

## Capítulo tercero

### **Análisis de la geopolítica mundial mediante inteligencia artificial (IA) y big data**

*Álvaro Ortiz, Tomasa Rodrigo*

#### **Resumen**

La digitalización de la información y el desarrollo de la inteligencia artificial está propiciando un cambio sin precedentes en la disponibilidad de nueva información de manera no-estructurada (texto, vídeo...). Gracias al desarrollo de los modelos de procesamiento de lenguaje natural (NLP), el texto se ha convertido en una de las principales fuentes de información, y la posibilidad de trasladar «Texto a Números» se está convirtiendo en una poderosa herramienta de análisis en el campo de la ciencia política y las relaciones internacionales. En este artículo repasamos los distintos modelos y bases de datos que extraen y procesan texto de los medios de comunicación locales e internacionales para su uso en el análisis de las relaciones internacionales, incluyendo el seguimiento en tiempo real y la previsión de conflictos. Entre ellos, presentamos varios ejemplos desarrollados por el sistema de seguimiento geopolítico desarrollado en BBVA Research, incluyendo el análisis de flujos migratorios a Europa tras la crisis en Siria, la guerra entre Rusia y Ucrania, el conflicto armado Israel-Hamas y el impacto de los cuellos de botella y tensiones entre China y Taiwán en el sector estratégico de los semiconductores.

## Palabras clave

*Big data, Inteligencia artificial, Procesamiento del lenguaje natural, Geopolítica, Modelos de alerta temprana.*

## **Analysis of global geopolitics through artificial intelligence (AI) and big data**

### Abstract

*The digitisation of information and the development of artificial intelligence is bringing about an unprecedented change in the availability of new information in an unstructured form (text, video...). Thanks to the development of natural language processing (NLP) models, text has become one of the main sources of information, and the possibility of translating "Text to Numbers" is becoming a powerful tool for analysis in the field of political science and international relations. In this article we review the various models and databases that extract and process text from local and international media for use in the analysis of international relations, including real-time monitoring and conflict forecasting. Among them, we present several examples developed by the geopolitical monitoring system developed at BBVA Research, including the analysis of migration flows to Europe after the Syrian crisis, the war between Russia and Ukraine, the armed conflict between Israel and Hamas and the impact of shortgates and tensions between China and Taiwan in the strategic semiconductor sector.*

### Keywords

*Big data, Inteligencia artificial, Natural Language Processing, Geopolitics, Early warning models.*

## 1. La revolución en la inteligencia artificial y los modelos de procesamiento del lenguaje natural

El rápido desarrollo del proceso de digitalización y los crecientes avances en el campo de la inteligencia artificial (IA) están propiciando una revolución en el análisis de las ciencias sociales. Lo que conocemos como *big data* es el resultado de diferentes fuerzas que han venido operando simultáneamente desde hace tiempo y que se refuerzan entre sí: el rápido desarrollo de la capacidad de procesamiento de los ordenadores, el creciente volumen de información y el veloz desarrollo de los algoritmos.

La primera de ellas, la capacidad de procesar información de las computadoras, ha sido clave en el desarrollo de la IA. El rápido desarrollo de los ordenadores en las últimas décadas ha sido un fenómeno asombroso en el que el número de transistores, tal y como predijo Gordon Moore (1965: 114) hace seis décadas, ha venido casi doblándose cada dos años aproximadamente. Esto ha supuesto un aumento exponencial en el rendimiento y la capacidad de procesamiento, que ha venido acompañada de un descenso de tamaño considerable de los mismos (desde los primeros que ocupaban habitaciones enteras hasta los modernos dispositivos móviles que caben en la palma de nuestra mano), así como una disminución notable en su coste.

La mayor potencia de los ordenadores ha facilitado también el desarrollo de la gestión, procesamiento y análisis de cantidades masivas de datos a una velocidad antes inimaginable. Esto ha sido fundamental para la evolución y expansión de internet, que ha actuado como catalizador de la digitalización de la información. Entre otras cosas, internet ha supuesto una digitalización casi total de la información existente, facilitando el acceso a la misma, su distribución y su compartición a través de las redes globales.

El desarrollo de internet no solo ha mejorado la eficiencia, en términos de almacenamiento y distribución de la información, sino que también ha abierto nuevas vías para el análisis de datos como, por ejemplo, las tecnologías de computación en la nube. Estas han facilitado el acceso a una capacidad de procesamiento y almacenamiento de datos que hace años era simplemente inalcanzable. Todo ello, junto al rápido descenso de los costes de los ordenadores y dispositivos, ha permitido una democratización del acceso a tecnologías avanzadas sin precedentes.

Por último, el desarrollo de nuevos algoritmos está contribuyendo también al rápido avance de la inteligencia artificial (IA). Los algoritmos proporcionan las reglas y procedimientos que guían el aprendizaje y la toma de decisiones de las computadoras. A medida que evolucionan, también lo hace su capacidad de realizar tareas cada vez más complejas. Los algoritmos de aprendizaje profundo, por ejemplo, han permitido avances significativos en áreas como el reconocimiento de voz, de imágenes, la conducción autónoma... Los recientes desarrollos en los grandes modelos del lenguaje (LLM) e inteligencia generativa aventuran avances sin precedentes en el ámbito de análisis de las ciencias sociales, incluyendo el análisis geopolítico.

Todo ello forma parte de lo que, en este artículo, entendemos como *big data*. Esto es, el acceso al procesamiento de datos masivos gracias al desarrollo de la capacidad de procesamiento de los ordenadores y a los nuevos algoritmos que nos permite convertir esta información, muchas veces no estructurada, en datos que podemos utilizar para analizar cuestiones complejas en las ciencias sociales.

Dentro del amplio espectro de información no estructurada con potencial para ser utilizada en el campo del análisis de las ciencias sociales (imágenes, vídeo, audio etc.), y en particular de las relaciones internacionales, la información proveniente de los textos ha jugado un papel relevante. Aquí, el desarrollo de un conjunto de algoritmos conocidos como modelos de procesamiento del lenguaje natural (NLP) han permitido extraer texto en forma no estructurada y convertirlo en información estructurada numérica susceptible de ser procesada para el análisis de las relaciones internacionales.

El avance en los modelos de procesamiento del lenguaje natural ha sido también relativamente rápido y ha ido evolucionando en complejidad, desde el análisis de palabras individuales y su distribución en temas, a encontrar sentido a las palabras según el contexto que los acompaña; hasta inferir o predecir texto como en el caso de los nuevos modelos de procesamiento del lenguaje natural.

Los primeros modelos se centraron básicamente en palabras o *tokens* individuales. Estos primeros modelos se caracterizan por analizar la frecuencia de las palabras en las búsquedas booleanas. Aunque simples, estos algoritmos nos permiten, por ejemplo, analizar cuestiones como el grado de incertidumbre, simplemente

analizando búsquedas de palabras como «Incertidumbre» y el conjunto del campo de palabras asociadas al fenómeno que queremos aplicar. En el caso de la incertidumbre de política económica, un buen ejemplo de ello es el trabajo de Baker, Bloom y Davis (2016), mientras que, en el ámbito de las relaciones internacionales, Caldara y Iacovello (2022) han desarrollado índices de incertidumbre geopolítica al estilo de los que describiremos en los siguientes apartados.

