



**BANCO CENTRAL DE RESERVA DEL PERÚ**

# **Análisis textual y legibilidad de las Notas Informativas de la política monetaria en el Perú: 2001-2021**

**Erick Lahura\*, Luz Pumacahua\*\* y Rebeca Sosa\*\***

\* Banco Central de Reserva del Perú.

\*\* Universidad Científica del Sur.

**DT. N°. 2023-004**

**Serie de Documentos de Trabajo**

**Working Paper series**

**Agosto 2023**

Los puntos de vista expresados en este documento de trabajo corresponden a los de los autores y no reflejan necesariamente la posición del Banco Central de Reserva del Perú.

The views expressed in this paper are those of the authors and do not reflect necessarily the position of the Central Reserve Bank of Peru

# Análisis textual y legibilidad de las Notas Informativas de la política monetaria en el Perú: 2001-2021<sup>\*</sup>

ERICK LAHURA<sup>\*\*</sup>

Banco Central de Reserva del Perú  
erick.lahura@bcrp.gob.pe

LUZ MARIA PUMACAHUA<sup>\*\*\*</sup>

Universidad Científica del Sur  
180000591@cientifica.edu.pe

REBECA SOSA<sup>\*\*\*\*</sup>

Universidad Científica del Sur  
rsosab@cientifica.edu.pe

## Resumen

*El objetivo de esta investigación es analizar la legibilidad de la comunicación escrita de la política monetaria del Banco Central de Reserva del Perú (BCRP) y explorar sus determinantes. Para esto, se utilizan las notas informativas, las cuales son publicaciones mensuales en las que se comunica oficialmente las decisiones de política monetaria. El estudio de la legibilidad se realiza a través de la aplicación de técnicas de análisis textual a 252 notas informativas publicadas mensualmente desde febrero del 2001 hasta diciembre 2021. Con esta información se calcula el índice de legibilidad de Szigriszt Reading Ease (SRE) ajustado al idioma español por Szigriszt (1993). Luego, se exploran algunos determinantes potenciales de la legibilidad de las notas informativas a través de la estimación de modelos regresión dinámicos. La evolución del índice SRE indica que la legibilidad de las notas informativas se ha mantenido estable alrededor de un nivel normal durante el periodo de análisis. Por otro lado, los resultados de los modelos de regresión estimados muestran que la legibilidad es persistente, lo cual implica que esta no ha sufrido cambios drásticos en el tiempo, y que es relativamente mayor durante el primer semestre de cada año. Además, como es de esperarse durante coyunturas difíciles para cualquier banco central, se encuentra que la pandemia del Covid-19 y la aceleración de las expectativas de inflación y de la tasa de interés de referencia están asociadas a una reducción de la legibilidad de las notas informativas; sin embargo, la magnitud de dichos efectos ha sido pequeña y no ha desviado el nivel de legibilidad del rango normal.*

Palabras clave : Política monetaria, análisis textual, legibilidad, nota informativa, expectativas de inflación, tasa de interés de referencia, Covid-19.

Clasificación JEL : C32, C82, E52, E58

---

<sup>\*</sup> Los autores agradecen los comentarios y sugerencias de Adrián Armas, Carlos Montoro, César Ordoñez, Zenón Quispe, Víctor Roca, María Paula Vargas y Marco Vega, así como también a los participantes del XL Encuentro de Economistas 2022 del Banco Central de Reserva del Perú. Las opiniones expresadas en esta investigación corresponden a los autores y no reflejan necesariamente la posición de las instituciones a las cuales están afiliados.

<sup>\*\*</sup> Autor corresponsal. Subgerencia de Investigación Económica, Banco Central de Reserva del Perú (BCRP). Profesor del Departamento de Economía de la Pontificia Universidad Católica del Perú y del Departamento de Ingeniería Económica y Contabilidad Corporativa de la Universidad Científica del Sur.

<sup>\*\*\*</sup> Ingeniera Economista, Universidad Científica del Sur.

<sup>\*\*\*\*</sup> Ingeniera Economista. Docente del Departamento de Ingeniería Económica y Contabilidad Corporativa de la Universidad Científica del Sur.

# 1. Introducción

Desde mediados de los años 90, la comunicación de los bancos centrales ha evolucionado en términos de transparencia y rendición de cuentas, pasando de ser poco frecuente y escueta a ser más regular y extensa (Woodford, 2005; Issing, 2019). La publicación de comunicados de política monetaria, reportes de inflación y minutas de las reuniones de directorio forman parte del conjunto de medios que actualmente usan los bancos centrales para transmitir sus decisiones e información relevante para guiar las expectativas del mercado (Blinder et al., 2008). En este contexto, no es extraño que el interés por el estudio de la comunicación de los bancos centrales haya crecido en los últimos años.

Una característica común de las investigaciones recientes sobre la comunicación de los bancos centrales es el uso de técnicas de análisis textual o minería de textos,<sup>1</sup> las cuales permiten extraer y analizar información de documentos a través de herramientas estadísticas, computacionales y lingüísticas de una manera más objetiva y rápida en comparación al análisis que podría realizar un ser humano (Bholat et al., 2015). El análisis textual ha sido utilizado para estudiar diferentes aspectos de la política monetaria, como por ejemplo sus efectos sobre variables económicas y financieras (Lucca y Tebbi, 2009; Hendry y Madeley, 2010; Hansen y McMahon, 2016; Lahura y Vega, 2020; Aruoba et al., 2022; Doh et al., 2022, entre otros), las preferencias del banco central (Shapiro y Wilson, 2019), los efectos de la transparencia sobre las decisiones de política monetaria (Hansen et al., 2018) y la legibilidad de sus comunicaciones (por ejemplo Hernandez y Shell, 2014; Taborda, 2015; Tumala y Omotosho, 2019).

El objetivo de esta investigación es analizar la legibilidad de la comunicación escrita de la política monetaria del Banco Central de Reserva del Perú (BCRP) y explorar sus determinantes. Para esto, se analiza la Nota Informativa, una publicación mensual en las que se comunica las decisiones de política monetaria del BCRP. El estudio de la legibilidad se realiza a través de la aplicación de técnicas de análisis textual a 252 notas informativas publicadas desde febrero del 2001 hasta diciembre 2021. Con esta información se calcula el índice de legibilidad de Szigriszt Reading Ease (SRE) ajustado al idioma español por Szigriszt (1993). Adicionalmente, se exploran los potenciales determinantes de la legibilidad de las notas informativas a través de la estimación de diferentes modelos regresión dinámicos.

La evolución del índice SRE indica que la legibilidad de las notas informativas se ha mantenido estable alrededor de un nivel “normal” a lo largo del periodo considerado, resultados que son consistentes con los que encuentran Castillo et al. (2019) para las notas informativas publicadas entre diciembre de 2018 y setiembre de 2019 en el Perú. Por otro lado, los resultados de los modelos de regresión estimados indican que la legibilidad es persistente, lo cual implica que esta no ha sufrido cambios drásticos en el tiempo, y que es relativamente mayor durante el primer semestre de cada año. Además, se encuentra que durante coyunturas difíciles para cualquier banco central, como la pandemia del Covid-19 y la aceleración de las expectativas de inflación y de la tasa de interés de referencia, la legibilidad de las notas informativas se reduce; sin embargo, dicha disminución es pequeña y no desvía el nivel de legibilidad del rango normal.

---

<sup>1</sup>En la práctica, se usan ambos términos de manera indistinta; sin embargo, es posible establecer diferencias entre ambas expresiones. El sitio web de IBM <https://www.ibm.com/topics/text-mining> detalla ciertos matices entre ellas.

Esta investigación tiene tres contribuciones. Primero, aporta a la escasa literatura sobre el análisis de la comunicación del BCRP, la cual incluye a Taborda (2015), Castillo et al. (2019) y Lahura y Vega (2020). En segundo lugar, aporta a la creciente literatura que aplica técnicas de análisis textual al estudio de temas en política monetaria, como por ejemplo Lucca y Tebbi (2009), Hernandez y Shell (2014), Hansen y McMahon (2016), (Hansen et al., 2018), Omotosho (2019), Shapiro y Wilson (2019), Tumala y Omotosho (2019), Ahokposi et al. (2020), (Kuncoro et al., 2021), Carotta et al. (2021), Aruoba et al. (2022), Doh et al. (2022), entre otros. Finalmente, es la primera investigación que explora los determinantes de la legibilidad de las notas informativas del BCRP.

El documento está organizado en seis secciones. En la sección 2, se presenta el marco conceptual en el cual se analiza la literatura sobre la comunicación de los bancos centrales con énfasis en su legibilidad y en la aplicación de la técnica de minería de textos. En la sección 4, se describen los datos utilizados. En la sección 3 se detalla la metodología empírica para analizar la legibilidad y sus determinantes. En la sección 5 se presentan y discuten los resultados. Finalmente, las conclusiones se presentan en la sección 6. Como información adicional, el índice de legibilidad de este documento es 79.8, lo que significa que es “algo fácil” de leer.

## 2. Marco conceptual y revisión de la literatura

### 2.1. Comunicación de los bancos centrales

Antes de los años 90, se creía que la política monetaria era más eficaz si los bancos centrales minimizaban la comunicación y sus decisiones eran sorprendidas. Actualmente, el consenso es que una mayor transparencia de la política monetaria, que se logra a través de una mayor y mejor comunicación, es importante para mantener niveles bajos y estables de inflación (Rozkrut et al., 2007; Blinder et al., 2008; Sturm y De Haan, 2011; Siklos, 2013). Específicamente, si la comunicación del banco central logra que los agentes económicos entiendan claramente sus objetivos y estrategia, entonces contribuirá al anclaje de las expectativas de inflación (Cosimano y Van Huyck, 1993; Blinder, 1999; Bernanke y Reinhart, 2004; Sturm y De Haan, 2011; Berger et al., 2011; Cermeno y Navarrete, 2011; Oshima y Matsubayashi, 2018). Como lo afirman Bernanke y Reinhart (2004) “en los últimos años los bancos centrales se han esforzado por mejorar la comunicación con el público; un objetivo clave de este esfuerzo es un mejor alineamiento de las expectativas del mercado a la política monetaria” (p. 86). Existe evidencia que la comunicación de los bancos centrales ha contribuido a alcanzar los objetivos de la política monetaria (Woodford, 2005), y tiene efectos sobre las expectativas del mercado financiero (Rosa y Verga, 2007) y los precios de los activos (Rosa, 2013; McMahon et al., 2018).

Durante los últimos años, la comunicación de los bancos centrales ha aumentado en volumen y frecuencia, a través de informes de política monetaria, actas de las reuniones de política monetaria, sesiones informativas posteriores a las reuniones de directorio, discursos de los miembros del Comité de Política Monetaria, entre otras formas (Dow et al., 2007; Issing, 2005; Bulir et al., 2013; Ernst y Merola, 2018). Además de la cantidad de información transmitida, la claridad del mensaje de la política monetaria es muy importante pues permite que el público tenga una mejor comprensión del objetivo del banco central y de los factores que motivan las decisiones de política monetaria, lo que a su vez facilita su ejecución (Duisenberg, 2000). Fracasso et al. (2021) encuentran que la claridad de la comunicación es crucial y que mucha información puede aumentar la complejidad de los comunicados,

desanimar a los lectores y reducir la comprensión del contenido de la información. Jansen (2011), encontró que una mayor claridad de los testimonios del presidente de la FED estaba asociada a una menor volatilidad de los mercados financieros. Sin embargo, existen casos en los que los bancos centrales no siempre han transmitido un mensaje claro al público Bulir et al. (2013).

