemas abordados y en los marcos teóricos o empíricos empleados. En este estudio, el análisis se desarrolló tomando como unidad las revistas científicas (*sources*), con el fin de identificar las publicaciones en las cuales se concentran las principales comunidades de conocimiento del campo, así como las afinidades temáticas existentes entre ellas.
3. **Co-citación de referencias:** Dos referencias están co-citadas cuando aparecen citadas conjuntamente en un mismo artículo del corpus. Este análisis permite identificar los trabajos fundacionales del campo, es decir, aquellas obras que constituyen la base intelectual común sobre la que se construye la literatura sobre transición energética en España.
4. **Redes de coautoría por países:** Este análisis permite representar las colaboraciones científicas internacionales a partir de la coautoría de artículos, con el propósito de identificar los países que ocupan una posición más destacada en la producción académica sobre transición energética en España, así como los patrones de colaboración internacional que estructuran este campo de investigación.

Capítulo 4

Resultados del Análisis Temático, Bibliométrico y de las Proyecciones

Este capítulo presenta los principales resultados derivados de la metodología descrita en el Capítulo 3. El análisis se estructura en cinco etapas sucesivas. En primer lugar, se describe el proceso de recopilación y exploración inicial de los datos, incluyendo la evolución temporal de la producción científica en el campo. A continuación, se exponen los resultados del preprocesamiento textual y del análisis exploratorio de n-gramas. Posteriormente, se presentan los resultados del modelado de tópicos mediante LDA, incluyendo la selección del número óptimo de tópicos, su caracterización temática y su distribución en el corpus. En cuarto lugar, se analiza la evolución temporal de cada tópico mediante modelos ARIMA, con proyecciones hasta el año 2030. Finalmente, se exponen los resultados del análisis bibliométrico realizado con VOSviewer. Todo el procesamiento y análisis de datos fue desarrollado en lenguaje Python, y el código empleado se encuentra disponible en el siguiente repositorio de GitHub para su consulta y reproducción: (Masotti, 2026)

4.1. Recopilación y exploración de los datos

Como se detalla en la Sección 3.1, la fase inicial de recopilación de datos consistió en el diseño y ejecución de una consulta estructurada en la base de datos Scopus, seleccionada por su amplia cobertura internacional de publicaciones científicas revisadas por pares y su capacidad de exportación de metadatos completos. El objetivo de esta consulta fue recopilar artículos y revisiones científicas que abordaran la transición energética, la política energética y la descarbonización en el contexto español. Para construir la búsqueda, se combinaron términos clave relacionados con estas temáticas aplicando la siguiente expresión booleana sobre los campos de título, resumen y palabras clave:

`TITLE-ABS-KEY("energy transition" OR "energy policy" OR "decarbonization" OR "decarbonisation") AND TITLE-ABS-KEY(Spain OR Spanish) AND PUBYEAR >1999`

**AND (LIMIT-TO(DOCTYPE, “ar”) OR LIMIT-TO(DOCTYPE, “re”)) AND
(LIMIT-TO(LANGUAGE, “English”))**

Con el fin de garantizar la calidad y homogeneidad del corpus, se aplicaron filtros directamente en la plataforma Scopus. Se incluyeron únicamente documentos clasificados como artículos científicos y revisiones (*articles* y *reviews*), redactados en inglés y publicados desde el año 2000. El corpus bruto resultante comprendió 1.262 documentos. Los registros fueron exportados desde Scopus en dos archivos CSV independientes, cada uno para análisis distintos. El primero, destinado al análisis bibliométrico con VOSviewer, se exportó con la totalidad de los campos disponibles en la plataforma, incluyendo afiliaciones institucionales, referencias citadas, fuente de publicación y métricas de citación, necesarios para construir las redes de coocurrencia, acoplamiento bibliográfico, co-citación y coautoría. El segundo archivo, destinado al modelado de tópicos mediante LDA y al análisis de tendencias temporales con ARIMA, se exportó con un conjunto reducido de campos:

- Authors
- Title
- Year
- Abstract
- Author Keywords
- Index Keywords

La Figura 4.1 presenta la evolución anual del número de publicaciones incluidas en el corpus durante el período 2000–2026, el año 2026, separado por una línea discontinua roja, por ser un año incompleto. Se observa una tendencia ascendente a lo largo del periodo analizado, con una producción científica más reducida durante los primeros años de la década de 2000, que experimenta un crecimiento sostenido a partir de 2010 y se acelera de forma notable desde 2015. Este patrón refleja el auge internacional del debate académico sobre transición energética, impulsado en buena medida por hitos regulatorios clave como el Acuerdo de París (2015) y el Plan Nacional Integrado de Energía y Clima (PNIEC) aprobado en España en 2021. El año 2025 registra el mayor volumen de publicaciones del periodo, confirmando que el campo se encuentra en plena expansión en el momento del análisis.

4.2. Procesamiento y preparación del texto

Una vez consolidado el corpus destinado al modelado de tópicos, se procedió a un proceso de preprocesamiento textual con el fin de reducir la complejidad del lenguaje natural

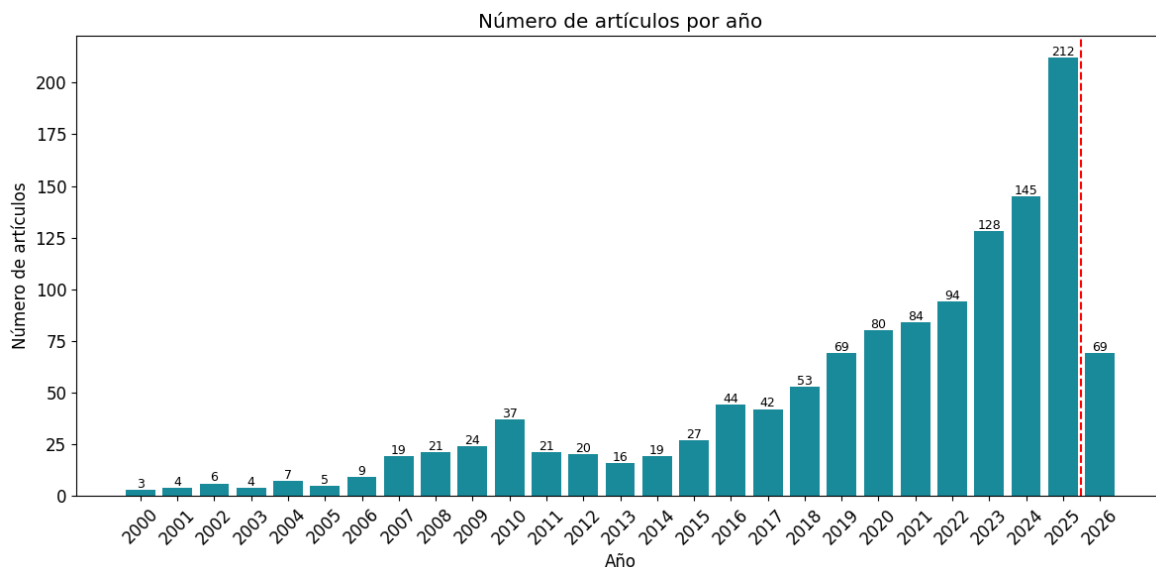


Figura 4.1: Evolución anual del número de publicaciones en el corpus (2000–2026). Elaboración propia

y optimizar la calidad de los datos para su posterior análisis. Este procedimiento se aplicó exclusivamente sobre los *abstracts* de los artículos con el propósito de unificar formas léxicas, eliminar elementos no sustantivos del texto y reducir ruido en el corpus, favoreciendo así una detección más precisa de coocurrencias y estructuras temáticas. En primer lugar, se eliminaron del dataframe las columnas no necesarias para el modelado de tópicos, así como los elementos que tenían el *abstract* vacío o duplicado, conservando únicamente el año de publicación y el texto del *abstract*. A continuación, se procedió a la eliminación del ruido editorial presente en los *abstracts* exportados desde Scopus, que habitualmente incluye textos de copyright al final de los resúmenes. Posteriormente, se llevó a cabo una fase de armonización léxica mediante expresiones regulares, unificando variantes compuestas de términos clave que el modelo LDA trataría como palabras independientes. Expresiones como *renewable energy*, *solar photovoltaic*, *wind energy*, se unificaron en una sola expresión, dando como resultado *renewable_energy*, *solar_photovoltaic* and *wind_energy*.