Otros algoritmos sencillos hacen uso de técnicas asistidas por diccionarios específicos para interpretar el sentimiento de los textos. Estos algoritmos supervisados pueden ser utilizados para clasificar el carácter positivo o negativo de los textos, la polarización o armonía, etc... Existen múltiples diccionarios, desde aquellos más generales a diccionarios con terminologías específicas (economía, relaciones internacionales, medicina, legales...) desarrollados por expertos. Dentro del campo de las relaciones internacionales, algunos artículos se han centrado en la polarización de los discursos políticos (Gennaro y Ash, 2021), mientras otros autores han mostrado, por ejemplo, cómo la combinación de técnicas y análisis del sentimiento pueden ser utilizado para diseñar un modelo de radicalización en las redes sociales (Bermingham *et al.*, 2009).

Con el tiempo, los algoritmos se sofisticaron y comenzaron a tener en cuenta el contexto global del texto a analizar. Nacieron así los modelos dinámicos de temas desarrollados por David Blei (2003), que resumen el texto en estructuras semánticas o temas latentes. Los analistas políticos o de relaciones internacionales han analizado varios textos como discursos, debates, iniciativas legales, contenido de los medios, etc. Son modelos probabilísticos cuyo resultado lo forman grupos de palabras con mayor o menor probabilidad de pertenecer a un grupo. La mayoría de esos modelos son de carácter no supervisado, pues es el analista el que debe etiquetar manualmente el conjunto de palabras incluidas en un grupo. Existen numerosos trabajos que utilizan los modelos temáticos dentro del ámbito de las relaciones internacionales o la ciencia política. Entre ellos cabe destacar los trabajos de Mueller y Rauh (2018), que utilizan la variación en los temas en periódicos nacionales para predecir conflictos armados y violencia política en distintos países, o Martin y McCrain (2019), que utilizan un modelo de temas para analizar cómo los cambios en la propiedad conducen a cambios en la cobertura de la política nacional o sesgos de orientación.

Un importante avance en los modelos de procesamiento del lenguaje natural se produjo con el desarrollo de los modelos de incrustación de palabras (*word embeddings*), que son el germen a los modelos generativos del lenguaje natural. Estos modelos están basados en la hipótesis de la distribución del lenguaje<sup>1</sup>. Esta hipótesis sostiene que las palabras que aparecen en contextos similares tienden a tener significados similares y es normalmente el contexto lo que proporciona a las palabras su significado. Esta propiedad se ha convertido en un concepto fundamental en el campo de la lingüística computacional y el procesamiento del lenguaje natural (NLP).

Una de sus aplicaciones más populares ha sido para el diseño de diccionarios y clasificación de sentimientos o temas de una manera semiautomática. De acuerdo con varios autores (Rodríguez y Stewart, 2023), los resultados en el campo de la ciencia política son satisfactorios y, en términos generales, estos modelos obtienen resultados relativamente buenos, a veces incluso mejores, que los que proporcionan los codificadores o etiquetadores humanos.

Si bien los modelos de incrustación de palabras avanzaron significativamente en la comprensión y análisis de las palabras con significados similares (sinónimos), tuvimos que esperar al desarrollo de los «Transformadores» (Vaswani *et al.*, 2017) y los grandes modelos de lenguaje para poder distinguir significados distintos para una misma palabra (polisemia). Para ello, estos modelos necesitan «prestar atención»<sup>2</sup>, lo que en términos computacionales significa poder valorar la importancia de cada vector de palabras incrustadas en cada caso. Estos modelos utilizan estructuras similares a las de las redes neuronales, capaces de valorar cuáles son las palabras clave del contexto para averiguar la siguiente palabra o aquella que queremos interpretar<sup>3</sup>.

---

<sup>1</sup> Este concepto se basa en el trabajo del lingüista Zellig Harris, quien en la década de 1950 propuso que la similitud semántica entre palabras podía ser determinada por sus patrones de ocurrencia conjunta en el lenguaje. La idea es que el significado de una palabra se construye y se refleja a través de su uso, y, por lo tanto, analizando los patrones de cómo las palabras se distribuyen en grandes cantidades de texto, se puede inferir su significado y sus relaciones semánticas con otras palabras.

<sup>2</sup> El título en inglés del famoso artículo de Vaswani (2017) es «Todo lo que necesitas es atención» (*Attention is all you need*).

<sup>3</sup> Un buen ejemplo de ello es el desarrollo y aplicación de los *Transformers* a la codificación automática de los modelos de alerta temprana de eventos de conflicto como «Polecat» que serán explicados en la próxima sección.

Los grandes modelos de lenguaje (LLM) han comenzado ya a utilizarse con éxito en el ámbito de las relaciones internacionales. Estos modelos han entrado recientemente en el debate público sobre inteligencia artificial, ya que posibilitan el uso de una nueva metodología fácil de usar para el estudio del lenguaje. Básicamente, los modelos LLM aprovechan técnicas de aprendizaje profundo, recursos computacionales a gran escala y enormes cantidades de datos de entrenamiento para generar textos coherentes y contextualmente relevantes. Su principal diferencia respecto a los anteriores modelos es que su objetivo es la predicción.

Mientras que los modelos LLM como GPT-3, BERT, LLaMA, BARD... han sido ampliamente utilizados en muchas aplicaciones, el reciente lanzamiento público de ChatGPT desarrollado por la empresa americana OpenAI ha abierto un debate sobre los posibles usos y abusos de los modelos de ciencia política y relaciones internacionales<sup>4</sup>.

## 2. Análisis geopolítico con modelos de procesamiento del lenguaje natural: clasificación de eventos y sistemas de alerta temprana

La utilización de la inteligencia artificial para el campo de las relaciones internacionales está íntimamente ligado al desarrollo de los modelos de alerta temprana para conflictos. Este es un campo que ha evolucionado significativamente a lo largo del tiempo, integrando avances en tecnología, análisis de datos y teoría de relaciones internacionales.

El concepto de sistemas de alerta temprana ganó relevancia durante la Guerra Fría, enfocándose principalmente en amenazas militares. Estos sistemas fueron diseñados inicialmente para detectar y poder responder con anticipación a ataques nucleares.

Uno de los primeros esfuerzos académicos en alerta temprana de conflictos fue el *Conflict and Peace Data Bank* (COPDAB), desarrollado por Edward Azar (1980), a comienzos de la década de 1970 y finalmente publicado en 1980. Su objetivo era crear una base de datos completa de eventos de conflicto y cooperación internacionales y domésticos. El proyecto recopilaba datos de informes de noticias y cobertura mediática, abarcando una amplia gama

---

<sup>4</sup> Para un resumen del estado de la cuestión véase Linnegard (2023).

de eventos políticos desde disputas diplomáticas hasta conflictos armados. COPDAB empleaba su propio sistema de codificación que clasificaba las interacciones entre actores (como Estados, organizaciones internacionales y grupos no estatales) en una escala de conflicto a cooperación.

A COPDAB le siguieron los proyectos como WEIS (McLelland, 1976) y KEDs (Schrodt, 1994), desarrollado por Philip Schrodt en la década de 1990. Ambos fueron significativos en la automatización del código de datos de eventos a partir de fuentes de noticias. A comienzos del siglo XXI se desarrollaron varios proyectos como el Programa de Datos de Conflictos de Uppsala (UCDP), desarrollado por Sundberg y Melander (2013), el sistema de alerta temprana integrado de crisis (ICEWS) financiado por DARPA en 2007 y desarrollado por la empresa Lockheed Martin, bajo la dirección de Paul O'Brien (2010) y que ha sido mejorado y rebautizado como POLECAT (Haltermann *et al.*, 2023) y actualmente financiado por la CIA, y el proyecto de la Base de Datos Global de Eventos, Lenguaje y Tono (GDELT), iniciado en 2011 y publicado en 2013 por Leetaru y Schrodt (2013), que utilizaremos posteriormente como base al sistema de seguimiento geopolítico desarrollado por BBVA Research y que utilizaremos en el resto del artículo.

