



Documento de Trabajo SECMCA-04-2026

Aplicaciones de minería de textos para Economía y Comunicaciones

Julio Bonilla Porras
Ángel Parajeles Paniagua
Juan Quiñonez Wu

Secretaría Ejecutiva

Octubre 2026 - San José, Costa Rica

Text mining applications for Economics and Communications

Julio Bonilla Porras
julio.bonillaporras@ucr.ac.cr
Estudiante de Economía de la
Universidad de Costa Rica

Ángel Parajeles Paniagua
angel.parajeles@ucr.ac.cr
Estudiante de Economía de la
Universidad de Costa Rica

Juan Quiñónez Wu
jquinonez@secmca.org
Economista Permanente de la
Secretaría Ejecutiva del
Consejo Monetario Centroamericano

Abstract

This note reviews the main applications of text mining and machine learning at the intersection of economics and communications. Drawing on a review of the literature, it analyzes three fundamental uses of these tools: the construction of sentiment and uncertainty indices, the identification of latent topics in large sets of documents, and the measurement of the complexity and readability of messages. The evidence shows that these methodologies make it possible to transform unstructured textual information into quantitative indicators useful for analyzing expectations, institutional priorities, and communication strategies. It concludes that, although they do not replace professional judgment, they constitute valuable tools for complementing economic analysis and evaluating the effectiveness of communication by central banks and other economic institutions.

The opinions contained in this document are the sole responsibility of the authors and do not necessarily reflect the views of SECMCA or CMCA.

Aplicaciones de minería de textos para Economía y Comunicaciones

Julio Bonilla Porras
julio.bonillaporras@ucr.ac.cr
Estudiante de Economía de la
Universidad de Costa Rica

Ángel Parajeles Paniagua
angel.parajeles@ucr.ac.cr
Estudiante de Economía de la
Universidad de Costa Rica

Juan Quiñónez Wu
jquinonez@secmca.org
Economista Permanente de la
Secretaría Ejecutiva del
Consejo Monetario Centroamericano

Resumen

Esta nota revisa las principales aplicaciones de la minería de textos (*text mining*) y del aprendizaje computacional (*machine learning*) en la intersección entre economía y comunicaciones. A partir de una revisión de la literatura, se analizan tres usos fundamentales de estas herramientas: la construcción de índices de sentimiento e incertidumbre, la identificación de temas latentes en grandes conjuntos de documentos y la medición de la complejidad y legibilidad de los mensajes. La evidencia muestra que estas metodologías permiten transformar información textual no estructurada en indicadores cuantitativos útiles para analizar expectativas, prioridades institucionales y estrategias de comunicación. Se concluye que, aunque no sustituyen el juicio profesional, constituyen herramientas valiosas para complementar el análisis económico y evaluar la efectividad de la comunicación de la banca central y otras instituciones económicas.

Las opiniones contenidas en este documento son responsabilidad exclusiva de los autores y no reflejan necesariamente los puntos de vista de SECMCA o CMCA.



I. Introducción

En la economía contemporánea, una parte creciente de la información relevante no se encuentra en bases de datos numéricas (fuentes de información muy preferidas por economistas), sino en documentos narrativos como noticias, informes corporativos, actas de reuniones, discursos oficiales y publicaciones en medios digitales. Estos textos contienen información sobre expectativas, percepciones, riesgos y prioridades institucionales que influyen sobre las decisiones de consumo, inversión y política económica, pero que tradicionalmente han sido difíciles de cuantificar de manera sistemática. De esta forma, se ha fortalecido la relación entre la Economía y la Comunicación, como ciencias.

El desarrollo de la minería de textos (*text mining*) y del aprendizaje computacional (*machine learning*) ha ampliado la capacidad de los investigadores para analizar este tipo de información. Estas herramientas permiten transformar grandes volúmenes de lenguaje no estructurado en datos susceptibles de análisis estadístico, facilitando el estudio de fenómenos que anteriormente solo podían examinarse mediante métodos cualitativos. En lugar de sustituir el análisis profesional, estas metodologías proporcionan mecanismos replicables para identificar patrones, clasificar información y extraer señales relevantes a partir del contenido textual.

La aplicación de estas herramientas ha dado lugar a una extensa literatura en economía y finanzas. Una primera línea de investigación se centra en la construcción de índices derivados de texto, utilizados para medir fenómenos difícilmente observables mediante indicadores convencionales, como el sentimiento de los agentes económicos o la incertidumbre asociada al entorno macroeconómico, financiero o político. Una segunda línea utiliza modelos de temas latentes para identificar automáticamente los asuntos predominantes dentro de grandes colecciones de documentos, permitiendo estudiar la evolución de las prioridades institucionales, la dispersión temática del debate público y distintos aspectos relacionados con la transparencia de la comunicación. Finalmente, una tercera línea también de gran relevancia analiza la complejidad y legibilidad de los mensajes, evaluando hasta qué punto la forma en que se

comunica una decisión económica afecta su comprensión e incluso su impacto sobre los mercados.

Estas aplicaciones revisten una importancia particular para los bancos centrales y otras instituciones económicas. En regímenes modernos de política monetaria, la comunicación constituye un instrumento complementario de acción, capaz de influir sobre las expectativas de hogares, empresas y otros agentes de la economía. En consecuencia, comprender qué mensajes transmiten las instituciones, cómo evolucionan sus prioridades comunicativas y qué tan accesibles resultan sus comunicaciones se ha convertido en un elemento central para evaluar la efectividad de la política económica.

El objetivo de esta nota es analizar las principales aplicaciones de la minería de textos en la intersección entre economía y comunicaciones mediante una revisión de la literatura especializada. Se examina cómo estas herramientas permiten convertir datos narrativos en evidencia cuantitativa para estudiar el sentimiento y la incertidumbre económica, identificar temas latentes en el discurso institucional y evaluar la complejidad de los mensajes. El documento se organiza en tres secciones. Después de esta introducción, la segunda sección revisa las principales aplicaciones de la minería de textos agrupadas en esas tres dimensiones. Finalmente, la tercera sección presenta las conclusiones y principales consideraciones derivadas de la literatura analizada.