## 2.2. Análisis textual y comunicación de los bancos centrales

Las técnicas de análisis textual o minería de textos permiten extraer y analizar información de documentos a través de herramientas estadísticas, computacionales y lingüísticas de una manera más objetiva y rápida en comparación al análisis que podría realizar un ser humano (Bholat et al., 2015). Se asemeja a la lectura en el sentido de que ambas actividades consisten en extraer el significado de cadenas de letras. Sin embargo, el análisis textual tiene dos ventajas: (i) permite procesar y resumir mucho más texto del que cualquier persona podría hacerlo, y (ii) puede extraer significados no identificados por los lectores humanos, quienes pueden pasar por alto ciertos patrones porque no se ajustan a sus creencias y expectativas previas (Bholat et al., 2015; Tumala y Omotosho, 2019). Así esta técnica permite acceder a información no estructurada y hace posible que las búsquedas de información sean más rápidas y eficientes (Consuegra et al., 2016).

El análisis textual de las comunicaciones escritas de los bancos centrales ha sido utilizado para estudiar diferentes aspectos de la política monetaria, como por ejemplo sus efectos sobre variables económicas y financieras (Lucca y Tebbi, 2009; Hendry y Madeley, 2010; Hansen y McMahan, 2016; Oshima y Matsubayashi, 2018; Bennani, 2019; Park et al., 2019; Lahura y Vega, 2020; Aruoba et al., 2022; Doh et al., 2022, entre otros), las preferencias del banco central (Shapiro y Wilson, 2019), los efectos de la transparencia sobre las decisiones de política monetaria (Hansen et al., 2018) y la legibilidad de sus comunicaciones (por ejemplo Hernandez y Shell, 2014; Taborda, 2015; Tumala y Omotosho, 2019; Kuncoro et al., 2021).

## 2.3. Legibilidad de la comunicación de los bancos centrales

Según Ballesteros y Fernández (2013), la palabra legibilidad en el idioma español se utiliza para hacer referencia a dos términos ingleses: “*legibility*” y “*readability*”. A pesar de que estos términos son sinónimos desde el punto de vista semántico, el uso les ha otorgado significados diferentes. En la presente investigación, la palabra “legibilidad” hará referencia a la legibilidad lingüística del texto, es decir, al grado de complejidad inherente del mensaje escrito y al grado de facilidad para ser leído. En el anexo A se analiza la definición de legibilidad con más detalle.

Los trabajos que aplican el análisis textual para evaluar la legibilidad de la comunicación de la política monetaria son escasos e incluyen a Hernandez y Shell (2014), Taborda (2015), Omotosho (2019), Tumala y Omotosho (2019), Ahokpossi et al. (2020), Carotta et al. (2021) y (Kuncoro et al., 2021). Hernandez y Shell (2014) utilizan el índice Flesch-Kincaid para evaluar la claridad de las declaraciones del Comité Federal de Mercado Abierto (FOMC por sus siglas en inglés) durante un periodo 1994-2014 y encuentran que las declaraciones se han vuelto más complejas con el tiempo.

Tumala y Omotosho (2019) analizan 87 comunicados del Banco Central de Nigeria (BCN) en el periodo 2004-2019 y calculan el índice de legibilidad de Coleman y Liau (1975). Sus resultados muestran que el nivel de legibilidad de los comunicados ha disminuido ligeramente

a lo largo de los años, pasando de una media de 23.3% en el 2004 a 17.4% en el 2019, lo que indica una mayor complejidad de las estructuras de palabras y frases. Omotosho (2019) utiliza el mismo índice de legibilidad para analizar los comunicados de prensa del comité de política monetaria del Banco Central de Ghana (BCG) para el periodo 2018-2019 y encuentran que su lectura se ha hecho más fácil. Ahokpossi et al. (2020) usan el índice de legibilidad de Flesh-Kincaid para evaluar la claridad de 158 comunicados de prensa del Banco de Indonesia (BI) durante el periodo 2006-2018. Sus resultados revelan que la legibilidad ha disminuido en el tiempo debido a la presencia de oraciones largas y a estructuras complejas de redacción y su comprensión requiere al menos once años de educación.

Para el caso de Latinoamérica, Taborda (2015) analiza la legibilidad de los comunicados de política monetaria de los bancos centrales de Brasil, Chile, Colombia, México y Perú. La muestra estudiada abarca el período de metas explícitas de inflación en cada país hasta el 2013. Dado que los documentos analizados están publicados en español (Chile, Colombia, México y Perú) y en inglés (Brasil), se utilizaron los índices de legibilidad Szigriszt Reading Ease (SRE) y Flesch Reading Ease (FRE). Los resultados muestran que el índice de legibilidad se ha mantenido en un nivel estándar en todos los países durante periodo analizado, excepto en algunos casos específicos. Por ejemplo, Colombia alcanzó la categoría “bastante fácil” entre 2009 y 2011, a partir de entonces se mantuvo en niveles estándar.

Carotta et al. (2021) analizan la legibilidad de los comunicados del Banco Central de Uruguay utilizando tres índices: Fernández-Huerta, Pazos y Cantalejo y otros. Los resultados indican que la legibilidad de dichos comunicados se ubican principalmente en el rango “difícil” de leer (un comunicado se ubica en la categoría “muy difícil” y cuatro en la categoría de “algo difícil”). Si bien la dificultad ha disminuido en el tiempo, el tiempo de lectura ha aumentado. Además, encuentran que una mayor legibilidad de la comunicación refuerza el efecto del tono de la política monetaria.

Kuncoro et al. (2021) utiliza el índice de Flesch para medir la legibilidad de las comunicaciones del Banco Central de Indonesia y construye un índice de tono del mensaje. Al correlacionar estos indicadores con las expectativas de inflación (a uno, tres y doce meses), encuentra que una mayor legibilidad de la política monetaria está asociada a menores expectativas de inflación, mientras que una mayor claridad del tono esta asociada a mayores expectativas de inflación.

Por su parte, Castillo et al. (2019) analizan la comunicación de los bancos centrales de América del Sur, específicamente para 6 países (Brasil, Chile, Colombia, Uruguay, Paraguay y Perú). Como parte del estudio, analizan la legibilidad de los comunicaciones de política monetaria entre diciembre 2018 y setiembre 2019 usando los índices de Fernández-Huerta y Flesch-Szigriszt. Sus resultados muestran que la comunicación de Perú tiene los índices más altos de legibilidad, equivalentes a una exigencia de lectura de 7 años de escolaridad.

Como muestra la revisión de la literatura, los estudios sobre el análisis de la comunicación del BCRP son relativamente escasos, siendo Taborda (2015) el único que analiza la legibilidad de las notas informativas para una muestra larga. De esta forma, la presente investigación tiene tres contribuciones. Primero, aporta a la escasa literatura sobre el análisis de la comunicación del BCRP, la cual incluye a Taborda (2015), Castillo et al. (2019) y Lahura y Vega (2020). En segundo lugar, aporta a la creciente literatura que aplica técnicas de análisis textual para analizar la comunicación de la política monetaria como por ejemplo

Hernandez y Shell (2014), Tumala y Omotosho (2019), Omotosho (2019), Ahokpossi et al. (2020), Carotta et al. (2021), entre otros. Finalmente, es la primera investigación que explora los determinantes de la legibilidad de las notas informativas del BCRP.

### 3. Metodología empírica

#### 3.1. Análisis textual

Las técnicas de análisis textual que se describen a continuación fueron implementadas el Software Estadístico R (v4.2.3; R Core Team (2023)).

##### 3.1.1. Índice de legibilidad automatizado SRE

La legibilidad de los documentos publicados por el banco central contribuye en el nivel de transparencia de su comunicación, pues proporciona una medida del esfuerzo que requiere un lector para comprender un texto (Bruno, 2016). En este estudio utilizamos el índice Szigriszt Reading Ease (en adelante SRE) también conocido como el índice de perspicuidad, propuesto por Szigriszt (1993). Este índice, que es una readaptación de la fórmula de Flesch al español, permite medir el nivel de comprensibilidad de un texto escrito en español basándose en la cantidad de las palabras, oraciones y sílabas que tiene un documento. La fórmula es similar a la del índice Flesch Reading Ease pero con distintas ponderaciones.<sup>2</sup> El índice SRE toma valores en el intervalo 0 - 100 y se calcula de la siguiente manera:

$$SRE = 206.835 - 62.3 \times \frac{S}{P} - \frac{P}{O} \quad (1)$$

donde,  $S$  denota el número de sílabas,  $O$  el número de oraciones y  $P$  el número de palabras del documento. De la ecuación 1 se deduce que a mayor (menor) valor del índice SRE, mayor (menor) facilidad de lectura o mayor (menor) legibilidad del documento. El Cuadro 1 muestra Escala INFLESZ propuesta por Barrio et al. (2008), la cual asocia rangos de valores del índice SRE a niveles de legibilidad (facilidad de lectura), tipos de publicación y niveles de educación.<sup>3</sup>

**Cuadro 1.** Escala INFLESZ para la fórmula de Flesch-Szigriszt

Valor	Grado	Tipo de publicación
0 – 40	Muy difícil	Universitario, Científico
40 – 55	Algo difícil	Bachillerato, Divulgación científica
55 – 65	Normal	Prensa general, Prensa deportiva
65 – 80	Algo fácil	Educación primaria, Novelas de éxito
85 – 100	Muy fácil	Educación primaria, Comic

Fuente: Barrio et al. (2008).

Existen varios trabajos que calculan el índice de legibilidad de forma manual o utilizando softwares libres del internet. Sin embargo, si se quiere analizar una gran cantidad de textos

<sup>2</sup>El índice de Flesch Reading Ease (FRE) permite medir la legibilidad (facilidad de lectura) de un documento y ha sido diseñado para el idioma inglés.

<sup>3</sup>Según Gallego (2017), esta es la escala más utilizada.

(corpus), estas opciones se vuelven costosas. En este trabajo, el índice de legibilidad SRE se calcula de forma automatizada usando las funciones disponibles en el paquete `koRpus` de R (v0.13.8; Michalke (2021)),<sup>4</sup> el cual también identifica el nivel de escolaridad necesaria para comprender un texto.

### 3.1.2. Preprocesamiento de los datos textuales

El corpus o base de datos textual que se analiza está compuesto por 252 notas informativas publicadas por el BCRP en formato pdf, desde febrero 2001 hasta diciembre 2021. Antes de proceder con el análisis del texto, se realizó un preprocesamiento del corpus. El primer paso fue cambiar el formato de todas las notas informativas a “txt” para que el paquete `koRpus` pueda procesarlas.<sup>5</sup>

El siguiente paso fue realizar ajustes manuales para facilitar el análisis. Por un lado, se agregaron puntos a los títulos: un punto a las notas informativas antes de setiembre de 2003 (solo una oración) y dos puntos a las notas posteriores (dos oraciones). También se agregó punto final en la última oración de las notas que lo requerían y en algunas partes del contenido de la nota. Además, dado que casi todas las notas incluyen la fecha de la decisión de política monetaria, se incluyó la fecha a aquellas que no tenían para mantener un formato uniforme. Por otro lado, se eliminaron las secciones que contienen cuadros, lo que evita contabilizar palabras y números que se repiten; en total se eliminaron 10 cuadros. Además, se eliminaron algunos caracteres como las enumeraciones (por ejemplo 1. 2. 3. a. b. c. i. ii. iii., S/.) y puntos suspensivos para evitar que sean identificados como oraciones de una sola palabra y una sílaba. Finalmente, se borraron símbolos como % y los números de páginas que se incluían al final de algunas notas para evitar que se lean como palabras de una sílaba.