A continuación, se eliminaron las palabras vacías (*stopwords*) mediante un diccionario extendido que incorporó términos frecuentes en el discurso académico y en la propia query de búsqueda. En particular, se excluyeron palabras como *spain*, *spanish*, *study*, *paper*, *policy*, *energy*, *transition*, *result*, *model* o *assess*, entre otros. La exclusión de estos términos se justifica por su escaso poder discriminante dentro del corpus: varios de ellos formaban parte de las palabras clave utilizadas para recuperar los documentos en Scopus, por lo que su presencia era prácticamente transversal al conjunto de textos y, en consecuencia, aportaba una capacidad limitada para diferenciar tópicos. Del mismo modo, otros términos correspondían a vocabulario académico general de alta frecuencia, cuya recurrencia sistemática podía introducir ruido y sobrerrepresentar elementos contextuales o metodológicos, en detrimento

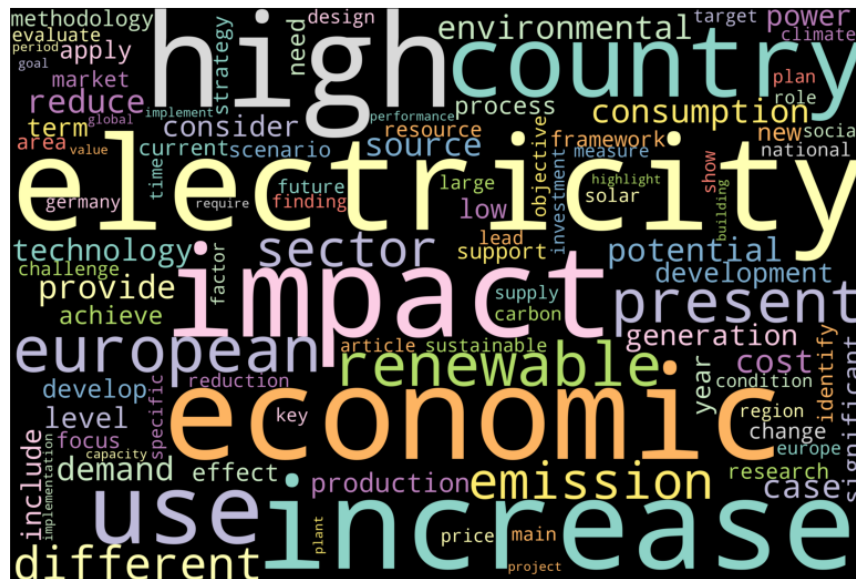
de conceptos temáticos más específicos. Finalmente, se aplicó la normalización del texto, que incluyó la conversión de todos los caracteres a minúsculas, la eliminación de signos de puntuación, caracteres especiales, dígitos y espacios redundantes, así como la eliminación de tildes mediante la librería `unidecode`. Sobre el texto normalizado se ejecutó la lematización. El resultado fue un corpus depurado de 1.040 documentos representados como listas de lemas, adecuado para su integración en el análisis exploratorio de n-gramas y el modelado de tópicos con LDA.

4.3. Análisis exploratorio

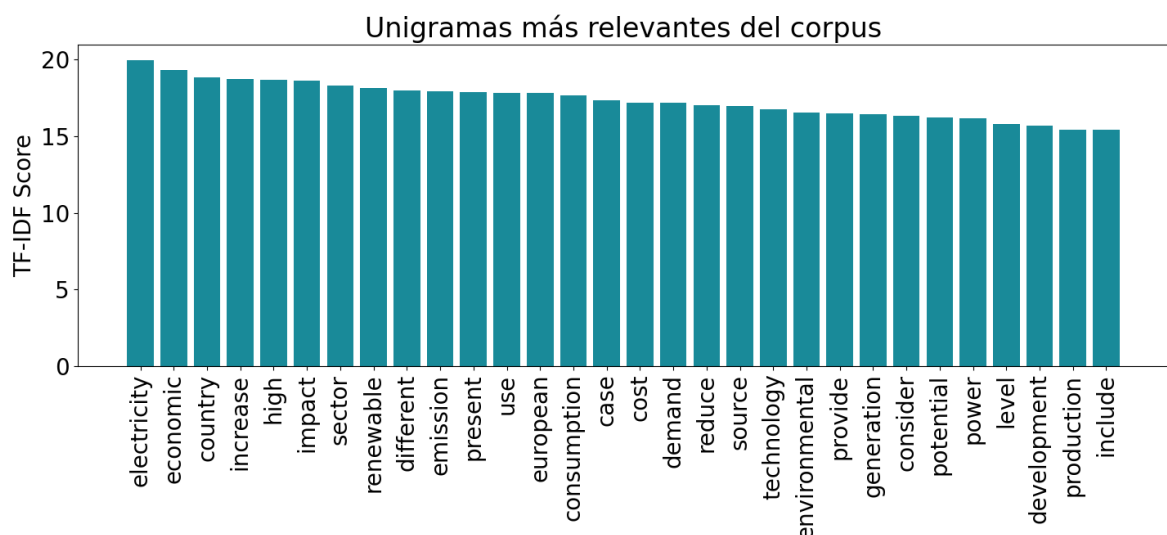
Antes de proceder al modelado de tópicos, se llevó a cabo un análisis exploratorio del contenido textual del corpus con el objetivo de identificar los términos y expresiones más representativos de la literatura sobre transición energética en España. Este análisis se estructuró en dos etapas: una visualización mediante nube de palabras y un análisis de n-gramas ponderado por TF-IDF. En primer lugar, se generó una nube de palabras a partir del corpus preprocesado, en la que el tamaño de cada término es proporcional a su frecuencia relativa en el conjunto de abstracts. La Figura 4.2 muestra la nube de palabras construida a partir de las frecuencias léxicas del corpus preprocesado. En términos relativos, los vocablos con mayor peso visual y, por tanto, con mayor frecuencia de aparición, son *electricity*, *economic*, *increase*, *impact*, *country*, *use*, *emission* y *renewable*. Esta distribución sugiere que el corpus se articula en torno a varios ejes temáticos dominantes, entre ellos una dimensión económica e institucional, reflejada en términos como *economic*, *country*, *european* y *sector*, una dimensión tecnológica y energética, asociada a vocablos como *electricity*, *renewable*, *technology*, *generation* y *consumption*, y una dimensión ambiental, evidenciada por la presencia destacada de términos como *emission*, *environmental* y *carbon*.

A continuación, se calcularon los valores TF-IDF para unigramas, bigramas y trigramas con el objetivo de identificar los términos más relevantes del corpus atendiendo no solo a su frecuencia absoluta, sino también a su capacidad discriminativa entre documentos. Esta métrica otorga mayor peso a aquellos términos que aportan contenido específico y diferenciador al corpus.

En línea con lo observado en la nube de palabras, la Figura 4.3 confirma la relevancia de términos como *electricity*, *economic*, *country*, *increase*, *impact*, *sector*, *renewable* y *emission*, lo que refuerza la idea de un corpus articulado en torno a dimensiones económicas, energéticas y ambientales. La coincidencia entre ambos resultados sugiere que estos vocablos no solo son frecuentes en el conjunto de documentos, sino que además poseen capacidad informativa para diferenciar contenidos dentro del corpus. No obstante, a diferencia de la nube de palabras, basada en frecuencias absolutas, el análisis mediante TF-IDF permite identificar aquellos términos cuya importancia deriva de su mayor poder discriminante, al concentrarse



de forma relativamente más intensa en determinados documentos y no únicamente de su recurrencia global. En este sentido, la presencia destacada de unigramas como *cost*, *demand*, *consumption*, *source*, *generation*, *technology*, *development* y *production* aporta una visión complementaria, al destacar componentes del debate más vinculados con dimensiones operativas, sectoriales y analíticas del estudio de la transición energética.



El análisis de bigramas, cuyos resultados se presentan en la Figura 4.4, permite capturar expresiones compuestas que aportan mayor precisión conceptual. Entre los bigramas con mayor puntuación TF-IDF destacan *european union*, *long term*, *life cycle*, *environmental impact* y *low carbon*, que apuntan hacia un discurso centrado en el marco regulatorio europeo,

la evaluación del ciclo de vida de las tecnologías energéticas y la transición hacia economías bajas en carbono. La presencia de *natural gas*, *electricity generation*, *power plant* y *feed tariff* refleja además el peso de los debates sobre el mix energético y los instrumentos de política de apoyo a las renovables, mientras que *self consumption* y *member state* destacan la dimensión tanto doméstica como comunitaria de la transición energética española.