II. Aplicaciones de la minería de texto

Índices derivados del texto

Los índices derivados de texto son herramientas estadísticas y computacionales cuyo objetivo es transformar grandes volúmenes de información cualitativa no estructurada (como artículos de prensa, minutas de bancos centrales o informes corporativos) en variables numéricas continuas. Su utilidad radica en que permiten medir fenómenos económicos intangibles, como el sesgo político de la prensa, el sentimiento del mercado o la incertidumbre en la política económica, aspectos que los indicadores macroeconómicos tradicionales no logran reflejar. Además, al construirse a partir de textos publicados de manera constante, ofrecen señales tempranas respecto a las estadísticas oficiales. Esta capacidad ha sido aprovechada por Nyman et al. (2021), cuyo índice se anticipa a los indicadores de confianza del mercado, y por Baker et al. (2016), cuyo índice de incertidumbre precede caídas en la inversión y el empleo.

La literatura destaca dos enfoques para la construcción de estos índices. El primer enfoque corresponde a los métodos de diccionario, de carácter deductivo. En ellos, el investigador define previamente una lista de palabras clave y calcula su frecuencia relativa en el texto. Loughran y McDonald (2011) advierten que para que un índice de este tipo resulte útil y no genere ruido estadístico, debe basarse en diccionarios ajustados a la jerga de la disciplina o complementarse con técnicas de ponderación de términos. De ahí que la mayoría de los trabajos que se revisan más adelante no empleen listas genéricas, sino diccionarios contruidos a la medida de su objeto de estudio, ya sea para captar la incertidumbre de política económica (Baker et al., 2016), el ánimo de los mercados (Nyman et al., 2021), la incertidumbre en las actas de un banco central (Moreno-Pérez y Minozzo, 2024) o el riesgo climático (Bua et al., 2024).

El segundo enfoque corresponde a los métodos de aprendizaje computacional, que identifican patrones y estructuras directamente en los datos. Un ejemplo es el trabajo de Sprenger y Welppe (2010), quienes clasificaron manualmente miles de mensajes de Twitter para entrenar un modelo que luego ordenó automáticamente el resto. También puede ser no supervisado, cuando el algoritmo detecta por sí mismo agrupaciones de vocabulario o temas sin categorías definidas de antemano, como en los estudios de Calvo-González et al. (2018) y Shugliashvili y Pirveli (2025), que se analizan en la sección dedicada a los temas latentes.

Ambos enfoques son maneras distintas de procesar el texto, no de decidir qué se mide con él. Un mismo método puede servir para captar el ánimo de un mercado o el grado de duda sobre el futuro, según las palabras que se decida rastrear. De las múltiples aplicaciones posibles, este documento se concentra en dos que resultan especialmente relevantes para la banca central: el sentimiento y la incertidumbre.

Índices de sentimiento

El primero de ellos es el índice de sentimiento o tono, una medida cuantitativa que refleja la postura o la carga emocional que los individuos expresan en un texto sobre un objeto o fenómeno económico. Para construirlo suele contabilizarse la proporción de términos con connotaciones positivas y negativas, o de vocablos vinculados a categorías emocionales concretas definidas en un diccionario.

El tono puede medirse en ámbitos distintos; en este documento se abordan dos: por un lado, el de los mercados financieros, donde interesa el ánimo de quienes observan y comentan la

economía; por otro, el de la política monetaria, donde lo que se mide es la orientación que la propia autoridad transmite en sus comunicados. En el ámbito financiero es importante medir el tono porque permite describir y anticipar el comportamiento del mercado y el estado de ánimo de sus participantes en tiempo real. Las investigaciones recurren a fuentes muy distintas, que van desde el lenguaje especializado de los analistas y las agencias de noticias, como en Nyman et al. (2021), hasta el contenido masivo y espontáneo de las redes sociales, examinado por Sprenger y Welppe (2010).

El trabajo de Nyman et al. (2021) busca construir un indicador de alerta temprana capaz de anticipar tensiones financieras a partir del tono emocional de las narrativas de mercado. Para ello, analizan tres fuentes de texto: noticias de Reuters, informes de analistas financieros y los comentarios internos de mercados del Banco de Inglaterra, institución con la que el estudio fue elaborado. El procedimiento consiste en comparar mensualmente la frecuencia de palabras asociadas con el entusiasmo y la ansiedad, para determinar cuál de las dos emociones predomina. La diferencia entre ambas (entusiasmo y ansiedad), dividida entre el tamaño del texto, da origen a su indicador: el Relative Sentiment Shift (RSS). Una decisión resulta clave en esa construcción, y es que, a diferencia del diccionario de Loughran y McDonald (2011), el de Nyman et al. (2021) excluye a propósito términos económicos como *crisis* o *recesión*, porque esas palabras aparecen en la prensa cuando el problema ya estalló, de modo que un índice que las contara reaccionaría tarde. Al observar solo emociones, en contraste, el indicador capta el cambio de ánimo antes de que la mala noticia se haga pública.

Los resultados respaldan la decisión, pues el índice se comporta como un termómetro emocional que, en retrospectiva, se desploma desde mediados de 2007 (antes de la caída de Bear Stearns y muy antes de la de Lehman Brothers) y que anticipa el movimiento de indicadores convencionales como la confianza del consumidor o el VIX¹. A ello se suma que una caída pronunciada del sentimiento deprime el empleo, la producción industrial y la bolsa londinense durante un período prolongado.

¹ El VIX es un índice que mide la volatilidad que los inversionistas esperan del mercado accionario estadounidense para los siguientes treinta días. Se le conoce como "el índice del miedo", porque tiende a subir en momentos de incertidumbre.

Sprenger y Welp (2010) se preguntan si los mensajes que los inversionistas publican en redes sociales contienen información con valor económico, es decir, si lo que se comenta sobre una empresa guarda relación con el comportamiento de su acción. Para responderlo estudian cerca de 250.000 mensajes de Twitter publicados durante el primer semestre de 2010 sobre empresas del índice S&P 100, identificados por la etiqueta que los usuarios emplean para referirse a una acción (por ejemplo, "\$AAPL" para Apple). A diferencia de los métodos de diccionario, aquí se emplea aprendizaje supervisado, pues los investigadores clasifican manualmente 2.500 mensajes como recomendaciones de compra, venta o mantenimiento y utilizan ese conjunto para entrenar un algoritmo capaz de etiquetar el resto de la muestra con una precisión cercana al 81%. A partir de esa clasificación construyen tres indicadores diarios: el optimismo de los mensajes, su volumen y el grado de acuerdo entre ellos.