### 3.1.3. Cálculo del índice SRE con el paquete `koRpus`

El siguiente paso fue cargar el corpus preprocesado a R y aplicar la función `readability()` del paquete `koRpus`, para así medir la legibilidad a través del conteo de sílabas, oraciones y palabras. Para contar el número de palabras y oraciones de manera automatizada, el paquete utiliza la función `tokenize()` y para el conteo de sílabas utiliza un algoritmo de guiones a través de la función `hyphen()` (Michalke, 2021).

La secuencia del proceso para calcular las métricas de legibilidad implementadas en el paquete `koRpus` se puede describir en cuatro pasos. Primero, se importa el corpus preprocesado en formato “txt” utilizando la función de `tm.plugin.korpus()`. Segundo, se etiqueta el corpus con la función `korpus.lang.es()`.<sup>6</sup> Tercero, se realiza la “tokenización” del corpus usando la función `tokenizer()`.<sup>7</sup> Finalmente, se calculan las métricas de legibilidad usando la función `readability()`.

---

<sup>4</sup>El paquete `koRpus` de R es una herramienta versátil para el análisis textual, pues cuenta con decenas de fórmulas que permiten medir la legibilidad y la diversidad léxica. Brinda 35 métricas de legibilidad, utilizando parámetros habituales como: número de sílabas, número de oraciones, uso de palabras polisilábicas, uso de palabras “difíciles”, utilizando la función `readability()` (Ballier et al., 2017).

<sup>5</sup>La conversión de las notas a un formato txt no genera ninguna pérdida de información ni tampoco distorsiona el contenido.

<sup>6</sup>Esta función permite leer los documentos en el idioma que se desee, en nuestro caso se especificó que las notas sean leídas al idioma español (text language: “es”).

<sup>7</sup>Esta es una función alternativa a `TreeTagger` (Schmid, 1994) que se utilizó debido a que se tuvo problemas con la instalación de `TreeTagger`.



El paquete koRpus ofrece 35 métricas de legibilidad, de las cuales solo tres son para el idioma español: Flesch Reading Ease (FRE), Fernandez-Huerta y Fórmula de Comprensibilidad de Gutiérrez. Siguiendo la literatura, se optó por utilizar el índice de Flesch Reading Ease (FRE), que adaptado al idioma español recibe el nombre de Szigriszt Reading Ease (SRE).

Un aspecto importante del conteo de oraciones es su identificación a través de, por ejemplo, un signo de puntuación que las separe. La función *tokenizer()* ofrece al usuario la posibilidad de definir el signo de puntuación para identificar el final de una oración a través de la opción *sentc.end = c(“.”, “!”, “¿?”, “;”, “: ”)*. Si bien lo tradicional es usar el “punto” para separar oraciones, la RAE (s.f) indica que el “punto y coma” se usa para separar oraciones “sintácticamente independientes entre las que existe una estrecha relación semántica”. En esta investigación se optó por usar el “punto” y así evitar el análisis semántico requerido. Además, dado que, los signos de exclamación y pregunta (“!”, “¿?”) no se suelen utilizar en las notas informativas, las opciones razonables para identificar oraciones son “punto” y “punto y coma”.

Respecto a la separación y conteo de las sílabas utilizando la función *hyphen()*, es importante mencionar que se identificó que dicha función no separa los hiatos y las sílabas de algunas palabras con una estructura específica.<sup>8</sup> El paquete ofrece la función *correct.hyph()* para realizar correcciones manuales. Dado que este problema afecta en promedio solo al 5 % de las palabras de cada nota informativa y que el ajuste manual es costoso, se decidió continuar con la contabilización de la función *hyphen()* sin ajustes.

#### 3.1.4. Ventajas y limitaciones del índice de legibilidad SRE

El índice de legibilidad SRE tiene la ventaja de ser un indicador objetivo, pues: (i) se basa en parámetros fijos como número de palabras, oraciones y sílabas, que se calculan de forma automatizada, y (ii) no está sujeto a juicios de valor o errores humanos. Además, es un indicador de bajo costo pues se calcula de forma automatizada.

Sin embargo, el índice solo es una medida del nivel de comprensibilidad lingüística del texto, basada en las características textuales de los documentos, y no de la comprensión del contenido o “la semántica léxica” del texto.<sup>9</sup> El índice de legibilidad no considera ciertos aspectos como el público objetivo, el interés del lector, la dificultad de los conceptos, la coherencia del texto y otros que podrían afectar la comprensibilidad de los comunicados. Por ejemplo, en los comunicados de política monetaria se usan palabras técnicas que dificultan la comprensión del público en general a pesar de que el nivel de legibilidad lingüística sea alto: “Alrededor del 95 % de todas las palabras que expresan los bancos centrales en sus comunicados escritos son inaccesibles para alrededor del 95 % de la población” (Haldane, 2018). Por lo tanto, es posible que el contenido del texto sea comprensible a pesar de que la estructura lingüística (sílabas, palabras, oraciones) sea compleja, y viceversa (oraciones cortas que no aporten ninguna información significativa).

Además, es importante observar que el índice de legibilidad no considera la extensión del documento (longitud). En circunstancias normales, el sentido común indica que no es igual

---

<sup>8</sup>Por ejemplo, si la palabra comienza con vocal seguida de una consonante, la función tiende a juntar ambas en una sola oración, como en “aprobó”: apro-bó.

<sup>9</sup>De acuerdo con la RAE (2021) la semántica léxica es la rama de la semántica que estudia el significado de las palabras (expresiones lingüísticas) como también a las diversas relaciones de sentido que se establecen entre ellas.

leer un documento de una sola página que leer un documento de 10 páginas. De hecho, como señalan Fracasso et al. (2021) la longitud excesiva de los documentos podría desanimar a los lectores.

### 3.2. Determinantes de la legibilidad

Para explorar los determinantes de la legibilidad de las notas informativas, se estima el siguiente modelo de regresión dinámico:

$$\log(\text{leg}_t) = c + \sum_{j=1}^p \phi_j \log(\text{leg}_{t-j}) + x_t' \beta + u_t \quad (2)$$

donde  $\log(\text{leg}(t))$  es el logaritmo natural del índice de legibilidad (leg) en el período  $t$  y  $u_t$  es un proceso ruido blanco. Este modelo asume que la legibilidad depende de la legibilidad de los periodos previos; es decir, la legibilidad (en logs) tiene persistencia la cual se puede medir aproximadamente por la suma de los coeficientes autorregresivos:  $\rho = \sum_{j=1}^p \phi_j$ . El vector  $x_t'$  representa a las variables que explican la legibilidad, como el cambio en la tasa de interés de referencia y en las expectativas de inflación previas a la decisión de la tasa de interés.

El modelo (2) asume que las variables están representadas por series temporales estacionarias; para verificar ese supuesto, se aplica la prueba eficiente de raíz unitaria DF-GLS propuesta por Elliot et al. (1996).<sup>10</sup> La estimación del modelo 2 se realiza usando el estimador mínimos cuadrados ordinarios (OLS). La validez de cada modelo estimado depende de que los errores sean homocedásticos y no tengan autocorrelación. Para determinar esto, se aplica la prueba de autocorrelación Breusch-Godfrey (BG), que permite evaluar la hipótesis de que no existe autocorrelación hasta el orden  $p = 1, 2, \dots, 12$ , y la prueba de heterocedasticidad condicional ARCH(r), que permite evaluar la presencia de heterocedasticidad condicional de orden  $r = 1, 2, \dots, 12$ . Además, la bondad de ajuste de cada modelo estimado se determina identificando el mayor R2 ajustado y los menores valores para los criterios de Akaike (AIC) y bayesiano de Schwarz (SIC).

## 4. Datos

Las notas informativas del programa monetario se publican el mismo día de la reunión mensual del directorio del banco. Según lo que señala el BCRP (s.f.) estas notas cumplen un papel importante, ya que contienen una descripción de la decisión sobre la tasa de interés de referencia, el sustento de dicha decisión, la explicación de la evolución de la inflación, así como las tasas de interés para las operaciones del BCRP con el sistema financiero.<sup>11</sup>

Para realizar el análisis se han utilizado datos textuales procedentes de los comunicados de la decisión de política monetaria del BCRP (notas informativas) que están publicados a partir del 2001. Los documentos analizados abarcan el periodo febrero 2001 a diciembre 2021, en total se descargaron 252 notas,<sup>12</sup> las mismas que están disponibles para este intervalo de

<sup>10</sup>De manera complementaria se aplican las pruebas Dickey-Fuller aumentada (ADF), Phillip-Perron (PP) y Kwiatkowski, Phillips, Schmidt y Shin (KPSS).

<sup>11</sup>Ver: <https://www.bcrp.gob.pe>

<sup>12</sup>Según el periodo de tiempo que se ha tomado, deberían ser 251 notas informativas para analizar, pero se descargaron 252 notas de <https://www.bcrp.gob.pe/politica-monetaria/notas-informativas-del-programa-monetario.html>, ya que para el periodo de marzo-2021 se publicó dos notas informativas.

tiempo y en formato descargable en el sitio web del Banco Central, estos documentos se encuentran en formato pdf y en idioma español.

Por otra parte, para la estimación de los determinantes de la legibilidad se ha utilizado el índice de legibilidad mensual SRE que se ha obtenido producto del análisis textual; también se ha utilizado las series mensuales de las expectativas de inflación a 12 meses, desde enero de 2002 a diciembre de 2021,<sup>13</sup> así como la serie de la tasa de interés de referencia desde setiembre de 2003, las dos últimas series se encuentran disponibles en el sitio web del BCRP.

## 5. Resultados

### 5.1. Legibilidad y sus componentes

El Cuadro 2 muestra los estadísticos descriptivos de la variable legibilidad y sus componentes. Se observa que en promedio una nota informativa tiene 478 palabras, 19 oraciones y 917 sílabas. Además, la longitud promedio de una palabra es 2 sílabas y la de una oración es 26 palabras aproximadamente. Como resultado, el índice de legibilidad (leg) de las notas informativas registra un promedio de 61.25 puntos y un rango intercuartil de 63.69 - 58.98 = 4.71 puntos, lo que implica que la legibilidad se ha mantenido dentro del nivel normal.

**Cuadro 2.** *Estadísticos descriptivos de la legibilidad y componentes*

Variables	Obs.	Prom.	Desv. Est.	Min.	1st Qu.	Med.	3rd Qu.	Max.
pal	252	477.79	139.28	221.00	397.25	451.50	509.50	1,002
orac	252	18.65	4.63	11.00	16.00	18.00	20.00	38.00
sil	252	917.46	252.63	416.00	759.75	872.00	988.25	1,845
S/P*	252	1.93	0.05	1.797	1.89	1.92	1.96	2.08
P/O**	252	25.56	3.01	18.65	23.41	25.35	27.52	33.60
leg***	252	61.25	3.38	50.98	58.98	61.49	63.69	69.47

\*Ratio de sílabas por palabras.

\*\*Ratio de palabras por oraciones.

\*\*\*Índice de legibilidad comprendido entre muy difícil (0) - muy fácil (100).

Fuente: BCRP. Elaboración propia.