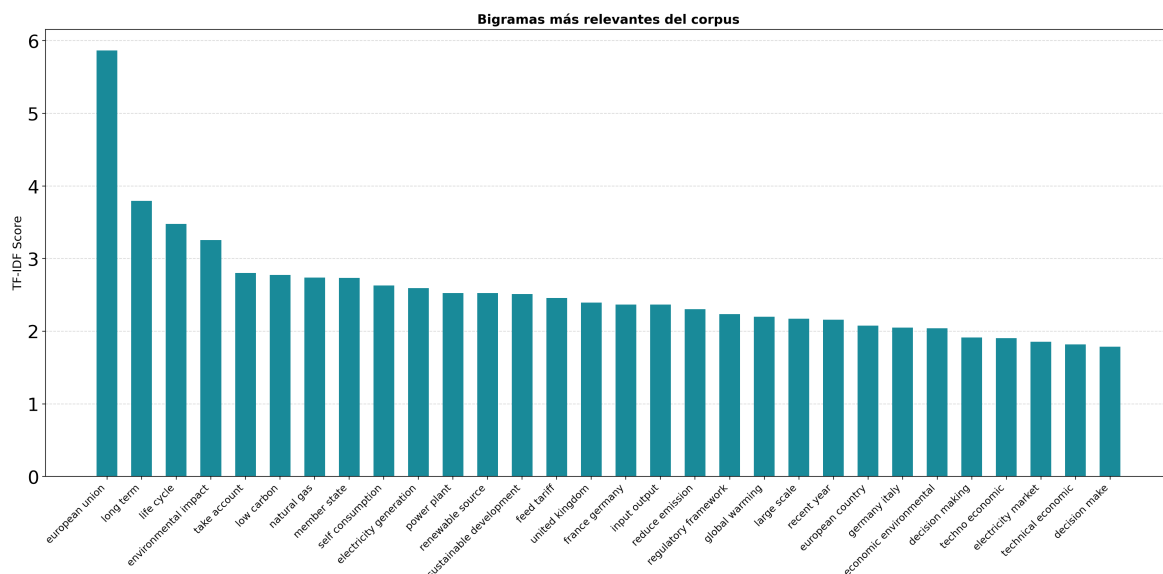


Figura 4.4: Bigramas más relevantes del corpus. Elaboración propia

Por último, el análisis de trigramas, presentado en la Figura 4.5, permite identificar expresiones técnicas de mayor granularidad conceptual. Los trigramas más destacados incluyen *play important role*, *play key role* y *play crucial role*, expresiones recurrentes en el discurso académico para enfatizar la relevancia de determinados factores en la transición energética. La presencia conjunta de *france germany italy* y *germany france italy* refleja la frecuente comparativa entre España y los principales países europeos en materia energética. Otros trigramas relevantes como *life cycle lca*, *carbon dioxide emission*, *feed tariff fit* o *net present value* apuntan hacia líneas de investigación centradas en la evaluación del impacto ambiental, los instrumentos regulatorios de apoyo a las renovables y el análisis de viabilidad económica de las tecnologías energéticas.

4.4. Modelado de tópicos (LDA)

Una vez finalizado el análisis exploratorio del corpus, se dio paso a la fase central del estudio, correspondiente al modelado de tópicos mediante el algoritmo LDA, de acuerdo con la metodología descrita en la Sección 3.4. Esta fase tiene como finalidad identificar de manera automatizada las principales líneas temáticas latentes en los *abstracts* académicos del corpus. Como paso previo al entrenamiento del modelo, se construyó el diccionario de tér-

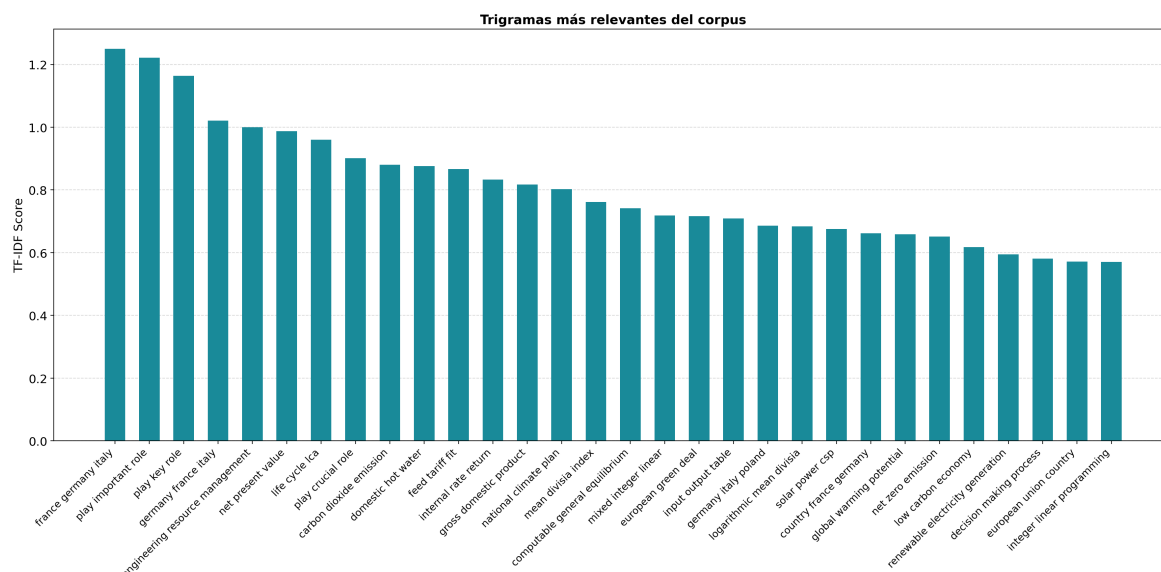


Figura 4.5: Trigramas más relevantes del corpus. Elaboración propia

minos a partir de las listas de lemas obtenidas en la fase de preprocesamiento, asignando un identificador único a cada término. Con el fin de filtrar términos poco informativos, se aplicaron dos umbrales sobre el diccionario: se eliminaron los términos que aparecían en menos de 5 documentos, por ser demasiado específicos, y los que aparecían en más del 75 % de los documentos, por ser excesivamente frecuentes y carecer de capacidad discriminativa. El diccionario resultante comprendió 2.313 términos únicos, a partir de los cuales se construyó la representación Bag-of-Words (BOW) de los 1.040 documentos del corpus.

La determinación del número óptimo de tópicos K representa uno de los aspectos más relevantes en la implementación del modelo LDA. Un valor de K demasiado bajo agrupa en un mismo tópico ideas conceptualmente distintas, mientras que un valor excesivamente alto tiende a fragmentar el contenido en tópicos difícilmente interpretables. Para abordar esta tarea de forma empírica, se entrenaron modelos LDA para valores de K comprendidos entre 2 y 8, calculando para cada uno el índice de coherencia C_v , métrica que evalúa la cohesión semántica interna de los tópicos generados y presenta una elevada correlación con evaluaciones humanas de interpretabilidad (Röder et al., 2015). La Figura 4.6 presenta la evolución del índice de coherencia para cada valor de K evaluado.

La coherencia alcanzó su valor máximo en la solución de $K = 3$, con un $C_v = 0.3758$, lo que sugiere que esta configuración proporciona la estructura temática más consistente entre los modelos considerados. Como complemento al análisis de coherencia, se empleó la herramienta pyLDavis para visualizar la distancia intertópica entre los tres tópicos identificados. La Figura 4.7 presenta el mapa de distancias intertópicas resultante, en el que cada tópico se representa como una circunferencia cuyo tamaño es proporcional a su peso relativo en el corpus. En dicha visualización se observa una separación clara entre los tres tópicos, lo que confirma que el modelo ha identificado estructuras temáticas diferenciadas y no redundantes.

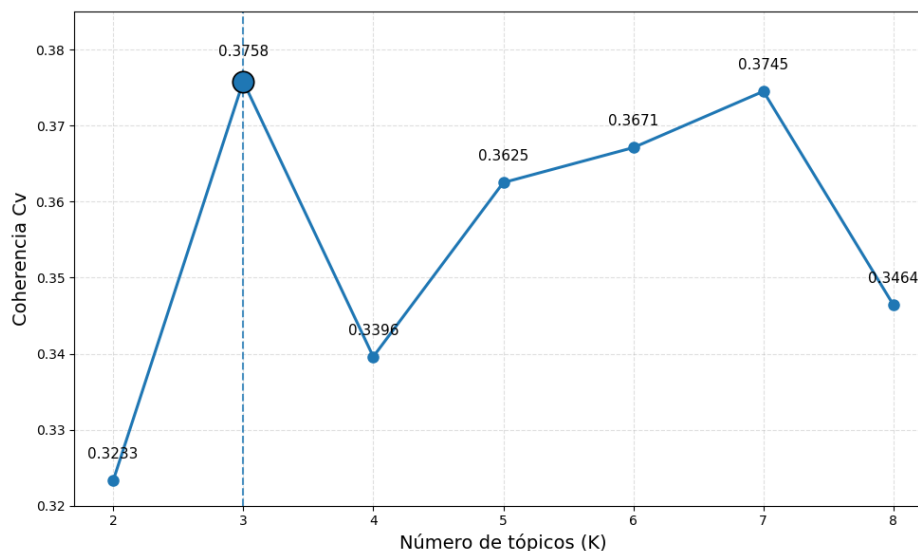


Figura 4.6: Evolución del índice de coherencia Cv en función del número de tópicos K. Elaboración propia

Según los datos de pyLDavis, el Tópico 1 es el más frecuente con un peso del 39,2 % del corpus, seguido del Tópico 2 con un 38,8 % y del Tópico 0 con un 22,0 %.