Los resultados muestran que el optimismo se asocia con retornos anormales de las acciones, que el volumen de mensajes anticipa el volumen de negociación del día siguiente y que un mayor acuerdo entre usuarios se relaciona con un menor volumen de negociación. El estudio documenta además un mecanismo de depuración dentro de la propia red, pues los mensajes no circulan todos por igual. Los usuarios cuyas recomendaciones resultan más acertadas ganan seguidores y son más retuiteados que quienes se equivocan, de modo que la red amplifica las voces de mayor criterio y el sentimiento agregado no equivale a una suma de opiniones al azar. Los autores advierten, no obstante, que la relación corre con más fuerza en sentido inverso al esperado, ya que los precios influyen sobre lo que se comenta más de lo que los comentarios influyen sobre los precios.

Los trabajos revisados en el ámbito financiero, tanto Nyman et al. (2021) como Sprenger y Welp (2010), miden el ánimo de quienes observan la economía desde fuera, sean analistas, agencias de noticias o inversionistas en redes sociales. El tono, sin embargo, también puede medirse en el ámbito monetario, esta vez en el lenguaje de la entidad que toma las decisiones. En un banco central la comunicación no describe el ánimo del mercado sino la orientación de la política monetaria, es decir, si el mensaje apunta hacia una postura más restrictiva o flexible respecto del rumbo futuro de las tasas de interés.

Apel y Blix Grimaldi (2012) se ocupan precisamente de medir ese tono monetario y examinan si las actas de las reuniones de política contienen información capaz de anticipar la siguiente

decisión de tasas. Para responderlo analizan las actas del banco central de Suecia entre 2000 y 2011.

El método se aparta de los diccionarios habituales, pues contar palabras sueltas resulta insuficiente, porque un término como "inflación" no indica por sí mismo hacia dónde se inclina el banco, ya que puede aparecer tanto en "inflación más alta" como en "inflación más baja", que son señales opuestas. Los autores buscan entonces combinaciones de dos palabras, formadas por un sustantivo propio de la política monetaria (inflación, crecimiento, precios, salarios o coyuntura) y un adjetivo que indique dirección. La combinación sí fija el sentido, pues "inflación más alta" señala una postura restrictiva, en la que el banco tenderá a subir la tasa, mientras que "crecimiento más lento" señala una postura flexible, en la que tenderá a bajarla. Al contar cuántas expresiones de cada tipo contiene cada acta y restar unas de otras, obtienen un índice neto del tono.

Los resultados muestran que ese índice anticipa la decisión siguiente. Cuando las actas resultan más restrictivas de lo habitual, la probabilidad de que la tasa suba en la reunión posterior asciende a 34%, mientras que cuando suenan más flexibles esa probabilidad cae a 8%. Los autores comprueban además que ese tono no se limita a reflejar lo que ya muestran las cifras. Un banco central decide subir o bajar la tasa mirando sobre todo la inflación y el crecimiento de la economía, de modo que es posible pensar que las actas suenan más restrictivas simplemente porque la inflación viene alta. Para verificarlo, repiten el ejercicio tomando en cuenta esos dos datos junto al índice de tono, y el tono conserva su capacidad de anticipar la decisión. Las actas, por tanto, aportan información sobre lo que hará el banco que no está contenida en los indicadores económicos.

Los índices presentados anteriormente, tanto en el ámbito financiero como en el monetario, miden la dirección del mensaje: si el mercado se muestra optimista o temeroso, o si el banco central adopta un tono restrictivo o flexible. Sin embargo, el lenguaje económico transmite algo más que una inclinación; también refleja el grado de desconocimiento sobre el futuro, y es precisamente esa dimensión la que buscan capturar los índices de incertidumbre.

Índices de incertidumbre

Un índice de incertidumbre es una métrica cuantitativa diseñada para medir el nivel de duda, falta de certeza o el riesgo que los agentes perciben respecto al futuro económico, financiero o

político. A diferencia del sentimiento, que registra si el ánimo predominante es favorable o adverso, la incertidumbre registra cuánto se habla de aquello que no se sabe.

Este trabajo distingue distintos tipos de índices de incertidumbre según el ámbito en que se miden. En el ámbito financiero y corporativo interesa la duda expresada por los directivos respecto al futuro de su empresa y cómo esa duda condiciona sus decisiones de inversión. En el ámbito macroeconómico la atención se dirige a la economía nacional; la incertidumbre se detecta en la prensa y en documentos oficiales y se relaciona con variables agregadas como inversión, producción y empleo. Por último, se incluye un enfoque dedicado a la incertidumbre climática, cuya discusión se reserva para el cierre de la sección.

En el ámbito corporativo, buena parte de la evidencia proviene de los informes anuales de las empresas, esta es justamente la fuente utilizada por Chen et al. (2025) y Hsu et al. (2022). De la Parra y Gallemore (2026) recurren en cambio a la comunicación oral de los directivos, atendiendo no a lo que escriben sino a cómo lo dicen.

Chen et al. (2025) se proponen medir la incertidumbre tal como la percibe la propia gerencia de la empresa y estimar su efecto sobre las decisiones de inversión. Por lo tanto, analizan los informes anuales de empresas japonesas, concretamente el apartado en que la gerencia examina el desempeño y las perspectivas del negocio, mediante un modelo de lenguaje capaz de interpretar oraciones completas en su contexto y no solo términos aislados. El procedimiento presenta dos ventajas frente al conteo de palabras. Reconoce, en primer lugar, expresiones que denotan inquietud, aunque no usen la palabra “incertidumbre”. En segundo lugar, permite identificar el motivo de esa incertidumbre, de modo que se separa la atribuible a la pandemia de COVID-19 de la vinculada a la política económica, incluso cuando ambas aparecen en una misma oración.

La validez del índice se contrasta con el error de pronóstico de ventas de la gerencia, definido como la diferencia entre las ventas efectivamente obtenidas y las proyectadas por la propia empresa al inicio del ejercicio, y que los autores identifican como una medida de incertidumbre de uso frecuente en la literatura. El índice construido a partir de oraciones muestra una correlación positiva con ese error de pronóstico, mientras que el índice tradicional de conteo de palabras muestra una correlación negativa, lo que sugiere que no captura el fenómeno que pretende medir. Una vez validada la medida, los autores encuentran que un aumento de la

incertidumbre percibida por la gerencia reduce la inversión de capital en cerca de un 3.57%, y que el efecto de la incertidumbre pandémica casi duplica al de la incertidumbre relacionada con la política económica.