Los programas UCDP (Programa de Datos de Conflictos de Uppsala), GDELT (Base de Datos Global de Eventos, Lenguaje y Tono), ICEWS (Sistema Integrado de Alerta Temprana de Crisis) y POLECAT son proyectos significativos en el campo de datos de eventos globales y análisis de conflictos, pero difieren en su enfoque, metodología y aplicaciones. Entre las principales diferencias:

- El programa de datos de conflictos de la Universidad de Uppsala (UCDP), se centra en datos de conflictos armados y eventos similares. Proporciona información detallada sobre guerras, conflictos no estatales y violencia unilateral y combina la información de expertos con la de los medios de comunicación para la recopilación y categorización de datos. Los datos son de frecuencia anual y se han utilizado ampliamente en investigaciones académicas sobre conflictos y relaciones internacionales. Mueller y Rauh (2018) utilizan por ejemplo la base de datos de UCDP en su trabajo sobre la relevancia de la información de los medios para la predicción de conflictos.
- La base de datos Global de Eventos, Lenguaje y Tono (GDELT), es una vasta base de datos que rastrea eventos, lenguaje y tono de los medios de comunicación globales. Cubre una

amplia gama de eventos, incluidos, pero no limitados a conflictos. Como en el caso de ICEWS utiliza técnicas computacionales avanzadas para procesar datos de medios a gran escala, incluyendo las técnicas de los modelos de procesamiento del lenguaje natural y análisis de big data. GDELT utiliza un sistema de codificación de eventos basado en el esquema CAMEO (del inglés Conflict and Mediation Event Observations), desarrollado por Gerner et al. (2002), para categorizar los eventos políticos. Sus aplicaciones son diversas, desde investigación académica en ciencias sociales hasta usos prácticos en periodismo, inteligencia empresarial y análisis de políticas.

- El Sistema Integrado de Alerta Temprana de Crisis (ICEWS), está diseñado como un sistema de alerta temprana para predecir crisis políticas. Cubre una gama más amplia de eventos políticos, no limitados a conflictos armados. Combina la recopilación de datos automatizada (usando PNL y aprendizaje automático) con análisis de expertos y, como GDELT, utiliza el esquema CAMEO para categorizar los eventos. Utiliza la base de datos para orientar a los policymaker en las políticas y para el desarrollo de modelos para anticipar y responder a crisis internacionales.
- El Sistema de clasificación de Eventos Políticos, Atributos y Tipos (POLECAT), ha reemplazado recientemente a ICEWS (Haltermann et al., 2023). Utiliza la ontología PLOVER (Political Language Ontology for Verifiable Event Records) para la codificación de eventos, en lugar de CAMEO. De acuerdo con sus autores, esta ontología es más flexible y adaptable, permitiendo una clasificación más detallada y variada de los eventos. POLECAT integra tecnologías avanzadas de procesamiento del lenguaje natural, incluyendo la base de los más avanzados modelos de lenguaje (Transformers) al análisis automatizado de eventos, buscando mejorar la precisión y la cobertura de los datos. Aunque también es útil en la formalización de políticas y análisis de Defensa, tiene un enfoque más amplio que puede ser de interés para investigadores académicos, organizaciones internacionales y otros usuarios interesados en el análisis político global.

La información subyacente a los diferentes sistemas es diferente. En el caso de UCDP se utilizan varios criterios para la inclusión de eventos, normalmente conflicto y violencia, junto a una amplia gama de fuentes de información proveniente de informes de organizaciones internacionales, gobiernos, ONGs, medios de comunicación y estudios académicos. Los datos son sometidos a

un riguroso proceso de verificación para asegurar su precisión. La base de datos se actualiza anualmente, proporcionando una perspectiva actualizada de los conflictos en curso y recientes.

En el caso de GDELT, ICEWS y POLECAT se utiliza el sistema de codificación de eventos de CAMEO. Este sistema está diseñado para clasificar y analizar eventos políticos y particularmente a los relacionados con conflictos y mediación. Utiliza una estructura jerárquica para clasificar eventos, dividida en categorías y subcategorías, que describen la naturaleza de la acción política o del conflicto. Las categorías de eventos en CAMEO varían desde niveles muy generales hasta más específicos. No solo identifica eventos, sino también actores y acciones. Los actores involucrados en eventos pueden ser estados, organizaciones internacionales, grupos no estatales, líderes políticos, grupos terroristas, etc.... mientras que las acciones que realizan estos actores son codificadas según la naturaleza del evento, como declaraciones diplomáticas, acciones militares, protestas, atentados, acuerdos, etc.

El Sistema de codificación de CAMEO está asociado a la escala de Goldstein (2002) y suelen utilizarse conjuntamente. El proceso de codificación en CAMEO se realiza en distintas fases. En una primera fase se analiza el texto, que normalmente proviene de noticias. En ellas, el sistema de codificación identifica los actores y las acciones descritas en el texto asignándoles el código correspondiente en la clasificación, mediante la utilización de técnicas de procesamiento de lenguaje natural. Esta clasificación tiene una estructura jerárquica, desde categorías generales hasta acciones más específicas. En su estructura más amplia consta de cuatro grupos o categorías que podemos ordenar desde cooperación material, cooperación verbal, conflicto verbal y conflicto material. Cada una de estas categorías consta de diferentes eventos, que, a su vez, pueden descomponerse en subcategorías con eventos más detallados.

El gráfico 1 nos ayuda a entender la clasificación. En la primera escala vemos las cuatro categorías ordenadas de mayor a menor relevancia en términos de violencia. En el ejemplo podemos apreciar cómo dentro del gran grupo de violencia verbal encontramos la agrupación de protesta. Esta se puede subdividir en varias categorías de protesta, ordenadas según su intensidad. Una vez los eventos y acciones están clasificadas, la escala de Goldstein asigna valores numéricos a diferentes tipos de eventos políticos, reflejando el potencial de cada tipo de evento para contribuir a la

estabilidad o inestabilidad política. Los valores en la escala varían desde mayores puntuaciones en las acciones cooperativas materiales y verbales (con valores positivos) a acciones con valores negativos asociados al conflicto verbal y al material. El conflicto material se asocia a los valores más negativos y, en su caso, por la utilización de armas de destrucción masiva.

Agrupaciones	Eventos	Codigo	Escala de Goldstein
Cooperacion Material	Establecer Cooperación Material (6)	6	6
	Proporcionar Ayuda (7)	7	7
	Producir (26)	8	5
	Investigar (6)	9	-2
Cooperación Verbal	Realizar Anuncio Público (11)	1	0
	Apelar (28)	2	3
	Expresar intencion de Cooperar (29)	3	4
	Consultar (8)	4	1
	Establecer Cooperación Diplomática (9)	5	3.5
Conflicto Verbal	Demandar (27)	10	-5
	Desaprobar (13)	11	-2
	Rechazar (27)	12	-4
	Amenazar (23)	13	-6
	Protestar (27)	14	-6.5
Conflicto Material	Exhibir Postura de Fuerza (6)	15	-7.2
	Reducir relaciones (14)	16	-4
	Coaccionar (13)	17	-7
	Asaltar (14)	18	-9
	Luchar (8)	19	-10
	Usar Armas de Destrucción Masiva (8)	20	-10

**Gráfico 1. Clasificación de CAMEO y escala de Goldstein. Fuente: GDELT y Goldstein**

Más allá de clasificar los eventos, poder evaluar la intensidad de estos y geolocalizarlos con precisión, las técnicas de procesamiento del lenguaje natural permiten desarrollar multitud de análisis en diferentes campos de las ciencias sociales. Entre ellas, nos permite evaluar el sentimiento con que se producen los eventos, con la ayuda de diccionarios especializados o la descomposición de grandes corpus de información de texto en los temas más relevantes. Los nuevos modelos de procesamiento de lenguaje natural ofrecen oportunidades que son difíciles de imaginar.