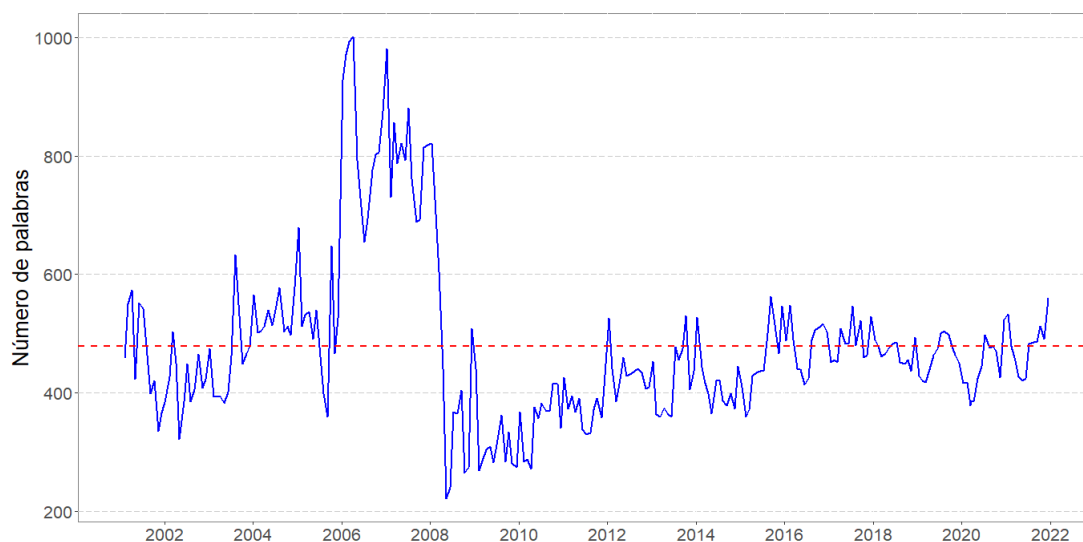
#### 5.1.1. Evolución del número de palabras y su complejidad

La Figura 1 muestra la evolución del número de palabras de las 252 notas informativas. La línea roja horizontal representa el número promedio de palabras por nota informativa, que es igual a 478. Como se puede observar, el comportamiento de esta serie ha sido muy variable y puede describirse en dos fases: (i) una tendencia creciente hasta antes de abr.2008, y (ii) una caída en el nivel promedio en abr.2008 (221 palabras, el promedio más bajo de la muestra), a partir del cual se observa una tendencia creciente que se estabiliza desde el año 2016. Sin embargo, cuando inició la pandemia del Covid-19 en mar.2020 esta serie registró una nueva caída (378 palabras, el menor número de palabras para dicho año), y luego fluctuó alrededor del promedio de la muestra. Según Taborda (2015), el BCRP es el que tiene las notas informativas más cortas (menor longitud de palabras) en comparación

<sup>13</sup>Las series de las expectativas de inflación se descargaron a partir del 2002 debido a la disponibilidad de datos.

con los bancos centrales que operan bajo un régimen de metas explícitas de inflación (Brasil, Chile, Colombia y México).

**Figura 1.** *Evolución del número de palabras de las notas informativas del BCRP (febrero 2001 - diciembre 2021)*



Fuente: BCRP. Elaboración propia.

El Cuadro 3 muestra la distribución de frecuencias de la longitud de las palabras medida por el número de sílabas. Como se observa, hay 120,401 palabras en todo el corpus, con una longitud mínima de 1 sílaba y máxima de 8 sílabas. En las notas informativas predominan las palabras de una sílaba (que incluyen “palabras vacías” o *stop words*),<sup>14</sup> con una frecuencia de 51.96 %. Le siguen las palabras de dos y tres sílabas con una frecuencia de 18.56 % y 17.81 % respectivamente. De esta forma, el 88.33 % de las palabras contenidas en las notas informativas tienen una longitud de entre 1 y 3 sílabas.

**Cuadro 3.** *Distribución de frecuencias de la longitud de las palabras medida en número de sílabas*

long. pal.	frec.	frec. acum.	%	% acum.
1	62,559	62,559	51.959 %	51.959 %
2	22,348	84,907	18.561 %	70.520 %
3	21,443	106,350	17.810 %	88.330 %
4	11,136	117,486	9.249 %	97.579 %
5	2,429	119,915	2.017 %	99.596 %
6	481	120,396	0.399 %	99.996 %
7	1	120,397	0.001 %	99.997 %
8	4	120,401	0.003 %	100.00 %

Fuente: BCRP. Elaboración propia.

<sup>14</sup>Las palabras vacías hacen referencia a artículos (el, la, etc.), preposiciones (de, en para, etc.) y pronombres (los, las, les, etc.)

La Figura 2 muestra nubes de palabras por longitud de palabras, a partir de las cuales se puede identificar las palabras más frecuentes. La lista es la siguiente: BCRP, mes, swap, PBI e IPC (una sílaba); ciento, tasa, meta, meses y rango (dos sílabas); inflación, interés, programa, temporal y créditos (tres sílabas); directorio, operaciones, monetaria, referencia y política (4 sílabas); indicadores, expectativas, determinantes, interbancario y volatilidad (5 sílabas); internacionales, metropolitana, transitoriamente, flexibilización y modificaciones (6 sílabas).

**Figura 2.** Nubes de palabras con diferentes longitudes



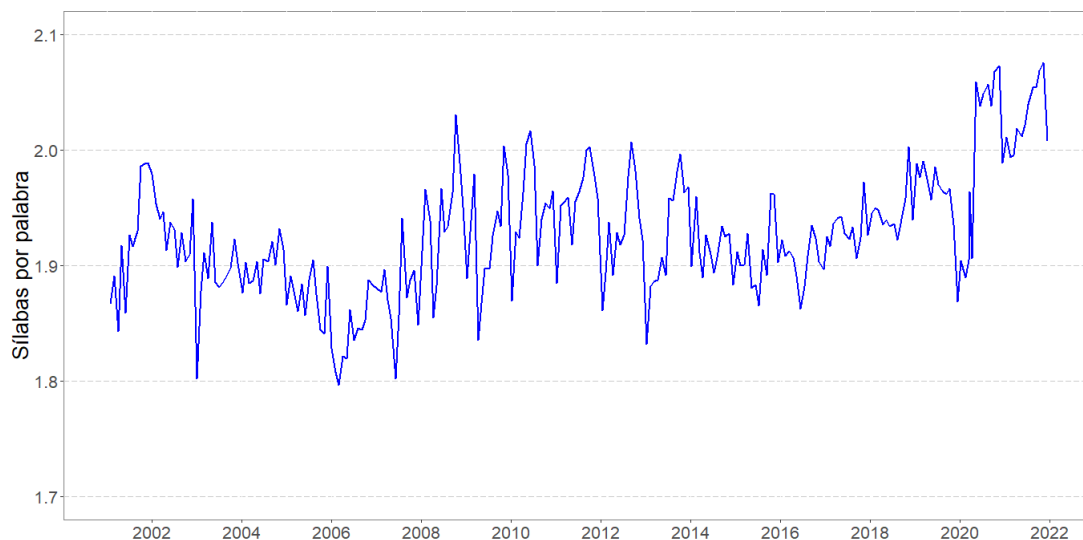
Fuente: BCRP. Elaboración propia.

Las palabras menos frecuentes tienen una longitud de 7 y 8 sílabas, y solo se han utilizado 5 veces en 20 años. Estas palabras son: (i) “considerablemente” (7 sílabas), que aparece en la nota informativa de mayo del 2020, (ii) “extraordinariamente” (8 sílabas), que aparece en la nota de abril del 2017 y “significativamente” (8 sílabas), que se encuentran en las notas de marzo, setiembre y octubre del 2020.

La Figura 3 muestra la evolución de la complejidad de las palabras, medidas como sílabas por palabra (SPP). Se observa una reducción de la complejidad de las palabras hasta inicios

del 2006, luego un comportamiento estable hasta inicios del 2016 y, posteriormente, una tendencia creciente. Además, se observa que la complejidad de las palabras se redujo entre los meses enero y abril del 2020.

**Figura 3.** *Evolución de la complejidad de las palabras (medida como sílabas por palabra)*

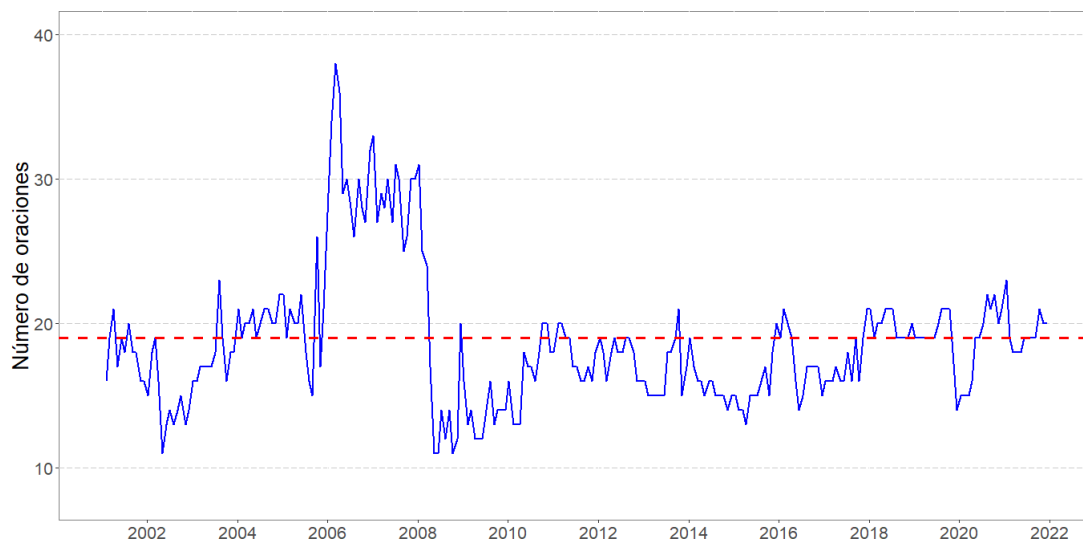


*Fuente: BCRP. Elaboración propia.*

### 5.1.2. Evolución del número de oraciones y su complejidad

La Figura 4 muestra la evolución del número de oraciones de las 252 notas informativas. La línea roja horizontal representa el número promedio de oraciones por nota informativa, que es aproximadamente igual a 19. Como se puede observar, la trayectoria está descrita por tres tramos: (i) hasta finales del 2005, un comportamiento fluctuante alrededor de un valor menor al promedio de la muestra, (ii) entre el 2006 e inicios del 2008, fluctuaciones alrededor de un valor mayor al promedio, y (iii) una caída en el nivel promedio desde inicios del 2008 y una ligera tendencia creciente desde entonces, la cual se estabiliza desde inicios del 2018. Sin embargo, se observa una caída temporal de la media entre los meses de enero y marzo del 2020.

**Figura 4.** *Evolución del número de oraciones de las notas informativas del BCRP (febrero 2001 - diciembre 2021)*



Fuente: BCRP. Elaboración propia.

El Cuadro 4 muestra la distribución de frecuencias de la longitud de las oraciones medida por el número de palabras. En todo el corpus existen 4,709 oraciones, con una longitud mínima de 4 palabras y máxima de 130 palabras.

**Cuadro 4.** *Distribución de frecuencia de la longitud de las oraciones medida en número de palabras*

Long. orac.	frec.	frec acum.	%	% acum.
4 - 10	660	660	14.02 %	14.02 %
11 - 20	1239	1899	26.31 %	40.33 %
21 - 30	1380	3279	29.31 %	69.63 %
31 - 40	799	4078	16.97 %	86.60 %
41 - 50	363	4441	7.71 %	94.31 %
51 - 60	162	4603	3.44 %	97.75 %
61 - 70	67	4670	1.42 %	99.17 %
71 - 80	19	4689	0.40 %	99.58 %
81 - 90	8	4697	0.17 %	99.75 %
91 - 100	5	4702	0.11 %	99.85 %
101 - 110	3	4705	0.06 %	99.92 %
111 - 120	2	4707	0.04 %	99.96 %
121 - 130	2	4709	0.04 %	100.00 %

Fuente: BCRP. Elaboración propia.