El lenguaje de los informes corporativos no solo revela cuánta incertidumbre percibe la gerencia, sino también qué riesgos concretos identifica. Hsu et al. (2022) se proponen determinar si los riesgos que una empresa declara en sus informes anuales guardan relación con su desempeño operativo posterior. Con el objeto de lograrlo, trabajan con los informes anuales de empresas de semiconductores cotizadas en Taiwán entre 2015 y 2019 e identifican de forma automática los textos donde la gerencia describe los riesgos que enfrenta y las medidas con que piensa afrontarlos. A partir de esos mismos textos construyen un diccionario propio, que distingue el vocabulario característico de las empresas más rentables del de las menos rentables.

El resultado confirma la hipótesis de partida, pues los riesgos declarados guardan una relación significativa con la eficiencia operativa posterior de la empresa, de modo que la forma en que una firma describe sus propias amenazas anticipa en alguna medida su desempeño futuro. Los autores determinan además el alcance de esa capacidad predictiva al comparar el valor del texto con el de los indicadores financieros convencionales. Un modelo que predice el desempeño empresarial únicamente con información textual acierta alrededor de un tercio de las veces, uno basado solo en indicadores financieros acierta cerca de dos tercios, y la combinación de ambas fuentes eleva el acierto por encima del ochenta por ciento. El texto, por consiguiente, no sustituye a los números, sino que los complementa, al aportar información sobre las expectativas de la empresa que los estados financieros, de carácter retrospectivo, no recogen.

Los trabajos anteriores se apoyan en texto escrito, que admite revisión y ajuste antes de su publicación, especialmente en documentos corporativos sometidos a escrutinio público. De la Parra y Gallemore (2026) exploran por ello un canal menos controlable: la forma en que los directivos hablan. Su objetivo es determinar si las vacilaciones del habla (las pausas, los titubeos y las interrupciones que se producen al responder sin guion) revelan la incertidumbre de quien habla y son captadas por quienes lo escuchan.

Los autores analizan miles de horas de audio de llamadas trimestrales de resultados para medir las vacilaciones de los directivos al responder a las preguntas de los analistas. Para aislar la incertidumbre de la forma habitual de hablar de cada directivo, comparan su expresión

espontánea con el fragmento donde leen el texto preparado. Este contraste permite identificar qué pausas se deben realmente a la duda y cuáles forman parte de su estilo personal de comunicación.

Los resultados muestran que, a medida que aumentan las vacilaciones del directivo, se eleva la dispersión en los pronósticos financieros de los analistas. Este efecto se mantiene incluso cuando se toma en cuenta el contenido de la respuesta, lo que indica que el impacto no depende de lo que se dice, sino de la manera de expresarlo. Ese desacuerdo entre quienes estudian a la compañía constituye una de las formas habituales de medir la incertidumbre que la rodea, de manera que la duda del directivo termina trasladándose a quienes evalúan su desempeño futuro, demostrando que la incertidumbre se manifiesta por vías que la comunicación escrita permite atenuar.

A escala macroeconómica, la incertidumbre deja de referirse a una empresa concreta y pasa a reflejar la duda sobre el rumbo de todo un país, es decir, hasta qué punto los agentes ignoran qué hará el gobierno o cómo evolucionará la economía. Como esa incertidumbre se expresa sobre todo en el debate público, los índices se construyen a partir de textos de alcance nacional. Las fuentes varían según el emisor, desde la prensa estadounidense que analizan Baker et al. (2016), hasta la prensa centroamericana estudiada por Diakonova et al. (2025), y las propias actas de un banco central examinadas por Moreno-Pérez y Minozzo (2024).

Baker et al. (2016) se proponen cuantificar la incertidumbre asociada a las decisiones de política económica y estimar sus efectos sobre la actividad. Construyen para ello el Índice de Incertidumbre de Política Económica (EPU) mediante un método de diccionario. Analizan diez de los principales periódicos estadounidenses y calculan la proporción de artículos que mencionan de forma simultánea tres elementos: incertidumbre, economía y alguna referencia a la acción gubernamental, como el Congreso, el déficit o la Reserva Federal. La exigencia de los tres términos a la vez permite aislar la incertidumbre atribuible a las decisiones del Estado y no la ansiedad económica en general.

La validez del índice se sostiene en dos comprobaciones. Sus máximos coinciden con episodios de conocida inestabilidad, como elecciones reñidas, las guerras del Golfo, el 11 de septiembre o la quiebra de Lehman Brothers. Además, una auditoría en la que lectores humanos clasificaron manualmente doce mil artículos arroja resultados muy próximos a los del cálculo automatizado,

y los índices contruidos por separado con periódicos conservadores y progresistas resultan prácticamente indistinguibles, lo que descarta que la medición recoja el sesgo editorial de la prensa. El aporte central del estudio es, sin embargo, económico, pues un aumento de la incertidumbre comparable al observado entre 2006 y 2012 precede caídas cercanas al 6% en la inversión, al 1.2% en la producción industrial y al 0.35% en el empleo, con resultados similares en un panel de doce economías. El enfoque fue adoptado posteriormente en otros países, aunque su aplicación a economías pequeñas encuentra un obstáculo práctico, pues sin archivos de prensa digitalizados y extensos no hay material suficiente que contabilizar.

Diakonova et al. (2025) enfrentan ese problema al construir índices de incertidumbre de política económica para los siete países de Centroamérica y República Dominicana. Por un lado, complementan la escasa prensa local disponible con agencias regionales que publican en español. Por otro lado, adaptan el diccionario a las particularidades del istmo, incorporando junto a los términos habituales de gobierno y déficit otros como tipo de cambio, dolarización, corrupción o aranceles, que son fuentes especialmente relevantes de incertidumbre en estas economías. Para verificar la validez del índice, los autores analizan individualmente sus picos y corroboran que corresponden a episodios reales (una rebaja en la nota de riesgo del país, una elección presidencial disputada o un escándalo de corrupción). Luego lo someten a una prueba estadística, de la cual resulta que un aumento inesperado de la incertidumbre deprime la actividad económica, la inversión extranjera directa y el turismo, siendo esta última donde el efecto se observa con mayor claridad. Los autores reconocen las limitaciones del análisis, puesto que en estos países las muestras son pequeñas, los datos ruidosos y no todas las respuestas alcanzan significancia estadística (Nicaragua y El Salvador, de hecho, apenas reaccionan).

Los índices de incertidumbre anteriores se construyen a partir de lo que la prensa dice sobre las decisiones de gobierno. Moreno-Pérez y Minozzo (2024) invierten esa perspectiva y se proponen medir la incertidumbre que el propio banco central expresa en sus comunicaciones, así como determinar si el asunto al que esa incertidumbre se refiere altera sus efectos sobre la economía.

