The Predictive Power of Central Bank Communication: Evidence from Mexico*

El Poder Predictivo de la Comunicación del Banco Central: Evidencia de México

Christian Admin De la Huerta Avila
Facultad de Economía, Universidad Nacional Autónoma de México
christiandlha@comunidad.unam.mx

* A previous version of the article has been published as a Working Paper on SSRN and is available at: https://ssrn.com/abstract=4595144. The author would like to thank Santiago Capraro, Iván Fernández, Jamel Kevin Sandoval, and two anonymous reviewers for valuable comments.
Abstract

Understanding the impact of central bank communication on the effective transmission and predictability of monetary policy is paramount. In this paper, we analyze the role of Banco de México’s qualitative communication and its potential scope for guidance by implementing Natural Language Processing. Using a Latent Dirichlet allocation model and dictionary-based sentiment analysis, we develop a Hawkish-Dovish Tone index to measure the bias of Banco de México’s monetary policy statements. We then incorporate this index into an ordinal logit regression to further evaluate the predictive power of Banco de México’s talk. Our findings show that a rate hike (cut) and a hawkish (dovish) message are associated with higher odds of observing a restrictive (accommodative) decision at the next monetary policy meeting, and communication captures information not reflected in traditional macroeconomic variables. Yet, when it comes to accurately anticipate future interest rate decisions, the information provided in policy statements has limited predictive value.

Keywords: central bank communication, monetary policy, natural language processing, ordinal logit, predictability.

Resumen

Comprender el impacto de la comunicación de los bancos centrales en la transmisión efectiva y la predictibilidad de la política monetaria es fundamental. En este artículo, analizamos el papel de la comunicación cualitativa del Banco de México y su alcance potencial para la orientación mediante la aplicación de Procesamiento de Lenguaje Natural. Utilizando un modelo de Asignación de Dirichlet Latente y un análisis de sentimiento basado en diccionarios, desarrollamos un índice de Tono Hawkish-Dovish para medir el sesgo de los comunicados de política monetaria del Banco de México. Seguidamente, se incorpora este índice en una regresión logística ordinal para evaluar el poder predictivo de las declaraciones del Banco de México. Los hallazgos muestran que un aumento (recorte) de tasas y un mensaje hawkish (dovish) se asocian con mayores probabilidades de observar una decisión restrictiva (expansiva) en la próxima reunión de política monetaria, y que la comunicación captura información no reflejada en variables macroeconómicas tradicionales. Sin embargo, cuando se trata de anticipar con precisión las próximas decisiones de tasas de interés, la información proporcionada en los comunicados tiene un valor predictivo limitado.

Palabras clave: comunicación del banco central, política monetaria, procesamiento del lenguaje natural, logit ordinal, predictibilidad.

JEL Classification: C25, C45, E52, E58

Fecha de recepción: 28 de septiembre de 2023
Fecha de aceptación: 1 de abril de 2024
1. Introduction

Over the years, central banks (CB) have acquired considerable power in managing monetary policy. These institutions historically grew up with a certain amount of mystery and opacity in the public perception, due to the technical nature of their functions, their ambiguous method of action, and their cautious communication.

However, in recent decades, many CBs have made efforts to increase transparency and accountability, recognizing the importance of greater clarity in their actions and communications. This evolution has led to a shift from the secretive, enigmatic, and ambiguous prototype of the old CB to a design that encourages openness, intelligibility, and honesty. Nowadays, in most CBs around the globe, communication, which includes policy statements, bulletins, minutes, press conferences, reports, forecasts, public speeches, among other channels, has become increasingly important in the central banker’s toolkit (Blinder et al., 2017). Therefore, communication and, more specifically, the language of central bankers has become an integral part of CB's day-to-day policymaking.

The importance of communication for CBs lies in its dual role. Firstly, effective communication may shape public expectations of future macroeconomic outcomes, influencing financial market behavior and, consequently, the broader economy. By providing guidance on future policy actions, CBs can anchor expectations and reduce noise in financial markets, thereby improving the transmission of monetary policy (Woodford, 2005; Blinder et al., 2008). Secondly, as independent institutions, central bankers operate beyond direct democratic control. Transparency through communication is essential for ensuring accountability to the public they serve. By sharing information about policy decisions, the rationale behind them, and the economic conditions influencing them, CB communication fosters transparency. While the literature has extensively examined transparency (see Fry et al., 2000; Blinder et al., 2001; Geraats, 2002, 2007; Eijffinger & Geraats, 2006; Cihák, 2007; Dincer & Eichengreen, 2008, 2010, 2014; and Dincer et al., 2019, 2022) leading to the creation of various indicators associated with these issues, and a consensus on information sharing, there is growing research exploring the potential power of CB communication as a monetary policy tool.