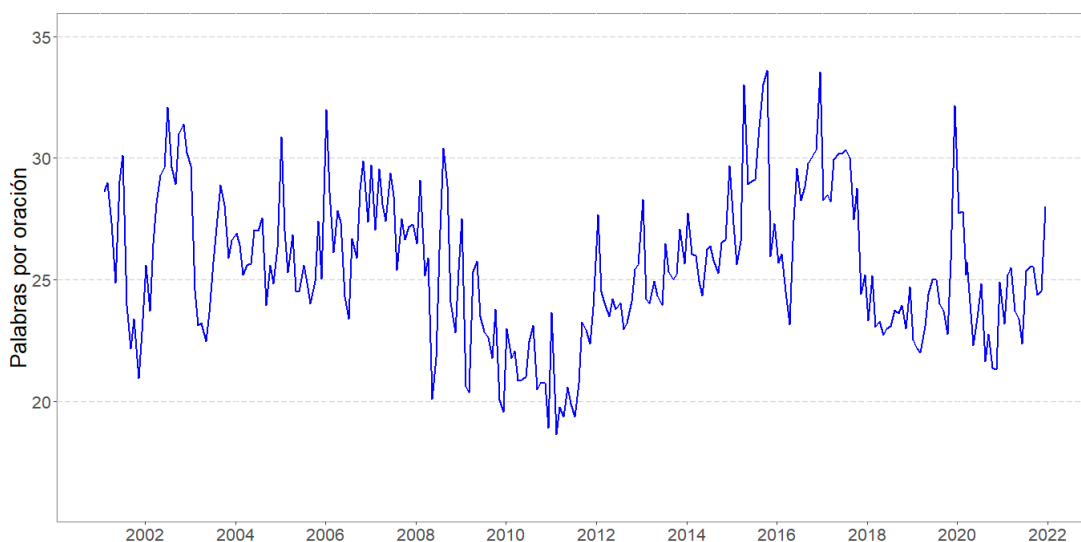
El 72.59 % de las oraciones tienen una longitud entre 11 y 40 palabras: 26.31 % tienen una longitud entre 11 y 20 palabras, 29.31 % entre 21 y 30 palabras, y 16.97 % entre 31 y 40 palabras. Las oraciones con menor número de palabras corresponden a los títulos de las notas informativas y a la fecha. Por ejemplo: “Programa Monetario: Marzo 2002” (4

palabras), “Programa Monetario de abril 2002” (5 palabras), “Lima 2 de febrero de 2001” (6 palabras). Por otra parte, encontramos oraciones muy largas con un rango de longitud entre 101 a 130 palabras; sin embargo, son poco frecuentes ya que solo representan un 0.14%. Por ejemplo, la nota informativa de abril del 2015 contiene una oración con 122 palabras y la de octubre del 2016 una oración con 130 palabras. El texto completo de estas oraciones se puede observar en el Anexo C.

Según Enago (s.f.), una buena regla general para mantener la longitud promedio de la oración es entre 20 y 25 palabras, las oraciones de más de 25 palabras no son accesibles y se vuelven cada vez más complejas de entender. Dado estos, los resultados indican que las notas informativas en promedio no utilizan oraciones complejas, pues casi el 70% de las oraciones tienen menos de 30 palabras.<sup>15</sup>.

La Figura 5 muestra la evolución de la complejidad de las oraciones, medida como palabras por oración (PPO). Se observa un comportamiento muy variable de la serie. Desde inicios de la muestra hasta mediados del 2008 la complejidad registró un comportamiento fluctuante alrededor de un valor promedio de 26.68 PPO. Luego la serie registró una tendencia decreciente, alcanzando su menor nivel en feb.2011 (18.6 PPO). Posteriormente, registró una tendencia creciente, alcanzando picos en oct.2015 (33.6 PPO, el mayor nivel en toda la muestra) y diciembre de 2016 (33.5 PPO); es decir, para estas fechas los comunicados se redactaron con oraciones largas. Luego la serie muestra una tendencia decreciente que se estabiliza desde inicios del 2018. Hacia el final de la muestra, se observa un pico en diciembre del 2019 (32.1 PPO).

**Figura 5.** *Complejidad de las oraciones (medida como palabras por oración)*



Fuente: BCRP. Elaboración propia.

<sup>15</sup>Cabe resaltar que el número varía según el campo, el público o la naturaleza de la redacción.



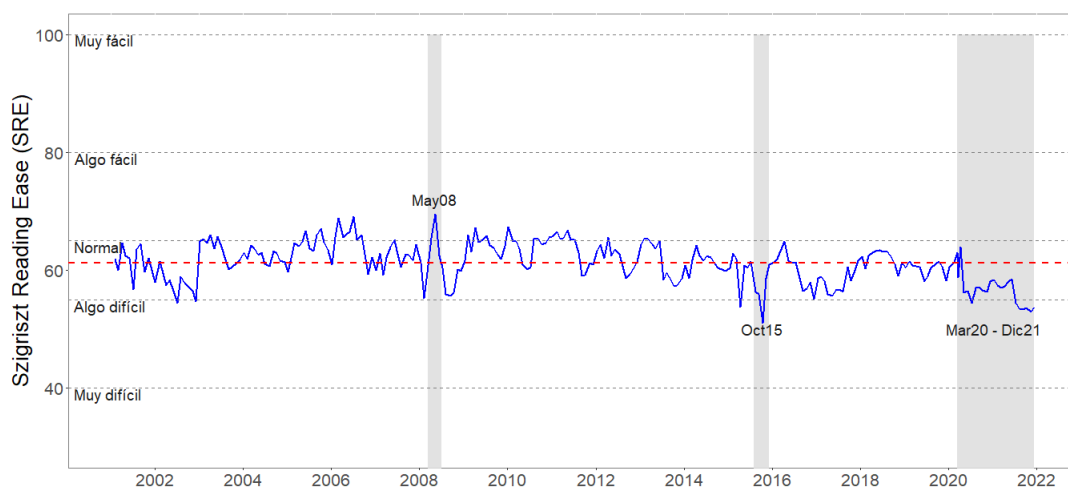
### 5.1.3. Evolución de la legibilidad de las notas informativas

Es importante recordar que el índice de legibilidad que se utiliza en este trabajo calcula el nivel de comprensibilidad de un documento basándose únicamente en las características textuales, es decir, en la dificultad de las palabras (medido como el número de sílabas por palabra) y en la dificultad de las oraciones (medido como el número de palabras por oración); sin embargo, no toma en cuenta la longitud de texto. Por ejemplo, si tenemos un documento con una sola oración versus otro documento con esa misma oración pero replicada 100 veces, el valor del índice va a ser el mismo para ambos documentos. Dentro del corpus existen notas que son largas (3 páginas); sin embargo, su índice es más alto en comparación al de una nota corta (1 página).

La fórmula del índice SRE (ver ecuación 1) captura la intuición de que muchas palabras por oración o muchas sílabas por palabra disminuyen la legibilidad (Bulir et al., 2013). En ese sentido si alguien tiene que procesar un texto con palabras u oraciones largas, será más difícil captar el mensaje.

En la Figura 6 se muestra la evolución del índice de legibilidad SRE para cada una de las notas informativas publicadas desde febrero del 2001 hasta diciembre del 2021. Se observa que el índice de legibilidad se ha mantenido estable a lo largo del tiempo, fluctuando alrededor de una media de 61.25 puntos (línea roja discontinua) y entre un rango de 59 y 63.7 puntos. Además, se observa una ligera tendencia decreciente que alcanza su nivel mínimo en octubre de 2015, y que luego se estabiliza. Finalmente, el promedio de legibilidad cae desde el inicio de la pandemia.

**Figura 6.** *Evolución del índice de legibilidad (febrero 2001 - diciembre 2021)*



*Nota:* El índice de SRE también puede interpretarse como el número de años de educación necesarios para comprender suficientemente un texto.

*Fuente:* BCRP. Elaboración propia.

La nota más fácil de leer es la del mes de mayo del 2008, cuyo índice es de 69.47 puntos (nivel algo fácil, se requiere en promedio de 8 a 9 años de escolaridad para comprender el texto) y es el más alto de la muestra. Por otro lado, la nota más difícil de leer fue la de

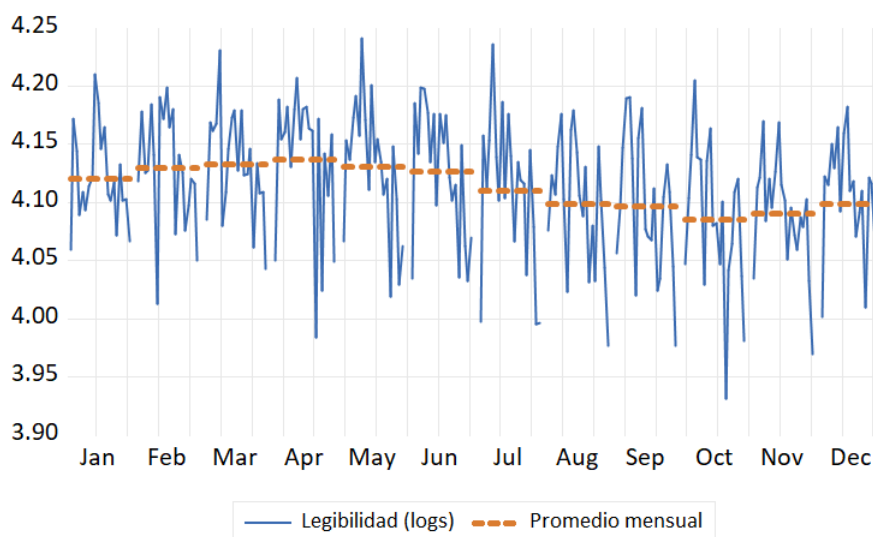
octubre del 2015, al registrar un nivel de legibilidad de 50.98 (algo difícil, requiere 10 años o más de escolaridad para comprender el texto), el más bajo de la muestra. Al analizar la fórmula del índice, se observa que el bajo nivel de legibilidad de esta nota se explica porque las PPO alcanzó el mayor nivel de la muestra (ver Figura 5); en promedio, las oraciones de esta nota contienen 34 palabras y es esta complejidad en las oraciones que hace que se reduzca el nivel de legibilidad tal y como señalan Bulir et al. (2013) y Ahokpossi et al. (2020).

A partir de la pandemia del Covid-19, se observa una caída en el nivel promedio de legibilidad, registrando valores entre normal y algo difícil desde mayo del 2020. Esta caída se explica por el hecho de que el número de SPP aumentó durante este periodo (ver Figura 3); es decir, los comunicados estaban redactados con palabras largas. Esto fue compensado por la disminución de las PPO durante el mismo periodo (ver Figura 5).

#### 5.1.4. Estacionalidad de la legibilidad

La Figura 7 muestra un gráfico de estacionalidad mensual de la serie de legibilidad (expresada en logaritmos), donde la línea horizontal representa el promedio histórico de la legibilidad para cada mes. Se observa que abril es el mes con el mayor nivel de legibilidad promedio, mientras que octubre es el mes que tiene el menor nivel promedio. Además, se evidencia que la legibilidad, en promedio, es mayor durante los primeros semestres del año (enero a junio) en comparación a los meses del segundo semestre.

**Figura 7.** Estacionalidad mensual de la serie de legibilidad  
(enero 2002 - diciembre 2021)



Fuente: BCRP. Elaboración propia.

## 5.2. Determinantes de la legibilidad

Los resultados de las pruebas de raíz unitaria aplicadas a las series temporales que representan a *logleg*, *dexp* y *dref* se detallan en el cuadro C-1 del anexo B. En todos los casos, las pruebas indican que las series son estacionarias, lo que valida la estimación del

modelo (2) usando mínimos cuadrados ordinarios. El cuadro 5 muestra los resultados de la estimación de diferentes especificaciones dinámicas para todo el periodo de metas explícitas de inflación (enero 2002 - diciembre 2021).