Analizan las actas de las reuniones de política monetaria del Banco Central de Brasil entre 2000 y 2019 y construyen su diccionario de incertidumbre a partir de esas mismas actas. Para ello emplean algoritmos que identifican palabras usadas en contextos similares y detectan los términos que, dentro de esas actas, funcionan como equivalentes de incertidumbre, incierto, o temores, en lugar de importar una lista prefabricada, como la de Loughran y McDonald (2011). El

diccionario resultante es menos sesgado y está mejor adaptado al texto analizado. A ello suman un segundo procedimiento, que identifica automáticamente el tema de cada párrafo de las actas, y que les permite separar dos incertidumbres distintas, la referida a las condiciones económicas generales y la referida a la inflación y a la decisión de política monetaria.

Los resultados muestran efectos relevantes sobre la economía, ya que un aumento inesperado de la incertidumbre en las actas se asocia con una depreciación del tipo de cambio y con caídas en la producción industrial y las ventas minoristas. Ambos tipos de incertidumbre inciden además sobre la inflación en direcciones opuestas, pues la incertidumbre referida a las condiciones generales la eleva, mientras que la relacionada con la política monetaria la reduce, lo que sugiere que no basta con saber si un banco central expresa incertidumbre, sino que importa sobre qué tema la expresa.

El clima constituye otra fuente de incertidumbre económica, tanto por los daños que provoca como por el costo de la transición hacia una economía más limpia. Bua et al. (2024) se proponen determinar si los mercados europeos incorporan ya esos riesgos en el precio de las acciones.

Los autores distinguen el riesgo físico, derivado de los daños materiales del cambio climático (sequías, inundaciones y olas de calor), del riesgo de transición, derivado del costo de adaptarse a una economía neutra en carbono (nuevas regulaciones, cambios tecnológicos y desplazamientos de las preferencias del público). Para medirlos, construyen dos diccionarios separados a partir de la literatura científica sobre el clima y los aplican al archivo de noticias de Reuters, con lo que obtienen dos índices diarios que registran cuánta atención dedica la prensa a cada tipo de riesgo.

Los resultados indican que el mercado europeo ya remunera la exposición al riesgo climático. Dado que un mayor riesgo demanda, por definición, una compensación adicional, los inversionistas exigen un rendimiento superior para mantener acciones de empresas expuestas, de alrededor de un 7% anual en el caso del riesgo de transición y un 6% en el del riesgo físico desde 2015, si bien los autores advierten que estas primas no alcanzan plena significancia estadística en el período analizado. Los índices permiten además identificar cómo responden diferentes tipos de empresas, ya que cuando el riesgo de transición se incrementa las acciones de las compañías más ecológicas suelen desempeñarse mejor, mientras que un aumento del riesgo físico perjudica el rendimiento de las más contaminantes.

Los índices derivados de texto, por lo tanto, resumen grandes volúmenes de información cualitativa no estructurada en variables numéricas susceptibles de análisis estadístico. Su utilidad reside en que permiten medir fenómenos que los indicadores macroeconómicos convencionales no recogen, como el ánimo de los mercados, la orientación de la política monetaria o el grado de incertidumbre que perciben empresas, gobiernos e inversionistas. Los trabajos revisados muestran que esas mediciones no son meramente descriptivas, pues se relacionan con decisiones de inversión, con variables agregadas como la producción y el empleo, y con la valoración de los activos financieros.

Temas latentes

Los índices de sentimiento e incertidumbre revisados hasta ahora permiten medir la orientación y el grado de duda presentes en un texto, pero no identifican con precisión sobre qué temas se construye esa orientación o esa incertidumbre. Esa es la pregunta que responden los temas latentes: estos son conceptos o temas recurrentes dentro de un conjunto de textos que se identifican aun cuando no hayan sido señalados previamente por el investigador. Generalmente no son directamente observables, por lo que se requiere el uso de modelos especializados para poder identificarlos. Uno de estos modelos es el método llamado Asignación Latente de Dirichlet (LDA, por sus siglas en inglés), formalizado por Blei et al. (2003): el modelo revisa qué palabras suelen aparecer juntas en los mismos documentos y agrupa el vocabulario en temas, sin que estos se determinen o definan previamente. El resultado son dos cosas: qué proporción de cada tema contiene cada documento, y qué palabras son más representativas de cada tema.

Otra herramienta para identificar estos temas es el Análisis Semántico Latente (LSA, por sus siglas en inglés) que toma una matriz de documentos y palabras, y a partir de álgebra lineal usa una técnica matemática de reducción de dimensiones. El resultado es una matriz de vectores numéricos, la cual revela la similitud entre párrafos de un mismo documento o de distintos documentos. Estas herramientas resultan de particular interés porque constituyen un nexo entre los datos y su comunicación. En última instancia, el objeto de estudio de estas herramientas es, qué se dice y cómo se dice. Para observar estas dimensiones es relevante analizar i) la entropía, ii) el cambio de prioridades y iii) la transparencia en la comunicación.

Entropía

Diferentes autores han convenido en que la entropía en este contexto es una medida de dispersión temática que hay en algún ámbito económico. Para el caso de los temas latentes resulta relevante

analizar qué tan variados están los temas dentro de un documento o un conjunto de documentos en un momento dado. Si la mayoría de los textos convergen hacia un mismo tema, la entropía es baja, si, por el contrario, están repartidos en distintos temas, la entropía es alta.

En el ámbito financiero, el mismo estudio de Nyman et al. (2021) muestra el potencial de LSA como herramienta de alerta temprana. En este caso, en lugar de enfocarse en el tono, se enfocan en qué tan concentrado o disperso es el discurso. Para esto agrupan los textos según su similitud (clustering) y calculan su entropía mes a mes, si los textos convergen a uno o pocos relatos, la entropía es baja y hay consenso; si se dispersan en muchos, la entropía es alta y hay desacuerdo. Encuentran que el consenso narrativo se mantuvo desde 2006 y alcanzó su punto máximo a mediados de 2007 (momento en que los agentes comienzan a prever debilidad crediticia e incertidumbre en la economía). Posteriormente cayó de forma abrupta (aumentó la entropía) en marzo de 2008, meses antes de la quiebra del banco de inversión Lehman Brothers, cuando el mercado dejó de coincidir en una misma lectura de lo que estaba ocurriendo.