### 3. De texto a números: seguimiento y análisis geopolítico en BBVA Research

En la sección anterior hemos mostrado cómo podemos utilizar las técnicas de procesamiento de lenguaje natural para examinar y analizar multitud de eventos. Esto nos ha ayudado a entender

mejor una situación geopolítica cambiante y cada vez más compleja. En BBVA Research hemos estado trabajando durante la última década con esta metodología. El proyecto comenzó con la idea de usar datos cuantitativos para observar y examinar conflictos geopolíticos, especialmente aquellos que podrían tener un impacto significativo en la economía mundial. Este sistema nos ha proporcionado una descripción de los eventos casi en tiempo real y en alta definición o granularidad. En este sentido, la combinación de *big data* de noticias mundiales y las técnicas de procesamiento del lenguaje natural nos han dado como resultado una herramienta de análisis muy valiosa para analizar las relaciones internacionales.

El sistema que hemos desarrollado está basado en la Base de Datos Global de Eventos, Lenguaje y Tono (GDELT) descrita en la sección anterior. GDELT es una fuente de código abierto que analiza noticias digitales en más de cien idiomas. Además, clasifica la información utilizando miles de taxonomías y temas, identificando emociones, organizaciones, ubicaciones y eventos, así como el tono promedio de las noticias, que varía de -100 a +100, indicando sentimientos negativos o positivos, respectivamente.

Para construir los índices capturamos tanto la cobertura como el sentimiento de los artículos de noticias diarios que mencionan conflictos y protestas. Hemos desarrollado un amplio conjunto de índices que incluyen el riesgo geopolítico, la estabilidad política, los índices de conflicto y protesta, el índice de incertidumbre de política económica y nuestro índice de sentimiento bilateral entre países. Para algunos de estos índices también podemos distinguir entre fuentes de medios extranjeros o locales. En esencia, capturamos cómo los medios de comunicación mundiales perciben la situación geopolítica<sup>5</sup>.

Durante estos años hemos utilizado este conjunto de indicadores para el seguimiento de diferentes conflictos. En la siguiente sección mostramos algunos ejemplos, como el seguimiento del impacto de la crisis de Siria, en la emigración hacia Europa, las guerras entre Rusia y Ucrania y, la más reciente, entre Israel y Hamas, así como la evolución del sentimiento mundial del mercado de los semiconductores y su relación con las tensiones entre China y Taiwán.

---

<sup>5</sup> Todos nuestros índices están normalizados, lo que significa que muestran su rendimiento relativo en comparación con su propio historial, a lo que definimos como «señales». Además, les aplicamos una media móvil geométrica de 28 días para reducir el ruido, dando más peso a la información diaria más reciente.

### 3.1. El conflicto sirio y la crisis migratoria a Europa

Una de las primeras aplicaciones que llevamos a cabo fue el impacto de la guerra civil en Siria en el flujo de emigrantes hacia Europa. Una de las fuentes que se han utilizado para el seguimiento de los flujos migratorios son las noticias y su geolocalización (Ahmed, 2016). Para capturar información sobre el impacto de conflictos, cambios en políticas (por ejemplo, el cierre de fronteras) y otros eventos externos, los datos de noticias proporcionados por GDELT son muy útiles.

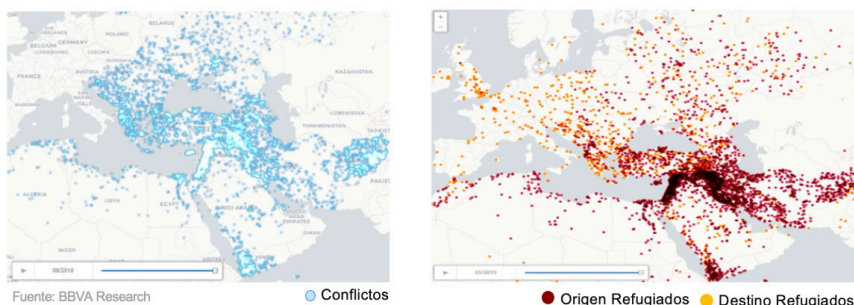
La geolocalización de los eventos permite al analista monitorear las noticias digitalizadas en todo el mundo y extraer información valiosa de los textos. Esta incluye información sobre entidades como, personas, lugares, organizaciones etc., en más de cien idiomas diferentes. Los documentos están anotados, aplicando técnicas de procesamiento de lenguaje natural de última generación. Entre las categorías monitorizadas en la base de datos de GDELT se puede realizar un conteo de arrestos, muertes, protestas, heridos y refugiados entre otros. Además, para cada artículo, se proporcionan las ubicaciones del evento y se utilizan algoritmos para corregir el conteo múltiple. Procesada correctamente, esta información es muy valiosa para el seguimiento de los flujos de refugiados.

La guerra civil siria comenzó en marzo de 2011 como parte de la Primavera Árabe. Inició con protestas a gran escala y manifestaciones prodemocracia contra el gobierno de Bashar al-Assad, que rápidamente se reprimieron violentamente, llevando a la formación de grupos rebeldes armados y escalando hacia una guerra civil completa. Los rebeldes lograron hacer avances significativos, pero la intervención militar de Rusia, en 2015, cambió el equilibrio del conflicto a favor del gobierno. Además, el Estado Islámico tomó control de grandes partes de Siria, lo que llevó a una campaña de bombardeos liderada por EE. UU. y el apoyo a las milicias kurdas YPG y sus aliados. Turquía, preocupada por la influencia kurda en sus fronteras, lanzó operaciones transfronterizas para combatir tanto al Estado Islámico como a las fuerzas kurdas.

El conflicto generó una crisis de refugiados masiva, con millones de sirios desplazados, tanto internamente como hacia países vecinos como Turquía, Líbano y Jordania. Muchos también han buscado asilo en Europa, contribuyendo a la crisis migratoria europea. La crisis ha sido descrita como una de las mayores

crisis de refugiados en la historia, exacerbada por violaciones de derechos humanos y condiciones severas en los campos de refugiados. La situación sigue siendo una crisis humanitaria significativa con millones de personas desplazadas y en necesidad de asistencia.

La figura 2 muestra la magnitud de esta crisis de refugiados en relación con la intensificación del conflicto. Utilizando GDELT hemos desarrollado dos gráficos. El primero de ellos muestra geográficamente aquellas zonas donde los conflictos, en relación con el total de eventos, han sido más intensos. Dentro de Europa y Oriente Medio, los focos de conflicto más intenso se concentran en Siria, Irak y Yemen, con focos de conflicto aislados en Egipto y algunas de las ciudades importantes del Norte de África y menor tensión acumulada en Turquía y Grecia. En el segundo, acumulamos las noticias sobre refugiados discriminando entre origen de los refugiados (color granate) y destino (amarillo) para visualizar el origen y destino de los refugiados.



**Gráfico 2. La guerra civil en siria y la emigración a Europa**

Los principales destinos de los refugiados de la guerra civil siria han sido países vecinos y más allá, enfrentando una de las mayores crisis de refugiados en la historia. Turquía ha sido el país anfitrión más grande, albergando a más de 3,7 millones de refugiados sirios. Además, Líbano y Jordania recibieron incluso más en términos de porcentaje de la población, asentados, sobre todo, en campos de refugiados y comunidades urbanas.