**Cuadro 5.** Resultados de las estimaciones por OLS (2002m2 - 2021m12)

<b>Variable dependiente:</b> Logaritmo de la legibilidad (logleg)						
	<b>Modelo1</b>	<b>Modelo2</b>	<b>Modelo3</b>	<b>Modelo4</b>	<b>Modelo5</b>	<b>Modelo6</b>
logleg(-1)	0.695*** (0.045)	0.622*** (0.048)	0.622*** (0.048)	0.575*** (0.049)	0.597*** (0.051)	0.590*** (0.051)
logleg(-11)	0.122** (0.048)	0.091** (0.048)	0.091** (0.048)	0.104** (0.047)	0.111** (0.050)	0.116** (0.050)
esta	0.013** (0.005)	0.015*** (0.005)	0.015*** (0.005)	0.017*** (0.005)	0.012** (0.005)	0.013** (0.005)
covid		-0.033*** (0.009)	-0.033*** (0.009)	-0.030*** (0.009)	-0.030*** (0.009)	-0.029*** (0.009)
dvel			-0.015*** (0.005)	-0.016** (0.006)	-0.015** (0.006)	-0.015** (0.006)
dexp(-1)				-0.036** (0.016)		-0.020 (0.017)
dref					-0.021* (0.011)	-0.018 (0.011)
Observ.	229	229	229	229	219	219
R2	0.590	0.613	0.613	0.633	0.637	0.640
R2 aj.	0.584	0.606	0.606	0.623	0.627	0.628
AIC	-3.778	-3.826	-3.826	-3.862	-3.881	-3.878
SIC	-3.718	-3.751	-3.751	-3.757	-3.772	-3.754
<b>Autocorrelación</b>						
BG(1)	0.480	0.530	0.530	0.650	0.970	0.910
BG(2)	0.640	0.630	0.630	0.490	0.600	0.500
BG(3)	0.780	0.790	0.790	0.670	0.770	0.690
<b>Heterocedasticidad</b>						
ARCH(1)	0.480	0.540	0.540	0.650	0.970	0.920
ARCH(2)	0.650	0.630	0.630	0.500	0.610	0.510
ARCH(3)	0.790	0.800	0.800	0.680	0.780	0.710

*Nota:* Estimaciones de los autores. Resultados de la regresión por mínimos cuadrados con información desde enero de 2002 para los modelos del 1 al 4 y desde setiembre de 2003 para los modelos 5 y 6.

Niveles de significancia: \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Errores estándar robustos entre paréntesis.

Además de los rezagos del logaritmo de la legibilidad, *logleg*, los modelos estimados incluyen otras variables relacionadas a la política monetaria en el Perú: *esta*, *covid*, *dvel*, *dexp(-1)* y *dref*. La primera variable es una dummy que captura la estacionalidad de la serie de legibilidad, toma el valor de 1 en todos los meses que pertenecen al primer semestre de cada año. Por su parte, la variable *covid* es una dummy que toma el valor de 1 desde abril 2020 en adelante, y captura el efecto de la pandemia del Covid-19. *dvel* es también una variable dummy igual a 1 desde enero del 2007 hasta diciembre 2021, y captura el efecto de tener una nueva meta de inflación, al pasar de un rango de 1.5% - 3.5% al rango actual de 1% - 3%; además, coincide con las tres gestiones consecutivas de Julio Velarde como

presidente del directorio del BCRP.<sup>16</sup>  $dexp(-1)$  representa la aceleración de las expectativas de inflación observada el periodo anterior, definida como  $\pi_{t-1}^e - \pi_{t-2}^e$ .  $dref$  representa el cambio mensual en la tasa de interés de referencia.

En todos los modelos estimados, la prueba de autocorrelación Breusch-Godfrey (BG) y la prueba de heterocedasticidad condicional autorregresiva (ARCH) indican que no hay indicios de autocorrelación hasta el orden  $p = 1, 2, \dots, 12$  o de heterocedasticidad condicional de orden  $r = 1, 2, \dots, 12$ .<sup>17</sup> Si bien la muestra completa está compuesta por 240 observaciones, solo se usan 229 observaciones debido a que se considera el rezago 11 de la variable legibilidad como regresor. Además, los modelos 5 y 6 incluyen solo 219 observaciones pues la tasa de interés de referencia se empieza a usar como instrumento de política desde setiembre del 2003.

El modelo 1 representa la mejor especificación autorregresiva, la cual incluye los rezagos 1 y 11 del logaritmo de la legibilidad. En este caso, se observa que la legibilidad tiene un alto grado de persistencia ( $0.70 + 0.12 = 0.82$ ). Esto significa que la legibilidad de cada nota informativa está determinada en gran medida por la legibilidad de los meses previos, especialmente la del mes previo (0.70). Al incluir otros regresores (modelos 2 al 6), se observa que la persistencia del logaritmo de la legibilidad se mantiene por encima de 0.70 y es estadísticamente significativa. Adicionalmente, en este modelo se incluye la variable de estacionalidad *esta* que es igual a 0.013 y es estadísticamente significativa al 5%. Esto indica que la legibilidad de la nota informativa aumentó en 1.3 por ciento en los meses del primer semestre a comparación de los meses del segundo semestre. Al incluir otros regresores (modelos 2 al 6), se observa que el efecto marginal de esta variable se mantiene estable y estadísticamente significativa.

El modelo 2 incluye la variable dummy *covid*, la cual es igual a -0.033 y estadísticamente significativa al 1%. Esto significa que durante la pandemia la legibilidad de la nota informativa disminuyó en 3.3 por ciento. Al incluir otros regresores (modelos 3 al 6), se observa que el efecto marginal de *covid* se mantiene estable entre 2.9% y 3.3% y es estadísticamente significativo al 1% en todos los modelos.

El modelo 3 incluye la variable dummy *dvel*, la cual es igual a -0.015 y estadísticamente significativa al 1%, lo que indica que la legibilidad de la comunicación disminuyó en 1.5% en comparación al periodo previo a enero 2007. Este efecto se mantiene estable y estadísticamente significativo al incluir el cambio en las expectativas de inflación (1.6% y estadísticamente significativo al 5%) o el cambio en la tasa de interés de referencia (1.5% y estadísticamente significativo al 5%).

El modelo 4 incluye la variable  $dexp(-1)$ , la cual es igual a -0.036 y estadísticamente significativa al 5%, lo que implica que una aceleración de las expectativas de inflación equivalente a 0.15 puntos porcentuales (una desviación estándar) reduce la legibilidad en  $(0.15) \times (0.36) = 0.54\%$ . Este efecto es relativamente pequeño pues, dado que la legibilidad promedio es 61.25 y la desviación estándar 3.38, la reducción de la legibilidad es de 0.33 puntos, equivalente a menos del 10% de una desviación estándar.

El modelo 5 incluye la variable *dref*, la cual es igual a -0.021 y estadísticamente significativa al 10%, lo que indica que una aceleración de la tasa de referencia equivalente a 0.25

---

<sup>16</sup>Julio Velarde juramentó como presidente del BCRP el 17 de octubre del 2006.

<sup>17</sup>Solo se muestran los resultados de los rezagos 1, 2 y 3 para ambas pruebas.

puntos porcentuales (una desviación estándar) reduce la legibilidad en  $0.25 \times (0.021) = 0.5\%$ , efecto similar al de la expectativas de inflación.

Finalmente, en el modelo 6 se incluyen simultáneamente las variables  $dexp(-1)$  y  $dref$  las cuales son individualmente significativas solo al 10% y a una cola, al igual que la variable  $dvel$  es significativa al 5%. Sin embargo, al analizar la significancia conjunta de estas dos variables, se obtiene que en conjunto son estadísticamente significativas al 5% ( $F(2,211) = 3.26$  y  $p\text{-value} = 0.023$ ). Los estimados indican que una aceleración de la inflación y de la tasa de interés de referencia reducen la legibilidad en  $(0.15) \times 0.02 = 0.3\%$  y  $0.25 \times (0.018) = 0.5\%$ , respectivamente. Estos resultados muestran que si bien la pandemia del Covid-19, la aceleración de las expectativas de inflación y de la tasa de interés de referencia tienen un efecto negativo sobre la legibilidad de las notas informativas, la magnitud de dichos efectos es pequeña.

Durante coyunturas difíciles para un banco central como un choque exógeno severo como la pandemia del Covid-19 o una aceleración de las expectativas de inflación o de la tasa de interés, es razonable observar una reducción de la legibilidad de la comunicación. Esto se debe que el banco central tiene que transmitir y explicar más, lo que naturalmente conduce a un aumento del número de palabras por oración (oraciones más complejas). También es posible que aumente el número de sílabas por palabra (complejidad de las palabras), dado que se tiene que introducir palabras “complejas”, como por ejemplo “pandemia”. La disminución de la legibilidad en periodos de crisis o incertidumbre son consistentes con los resultados obtenidos por Tumala y Omotosho (2019).

También es importante señalar que para los lectores podría ser más “comprensible” la explicación de una situación difícil por parte del banco central si este usa más palabras por oración o más sílabas por palabra; sin embargo, esto no es capturado por el índice de legibilidad utilizado. Por ello, sería importante complementar este índice con medidas que capturen la semántica.

Al comparar los modelos con 229 observaciones (modelos 1 al 4), el modelo con mejor ajuste es el modelo 4 debido a que presenta el mayor  $R^2$  ajustado y los menores AIC y SIC. Al comparar los modelos 5 y 6, se observa que el modelo 6 es mejor según el  $R^2$  ajustado y el modelo 5 es mejor según AIC y SIC. Como se observa en el Anexo C-2, cuando se estiman todos los modelos con información desde setiembre de 2003, los resultados no cambian significativamente y el modelo con el mejor ajuste es el modelo 5, seguido del modelo 4.

En resumen, los resultados muestran que la legibilidad del BCRP está explicada principalmente por el componente de persistencia, lo cual implica que esta no ha sufrido cambios drásticos en el tiempo, y que es relativamente mayor durante el primer semestre de cada año. También se encuentra que la pandemia del Covid-19 y la aceleración de las expectativas de inflación y de la tasa de interés de referencia están asociadas a una reducción de la legibilidad de las notas informativas, resultado esperado para cualquier banco central que enfrenta coyunturas difíciles similares. Sin embargo, el nivel de legibilidad del rango normal se ha mantenido dentro del rango normal pues dicha reducción fue pequeña.

## 6. Conclusiones

El objetivo de esta investigación fue analizar la legibilidad de la comunicación escrita de la política monetaria del Banco Central de Reserva del Perú (BCRP) y explorar sus determinantes. Para ello, se aplicaron técnicas de análisis textual a 252 notas informativas publicadas mensualmente desde febrero del 2001 hasta diciembre 2021 y se exploraron sus determinantes a través de la estimación de diferentes modelos de regresión dinámicos.

El análisis del índice de legibilidad SRE de las notas informativas muestra que la legibilidad se ha mantenido estable a lo largo del tiempo y dentro del nivel normal, fluctuando entre un rango de 63.7 a 59 puntos. Por su parte, los resultados de la estimación de los diferentes modelos dinámicos indican que la legibilidad del BCRP ha sido persistente, es decir, la legibilidad de cada nota informativa ha estado determinada en gran medida por la legibilidad de los meses previos, lo que implica que la legibilidad de su comunicación no ha sufrido cambios drásticos en el tiempo. Además, se encuentra que la legibilidad es relativamente mayor durante el primer semestre de cada año.