In this regard, studying the predictive effects of CB communication on upcoming policy rate decisions holds crucial significance in the field of monetary
policy. Communication serves as a vital channel through which policymakers convey their intentions, insights, and assessments to financial markets, economic agents, and the public at large. Consequently, understanding and gauging the information contained in this communication provides valuable insights to capture the forward-looking perspective of CB policies. This information can be leveraged to assess the relationship between CB statements and subsequent interest rates decisions, gaining a deeper understanding of the predictability of monetary policy, the effectiveness of communication strategies, the accuracy of market interpretations, and the potential scope for forward guidance.

The purpose of this paper is to analyze the role of CB qualitative communication in the Mexican economy. We are interested in answering the question of whether the information communicated by Banco de México (Banxico) can be used as a potential source of guidance for upcoming interest rate decisions. Specifically, we examine the language used in Banxico’s monetary policy statements (mps) during the period in which it has implemented its monetary policy through an operational target for the overnight interbank interest rate. Since mps are often accompanied by changes in the monetary stance (whether hikes, cuts, or no changes), they are perceived as a potential tool for generating “monetary policy news” (Bomfim, 2003). Consequently, in these publications, CBs can disseminate relevant information regarding their intended monetary policies, thereby potentially influencing financial adjustments and market expectations about the future path of short-term interest rates (Blattner et al., 2008; Milani & Treadwell, 2012).

To assess the potential power of Banxico’s communication as a policy tool, we first focused on extracting qualitative information from the mps. We implemented Natural Language Processing (NLP) techniques over a set of 131 documents, starting on Banxico’s February 2008 decision and ending in December 2022. To do so, we used a Latent Dirichlet Allocation (LDA) model, first proposed by Blei et al. (2003), to analyze latent topics contained in the set of Banxico’s publications. We identified the semantic structure of the mps and labeled subjects on which Banxico has focused its communication, such as monetary policy, economic activity, international affairs, among others. Subsequently, we opted to leverage the insights gained from the LDA model to develop our Hawkish-Dovish Tone (HDT) weighted index, by applying a dictionary-based sentiment analysis method.
With the information gathered, we focused on determining whether the tone of the documents may help to predict upcoming policy decisions. In particular, we tested if the qualitative information provided by Banxico contains relevant “monetary policy news” by including the HDT weighted index on an ordinal logit regression using discrete changes in the policy rate as the dependent variable. Likewise, we incorporated other relevant explanatory variables in the regression, such as inflation and output gaps and the federal funds rate. Throughout the paper, it is argued that current interest rate decisions and the verbal policy stance are associated with the probability of observing a restrictive or accommodative decision at the next monetary policy meeting, and communication captures information not reflected in traditional macroeconomic variables. However, when it comes to effectively anticipating the next monetary policy decisions, information conveyed through policy statements contains limited predictive ability.

These findings are in line with other relevant papers (see Picault & Renault, 2017; Baranowski et al., 2021; Priola et al., 2022; and Astuti et al., 2022) that evaluate the predict effects of CB communication on future policy decisions. This study contributes to the existing literature by focusing on these aspects for the Mexican CB, utilizing recent NLP techniques outlined in the literature.

The rest of the paper is divided as follows. Firstly, a brief review of related literature is provided in Section 2. In Section 3, the evolution of Banco de México’s monetary and communication strategy over the last few decades is described, along with an overview of the structure of monetary policy statements. Section 4 presents the main analysis of this paper, including LDA modeling, sentiment analysis, and econometric estimations. Finally, Section 5 concludes with a summary of the results and implications for CB communication.