Más allá de los países vecinos, los refugiados sirios buscaron asilo sobre todo en Europa, contribuyendo a la crisis migratoria europea. Estos países, y otros, han enfrentado desafíos significativos al tratar de acomodar y asistir a los refugiados sirios, lidiando con cuestiones de infraestructura, empleo, educación y tensiones sociales. Dentro de Europa, los principales destinos de los

refugiados sirios fueron aquellos países que ofrecieron políticas de asilo más accesibles y mejores condiciones de vida. Entre ellos destacan los países centroeuropeos (Alemania, Austria, Países Bajos y Francia), Escandinavia y el Reino Unido. Dentro de los países mediterráneos, Grecia e Italia experimentaron los mayores flujos de refugiados.

Turquía ha sido el principal receptor de refugiados sirios, con millones de sirios buscando refugio allí desde el comienzo del conflicto. El gobierno turco ha establecido campamentos de refugiados y ha ofrecido diferentes grados de apoyo y servicios a los refugiados, aunque las condiciones y el acceso a los derechos y servicios pueden variar considerablemente. En marzo de 2016, Turquía y la Unión Europea firmaron un acuerdo para frenar el flujo de migrantes hacia Europa. Según este acuerdo, Turquía aceptaría la devolución de todos los migrantes y refugiados que llegan a Grecia desde Turquía y, a cambio, la UE reaceptaría a un sirio de Turquía por cada sirio devuelto de las islas griegas.

Aunque la crisis se estabilizó a medida que el número de nuevos refugiados fue descendiendo en Siria, y que Isis comenzaba a perder influencia, muchos de los refugiados continúan en campos de refugiados o establecidos en países fuera de sus fronteras de origen.

### 3.2. El conflicto Rusia-Ucrania

Otra de las aplicaciones que se pueden desarrollar con GDELT es un análisis detallado de la evolución de los diferentes eventos dentro de un conflicto. En particular es posible monitorear, en tiempo real, las noticias sobre el enfrentamiento, tanto en términos generales o agregados como el análisis más detallado de eventos, de acuerdo con la clasificación de CAMEO, en términos de cooperación y conflicto material y verbal. Además, podemos monitorear la intensidad del conflicto a través de la escala de Goldstein.

Un ejemplo de esta aplicación en BBVA Research ha sido el monitoreo de eventos y su intensidad durante la guerra entre Rusia y Ucrania. Este ha sido un conflicto prolongado y complejo que ha evolucionado significativamente desde su inicio hace casi una década. En 2014, la tensión en la región se intensificó dramáticamente después de la anexión de Crimea por parte de Rusia y el estallido de la guerra en el este de Ucrania, en las regiones

de Donetsk y Luhansk. Estos eventos marcaron un cambio drástico en las relaciones entre Rusia y Ucrania, con un aumento en la intervención militar y una serie de sanciones internacionales impuestas a Rusia.

A lo largo de los años siguientes, el conflicto en el este de Ucrania se mantuvo en un estado de conflicto latente e intermitente. A pesar de los acuerdos de paz y los esfuerzos diplomáticos, como los Acuerdos de Minsk, las violaciones al alto el fuego fueron constantes y la región permaneció en un estado de inestabilidad crónica.

A lo largo de 2021 y principios de 2022, la situación se agravó significativamente cuando Rusia comenzó a acumular tropas en la frontera con Ucrania. Este movimiento generó una alarma internacional y temores de una invasión a gran escala. Las tensiones alcanzaron su punto máximo en febrero de 2022, cuando Rusia lanzó una operación militar a gran escala en Ucrania, marcando una escalada dramática en el conflicto. Esta acción fue condenada por la comunidad internacional y llevó a una nueva ola de sanciones contra Rusia, así como a un aumento significativo en el apoyo militar y humanitario a Ucrania por parte de países occidentales.

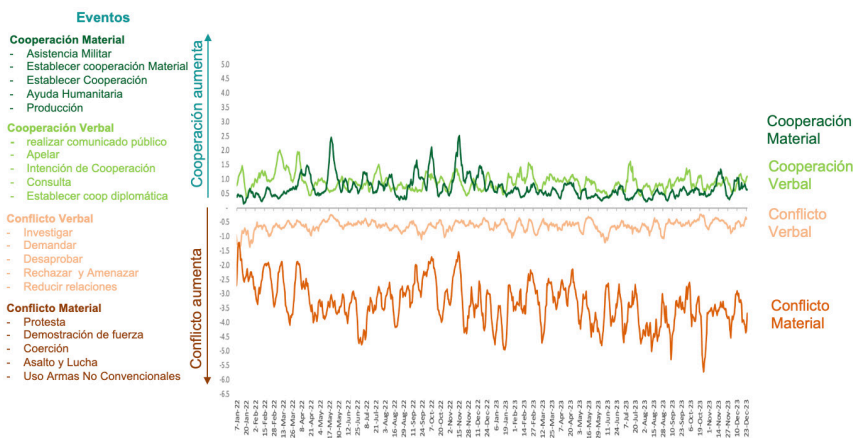
El siguiente gráfico resume la evolución de los distintos tipos de eventos en el conflicto desde la invasión de Rusia en febrero de 2022. Los cuatro índices reflejan la evolución ponderada de las cuatro grandes categorías de la escala de Goldstein (cooperación material y verbal y conflicto verbal y material). De acuerdo con esta clasificación, su intensidad y su evolución, las fases del conflicto han sido las siguientes:

- Inicio de la invasión (febrero a mayo 2022): con la invasión a gran escala de Ucrania en febrero de 2022 por parte de Rusia, el conflicto se disparó a los niveles más deteriorados de la escala de Goldstein, a medida que el número de eventos de conflicto material y su intensidad aumentaban. La reacción internacional fue en un principio de carácter verbal y pasó rápidamente a convertirse en material tras la ayuda, sobre todo, por EE. UU., pero también de Polonia, Reino Unido, Canadá, Noruega, Estonia y Letonia.
- Estabilización del conflicto y guerra de desgaste (primavera 2022 a finales de 2022): a medida que avanzaba la primavera de 2022, el conflicto se estabilizó en una guerra de desgaste, con un aumento inicial del índice de conflicto material y vaivenes en la intensidad de los enfrentamientos en la segunda mitad del año. Como refleja el índice de cooperación material,

se produjo una segunda ola de asistencia material a Ucrania y se intensificaron las sanciones a Rusia. A pesar de los esfuerzos diplomáticos internacionales, las negociaciones de paz no lograron detener el conflicto.

- Niveles de conflicto máximo (invierno de 2023): los primeros meses de 2023 fueron duros. El apoyo y la cooperación verbal sustituyó al material, mientras que el número e intensidad de los conflictos materiales alcanzaron niveles de máximo deterioro.
- Estancamiento del conflicto (finales 2023): los niveles de conflicto material continuaron en niveles de máxima tensión mientras comenzaba a apreciarse cierta mejora en los índices de cooperación material, tras haber permanecido en niveles mínimos durante todo el invierno.

En resumen, el conflicto se ha mantenido intenso, con esporádicos y limitados intentos de diálogo. En términos materiales, el continuo enfrentamiento militar y los ataques indican un sostenido nivel de conflicto material con oscilaciones alrededor de niveles de tensión elevados (aunque sin llegar a los niveles máximos de armamento no convencional). Verbalmente, aunque se hayan producido declaraciones ocasionales que sugieren la posibilidad de negociaciones, estas han estado a menudo acompañadas de condiciones previas y acusaciones mutuas, manteniendo un nivel bajo en la escala verbal de Goldstein. La guerra continuaba siendo un claro ejemplo de conflicto material y verbal, con pocas señales de cooperación significativa en cualquiera de los frentes.