Shugliashvili y Pirveli (2025) analizan el comportamiento del tipo de cambio euro-dólar con base en noticias. Los autores emplean LDA para extraer cinco temas noticiosos sobre el dólar, y utilizan la entropía para medir cuánta atención pública y mediática recibe cada tema mes a mes. Si en un mes la información asociada a un tema es más abundante o variable, la entropía aumenta; en caso contrario, cuando la atención se concentra, disminuye. Con la serie mensual encuentran que las noticias explican el 29% de la variabilidad de largo plazo del tipo de cambio, con el desarrollo económico y las noticias microeconómicas como los temas de mayor peso.

En ambos casos, el tema latente deja de ser solo una etiqueta descriptiva y se convierte en una variable numérica que se integra a modelos econométricos.

Cambios de prioridades

El cambio de prioridades mide qué tan estable es el enfoque temático de la comunicación a lo largo del tiempo. Una agenda persistente puede contribuir a formar expectativas claras, mientras que una agenda que cambia de forma errática puede introducir incertidumbre sobre los objetivos que se mantendrán en el futuro. Esta dimensión es relevante porque los agentes toman decisiones con base en la información disponible y en la credibilidad de los compromisos comunicados. Sin embargo, no toda variación temática es señal de inestabilidad, en algunos casos, puede reflejar

una adaptación del discurso ante un choque en la economía, y distinguir entre ambos casos resulta de gran relevancia.

Analizando el cambio de prioridades en terreno de la política pública y su comunicación, Calvo-González et al. (2018) construyen, a partir de LDA aplicado a 953 discursos presidenciales de tipo "informe de la nación" en diez países latinoamericanos (incluidos Costa Rica y República Dominicana) y España, un indicador de qué tan estables son las prioridades de política de un gobierno en el tiempo: si los mismos temas dominan año tras año, o si cambian de manera impredecible. Su hallazgo central es que esa volatilidad temática está asociada negativamente con el crecimiento económico de largo plazo. Los países cuyos discursos cambiaron menos de tema (España, Costa Rica) convergieron más con Estados Unidos, mientras que los de mayor volatilidad temática (Argentina, Venezuela) se rezagaron.

A una escala distinta, observando la comunicación cotidiana de un banco central, no la de un presidente, Roldan Ferrin (2023) usa LDA sobre las minutas de la Junta Directiva del Banco de la República de Colombia (JDBR) para mostrar que la pandemia de COVID-19 generó un tema completamente nuevo dentro de las discusiones internas del banco y desplazó el peso relativo de la inflación hacia la liquidez y el desempleo, evidenciando que el discurso interno se adapta a la coyuntura incluso antes de que eso se refleje en anuncios formales.

A diferencia de una agenda que cambia de forma inesperada, el caso de Colombia ilustra la otra cara de un cambio en las prioridades, donde la pandemia fue un choque atípico en la economía que llevó a la junta del banco a desplazar su objetivo tradicional (la inflación) para responder a la crisis. La diferencia no está directamente en si cambiaron las prioridades, sino en la razón de ese cambio: mientras en Calvo-González et al. (2018) la volatilidad temática refleja la falta de dirección sostenida en el tiempo, en Roldan Ferrin (2023) refleja la capacidad de respuesta ante un evento inesperado.

Transparencia en la comunicación

La transparencia mide cuánta información revelan las instituciones sobre sus decisiones y sobre el proceso que conduce a ellas. En documentos como informes, actas o transcripciones, esta dimensión permite observar no solo qué decisión se comunica, sino también cómo se justifica y qué deliberación la antecede. En principio, con una mayor transparencia, se esperaría una mejor asimilación de la información por parte de los agentes para la toma de decisiones; sin embargo,

la evidencia revisada en esta sección muestra que la apertura informativa también puede modificar los incentivos de quienes participan en la deliberación.

La tesis de Acosta (2014) sobre el Comité Federal de Mercado Abierto de Estados Unidos (FOMC, por sus siglas en inglés) introduce un matiz importante sobre la transparencia misma. Usando LSA, muestra que la publicación de transcripciones casi literales de las reuniones en 1993, sí aumentó la transparencia hacia el público, pero tuvo un costo interno, donde los miembros del comité empezaron a disentir menos abiertamente entre ellos una vez que supieron que sus palabras exactas serían públicas. Conviene destacar este hallazgo, pues más comunicación o comunicación más transparente no pareciera ser automáticamente mejor comunicación.

Hansen et al. (2018) usan el mismo experimento de la FOMC que analizó Acosta (2014) para observar los efectos de la publicación obligatoria de las transcripciones (con un rezago de cinco años) de las reuniones de la FED. Estas transcripciones comenzaron en los años setenta, pero fue hasta 1993 que se le obligó a la FED a publicarlas. Con esto, los autores disponen de dos conjuntos de datos claramente diferenciados: el conjunto de transcripciones previas a 1993 que existían, pero no eran públicas hasta el momento, y el conjunto de aquellas posteriores a la fecha en que los miembros ya sabían que serían públicas. Dado esto, plantean que la teoría predice dos efectos que operan en direcciones opuestas, un efecto disciplina (positivo): saber que serán observados motiva a los miembros a prepararse mejor y trabajar más; y un efecto conformidad (negativo): saber que sus palabras serán públicas los vuelve más cautelosos y menos dispuestos a disentir abiertamente.

Para medir estos efectos, usan LDA sobre las transcripciones del período de gestión de Alan Greenspan (1987-2006) y separan cada reunión en dos partes: la discusión de la situación económica (FOMC1, cada miembro da su lectura del panorama): se espera que al ser preparada sea más propensa a mostrar disciplina; y la discusión de estrategia de política monetaria (FOMC2, reacción a la propuesta del presidente): al ser más espontánea se espera que sea más propensa a mostrar conformidad. Implementan un modelo de diferencias en diferencias para aislar el efecto de transparencia comparando cómo reaccionan los miembros según su nivel de experiencia antes y después de 1993, bajo el supuesto de que los de menor experiencia (novatos) tienen más en juego para ganar reputación y por tanto deberían reaccionar más fuerte a la transparencia.

Encuentran que tras la apertura al público las reuniones se vuelven menos interactivas, parecen más guionizadas y con más uso de lenguaje cuantitativo. En cuanto al efecto disciplina (FOMC1) encuentran que los miembros con menos experiencia empiezan a usar significativamente más números y datos en sus intervenciones, además de que empiezan a cubrir una gama más amplia de temas, evidenciando que se preparan más al saber que ahora son reuniones públicas. Por otra parte, en cuanto al efecto conformidad (FOMC2), esos mismos miembros novatos, en cambio, hacen menos intervenciones, discuten un rango más pequeño de temas (al momento de la lectura aportan nuevos temas, pero en la discusión se apegan a los de sus compañeros) y se vuelven más cautelosos al momento de disentir abiertamente de lo que dice el presidente. Para observar la influencia sobre otros miembros, construyen un indicador que mide como incide lo que dice un miembro sobre lo que dicen después sus colegas, incluido el presidente. Encuentran que, después de la apertura, los miembros novatos se vuelven más influyentes sobre el resto del comité y sobre Greenspan. Por tanto, aunque el efecto conformidad se da (novatos disienten menos), el hecho de que su discurso se vuelva más influyente sugiere que el efecto disciplina domina, ya que comunican con más contenido informativo, aunque lo hagan de forma menos confrontativa.