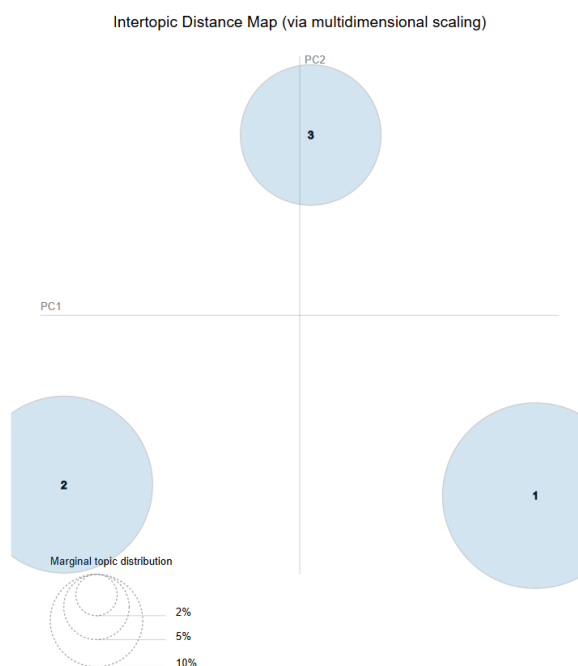


Figura 4.7: Mapa de distancia intertópica del modelo LDA con K=3 (pyLDavis). Elaboración propia

Nota: La numeración de tópicos mostrada en la figura corresponde al orden asignado automáticamente por pyLDavis, basado en la frecuencia relativa de cada tópico en el corpus. Esta numeración difiere de la empleada en el texto y en la Tabla 4.1, donde se mantiene el orden original del modelo LDA. Concretamente, el Tópico 1 del modelo aparece como Tópico 1 en la visualización, el Tópico 2

como Tópico 2, y el Tópico 0, al ser el menos frecuente, aparece como Tópico 3 en la visualización.

A partir de las diez palabras con mayor peso probabilístico en cada tópico y del análisis de bigramas y trigramas por tópico mediante TF-IDF, se procedió a interpretar el contenido temático de cada uno de los tres tópicos identificados. Para la asignación de etiquetas interpretativas a cada tópico, se empleó un modelo de lenguaje de gran escala (LLM), concretamente Claude, siguiendo la aproximación propuesta por Kozlowski, Pradier, y Benz (2024) y Rijcken et al. (2023), quienes demuestran el potencial de los LLMs para interpretar y etiquetar los tópicos generados mediante LDA, al ser capaces de integrar simultáneamente las palabras clave, los n-gramas y el contexto temático del corpus para producir descripciones coherentes y representativas que complementan el juicio del investigador.

El **Tópico 0**, etiquetado como “*Mercado eléctrico, inversión renovable y marcos regulatorios*” (visualizado como Tópico 3 en la Figura 4.7), representa el 22,0 % del corpus. Este tópico recoge la línea de investigación de carácter más técnico-regulatorio, centrada en el análisis del despliegue de tecnologías de generación eléctrica renovable en España y los instrumentos de política energética que lo condicionan. Las palabras con mayor peso, como *electricity*, *power*, *generation*, *framework* y *cost*, apuntan a estudios que evalúan la viabilidad técnica y económica de tecnologías renovables en el marco de mecanismos regulatorios concretos. Los bigramas más relevantes del tópico refuerzan esta interpretación: *self consumption*, *feed tariff*, *regulatory framework*, *electricity market* y *levelized cost lcoe* son expresiones recurrentes en estudios que analizan el impacto de los sistemas de retribución regulada, el autoconsumo o el acceso al mercado eléctrico sobre la inversión renovable. La presencia de trigramas como *feed tariff fit*, *wholesale electricity market* o *red electrica españa* confirma el anclaje de este tópico en el contexto institucional español.

El **Tópico 1**, etiquetado como “*Análisis de ciclo de vida, huella de carbono y descarbonización sectorial*”, es el más representado el 39,2 % del corpus. Agrupa investigaciones orientadas a cuantificar el impacto ambiental de tecnologías y estrategias energéticas, con un enfoque marcadamente analítico y sectorial. Las palabras dominantes *emission*, *reduce*, *demand* y *potential* reflejan estudios que buscan medir y proyectar el potencial de reducción de emisiones de distintas opciones de descarbonización. El bigrama más relevante del tópico es *life cycle*, con diferencia respecto al resto, lo que evidencia el peso de los estudios de análisis de ciclo de vida (ACV) en esta línea de investigación. Le siguen *environmental impact*, *global warming*, *reduce emission*, *carbon footprint* y *green hydrogen*, que completan una imagen coherente de estudios centrados en la huella de carbono de distintos vectores energéticos y procesos industriales. A nivel sectorial, los trigramas *domestic hot water*, *residential building stock* y *heating cool demand* señalan al sector residencial como uno de los focos principales de análisis, mientras que *wastewater treatment plant* apunta al sector industrial.

El **Tópico 2**, etiquetado como “*Gobernanza energética europea, análisis macroeconómico y política climática*”, agrupa 403 artículos (38,8 % del corpus). Este tópico recoge la literatura de carácter comparado e institucional, centrada en el posicionamiento de España en el marco de la gobernanza energética europea y en los efectos macroeconómicos de las políticas climáticas. Las palabras dominantes *country*, *european*, *germany*, *development* y *effect* reflejan estudios que analizan la transición energética desde una perspectiva supranacional, comparando a España con otras economías europeas de referencia. Los bigramas *european union*, *member state*, *france germany* y *economic growth* apuntan a esta dimensión comparada, que los trigramas confirman con expresiones recurrentes como *france*

germany italy o *sweden united kingdom*. Destacan también metodologías econométricas específicas como el índice de Divisia (*mean divisia index, logarithmic mean divisia*), los modelos de equilibrio general computable (*computable general equilibrium*) y los análisis input-output (*input output table*), lo que evidencia el rigor cuantitativo de esta línea de investigación. Los trigramas *european green deal, kyoto protocol* y *merit order effect* anclan el tópico en el marco normativo de la política climática europea.

La Tabla 4.1 presenta un resumen comparativo de los tres tópicos identificados, incluyendo su etiqueta interpretativa, el número de artículos asignados, las palabras clave y los bigramas más representativos.

Tabla 4.1: Resumen de los tópicos identificados mediante LDA. Fuente: elaboración propia.

Tópico	Etiqueta	N (%)	Palabras clave	Bigramas clave
T0	Mercado eléctrico, inversión renovable y marcos regulatorios	229 (22,0 %)	<i>case, electricity, power, generation, renewable, cost, framework, technology</i>	<i>self consumption, feed tariff, electricity market, power plant, regulatory framework</i>
T1	Análisis de ciclo de vida, huella de carbono y descarbonización sectorial	408 (39,2 %)	<i>emission, reduce, economic, demand, potential, cost, electricity</i>	<i>life cycle, environmental impact, low carbon, reduce emission, net zero emission</i>
T2	Gobernanza energética europea, análisis macroeconómico y política climática	403 (38,8 %)	<i>country, european, impact, germany, sector, development, economic, effect</i>	<i>european union, member state, economic growth, european green deal, kyoto protocol</i>

La Figura 4.8 muestra la distribución de los tópicos dominantes en el corpus y permite apreciar una estructura relativamente concentrada en torno a dos grandes líneas de investigación. En términos cuantitativos, el Tópico 1 constituye la categoría más representada, con 408 artículos (39,2 %), seguido muy de cerca por el Tópico 2, con 403 artículos (38,8 %). En contraste, el Tópico 0 reúne 229 documentos (22,0 %), situándose claramente por debajo de los otros dos. Esta distribución indica que la literatura sobre transición energética en España se organiza principalmente en torno a dos ejes de similar peso relativo, por un lado el análisis de ciclo de vida, la huella de carbono y la descarbonización sectorial, y por otro la gobernanza energética europea, el análisis macroeconómico y la política climática. La diferencia mínima entre los Tópicos 1 y 2, de apenas cinco documentos, sugiere un equilibrio notable entre una orientación más aplicada y ambiental, centrada en la cuantificación de impactos y estrategias de reducción de emisiones, y una orientación más institucional, enfocada en la dimensión macroeconómica y regulatoria de la transición energética.