Fuente: BBVA Research & Gdelt Project y Goldstein

**Gráfico 3. Escala de eventos de la guerra Rusia-Ucrania (enero 2002-diciembre 2003). Número de eventos en cada categoría x la intensidad en escala de Goldstein**

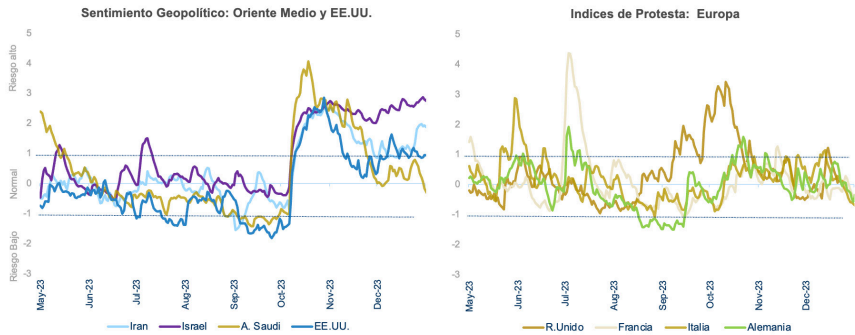
### 3.3. El conflicto entre Hamas e Israel

Una de las propiedades del sistema de alerta de señales geopolítico de BBVA Research es que incluye multitud de países y de índices de sentimiento extraídos de la base de datos GDELT. Esto supone una ventaja en aquellos conflictos en los que su situación estratégica puede tener implicaciones globales. Para capturar estas implicaciones, y su grado de transmisión o contagio, el sistema incluye índices de sentimiento geopolíticos, políticos, de incertidumbre de política económica, protesta y conflicto. Todo ello nos permite llevar a cabo un análisis detallado de las implicaciones de determinados conflictos en varias dimensiones y países, capturado en los diferentes medios de comunicación. Como hemos explicado anteriormente, este sistema de alerta nos permite analizar el sentimiento mediático según el origen de los medios de comunicación (local o extranjero), así como las relaciones bilaterales de los países.

Cuando Hamas atacó por sorpresa a Israel el pasado 7 de octubre de 2023, el sentimiento geopolítico mundial en los medios de comunicación era relativamente tranquilo. Desde el punto de vista geopolítico y en relación con el pasado reciente, el conflicto entre Rusia-Ucrania estaba todavía presente, pero su impacto mediático se había relajado notablemente. Como se aprecia en el gráfico 4, la mayoría de nuestros indicadores de sentimiento (geopolítico, político, incertidumbre de política económica, conflicto y protesta) mostraban una anormal neutralidad o tranquilidad tras meses de elevado riesgo durante los períodos más tensos del conflicto entre Rusia y Ucrania.

El efecto sorpresa también fue notorio en las noticias mundiales y, salvo una situación de conflicto histórico entre Israel y Hamas y los acuerdos Abrahams, que acercaban posturas entre Israel y Arabia Saudí, nada hacía presagiar un desenlace tan crítico.

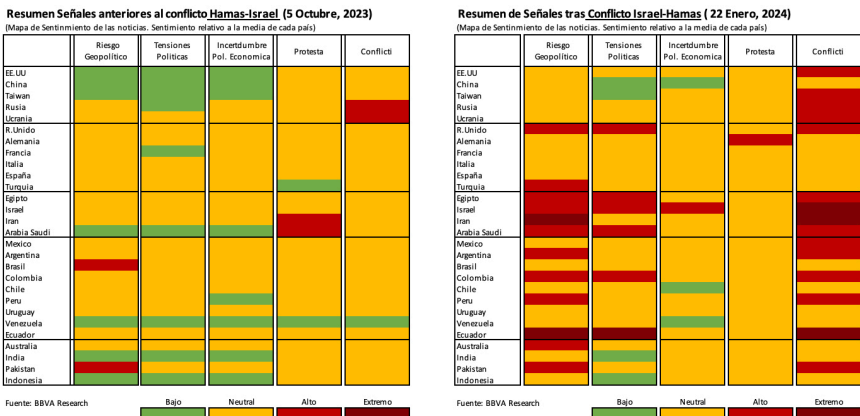
El gráfico 4 muestra cómo los índices de sentimiento geopolítico de los países de Oriente Medio y de EE. UU. aumentaron súbitamente y alcanzaron niveles de riesgo extremo en cuestión de días. A excepción de EE. UU., cuyo índice de sentimiento geopolítico se fue moderando hacia un sentimiento más neutral a medida que las noticias comenzaron a digerirse tras la sorpresa inicial, la mayoría de los países del golfo han permanecido en zona de riesgo con algunos vaivenes alrededor de las treguas temporales.



**Gráfico 4. Indicadores de sentimiento: geopolítico y protesta (unidades de desviación típica)**

Los efectos secundarios del conflicto en Europa también fueron evidentes desde el comienzo de la crisis, con estallidos de protesta en muchos países europeos, como se puede observar en el segundo gráfico 4b. El sentimiento se polarizó y el desencanto social comenzó a hacerse eco en las calles. Esto ha sido particularmente relevante en el Reino Unido y, en menor medida, en Francia, donde hemos asistido a una cierta polarización en el apoyo a los contendientes.

Pero no todos los indicadores reaccionaron de igual manera. Los siguientes diagramas en el gráfico 5 muestran un resumen gráfico de cómo evolucionaron los diferentes índices de sentimiento, incluyendo riesgo geopolítico, tensiones políticas, incertidumbre de política económica y protesta y conflicto desde el comienzo del conflicto.



**Gráfico 5. Indicadores de sentimiento: geopolítica, política, incertidumbre de política económica, protesta y conflicto antes y después del enfrentamiento Israel-Hamas**

El riesgo geopolítico, la inestabilidad política, el conflicto y la protesta aumentaron en los países de Oriente Medio a niveles altos o extremos. Una excepción importante ha sido el sentimiento de protesta en Irán, que se mantuvo neutral. Esto es particularmente importante, ya que algunos analistas comentaron sobre la posibilidad de que la crisis movilizara al pueblo iraní y desestabilizara el régimen.

La crisis se extendió rápidamente a algunos de los índices occidentales. De hecho, los indicadores geopolíticos y de conflicto se movieron de prisa hacia el área de alto riesgo en EE. UU. y algunos países europeos.

Afortunadamente, el índice de incertidumbre de política económica ha permanecido resiliente. Este ha sido un patrón generalizado. Hay algunas explicaciones aquí. Primero, el lenguaje del conflicto ha desplazado al económico. Segundo, el conflicto ha coincidido en tiempo con los Bancos Centrales clave, manteniéndose al margen después de un ciclo de endurecimiento muy agresivo y condiciones económicas razonables, lo que también ha impulsado un momento muy positivo en el mercado de valores.