Respecto a los determinantes de la legibilidad, los resultados indican que la pandemia del Covid-19 y la aceleración de las expectativas de inflación y de la tasa de interés de referencia tienden a reducir la legibilidad de las notas informativas; sin embargo, el nivel de legibilidad no se desvía del rango normal. Además, el sentido de dichos efectos es el esperado, ya que cuando un banco central tiene que anunciar situaciones difíciles como las anteriores, es probable que se reduzca la legibilidad en la redacción debido a que el banco central tiene que explicar más, lo que naturalmente puede inducir un aumento del número de palabras por oración (oraciones más complejas) y número de sílabas por palabra (complejidad de las palabras). Hasta donde se tiene conocimiento, estos resultados son novedosos para el caso peruano.

Finalmente, es importante señalar tres limitaciones del trabajo que naturalmente motivarán nuevas investigaciones. La primera es que el análisis no considera la extensión del documento (longitud) para evaluar la legibilidad. En circunstancias normales, el sentido común indica que no es igual leer un documento de una sola página que leer un documento de 10 páginas. Además, como señalan Fracasso et al. (2021) la longitud excesiva de los documentos podría desanimar a los lectores. Por ello, una extensión natural de la presente investigación sería incorporar la longitud de las notas. La segunda limitación es que el trabajo no analiza la contribución de la legibilidad de las notas informativas en la efectividad de la política monetaria, lo cual permitiría evaluar la importancia de la comunicación escrita en la estrategia de la política monetaria. La tercera es que para los lectores podría ser más “comprensible” la explicación de una situación difícil por parte del banco central si este usa más palabras por oración o más sílabas por palabra; sin embargo, esto no es capturado por el índice de legibilidad. Por ello, sería importante complementar este índice con medidas que capturen la semántica.

Como información adicional, el índice de legibilidad de este documento es 79.8, lo que significa que es “algo fácil” de leer.

## REFERENCIAS

- Ahokpossi, C., Isnawangsih, A., Naoaj, S. and Yan, T. (2020). The Impact of Monetary Policy Communication in an Emerging Economy: The Case of Indonesia. Working Paper No. 20, 109.
- Aruoba, S. B. and Drechsel, T. (2022). Identifying Monetary Policy Shocks: A Natural Language Approach. CEPR Discussion Paper No. DP 17133. <https://ssrn.com/abstract=4069942>
- Ballesteros, S. y Fernández, I. (2013). Análisis de la legibilidad lingüística de los prospectos de los medicamentos mediante el índice de Flesch-Szigriszt y la escala Inflesz. In *Anales del Sistema Sanitario de Navarra*, 36(3), 397-406. Gobierno de Navarra. Departamento de Salud. <https://dx.doi.org/10.4321/S1137-66272013000300005>
- Ballier, N., Lissón, P. y Trujillo-González, V. (2017). Introducción al uso del paquete koRpus de R para el tratamiento de la diversidad léxica y la legibilidad en corpus lingüísticos. Université Paris Diderot (USPC). <https://hal.archives-ouvertes.fr/hal-01673712/document>
- Baquadano, M. M. (2006). Legibilidad y variabilidad de los textos. *Boletín de Investigación Educativa* [Artículo de Revista], 21, 13-25. <http://legibilidadmu.cl/1.pdf>
- Barrio, I., Simón, P., Melguizo, M., Escalona, I., Marijuán, M., y Hernando, P. (2008). Validación de la escala INFLESZ para evaluar la legibilidad de los textos dirigidos a pacientes. *Anales del Sistema Sanitario de Navarra*, 31(2), 135-152.
- Banco Central de Reserva del Perú (s.f.). Notas Informativas del Programa Monetario. <https://www.bcrp.gob.pe/politica-monetaria/notas-informativas-del-programa-monetario.html>
- Bennani, H. (2019). Does People's Bank of China communication matter? Evidence from stock market reaction. *Emerging Markets Review*, 40, 100617. <https://doi.org/10.1016/j.ememar.2019.05.002>
- Berger, H., Ehrmann, M. and Fratzscher, M. (2011). Monetary Policy in the Media. *Journal of Money, Credit and Banking*, 43(4), 689-709. <https://www.jstor.org/stable/20870072>
- Bernanke, B. S. and Reinhart, V. R. (2004). Conducting Monetary Policy at very Low Short-Term Interest Rates. *American Economic Review*, 94(2), 85-90. <https://www.jstor.org/stable/3592862>
- Bholat, D., Hansen, S., Santos, P. and Schonhardt-Bailey, C. (2015). Text mining for central banks: handbook. *Centre for Central Banking Studies, Bank of England*, (33). pp. 1-19. <http://eprints.lse.ac.uk/62548/>
- Blinder, A. S. (1999). *Central Banking in Theory and Practice*. MIT Press paperback edition. <http://www.artsrn.ualberta.ca/econweb/landon/1999%20Blinder%20Central%20Banking.pdf>
- Blinder, A. S., Ehrmann, M., Fratzscher, M., de Haan, J. and Jansen, D.-J. (2008). Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence. *Journal of Economic Literature*, 46(4), 910-945. <http://www.jstor.org/stable/27647085>

- Bruno, G. (2016). Text Mining and Sentiment Extraction in Central Bank Documents. *IEEE International Conference on Big Data (Big Data)*. <https://ieeexplore.ieee.org/document/7840784>
- Bulír, A., Čihák, M. and Jansen, DJ. (2013a). What Drives Clarity of Central Bank Communication About Inflation? *Open Economies Review*, 24, 125–145. <https://doi.org/10.1007/s11079-012-9259-z>
- Bulír, A., Čihák, M. and Šmídková, K. (2013b). Writing Clearly: The ECB's Monetary Policy Communication. *German Economic Review*, 14(1), 50-72. <https://doi.org/10.1111/j.1468-0475.2011.00562.x>
- Castillo, P., Herrada, R., Montoro, C. y Perez, F. (2019). La comunicación de la política monetaria en los bancos centrales de América del Sur. Documento de trabajo del BCRP. <https://www.bcrp.gob.pe/docs/Publicaciones/Documentos-de-Trabajo/2019/documento-de-trabajo-011-2019.pdf>
- Carotta, G., Mello, M. and Ponce, J. (2021). Monetary policy communication and inflation expectations: new evidence about tone and readability. <https://www.cemla.org/actividades/2021-final/2021-11-xxvi-meeting-of-the-central-bank-researchersnetwork/5B.2%20PAPER%20Carotta%20Mello%20Ponce.pdf>
- Cermeno, R. y Navarrete, C. (2011). ¿Que tan predecible es la política monetaria? Un análisis econométrico del contenido informativo de los comunicados del Banco de México. Documento de trabajo No 514.
- Consuegra, A. A., Salazar, Y. M., Garc, J. H., y Vizcaino, D. H. (2016). Minería de texto como una herramienta para la búsqueda de artículos científicos para la investigación. *Investigación y desarrollo en TIC*, 7(1), 14-20. <http://revistas.unisimon.edu.co/index.php/identic/article/view/2502>
- Cosimano, T. and Van Huyck, J. (1993). Central Bank Secrecy, Interest Rates, and Monetary Control. *Economic Inquiry, Western Economic Association International*, 31(3), 370-382.
- Doh, T., Song, D. and Yang, S. (2022). Deciphering Federal Reserve Communication via Text Analysis of Alternative FOMC Statements. Federal Reserve Bank of Kansas City, Research Working Paper No. 20-14, October. <https://doi.org/10.18651/RWP2020-14>
- Dow, S., Klaes, M. and Montagnoli, A. (2007). Monetary Policy by Signal. In: Mayes D and Toporowski J (eds.) *Open Market Operations and the Financial Markets*. First ed. Routledge International Studies in Money and Banking, 40. Abingdon, UK: Routledge, pp. 264-280. <http://www.routledge.com/books/details/9780415417754/>
- Duisenberg F. (2000). Making monetary policy in a broad monetary union. *BIS Review*, 88. <https://www.bis.org/review/r001018a.pdf>
- Elliott, G., Rothenberg, T. J. y Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64, 813-836. <https://doi.org/10.2307/2171846>
- Enago (s.f.). Los 6 mejores consejos para optimizar la longitud de las oraciones en su trabajo de investigación. <https://www.enago.com/es/academy/how-to-optimize-sentence-length-in-academic-writing/>



- Ernst, E. and Merola, R. (2018). Central bank communication: A quantitative assessment. Research Department, International Labour Organization. Working Paper (33).
- Fracasso, A., Genberg, H. and Wyplosz, C. (2003). How do Central Banks Write? An Evaluation of Inflation Targeting Central Banks. Geneva Reports on the World Economy Special Report 2. [https://cepr.org/sites/default/files/geneva\\_reports/GenevaP161.pdf](https://cepr.org/sites/default/files/geneva_reports/GenevaP161.pdf)
- Gallego, A. (2017). Análisis de algoritmos para determinar el nivel de complejidad de textos sanitarios y recomendaciones para mejorar el empoderamiento de un paciente. Documento de trabajo. Universidad Politécnica de Valencia.
- Hansen, S. and McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99, 114-133. <https://doi.org/10.1016/j.jinteco.2015.12.008>
- Hansen, S., McMahon, M. and Prat, A. (2018). Transparency and deliberation within the FOMC: a computational linguistics approach. *The Quarterly Journal of Economics*, 133(2), 801-870. <https://doi.org/10.1093/qje/qjx045>
- Haldane, A. (2018). Climbing the public engagement ladder. Speech at the Royal Society for the encouragement of Arts, Manufactures and Commerce (RSA). <https://www.bis.org/review/r180323b.pdf>
- Hendry, S. and Madeley, A. (2010). Text mining and the information content of Bank of Canada communications. <http://dx.doi.org/10.2139/ssrn.1722829>
- Hernandez, R. and Shell, G. (2014). The Rising Complexity of the FOMC Statement. Economic Research, Federal Reserve Bank of St. Louis. Economic Synopses, No. 23(2). <https://research.stlouisfed.org/publications/economic-synopses/2014/11/05/the-rising-complexity-of-the-fomc-statement/>
- Issing, O. (2005). Communication, Transparency, Accountability: Monetary Policy in the Twenty-First Century. *Canadian Parliamentary Review*, 87(2), 65-83. <https://doi.org/10.20955/R.87.65-83>
- Issing, O. (2019). *The Long Journey of Central Bank Communication*. The MIT Press.
- Jansen, D. J. (2011). Does the Clarity of Central Bank Communication Affect Volatility in Financial Markets? Evidence from Humphrey-Hawkins Testimonies. *Contemporary Economic Policy*, 29(4), 494-509. <https://doi.org/10.1111/j.1465-7287.2010.00238.x>
- Kuncoro, H., Nazir, G. and Sebayang, D. (2021). A textual analysis of central bank communication the case of Indonesia. *Economics Bulletin*, 41(3), 2158-2172. <http://www.accessecon.com/Pubs/EB/2021/Volume41/EB-21-V41-I3-P186.pdf>
- Lahura, E. and Vega, M. (2020). Assessing central bank communication through monetary policy statements: Results for Colombia, Chile and Peru (No. 2020-017). Banco Central de Reserva del Perú.
- Lucca, D.O. y Trebbi, F. (2009). Measuring central bank communication: an automated approach with application to FOMC statements. NBER Working Papers 15367. <https://ideas.repec.org/p/nbr/nberwo/15367.html>