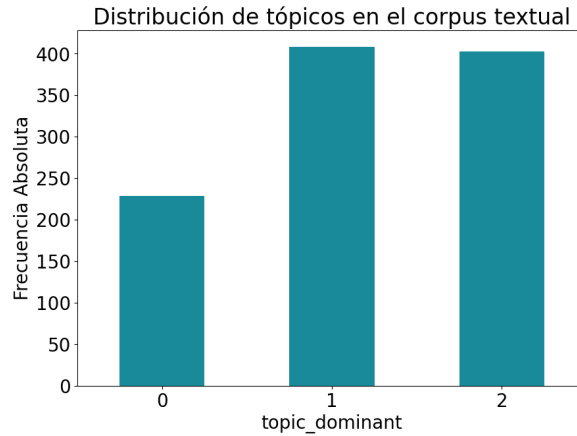


Figura 4.8: Distribución de tópicos dominantes en el corpus (N=1.040). Elaboración propia

4.5. Análisis de tendencias temporales con ARIMA

Con el objetivo de proyectar la evolución futura de cada línea temática identificada mediante LDA, se construyó una serie temporal con el número de publicaciones anuales asignadas a cada tópico dominante y se ajustaron modelos ARIMA individuales para cada una de ellas, siguiendo la metodología descrita en la Sección 3.5. El análisis se realizó sobre el período 2000–2025, excluyendo el año 2026 por estar incompleto en el momento de la extracción, lo que habría introducido un sesgo a la baja en la tendencia. El horizonte de proyección se fijó hasta el año 2030. Tras excluir el año 2026, el corpus quedó reducido a 988 artículos, distribuidos de la siguiente forma: 221 para el Tópico 0, 383 para el Tópico 1 y 384 para el Tópico 2.

Como paso previo al ajuste de los modelos, se evaluó la estacionariedad de cada serie temporal mediante el test de Dickey-Fuller Aumentado (ADF), cuya hipótesis nula plantea la existencia de una raíz unitaria, es decir, la no estacionariedad de la serie. Dado que las series de publicaciones científicas agregadas por año pueden presentar tendencia, se aplicó el contraste de forma secuencial sobre la serie original y, cuando no era posible rechazar la hipótesis nula ($p > 0,05$), sobre su primera y segunda diferencia. El contraste se aplicó de forma iterativa siguiendo el procedimiento descrito en la metodología: en primer lugar sobre la serie original, si el p-valor era superior a 0,05 (indicando que no se puede rechazar la hipótesis nula de no estacionariedad), se calculaba la primera diferencia de la serie ($\Delta y_t = y_t - y_{t-1}$) y se repetía el test, de seguir siendo no estacionaria, se aplicaba una segunda diferenciación ($\Delta^2 y_t$) y se efectuaba el contraste por tercera vez. Los resultados del proceso iterativo se recogen en la Tabla 4.2.

Los resultados del test ADF muestran comportamientos diferentes entre los tres tópicos. El Tópico 0 alcanzó la estacionariedad tras una única diferenciación ($d = 1$, $p < 0,05$), mientras que el Tópico 2 requirió dos diferenciaciones ($d=2$, $p < 0,05$). En cambio, el Tópico 1 no presentó evidencia suficiente de estacionariedad ni en niveles ni tras la primera y segunda diferenciación. Este resultado sugiere que la serie asociada a este tópico presenta una dinámica temporal más compleja, por lo que cualquier proyección derivada de su modelización debe interpretarse con especial cautela. En este caso, los resultados tienen un carácter más exploratorio que confirmatorio y reflejan principalmente una

Tabla 4.2: Resultados del test ADF iterativo por tópico. Fuente: elaboración propia.

Tópico	Serie testada	p-valor ADF	Criterio	Resultado
T0	Original	0,9989	$> 0,05$	No estacionaria
T0	1ª diferencia	0,000	$< 0,05$	Estacionaria ($d = 1$)
T1	Original	0,9981	$> 0,05$	No estacionaria
T1	1ª diferencia	0,9970	$> 0,05$	No estacionaria
T1	2ª diferencia	0,645	$> 0,05$	No estacionaria
T2	Original	1,000	$> 0,05$	No estacionaria
T2	1ª diferencia	0,9948	$> 0,05$	No estacionaria
T2	2ª diferencia	0,0220	$< 0,05$	Estacionaria ($d = 2$)

extrapolación de la evolución observada.

Tras haber evaluado la estacionariedad de cada serie, la selección de los parámetros p y q se llevó a cabo mediante un procedimiento automatizado de búsqueda y comparación entre especificaciones alternativas, utilizando el criterio de información de Akaike (AIC) como criterio de selección y restringiendo la exploración a un rango de valores $p \in [1,3]$ y $q \in [0,2]$. Este procedimiento evaluó distintas combinaciones posibles de los componentes autorregresivo, de diferenciación y de media móvil, seleccionando en cada caso la especificación con mejor ajuste relativo según el AIC. Para el parámetro d , se seleccionó el número de diferenciaciones logradas anteriormente, asignando $d=2$ para el tópico 1 a pesar de su no estacionariedad. Los modelos finalmente ajustados, junto con sus parámetros y proyecciones para 2030, se recogen en la Tabla 4.3.

Tabla 4.3: Resumen de modelos ARIMA ajustados por tópico. Fuente: elaboración propia.

Tópico	ARIMA(p,d,q)	AIC	Proyección 2030
T0	(1,1,0)	144,94	28 arts.
T1	(1,2,0)	161,02	125 arts. (exploratorio)
T2	(2,2,1)	162,00	89 arts.

La Figura 4.9 presenta la evolución histórica (2000–2025) y la proyección ARIMA (2026–2030) de los tres tópicos de forma comparada, con el intervalo de confianza al 95 % representado como banda sombreada.

El **Tópico 2** (Gobernanza energética europea, análisis macroeconómico comparado y política climática) muestra también una proyección de crecimiento sostenido, con un modelo ARIMA(2,2,1) que estima 89 publicaciones anuales en 2030. La estructura más compleja del modelo, con dos términos autorregresivos y un término de media móvil, refleja una dinámica de crecimiento con cierta inercia y corrección de corto plazo, característica de una línea de investigación consolidada que responde a estímulos regulatorios e institucionales como la aprobación del European Green Deal o las revisiones del marco energético europeo.

El **Tópico 0** (Mercado eléctrico, inversión renovable y marcos regulatorios) es el que presenta la proyección más moderada, con un modelo ARIMA(1,1,0) que estima 28 publicaciones anuales en 2030, prácticamente sin crecimiento respecto a los niveles de 2025. Esta trayectoria plana no debe interpretarse como una pérdida de relevancia de esta línea de investigación, sino como una limitación



Figura 4.9: Evolución histórica y proyección del número de publicaciones por tópico (2000–2030). Elaboración propia.

estadística inherente al tamaño reducido de la serie: con solo 221 artículos distribuidos en 26 años, y con los primeros quince años de la serie presentando valores prácticamente nulos, el modelo ARIMA dispone de escasa información para proyectar una tendencia creciente. El orden de diferenciación simple ($d = 1$) indica que la serie crece a una tasa aproximadamente constante, lo que se traduce en una proyección lineal plana cuando el crecimiento reciente no es lo suficientemente pronunciado para que el modelo lo extrapole.

El **Tópico 1** (Análisis de ciclo de vida, huella de carbono y descarbonización sectorial) es el que registra la proyección más dinámica en términos numéricos, con un modelo ARIMA(1,2,0) que anticipa 125,6 publicaciones anuales en 2030. No obstante, esta proyección debe interpretarse con cautela. Como se ha señalado, la serie del Tópico 1 no alcanzó la estacionariedad en ninguna de las tres iteraciones del test ADF, lo que indica que el supuesto estadístico que sustenta la validez del modelo ARIMA no queda satisfecho. Por tanto, la cifra proyectada tiene valor orientativo y exploratorio, no puede considerarse una predicción estadísticamente robusta en el mismo sentido que las de los Tópicos 0 y 2.

4.6. Análisis bibliométrico con VOSviewer

Los resultados del análisis bibliométrico realizado con VOSviewer se presentan en cuatro dimensiones complementarias: coocurrencia de palabras clave de autor, coocurrencia de la totalidad de palabras clave, acoplamiento bibliográfico por revistas, co-citación de referencias y redes de coautoría por países. Estos análisis permiten contextualizar los resultados del modelado de tópicos respecto a un estudio de la literatura, identificando los núcleos conceptuales del campo, las comunidades de conocimiento dominantes y los patrones de colaboración científica internacional.

En una primera etapa se construyó un mapa de coocurrencia a partir de las palabras clave asignadas por los propios autores de los artículos (*author keywords*), aplicando un umbral mínimo de cinco

ocurrencias. La Figura 4.10 presenta la red resultante, en la que los nodos más prominentes son *energy transition*, *renewable energy*, *energy policy* y *energy efficiency*, que actúan como ejes articuladores del campo y presentan conexiones con prácticamente todos los demás clústeres identificados. A partir de esta red se distinguen varios clústeres temáticos bien diferenciados. El clúster rojo agrupa términos relacionados con la descarbonización del sector edificatorio y las tecnologías de almacenamiento térmico, incluyendo *decarbonization*, *buildings* y *carbon footprint*, lo que se alinea con el Tópico 1 identificado en el LDA. El clúster amarillo-verde recoge la dimensión social y de justicia energética, con términos como *energy poverty*, *just energy transition*, *energy vulnerability* y *energy community*, reflejando una línea de investigación en auge sobre las implicaciones distributivas de la transición energética. El clúster azul-morado central agrupa términos de mercado y regulación como *energy policy*, *sustainability*, *energy efficiency* y *economic growth*, mientras que el clúster inferior recoge la dimensión técnico-económica del sistema eléctrico, con *electricity generation*, *renewable electricity*, *feed-in tariff* y *power markets*. La estructura de esta red valida los tres tópicos identificados mediante LDA, mostrando que los mismos grandes ejes temáticos emergen de forma independiente tanto en el análisis textual de abstracts como en el análisis de coocurrencia de palabras clave.