#### 3.4. Las relaciones bilaterales entre países y los semiconductores: China, Taiwán y EE. UU.

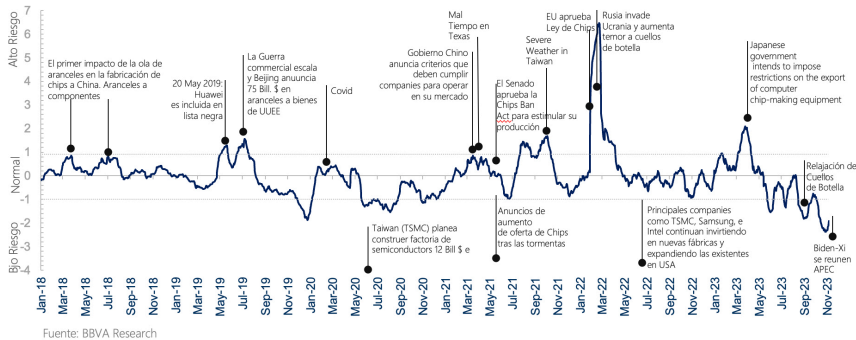
Dentro del Sistema de Alerta diseñado en BBVA Research, una herramienta de análisis interesante es la posibilidad de analizar las tensiones en las relaciones entre países. Un ejemplo de ello es el análisis llevado a cabo sobre la crisis de los semiconductores provocada, entre otros factores, aunque no fue el único, por las tensas relaciones entre China y Taiwán y, por ende, entre China y EE. UU.<sup>6</sup>

La crisis de los semiconductores, que se extendió desde 2018 hasta finales de 2023, ha estado originada por varios factores a lo largo del tiempo. Para analizar el sentimiento mundial hacia la industria de semiconductores desarrollamos un indicador para monitorear el sentimiento mediático sobre los semiconductores a nivel mundial utilizando la base de datos GDELT para construir el índice de sentimiento de semiconductores de BBVA Research. Capturamos tanto la cobertura como el sentimiento de los artículos de noticias por día que incluyen cualquier mención de este tema.

---

<sup>6</sup> Para un análisis detallado de este apartado véase Hsu *et al.* (2023).

La evolución del Índice Global de Sentimiento de Semiconductores de BBVA Research se resume en el gráfico 6. Como se puede observar, la crisis de semiconductores y su posterior normalización han sido impulsadas por varios factores, como la guerra comercial entre EE. UU. y China, las escaseces relacionadas con el COVID-19 y los problemas de capacidad, el mal tiempo, las tensiones geopolíticas, pero también los avances en la diplomacia internacional que han operado con distinta intensidad durante los últimos años.

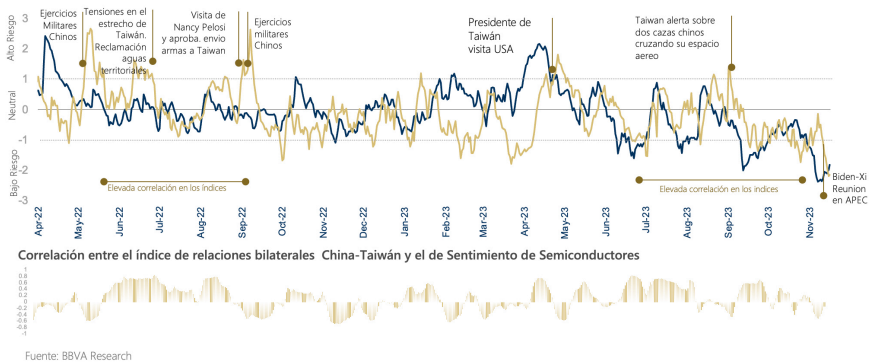


**Gráfico 6. Índice Global de Sentimiento de Semiconductores (unidades de desviación típica)**

Para entender mejor el impacto de los factores geopolíticos en la industria global de semiconductores, examinamos las interacciones entre China, Taiwán y Estados Unidos en el período entre 2022 y 2023. Este marco temporal nos ayuda a separar la influencia de estas relaciones de otros problemas como los problemas de capacidad relacionados con el COVID-19 y las interrupciones climáticas experimentadas durante 2020 y 2021. Los años 2022 y 2023 se caracterizaron por tensiones crecientes y complejas interacciones entre China, Taiwán y Estados Unidos, impulsadas por conflictos políticos arraigados e intereses estratégicos.

La relación entre China, Taiwán y Estados Unidos experimentó una tensión considerable en 2022. En abril, China intensificó su presencia militar cerca de Taiwán. Para junio, las tensiones escalaron aún más cuando el Ministerio de Relaciones Exteriores de Taiwán criticó a Pekín por reclamar el Estrecho de Taiwán como parte de su zona económica exclusiva. La situación empeoró con el apoyo de Estados Unidos a Taiwán mediante ventas de armas y la visita de la presidenta de la Cámara de Representantes, Nancy

Pelosi, en agosto, lo que llevó a China a realizar ejercicios militares en respuesta. Durante este período, el índice de relaciones bilaterales entre China y Taiwán pasó de neutral a una zona de mayor riesgo, mostrando una estrecha correlación con el índice global de semiconductores.



**Gráfico 7. Relaciones bilaterales China-Taiwán e índice global semiconductores (unidades de desviación típica)**

El año 2023 fue testigo de variados niveles de tensión. A principios de año, se observaron picos de tensión, particularmente alrededor de la visita de la presidenta Tsai a Estados Unidos. Las tensiones militares resurgieron en septiembre. Sin embargo, asistimos a una notable mejora en las relaciones más adelante en el año, a medida que los esfuerzos diplomáticos entre China y Estados Unidos se intensificaron en preparación para la reunión de Xi Jinping y Biden en la cumbre de APEC en noviembre.

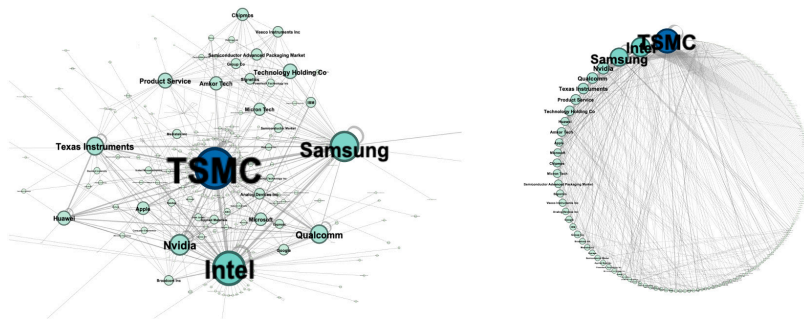
Estos desarrollos indican una fuerte correlación entre las relaciones China-Taiwán-EE. UU. y la estabilidad de la industria global de semiconductores. Comprender esta relación es crucial a la hora de diseñar políticas industriales que puedan limitar el impacto en una industria tan estratégica.

Los analistas de Ciencias Geopolíticas y Economía se han centrado en las posibles implicaciones económicas de un hipotético conflicto entre China y Taiwán. Un tema central en el análisis ha sido la relevancia de Taiwán en la industria mundial de semiconductores a través de la empresa Taiwán Semiconductor Manufacturing Company (TSMC), lo cual es, en particular, relevante para producir chips para los dispositivos más sofisticados.

Más que la evolución de las noticias de forma aislada para analizar el papel sistémico de TSMC en la industria de semiconductores,

construimos una red de noticias globales. Esta estrategia nos permite analizar las posibles relaciones en las noticias de la crisis de semiconductores de Taiwán con el resto de la industria. En esencia, analizamos quién está especialmente vinculado a TSMC en la industria de semiconductores. Para hacer esto, utilizamos la base de datos GDELT para analizar la co-ocurrencia de artículos de noticias de empresas de semiconductores, juntamente con TSMC.