En concordancia con Acosta (2014) encuentran que mayor transparencia sí cambia como se comunican, pero no necesariamente empeora la calidad de la discusión interna.

A modo de síntesis, se puede concluir que los temas latentes sirven simultáneamente para dos propósitos, por un lado, son una forma de convertir texto en una variable cuantitativa que alimenta modelos econométricos, y, por otro lado, son una forma de auditar el propio discurso institucional (detectar cuándo cambia, qué tan disperso o consensuado está, y qué efectos secundarios puede tener la manera en que se comunica). La principal limitación de estos modelos es la misma que ya señalaba el propio Blei et al. (2003) para LDA y que aplica igual para LSA: ninguno de los dos piensa por sí solo, sino que organiza la evidencia para que alguien la interprete. LDA asume una estructura fija de temas y depende de que el investigador interprete correctamente qué representa cada uno; LSA por su parte, solo mide que tan parecidos son los documentos entre sí, dejando al investigador la tarea de darle sentido a esa similitud.

Complejidad

Los temas latentes revelan de qué habla un texto y qué tan estable es su agenda en el tiempo, pero dejan abierta una pregunta igual de importante: si ese mensaje realmente se entiende. De

eso se ocupa la complejidad: una medida de qué tan fácil o difícil resulta comprender un texto, es decir, qué tanto conocimiento o nivel educativo necesita un lector para entenderlo. A diferencia de los temas latentes, que capturan qué dice un texto, la complejidad captura cómo lo dice. Esta se calcula a partir de características observables del texto, como la longitud de las oraciones y de las palabras. Usando fórmulas estándar (basadas en el número de sílabas, palabras y oraciones) convierten esas características en un puntaje que equivale, aproximadamente, a los años de educación formal necesarios para comprender el texto. La complejidad de un documento reviste particular relevancia para los comunicadores (bancos centrales, gobiernos, instituciones), cuyo mensaje solo cumple su función si el público al que va dirigido realmente lo entiende.

Taborda (2015) aplica esta idea a bancos centrales latinoamericanos, calculando la legibilidad de 564 minutas de los bancos centrales de Brasil, Chile, Colombia, México y Perú. Encuentra diferencias notables entre países, donde México y Perú resultan los más legibles, y Brasil el menos, y más preocupante. Además, encuentra que la legibilidad de la región no mejoró de forma clara con el tiempo, lo que sugiere que los bancos no han priorizado este aspecto de su comunicación aun cuando publican cada vez más información. Benchimol et al. (2023) añaden una pieza clave a esta discusión al comparar al Banco de Israel con la Reserva Federal de Estados Unidos (FED) y el Banco Central Europeo (BCE). Encuentran que los comunicados del banco israelí requieren varios años menos de educación para comprenderse que los de esas otras dos instituciones, sin que eso implique sacrificar contenido informativo. Esto es relevante porque responde a una objeción común de que simplificar el lenguaje de un banco central equivale a diluir su mensaje técnico. La evidencia no respalda de forma concluyente esta premisa.

Vyshnevskyi et al. (2024) llevan esta pregunta un paso más allá al cuestionarse si la complejidad del texto tiene consecuencias reales sobre los mercados, no solo sobre la comprensión del lector. Con datos de 21 países en desarrollo entre 2010 y 2021, encuentran que los comunicados más complejos están asociados con mayor volatilidad del tipo de cambio en los días posteriores a su publicación, mientras que no encuentran el mismo efecto sobre la volatilidad de las tasas de interés, posiblemente porque el tipo de cambio puede reaccionar de inmediato a la incertidumbre generada por un mensaje ambiguo, mientras que las tasas se ajustan más gradualmente. Este hallazgo conecta directamente con el de Benchimol et al. (2023) de que la claridad no es solo una cortesía hacia el público, sino una variable con efectos medibles sobre la estabilidad financiera, particularmente relevante para economías en desarrollo como las de la región

centroamericana y la República Dominicana, donde el tipo de cambio suele ser una variable especialmente sensible.

Estos hallazgos ofrecen evidencia empírica de que la forma del lenguaje de un banco central u otras instituciones y, no solo su contenido, condiciona la manera en que dicha información llega a los agentes que la utilizan para la toma de decisiones. Estas decisiones a su vez se reflejan en el comportamiento de los mercados, por eso la claridad o complejidad de un comunicado puede llegar a predecir o dar indicios de la volatilidad que tienen estos. Otra afirmación contundente que se desprende de estos hallazgos es que, la claridad se puede medir, se puede comparar entre instituciones y en el tiempo, y no compite con la precisión técnica del mensaje, de hecho, parece reducir la incertidumbre que el propio mercado percibe.

III. Consideraciones finales

La evidencia revisada en esta nota muestra que la minería de texto no es una herramienta especulativa, sino que los índices de sentimiento e incertidumbre, los temas latentes y las métricas de complejidad se validan con resultados observables (errores de pronóstico, episodios históricos conocidos, movimientos del tipo de cambio o la tasa de interés) antes de emplearse con fines predictivos. Esto es lo que distingue a los índices de un simple conteo de palabras: no describen el texto, sino lo que el texto anticipa.

Un hallazgo que conviene subrayar para la región es que estos métodos no exigen economías grandes ni archivos de prensa para funcionar. Diakonova et al (2025). adaptaron con éxito la metodología de Baker et al (2016). a Centroamérica y República Dominicana pese a contar con menos prensa digitalizada, ajustando el diccionario a particularidades propias de la región (tipo de cambio, dolarización, corrupción, aranceles). Esto sugiere que la limitante no es la escala de la economía, sino que el instrumento se adapte al entorno que pretende medir.