- MacKinnon, J. G. (1996). Numerical Distribution Functions for Unit Root and Cointegration Tests. *Journal of Applied Econometrics*, 11, 601-618. <https://www.jstor.org/stable/2285154>
- McMahon, M., Schipke, A. and Li, X. (2018). China's Monetary Policy Communication: Frameworks, Impact, and Recommendations. IMF Working Paper No. 2018 (244), 32. <https://doi.org/10.5089/9781484385647.001>
- Michalke, M. (2021). Using the koRpus Package for Text Analysis. [https://cran.r-project.org/web/packages/koRpus/vignettes/koRpus\\_vignette.html](https://cran.r-project.org/web/packages/koRpus/vignettes/koRpus_vignette.html)
- Michalke, M. (2021). koRpus: Text Analysis with Emphasis on POS Tagging, Readability, and Lexical Diversity (Version 0.13-8). <https://reaktanz.de/?c=hacking&s=koRpus>
- Omosho, B. (2019). Central Bank Communication in Ghana: Insights from a Text Mining Analysis. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3526451>
- Oshima, Y. and Matsubayashi, Y. (2018). Monetary Policy Communication of the Bank of Japan: Computational Text Analysis. Graduate School of Economics Working paper, Kobe University. <https://EconPapers.repec.org/RePEc:koe:wpaper:1816>
- Park, K. Y., Lee, Y. J. and Kim, S. (2019). Deciphering monetary policy board minutes through text mining approach: the case of Korea. Bank of Korea WP, 1. <http://dx.doi.org/10.2139/ssrn.3312561>
- R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Real Academia Española (s.f). ¿Cuándo se usa el punto y coma? <https://www.rae.es/duda-linguistica/cuando-se-usa-el-punto-y-coma>
- Real Academia Española (2021). Semántica léxica. <https://dle.rae.es/sem%C3%A1ntico>
- Rosa, C. (2013). The Financial Market Effect of FOMC Minutes. *Economic Policy Review*, 19(2), 67-81. <https://www.newyorkfed.org/research/epr/2013/0913rosa.html>
- Rosa, C. and Verga, G. (2007). On the consistency and effectiveness of central bank communication: Evidence from the ECB. *European Journal of Political Economy*, 23 (1), 145–175. <https://doi.org/10.1016/j.ejpoleco.2006.09.016>
- Rozkrut, M., Rybiński, K., Sztaba, L. and Szwaja, R. (2007). Quest for central bank communication: Does it pay to be “talkative”? *European Journal of Political Economy*, 23(1), 176-206. <https://doi.org/10.1016/j.ejpoleco.2006.09.011>
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. Proceedings of International Conference on New Methods in Language Processing, Manchester, UK. <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>
- Shapiro, A.H. y Wilson, D. (2021). Taking the Fed at its Word: A New Approach to Estimating Central Bank Objectives using Text Analysis. Federal Reserve Bank of San Francisco Working Paper 2019-02. <https://doi.org/10.24148/wp2019-02>

- Sigaud, P. (2010). Aproximación a los conceptos de legibilidad y lecturabilidad: aplicación a la lectura de textos digitales. Pamplona: Universidad de Navarra. Documento de trabajo. [https://www.academia.edu/4137854/Aproximaci%C3%B3n\\_a\\_los\\_conceptos\\_de\\_legibilidad\\_y\\_readability\\_aplicaci%C3%B3n\\_a\\_la\\_lectura\\_de\\_textos\\_digitales](https://www.academia.edu/4137854/Aproximaci%C3%B3n_a_los_conceptos_de_legibilidad_y_readability_aplicaci%C3%B3n_a_la_lectura_de_textos_digitales)
- Siklos, P. L. (2013). The Global Financial Crisis and The Language of Central Banking: Central Bank Guidance in Good Times and in Bad. CAMA Working Paper Series Paper 58/2013. <https://dx.doi.org/10.2139/ssrn.2317062>
- Sturm, J. and De Haan, J. (2011). Does Central Bank Communication Really Lead to Better Forecasts of Policy Decisions? New Evidence Based on a Taylor Rule Model for the ECB. *Review of World Economics*, 147(1), 41–58. <https://www.jstor.org/stable/41485746>
- Szigriszt Pazos, F. (1993). Sistemas predictivos de legibilidad del mensaje escrito: Formula de perspicuidad. (Doctoral dissertation, Universidad Complutense de Madrid. Servicio de reprografía).
- 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, 76-92. <http://dx.doi.org/10.1016/j.espe.2015.01.002>
- Tumala, M. M. and Omotosho, B. S. (2019). A text mining analysis of central bank monetary policy communication in Nigeria. *CBN Journal of Applied Statistics*, 10(2). <http://dx.doi.org/10.2139/ssrn.3545508>
- Woodford, M. (2005). Central bank communication and policy effectiveness. <http://www.columbia.edu/~mw2230/JHole05.pdf>

## ANEXOS

### A. LEGIBILIDAD

Según Ballesteros y Fernández (2013), la palabra legibilidad en el idioma español se utiliza para hacer referencia a dos términos ingleses: “*legibility*” y “*readability*”. A pesar de que estos términos son sinónimos desde el punto de vista semántico, el uso les ha otorgado significados diferentes. El término “*legibility*” hace referencia a la legibilidad tipográfica, es decir, a la cualidad de que un texto pueda ser leído sin dificultad desde un punto de vista mecánico. Para ello, se toma en cuenta aspectos como la correcta aplicación de los estilos asociados al tipo, color, tamaño de letra o del papel, la presencia o ausencia de figuras, entre otros. Por otra parte, el término “*readability*” se refiere a la legibilidad lingüística, también denominada comprensibilidad, perspicuidad o lecturabilidad. Este parámetro mide el grado la facilidad con que el texto puede ser leído, basado en complejidad gramatical de la construcción del mensaje (número de palabras y oraciones, promedio de sílabas por palabra, promedio de palabras por oración) y de factores personales del lector (nivel cultural e intelectual, interés por el tema, entre otros aspectos).

Otros autores hacen una clara diferenciación entre estos términos. Así, Sigaud (2010) recurriendo al idioma anglosajón refiere que el término “*legibility*” se traduce como legibilidad al idioma español y el término “*readability*” se traduce como lecturabilidad, comprensibilidad o perspicuidad. Baquedano (2006) también sostiene conceptos separados de estos términos.

## B. ESTIMACIONES ADICIONALES

**Cuadro C-1.** *Resultados de las pruebas de raíz unitaria*

	ADF	PP	KPSS	DF-GLS
logleg	-5.60	-5.48	0.93	-4.27
dref	-8.06	-6.19	0.06	-6.14
dexp	-6.13	-11.85	0.04	-7.93
Valores críticos				
1 %	-3.46	-3.46	0.74	-2.57
5 %	-2.87	-2.87	0.46	-1.94
10 %	-2.57	-2.57	0.35	-1.62

*Nota: La parte superior del cuadro muestra el valor de los estadísticos de cada prueba para el logaritmo de la legibilidad. La parte inferior del cuadro muestra los valores críticos propuestos por MacKinnon (1996). La muestra de análisis es 2002m1-2021m12, con excepción de dref que solo está disponible desde octubre del 2003. Elaboración propia.*

Los resultados de las pruebas de raíz unitaria para las variables *logleg*, *dexp* y *dref* se presentan en el cuadro C-1. Para el caso del *logleg* la prueba ADF, PP y DF-GLS indican que la serie es estacionaria al 1 % de significancia, pues los estadísticos calculados correspondientes, son menores a los valores críticos. Por su parte, la prueba KPSS indica que la serie es raíz unitaria, pues se rechaza la hipótesis nula de estacionariedad dado que el estadístico calculado es mayor al valor crítico. En el caso de las variables *dref* y *dexp*, todas las pruebas coinciden en que ambas series son estacionarias. Tomando en cuenta estos resultados y que la prueba DF-GLS es más eficiente, se puede afirmar que las tres series son estacionarias.

**Cuadro C-2.** Resultados de las estimaciones por OLS (2003m10 - 2021m12)

<b>Variable dependiente:</b> Logaritmo de la legibilidad (logleg)						
	<b>Modelo1</b>	<b>Modelo2</b>	<b>Modelo3</b>	<b>Modelo4</b>	<b>Modelo5</b>	<b>Modelo6</b>
logleg(-1)	0.707*** (0.046)	0.637*** (0.050)	0.609*** (0.051)	0.596*** (0.052)	0.597*** (0.051)	0.590*** (0.051)
logleg(-11)	0.138** (0.050)	0.107** (0.050)	0.095* (0.050)	0.106** (0.050)	0.111** (0.050)	0.116** (0.050)
esta	0.010* (0.005)	0.012** (0.005)	0.013** (0.005)	0.013** (0.005)	0.012** (0.005)	0.013** (0.005)
covid		-0.031*** (0.009)	-0.031*** (0.009)	-0.029*** (0.009)	-0.030*** (0.009)	-0.029*** (0.009)
dvel			-0.014** (0.006)	-0.014** (0.006)	-0.015** (0.006)	-0.015** (0.006)
dexp(-1)				-0.028* (0.016)		-0.020 (0.017)
dref					-0.021* (0.011)	-0.018 (0.011)
Observ.	219	219	219	219	219	219
R2	0.603	0.623	0.631	0.636	0.637	0.640
R2 aj.	0.598	0.616	0.622	0.625	0.627	0.628
AIC	-3.817	-3.860	-3.871	-3.875	-3.881	-3.878
SIC	-3.755	-3.783	-3.778	-3.767	-3.772	-3.754
<b>Autocorrelación</b>						
BG(1)	0.440	0.540	0.700	0.610	0.970	0.910
BG(2)	0.700	0.720	0.750	0.640	0.600	0.500
BG(3)	0.600	0.710	0.760	0.690	0.770	0.690
<b>Heterocedasticidad</b>						
ARCH(1)	0.450	0.550	0.710	0.610	0.970	0.920
ARCH(2)	0.710	0.720	0.750	0.650	0.610	0.510
ARCH(3)	0.610	0.720	0.770	0.700	0.780	0.710

*Nota:* Estimaciones de los autores. Resultados de la regresión por mínimos cuadrados con información desde setiembre de 2003 para todos los modelos.

*Niveles de significancia:* \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Errores estándar robustos entre paréntesis.

## C. NOTAS INFORMATIVAS CON ORACIONES LARGAS

### C.1. Nota Informativa de abril 2015

“El Directorio del Banco Central de Reserva del Perú acordó mantener la tasa de interés de referencia de la política monetaria en 3,25 por ciento Este nivel de la tasa de referencia es compatible con una proyección de inflación que converge a 2,0 por ciento en el horizonte de proyección 2015 - 2016 y toma en cuenta que: la actividad económica continúa por debajo de su potencial; las expectativas de inflación permanecen ancladas dentro del rango meta; los indicadores internacionales recientes muestran señales mixtas de recuperación de la economía mundial, así como una volatilidad alta en los mercados financieros y cambiarios externos; y la inflación nacional ha sido afectada por factores temporales del lado de la oferta, que se irán revirtiendo”. **(122 palabras)**

*Fuente: Nota Informativa del Programa Monetario para abril de 2015.*

### C.2. Nota Informativa de noviembre 2016

“Esta decisión es consistente con una proyección de inflación que converge gradualmente a 2,0 por ciento en el horizonte de efectividad de la política monetaria, y toma en cuenta que: Las expectativas de inflación a 12 meses se encuentran dentro del rango meta; Los efectos de los aumentos de los precios de algunos alimentos y combustibles sobre la inflación de setiembre y octubre han sido transitorios, por lo que se espera una pronta convergencia de la inflación al rango meta; La actividad económica local viene creciendo a un ritmo cercano a su potencial; y, La economía mundial continúa registrando señales mixtas de recuperación en la producción y el empleo, así como una mayor incertidumbre en los mercados financieros internacionales, frente a lo cual la economía peruana mantiene sólidos fundamentos”. **(130 palabras)**

*Fuente: Nota Informativa del Programa Monetario para noviembre de 2016.*