Un gráfico de red o *Network* es una forma sencilla de representar datos de red donde los nodos (vértices) representan unidades (es decir, empresas) y una arista entre dos nodos indica que existe una relación entre ellos. Hay varias medidas de centralidad, incluyendo grados, cercanía e intermediación. Estas medidas evalúan el grado en que cada nodo juega un papel central en un gráfico. Como en este caso estamos interesados en analizar el papel sistémico de TSMC, implementamos un algoritmo para describir el tamaño de cada nodo, no solo en el número de aristas, sino también para considerar la relevancia de estas conexiones de aristas. La medida que usamos es la centralidad de auto vector, que tiene en cuenta la influencia transitiva de los nodos. Una alta puntuación de auto vector significa que un nodo está conectado a muchos nodos, que a su vez tienen altas puntuaciones, reflejando así mejor la naturaleza sistémica.



\* La centralidad de vector es un algoritmo que mide la influencia transitiva de los nodos. Las relaciones que provienen de nodos con puntuaciones altas contribuyen más a la puntuación de un nodo que las conexiones de nodos con puntuaciones bajas. Una puntuación alta de vector propio significa que un nodo está conectado a muchos nodos que a su vez tienen puntuaciones altas

Fuente: BBVA Research

**Gráfico 8. Red de relaciones industria de semiconductores**

Los resultados de las noticias de 2022 revelan la importancia sistémica de TSMC y su relación con otras corporaciones en la industria de semiconductores. Grandes IDMs como Intel, Samsung, Texas Instruments y Micron Tech lideran, en términos de relevancia, subcontratando parte de su producción a TSMC

para tecnologías avanzadas. Empresas *fabless* como Nvidia y Qualcomm, seguidas por Broadcom y AMD, son clientes significativos de TSMC, diseñando y vendiendo *hardware* y chips mientras externalizan la fabricación. Grandes tecnológicas como Apple, Huawei y Microsoft dependen de TSMC para chips personalizados, aunque Huawei ha sido afectada por restricciones de EE. UU. Empresas OSAT<sup>7</sup> como Amkor Tech y Chipmos proporcionan servicios de empaquetado y pruebas para los semiconductores de TSMC, con Amkor recientemente anunciando que empaquetará chips fabricados en Arizona.

En resumen, la red de noticias muestra la centralidad de TSMC en la industria de semiconductores y su interconexión con importantes actores globales. La relación entre TSMC y estas empresas refleja la complejidad y la interdependencia del sector, destacando la influencia de TSMC en la cadena de suministro global y su papel crítico en la producción de tecnologías avanzadas. Estos vínculos son fundamentales para entender la dinámica del mercado y las posibles repercusiones de cualquier perturbación en la industria.

#### 4. Conclusiones

En este artículo mostramos cómo la evolución de la digitalización de la información y el desarrollo de la inteligencia artificial, especialmente en el campo del procesamiento de lenguaje natural, están transformando radicalmente el análisis en las ciencias políticas y relaciones internacionales.

Hemos explorado diversos modelos que utilizan y procesan información no estructurada originada por los medios de comunicación internacionales y locales, a través de algoritmos de procesamiento de lenguaje natural y análisis de texto, que la convierten en valiosa información numérica, aplicable en el análisis geopolítico.

Adicionalmente, y basándonos en nuestra propia experiencia, mostramos cómo hemos utilizado este tipo de técnicas e información en BBVA Research a través de distintas herramientas aplicadas a diversos conflictos y eventos.

Desde el análisis de los flujos migratorios a Europa tras la crisis de Siria, al conflicto entre Rusia y Ucrania, las tensiones

---

<sup>7</sup> *Outsourced Semiconductor Assembly and Test.*

Israel-Hamas, y las implicaciones de las disputas entre China y Taiwán para el sector mundial de los semiconductores, estos ejemplos y otros en la literatura ilustran la capacidad de estos modelos para proporcionar análisis, realizar seguimientos detallados de los conflictos y, en algunos casos, llegar a poder a desarrollar sistemas de alerta temprana para anticipar conflictos.

La transformación del texto a números se perfila como una herramienta poderosa, ofreciendo una nueva perspectiva en el análisis político y de relaciones internacionales. Todo ello destaca la relevancia de las fuentes no estructuradas como herramientas analíticas en tiempo real. Este avance no solo permite un seguimiento en tiempo real de situaciones geopolíticas complejas, sino que también posibilita una comprensión más profunda de las dinámicas globales y regionales, contribuyendo significativamente a la formulación de políticas y estrategias más informadas y efectivas.

En conclusión, la integración de tecnologías avanzadas de IA y el procesamiento de lenguaje natural en el análisis de relaciones internacionales y las ciencias políticas abre un campo prometedor para investigaciones futuras y para la toma de decisiones estratégicas más informadas a nivel global.

## Bibliografía

- Ahmed, M. N. *et al.* (2016). A Multi-Scale Approach to Data-Driven Mass Migration Analysis. En: *SoGood@ECML-PKDD*.
- Azar, E. E. (1980). The Conflict and Peace Data Bank (COPDAB) Project. *Journal of Conflict Resolution*. Vol. 24, n.º 1, pp. 143-152.
- Baker, S. R., Bloom, N. y Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*. Vol. 131, n.º 4, pp. 1593-1636.
- Birmingham, A. *et al.* (2009). Combining Social Network Analysis and Sentiment Analysis to Explore the Potential for Online Radicalisation. En: *2009 International Conference on Advances in Social Network Analysis and Mining*.
- Blei, D. M., Ng, A. Y. y Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*. Vol. 3, pp. 993-1022.
- Caldara, D. y Iacoviello, M. (2022). Measuring Geopolitical Risk. *American Economic Review*. Vol. 112, n.º 4.

- Gennaro, G. y Ash, E. (2021). Emotion and Reason in Political Language. *The Economic Journal*. Vol. 132, n.º 643, pp. 1037-1059.
- Gerner, D. J. et al. (2002). Conflict and Mediation Event Observations (CAMEO): A New Event Data Framework for the Analysis of Foreign Policy Interactions. En: Goldstein, J. S. A Conflict-Cooperation Scale for WEIS Events Data. *Journal of Conflict Resolution*. Vol. 36, n.º 2, pp. 369-385.
- Halterman, A. et al. (2023). *PLOVER and POLECAT: A New Political Event Ontology and Dataset*. [Consulta: 2024]. Disponible en: <https://osf.io/preprints/socarxiv/rm5dw>.
- Leetaru, K. y Schrodt, P. (2013). GDELT: Global data on events, location, and tone. En: *ISA Annual Convention*.
- Martin, G. J. y McCrain, J. (2019). Local News and National Politics. *American Political Science Review*. Vol. 113, n.º 2, pp. 372-384.
- McLelland, C. A. (1976). World Event/Interaction Survey Codebook. *Ann Arbor: Inter-University Consortium for Political and Social Research*. ICPSR5211(4).
- Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*. Vol. 38, n.º 8, p. 114.
- Mueller, H. y Rauh, C. (2018). Reading Between the Lines: Prediction of Political Violence Using Newspaper Text. *American Political Science Review*. Vol. 112, n.º 2, pp. 358-375.
- O'Brien, S. P. (2010). Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research. *International Studies Review*. Vol. 12, n.º 1, pp. 87-104.
- Rodriguez, P. L., Spirling, A. y Stewart, B. M. (2023). Embedding Regression: Models for Context-Specific Description and Inference. *American Political Science Review*. Vol. 117, n.º 4, pp. 1255-1274.
- Schrodt, P. A., Davis, S. G. y Weddle, J. L. (1994). Political Science: KEDS—A Program for the Machine Coding of Event Data. *Social Science Computer Review*. Vol. 12, n.º 4, pp. 561-587.
- Sundberg, R. y Melander, E. (2013). Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research*. Vol. 50, n.º 4, pp. 523-532.
- Vaswani, A. et al. (2017). Attention is all you need. *Advances in neural information processing systems*. Vol. 30.