Los temas latentes, por su parte, aportan una lección con matices. Que la agenda de una institución cambie no es, por sí mismo, una mala señal: el caso de Colombia durante la pandemia por el covid-19 muestra que un cambio de prioridades puede reflejar una respuesta legítima ante un choque, mientras que el caso de los discursos presidenciales latinoamericanos muestra que la volatilidad temática sostenida sí se asocia con menor crecimiento de largo plazo. Distinguir entre ambos escenarios es lo que le da valor práctico a esta herramienta.

En cuanto a la transparencia, la evidencia igual tiene matices: publicar más información no mejora automáticamente la deliberación interna, y en el caso de la FED, la publicación de transcripciones rejudo el desacuerdo expresado entre sus miembros. Sin embargo, esa misma evidencia muestra que la cautela en el discurso convivió con una mayor preparación de quienes tenían más en juego, lo que indica que ser menos confrontativo no necesariamente implica estar menos informado.

Definir adecuadamente el público meta es una condición necesaria para que la comunicación económica cumpla su función. Un acta, informe o comunicado puede contener información técnicamente precisa y, aun así, fallar como instrumento de comunicación si exige un nivel de conocimiento que excede al de sus destinatarios. La literatura sugiere que mayor legibilidad no necesariamente implica pérdida de rigor técnico y que esta puede emplearse como un indicador de alerta.

En conjunto, la evidencia analizada muestra que la minería de textos funciona como un enlace entre el lenguaje y la medición económica. Como señalan Grimmer y Stewart (2013), ningún modelo del lenguaje captura completamente la ironía o el contexto de la comunicación humana. Sin embargo, aunque ninguna de estas metodologías sustituye el razonamiento de un profesional, ofrecen una forma sistemática y replicable de calibrar la propia comunicación institucional (qué tanto se dice, sobre qué y que tan bien se entiende), algo que hasta ahora dependía casi por completo de la lectura manual. En este entorno, la minería de textos se consolida como una herramienta estratégica para comprender lo que ocurre en la economía y como los actores la perciben, se comunican y reaccionan ante ella.

Referencias

- Acosta, J. M. (2014). *FOMC responses to calls for transparency: evidence from the minutes and transcripts using latent semantic analysis*. Stanford, CA: Stanford University, Department of Economics. Retrieved from <https://purl.stanford.edu/qs086rg2081>
- Apel, M., & Blix Grimaldi, M. (2012). The information content of central bank minutes. SSRN. doi:<https://doi.org/10.2139/ssrn.2092575>

- Baker, S. R., Bloom, N. J., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593-1636. doi:<https://doi.org/10.1093/qje/qjw024>
- Benchimol, J., Caspi, I., & Kazinnik, S. (2023). Measuring communication quality of interest rate announcements. *The Economists' Voice*, 20(1), 43-53. doi:10.1515/ev-2022-0023
- Blei, D., Ng, A., & Jordan, M. (2003). *Latent Dirichlet Allocation*. doi:<https://dl.acm.org/doi/10.5555/944919.944937>
- Bua, G., Kapp, D., Ramella, F., & Rognone, L. (2024). Transition versus physical climate risk pricing in European financial markets: A text-based approach. *The European Journal of Finance*, 30(17), 2076-2110. doi:<https://doi.org/10.1080/1351847X.2024.2355103>
- Calvo-González, O., Eizmendi, A., & Reyes, G. (2018). *Winners never quit, quitters never grow: using text mining to measure policy volatility and its link with long-term growth in Latin America*. Poverty and Equity Global Practice Group. Washington, D.C.: World Bank Group (Banco Mundial). Retrieved from <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/634591516387264234>
- Chen, Y., Kimura, Y., & Inoue, K. (2025). How does managerial perception of uncertainty affect corporate investment during the COVID-19 pandemic: A text mining approach. *Pacific-Basin Finance Journal*, 90, 102655. doi:<https://doi.org/10.1016/j.pacfin.2024.102655>
- De la Parra, D., & Gallemore, J. (2026). The sound of uncertainty: Examining managerial acoustic uncertainty in conference calls. *Journal of Accounting and Economics*, *En prensa*. doi:<https://dx.doi.org/10.2139/ssrn.4883575>
- Diakonova, M., Ghirelli, C., & Quiñónez Wu, J. (2026). Economic policy uncertainty in Central America and the Dominican Republic. *Latin American Journal of Central Banking*, 7(2), 100166. doi:<https://doi.org/10.1016/j.latcb.2025.100166>
- Grimmer, J., & Stewart, B. (2013). *Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts*. doi:10.1093/pan/mps028

- Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and deliberation within the FOMC: A computational linguistics approach. *The Quarterly Journal of Economics*, 133(2), 801-870. doi:10.1093/qje/qjx045
- Hsu, M.-F., Chang, C., & Zeng, J.-H. (2022). Automated text mining process for corporate risk analysis and management. *Risk Management*, 24(4), 386-419. doi:https://doi.org/10.1057/s41283-022-00099-6
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35-65. doi:https://doi.org/10.1111/j.1540-6261.2010.01625.x
- Moreno-Pérez, C., & Minozzo, M. (2024). 'Making text talk': The minutes of the Central Bank of Brazil and the real economy. *Journal of International Money and Finance*, 147, 103133. doi:https://doi.org/10.1016/j.jimonfin.2024.103133
- Nyman, R., Kapadia, S., & Tuckett, D. (2021). News and narratives in financial systems: Exploiting big data for systemic risk assessment. *Journal of Economic Dynamics and Control*, 127, 104119. doi:https://doi.org/10.1016/j.jedc.2021.104119
- Roldan Ferrin, F. (2023). Comunicación de la política monetaria en Colombia: un análisis a través de la minería de texto. *Semestre Económico*, 26(60), 1-26. doi:10.22395/seec.v26n60a4431
- Shugliashvili, T., & Pirveli, E. (2025). Unveiling news impact on exchange rates: a hybrid model using NLP and LDA techniques. *Applied Economic Analysis*, 33(99), 184-205. doi:10.1108/AEA-04-2024-0183
- Sprenger, T. O., & Welpe, I. M. (2010). Tweets and trades: The information content of stock microblogs. *SSRN*. doi:https://doi.org/10.2139/ssrn.1702854
- Taborda, R. (2015). Procedural transparency in Latin American central banks under inflation targeting schemes. A text analysis of the minutes of the Boards of Directors. *Ensayos sobre Política Económica*, 33(76), 76-92. doi:10.1016/j.espe.2015.01.002

Vyshnevskyi, I., Jombo, W., & Sohn, W. (2024). The clarity of monetary policy communication and financial market volatility in developing economies. *Emerging Markets Review*, 59(101121). doi:10.1016/j.ememar.2024.101121