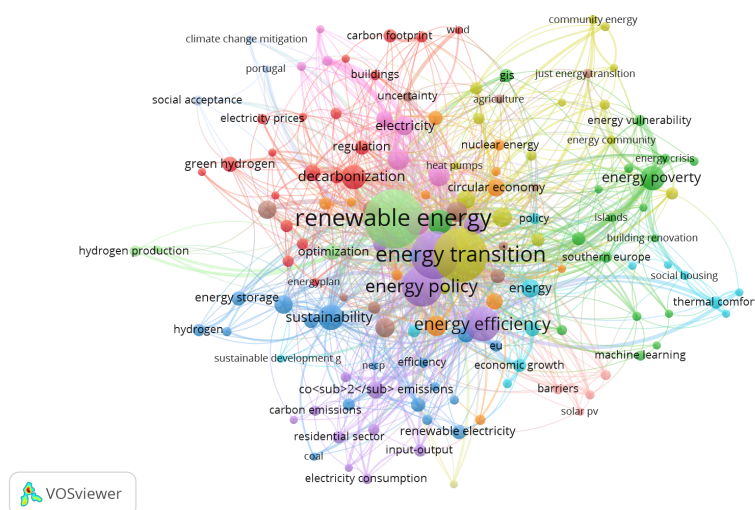


Figura 4.10: Red de coocurrencia de palabras clave de autor (umbral mínimo: 5 ocurrencias). Elaboración propia con VOSviewer

Con el objetivo de ampliar el alcance de este análisis, se generó un segundo mapa de coocurrencia incorporando la totalidad de las palabras clave disponibles, incluyendo las *index keywords* generadas automáticamente por Scopus, con un umbral mínimo de diez ocurrencias. La Figura 4.11 presenta esta red ampliada, en la que *energy policy* y *spain* emergen como los nodos centrales de mayor tamaño, confirmando el foco geográfico y temático del corpus. En este mapa se identifican con mayor precisión clústeres complementarios: el clúster rojo incorpora términos como *solar energy*, *investments*, *regulatory frameworks* y *cost reduction*, añadiendo profundidad a la dimensión tecnológica y regulatoria. El clúster verde recoge la dimensión comparada europea con términos como *germany*, *italy*, *sweden*, *economic development* y *environmental policy*. El clúster azul profundiza en la dimensión de mercado con *energy market*, *wind power*, *feed-in tariff* y *renewable resource*, y el clúster amarillo

refuerza la línea de eficiencia energética en edificación con *energy efficiency*, *energy use* y *carbon*.

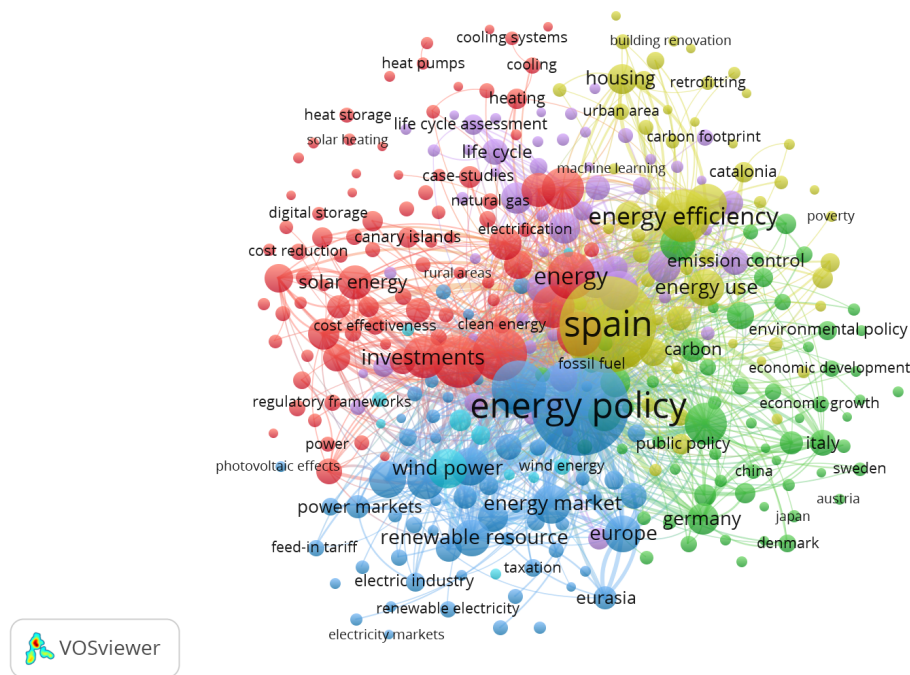


Figura 4.11: Red de coocurrencia de la totalidad de palabras clave (umbral mínimo: 10 ocurrencias). Elaboración propia con VOSviewer

El análisis de acoplamiento bibliográfico, realizado tomando como unidad de análisis las revistas científicas con un umbral mínimo de cinco documentos, permite identificar qué publicaciones concentran las principales comunidades de conocimiento del campo. La Figura 4.12 presenta la red resultante, en la que *Energy Policy* emerge como el nodo dominante con diferencia, seguido de *Renewable and Sustainable* y *Renewable Energy* como segundo clúster de peso, y *Sustainability* junto con *Energies* y *Applied Sciences* como tercer grupo. Un cuarto clúster claramente diferenciado agrupa *Energy and Buildings* y *Building and Environment*, revistas especializadas en la eficiencia energética del sector edificatorio. Destaca también la posición de *Energy Research and Social Science* como nodo periférico pero diferenciado, representando la dimensión social y de gobernanza de la investigación en transición energética. Esta estructura de revistas es coherente con los tópicos identificados: las revistas del clúster rojo (*Energies*, *Applied Sciences*) se asocian a estudios técnicos de carácter más aplicado, las del clúster verde a estudios de energías renovables, las del clúster azul a eficiencia energética en edificación y *Energy Research and Social Science* a la dimensión de gobernanza y política energética.

El análisis de co-citación, con un umbral mínimo de cinco citas, permite identificar los trabajos que constituyen la base común de la literatura sobre transición energética en España. La Figura 4.13 presenta la red resultante, en la que se distinguen varios clústeres de referencias bien diferenciados por línea temática. El nodo de mayor tamaño, corresponde a (Boardman, 1991), obra que introdujo la primera definición de la pobreza energética y estableció el vínculo entre eficiencia energética residencial y exclusión social. En torno a este nodo se articula un clúster azul de trabajos sobre pobreza

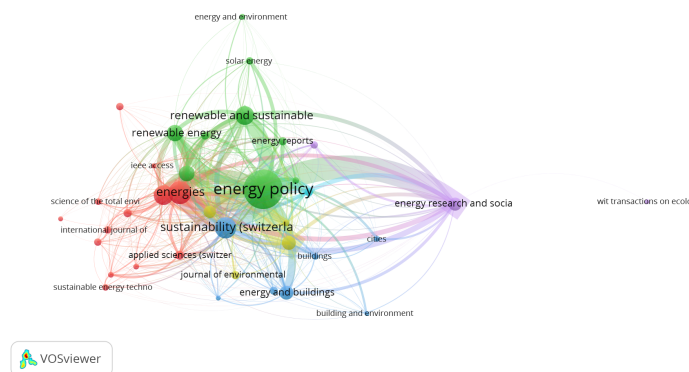
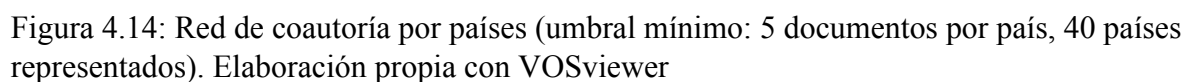
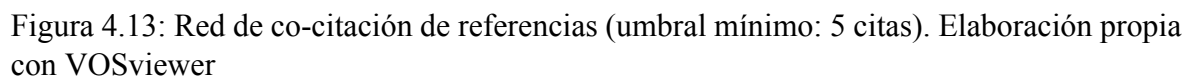


Figura 4.12: Red de acoplamiento bibliográfico por revistas (umbral mínimo: 5 documentos por revista). Elaboración propia con VOSviewer

y vulnerabilidad energética, con referencias como (Bouzarovski y Simcock, 2017), que introduce el concepto de justicia espacial en la comprensión de la pobreza energética, y (Thomson y Snell, 2013), que cuantifica la pobreza energética en los países de la Unión Europea, lo que confirma la relevancia de esta dimensión social en la literatura analizada y su conexión con los tópicos de gobernanza y transición justa identificados en el LDA. El clúster verde agrupa referencias de carácter técnico centradas en energías renovables y eficiencia energética, con (Capellán-Pérez, Campos-Celador, y Terés-Zubiaga, 2018) como nodo principal, trabajo que analiza las cooperativas de energía renovable como instrumento de transición en España. El clúster rojo recoge referencias orientadas a la política energética y la regulación fotovoltaica española, con *The European Green Deal* como nodo central, rodeado de trabajos sobre tecnología fotovoltaica como (Bernal-Agustín y Dufo-López, 2006), que analiza los sistemas fotovoltaicos conectados a la red eléctrica española desde una perspectiva económica y ambiental, y (Prol y Steininger, 2017), que estudia la regulación del autoconsumo fotovoltaico en España y su impacto sobre la rentabilidad de los inversores. Finalmente, el clúster amarillo agrupa referencias normativas y de planificación energética específicamente española, incluyendo la *Directiva (UE) 2019/944* sobre el mercado interior de la electricidad. La estructura de esta red confirma que la literatura sobre transición energética española se apoya en una base intelectual diversa y bien estructurada, con comunidades de referencia diferenciadas por línea temática.

El análisis de coautoría por países, con un umbral mínimo de cinco documentos, permitió identificar 40 países de los 92 presentes en el corpus. La Figura 4.14 presenta la red resultante, en la que España ocupa el nodo central de mayor tamaño, lo que es coherente con el foco geográfico del corpus. A partir de este nodo central se identifican varios clústeres de colaboración internacional. El clúster morado articula el eje de colaboración hispanoamericano, con conexiones de España hacia Chile, México, Perú, Ecuador y Colombia. El clúster verde agrupa el núcleo europeo continental, con Alemania, Países Bajos, Italia, Portugal, Francia, Bélgica, Noruega y Finlandia como países más conectados entre sí, formando una red densa de colaboración intraeuropea. El clúster azul recoge el grupo anglosajón, con Reino Unido, Irlanda, Canadá, Suecia y Suiza. En el extremo derecho del mapa aparecen clústeres más periféricos que incluyen economías emergentes y asiáticas como Polonia, Turquía, China,



Japón e India, con conexiones más tenues hacia el núcleo europeo. La conexión más gruesa del mapa, aparte de las europeas, es la que une España con Estados Unidos, lo que refleja la relevancia de la colaboración transatlántica en este campo. En conjunto, la red muestra que la investigación sobre transición energética española tiene una dimensión predominantemente europea, con una proyección iberoamericana notable y una integración creciente con economías emergentes.

Capítulo 5

Conclusiones y trabajo futuro

La transición energética se ha consolidado como uno de los ejes centrales del debate político, económico y científico de las últimas décadas. El compromiso de la Unión Europea con la neutralidad climática para 2050, materializado en instrumentos como el European Green Deal o el Plan Nacional Integrado de Energía y Clima español, ha generado un volumen creciente de investigación académica orientada a analizar los retos, las oportunidades y las implicaciones del proceso de descarbonización desde perspectivas muy diversas. En este contexto, comprender cómo se organiza y evoluciona el conocimiento científico sobre la transición energética en España resulta especialmente relevante, tanto para orientar futuras agendas de investigación como para informar el diseño de políticas públicas. El presente Trabajo de Fin de Grado se enmarca en esta necesidad, con el objetivo de identificar y analizar las principales tendencias de investigación en este campo a partir de un corpus de publicaciones académicas recientes.

La revisión de la literatura realizada en el Capítulo 2 permitió identificar las metodologías más adecuadas para abordar los objetivos del estudio. Se constató la utilidad creciente de las técnicas de minería de textos y modelado de tópicos como herramientas para mapear el conocimiento científico en campos con un elevado volumen de producción académica. En particular, el algoritmo LDA se consolidó como la técnica de referencia en la literatura para la identificación de estructuras temáticas latentes en grandes corpus textuales, dadas sus ventajas en términos de interpretabilidad. Asimismo, se estableció el modelo ARIMA para proyectar la evolución temporal de tendencias de investigación, siguiendo la aproximación de (Zou et al., 2024), y de herramientas de visualización bibliométrica como VOSviewer para complementar el análisis temático con una perspectiva estructural sobre las redes de conocimiento y colaboración del campo.

Este estudio adoptó una metodología estructurada en seis etapas principales. La primera consistió en la recopilación de 1.262 artículos y revisiones desde Scopus mediante una consulta booleana específica para la transición energética española. En la segunda etapa, los abstracts fueron sometidos a un proceso de preprocesamiento textual que incluyó la eliminación de ruido editorial, la armonización léxica, la eliminación de stopwords y la lematización, dejando un corpus compuesto por 1.040 abstracts. A continuación, se realizó un análisis exploratorio de n-gramas mediante TF-IDF para identificar los términos y expresiones más relevantes del corpus. La fase central del análisis consistió en la implementación del algoritmo LDA sobre el corpus preprocesado, seleccionando el número óptimo de

tópicos mediante el índice de coherencia Cv. Finalmente, se ajustaron modelos ARIMA para proyectar la evolución temporal de cada tópico y se realizó un análisis bibliométrico con VOSviewer centrado en la coocurrencia de palabras clave, el acoplamiento bibliográfico por revistas, la co-citación de referencias y las redes de coautoría internacional.

Los análisis realizados permiten identificar tres grandes líneas de investigación que concentran el interés académico sobre la transición energética española. En primer lugar, destaca una línea orientada al análisis de ciclo de vida, la huella de carbono y la descarbonización sectorial, que constituye la más amplia del corpus y refleja el interés por cuantificar el impacto ambiental de tecnologías y estrategias energéticas en sectores como la edificación y la industria. En segundo lugar, se identifica una línea centrada en la gobernanza energética europea, el análisis macroeconómico y la política climática, que agrupa estudios que sitúan a España en el contexto del marco normativo europeo y emplean metodologías econométricas avanzadas para medir el impacto de las políticas de descarbonización. En tercer lugar, emerge una línea más técnico-regulatoria, centrada en el mercado eléctrico, la inversión renovable y los marcos normativos que condicionan el despliegue de tecnologías como la solar fotovoltaica, la eólica o el autoconsumo en España. La validación cruzada de estos tópicos con los mapas de coocurrencia de palabras clave de VOSviewer refuerza la coherencia y robustez de los resultados.

En relación con las proyecciones obtenidas mediante modelos ARIMA, los resultados asociados al Tópico 0 y al Tópico 2 pueden considerarse metodológicamente más sólidos, dado que ambas series alcanzaron la estacionariedad tras el proceso de diferenciación aplicado. En ambos casos, las estimaciones sugieren una trayectoria de crecimiento sostenido de la producción académica hasta el año 2030. Por el contrario, la serie correspondiente al Tópico 1 no logró alcanzar la estacionariedad tras las diferenciaciones realizadas, por lo que la proyección derivada de su modelización debe interpretarse con mayor cautela. En este sentido, la estimación de 125 publicaciones anuales en 2030 apunta a una tendencia de crecimiento muy marcada, coherente con la creciente atención académica hacia los procesos de descarbonización sectorial, aunque su carácter debe entenderse como fundamentalmente exploratorio y no como una predicción estadísticamente robusta en términos comparables a los de los otros dos tópicos.

El análisis bibliométrico con VOSviewer complementa estos resultados al ofrecer una visión estructural del campo. Los mapas de coocurrencia de palabras clave revelan que los conceptos de *energy transition*, *renewable energy*, *energy policy* y *energy efficiency* actúan como ejes articuladores de toda la literatura, con clústeres bien diferenciados que recogen la dimensión tecnológica, la dimensión social y de justicia energética, y la dimensión de mercado y regulación. A nivel de colaboración científica, el análisis de coautoría por países confirma que España ocupa el nodo central de la red internacional, coherente con el foco geográfico del corpus, con una dimensión europea muy consolidada y una notable aportación hispanoamericana, lo que refleja tanto la integración del sistema académico español en el espacio europeo de investigación como sus vínculos con América Latina.

A partir de los resultados obtenidos, se abren diversas líneas de trabajo futuro que podrían enriquecer el alcance del presente estudio. En primer lugar, la incorporación de otras bases de datos como Web of Science, junto con Scopus, permitiría ampliar la cobertura del corpus y contrastar la robustez de los tópicos identificados en distintas fuentes. En segundo lugar, la construcción de series temporales

les mensuales mejoraría la fiabilidad de las proyecciones ARIMA y permitiría, en particular, abordar de forma más adecuada series como la del Tópico 1, cuya estructura temporal compleja no queda resuelta mediante diferenciación estándar sobre datos anuales. Por último, la extensión del análisis al conjunto de la literatura europea sobre transición energética, sin restricción geográfica a España, permitiría contextualizar el papel de la investigación española en el panorama científico internacional e identificar las áreas en las que la producción académica española presenta mayor especificidad o liderazgo relativo.

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que *ChatGPT* u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Nicola Masotti Villanueva, estudiante de E2-Analytics al presentar mi Trabajo Fin de Grado titulado “Tendencias de Investigación sobre la Transición Energética en España”, declaro que he utilizado la herramienta de Inteligencia Artificial Generativa *ChatGPT* u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. Brainstorming de ideas de investigación: Utilizado para idear y esbozar posibles áreas de investigación.
2. Crítico: Para encontrar contra-argumentos a una tesis específica que pretendo defender.
3. Referencias: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
4. Metodólogo: Para descubrir métodos aplicables a problemas específicos de investigación.
5. Interpretador de código: Para realizar análisis de datos preliminares.
6. Estudios multidisciplinarios: Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
7. Constructor de plantillas: Para diseñar formatos específicos para secciones del trabajo.
8. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
9. Generador previo de diagramas de flujo y contenido: Para esbozar diagramas iniciales.
10. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
11. Generador de datos sintéticos de prueba: Para la creación de conjuntos de datos ficticios.

12. Generador de problemas de ejemplo: Para ilustrar conceptos y técnicas.
13. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
14. Generador de encuestas: Para diseñar cuestionarios preliminares.
15. Traductor: Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado *ChatGPT* u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: Abril 2026

Firma: Nicola Masotti Villanueva

Referencias

- Ajibade, S.-S. M., Bekun, F. V., Adedoyin, F. F., Gyamfi, B. A., y Adediran, A. O. (2023). Machine learning applications in renewable energy (mlare) research: a publication trend and bibliometric analysis study (2012–2021). *Clean Technologies*, 5(2), 497–517.
- Alizadeh, M. I., Moghaddam, M. P., Amjady, N., Siano, P., y Sheikh-El-Eslami, M. K. (2016). Flexibility in future power systems with high renewable penetration: A review. *Renewable and Sustainable Energy Reviews*, 57, 1186–1193.
- Alvial-Palavicino, C. (2015). The future as practice. a framework to understand anticipation in science and technology. *Tecnoscienza–Italian Journal of Science & Technology Studies*, 6(2), 135–172.
- Aytac, S. (2025). Trends and insights in wind energy research: A bibliometric analysis. *Issues in Science and Technology Librarianship*(111).
- Balteanu, I., y Viani, F. (2023). La dependencia energética de la unión europea y de españa. *Boletín Económico*(2023/T3).
- Bernal-Agustín, J. L., y Dufo-López, R. (2006). Economical and environmental analysis of grid connected photovoltaic systems in spain. *Renewable energy*, 31(8), 1107–1128.
- Boardman, B. (1991). Fuel poverty is different. *Policy Studies*, 12(4), 30–41.
- Bouzarovski, S., y Simcock, N. (2017). Spatializing energy justice. *Energy Policy*, 107, 640–648.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., y Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Campbell, J. C., Hindle, A., y Stroulia, E. (2015). Latent dirichlet allocation: extracting topics from software engineering data. En *The art and science of analyzing software data* (pp. 139–159). Elsevier.
- Capellán-Pérez, I., Campos-Celador, Á., y Terés-Zubiaga, J. (2018). Renewable energy cooperatives as an instrument towards the energy transition in spain. *Energy Policy*, 123, 215–229.
- Das, M., Alphonse, P., y cols. (2023). A comparative study on tf-idf feature weighting method and its analysis using unstructured dataset. *arXiv preprint arXiv:2308.04037*.
- Diaconescu, M., Marinas, L. E., Marinoiu, A. M., Popescu, M.-F., y Diaconescu, M. (2024).

- Towards renewable energy transition: Insights from bibliometric analysis on scholar discourse to policy actions. *Energies*, 17(18), 4719.
- Ember. (2024). *Global electricity review 2024*. Descargado de <https://ember-energy.org/app/uploads/2024/05/Report-Global-Electricity-Review-2024.pdf> (Accedido: 2025-10-28)
- Epdata. (2023). *El cambio climático, en datos y gráficos*. Descargado de <https://www.epdata.es/datos/cambio-climatico-datos-graficos/447> (Accedido: 2025-10-28)
- Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., y Lachhab, A. (2018). Forecasting of demand using arima model. *International journal of engineering business management*, 10, 1847979018808673.
- Geels, F. W., Sovacool, B. K., Schwanen, T., y Sorrell, S. (2017). The socio-technical dynamics of low-carbon transitions. *Joule*, 1(3), 463–479.
- Harichandan, S., Kar, S. K., Bansal, R., Mishra, S. K., Balathanigaimani, M. S., y Dash, M. (2022). Energy transition research: A bibliometric mapping of current findings and direction for future research. *Cleaner Production Letters*, 3, 100026.
- IPCC. (2023). *Climate change 2023: Synthesis report of the ipcc sixth assessment report*. <https://www.ipcc.ch/report/sixth-assessment-report-cycle/>. (Intergovernmental Panel on Climate Change, Geneva)
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., y Zhao, L. (2019). Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia tools and applications*, 78(11), 15169–15211.
- Kern, F., y Rogge, K. S. (2016). The pace of governed energy transitions: Agency, international dynamics and the global paris agreement accelerating decarbonisation processes? *Energy Research & Social Science*, 22, 13–17.
- Kozłowski, D., Pradier, C., y Benz, P. (2024). Generative ai for automatic topic labelling. *arXiv preprint arXiv:2408.07003*.
- Kut, P., y Pietrucha-Urbanik, K. (2023). Bibliometric analysis of renewable energy research on the example of the two european countries: insights, challenges, and future prospects. *Energies*, 17(1), 176.
- Masotti, N. (2026). *TFG-BUSINESS-ANALYTICS: Repositorio de código y análisis de datos*. <https://github.com/nmasotti/TFG-BUSINESS-ANALYTICS>. GitHub.
- Ministerio para la Transición Ecológica y el Reto Demográfico. (2023). *Plan nacional integrado de energía y clima (pniec) 2023-2030*. <https://www.miteco.gob.es/es/energia/estrategia-normativa/pniec-23-30.html>. (Gobierno de España)
- Morte, I. B. B., Ofélia de Queiroz, F. A., Morgado, C. R., y de Medeiros, J. L. (2023). Electrification and decarbonization: a critical review of interconnected sectors, policies, and sustainable development goals. *Energy Storage and Saving*, 2(4), 615–630.
- Obaideen, K., Olabi, A. G., Al Swailmeen, Y., Shehata, N., Abdelkareem, M. A., Alami,

- A. H., ... Sayed, E. T. (2023). Solar energy: Applications, trends analysis, bibliometric analysis and research contribution to sustainable development goals (sdgs). *Sustainability*, 15(2), 1418.
- Our World in Data. (2024). *Global temperature anomaly (1850–2025)*. https://ourworldindata.org/explorers/climate-change?time=1850.2025&Metric=Temperature+anomaly&Long-run+series=false&country=ATA~Gulkana+Glacier~Lemon+Creek+Glacier~OWID_NAM~South+Cascade+Glacier~Wolverine+Glacier~Hawaii~Arctic+Ocean~OWID_WRL. (Data explorer: Climate Change, University of Oxford)
- Our World in Data, y. . . u. . h. n. . A., title = Modern Renewable Electricity Generation by Source. (s.f.).
- Prol, J. L., y Steininger, K. W. (2017). Photovoltaic self-consumption regulation in Spain: Profitability analysis and alternative regulation schemes. *Energy Policy*, 108, 742–754.
- Rijcken, E., Scheepers, F., Zervanou, K., Spruit, M., Mosteiro, P., y Kaymak, U. (2023). Towards interpreting topic models with ChatGPT. En *Proceedings of the 20th world congress of the international fuzzy systems association (ifsa 2023)*. Daegu, Republic of Korea.
- Röder, M., Both, A., y Hinneburg, A. (2015). Exploring the space of topic coherence measures. En *Proceedings of the eighth ACM international conference on web search and data mining* (pp. 399–408).
- Said, S. E., y Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3), 599–607.
- Thomson, H., y Snell, C. (2013). Quantifying the prevalence of fuel poverty across the European Union. *Energy policy*, 52, 563–572.
- UNFCCC. (2015). *The Paris Agreement. United Nations Framework Convention on Climate Change*. <https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement>. (Accessed: 2025-10-28)
- UNFCCC. (2023). *Cop28 final statements. United Nations Framework Convention on Climate Change*. <https://unfccc.int/cop28>. (Accessed: 2025-10-28)
- ur Rehman, A., Sanjari, M. J., Elavarasan, R. M., y Jamal, T. (2026). Sustainability-aligned pathways for energy transition: A review of low-carbon energy network solutions. *Renewable and Sustainable Energy Reviews*, 226, 116428.
- Van Eck, N., y Waltman, L. (2010). Software survey: Vosviewer, a computer program for bibliometric mapping. *scientometrics*, 84(2), 523–538.
- Zou, T., Guo, P., Li, F., y Wu, Q. (2024). Research topic identification and trend prediction of China's energy policy: A combined LDA-ARIMA approach. *Renewable Energy*, 220, 119619.