2. Literature review

Central banking has undergone significant changes over the past few decades. Today, the era of “mumbling with great incoherence” is over\(^1\), and CBs around the world are more independent and transparent than ever before. This wasn’t always the case. In the past, the consensus was that central bankers should

---

\(^1\) The words “mumbling with great incoherence” were uttered by Alan Greenspan during his appearance before the Senate in 1987, referring to the ambiguous central banking practices of the time. As Gabriel Makhlouf, Governor of the Central Bank of Ireland, said, those days are long gone (see Makhlouf, 2020).
be secretive and mysterious authorities with a high degree of opacity. It was believed that central banks should maintain little or no communication with the public and markets, or if so, convey coded messages extremely difficult to read and understand.²

Even though this view of central banking as full of mysticism, secrecy, and opacity has remained at the heart of some scholars and central bankers, the experience of the “Great Inflation” and developments in Time Inconsistency and Institutional Design Theory highlighted the need for central bankers with a high degree of independence and credibility (see Kydland & Prescott, 1977; Calvo, 1978; Barro & Gordon, 1983; and Rogoff, 1985). The changes that originated in the last decades of the twentieth century, both in theory and practice, were a significant departure from the past, fostering greater central bank independence, promoting inflation targeting schemes, rethinking monetary policy methods, and prioritizing transparency and communication. On the other hand, these changes also led to the central banks being in the spotlight and monetary policy gradually evolving into the main tool of macroeconomic stability (see Bernanke, 2022). Alan Blinder, former Vice Chairman of the Federal Reserve, termed the shift towards greater openness and transparency in central banking as the “Quiet Revolution” (see Blinder, 2004).

The study of central bank communication dates to the research conducted by Christina and David Romer (1989), which is a significant contribution to this field. They analyzed a series of time periods during which the Federal Reserve’s language was restrictive and found that narratives, which include historical records, explanations of the decision-making process, and accounts of the sources of monetary disruptions, are important sources of information, in addition to those derived purely from statistical evidence. A large number of theoretical and empirical papers have focused on communication in the past two decades. For example, some

² For example, during the 1960s, economist Milton Friedman underscored the pivotal elements of a central bank’s function, emphasizing the avoidance of accountability and the attainment of public prestige (see Faust & Svensson, 2001). The principle of “never excuse, never explain” is attributed to Montagu Norman, former Bank of England chairman, who steadfastly refrained from parliamentary testimonies (see Boyle, 1967). Mervyn King encapsulates the late 20th-century perspective on central bank communication, recalling a dinner with Paul Volcker where the advice “mystique” was bestowed, epitomizing the enigmatic tradition and wisdom of central banking during that era (Lindsey et al, 2005).
studies stress that central bank (CB) communication could shape market expectations, making central bank behavior more predictable and market reactions more consistent with macroeconomic goals (for instance, see Blinder, 1999; Blinder et al., 2001; Woodford, 2001; Kohn & Sack, 2004; Bernanke et al., 2004; Blinder et al., 2008; Bernanke et al., 2019).

Initially perceived as a credibility enhancer or support for limited transmission channels, CB communication has evolved into a consensus-driven tool integral to monetary policy, particularly in influencing long-term interest rates, bolstered by a deeper understanding of market effects on financial stability and prices (see Ehrmann & Fratzscher, 2007). This transformation has reshaped monetary policy strategy and the central banker toolkit in the context of growing authority and market expectations influence. With the development of communication tools such as policy statements, bulletins, minutes, press conferences, reports, forecasts, public speeches, among others, a lot of literature has emerged trying to understand the impact of communication, both theoretically and empirically, on market expectations, signals on the yield curve and the effective transmission of monetary policy (see Blattner et al., 2008; Campbell et al., 2012; Campbell, 2013; Charbonneau & Rennison, 2015; Smith & Becker, 2015; McKay et al., 2016; Angeletos & Lian, 2018; Dell’Ariccia et al., 2018; Altavilla et al., 2019; Bernanke et al., 2019; and Bernanke, 2020). The findings of these studies suggest that, in general, communication is seen as a tool to prepare markets for upcoming decisions and can be used as a potential source of “monetary policy news”, as it provides relevant information to financial markets and the public at large so that they become increasingly familiar with the way CBs think and act. This, in turn, contributes to making actions more predictable and enhancing credibility.

Interest in CB’s language has increased especially after the episodes of the Global Financial Crisis and the COVID-19 pandemic. Some empirical works have constructed indicators by identifying certain relevant words in CB communications, and then investigating how the tone bias impacts the predictability of monetary policy³, and other set of macroeconomic and financial variables (for example, Blattner et al. (2008) distinguish between two types of predictability of monetary policy: short-term and long-term predictability. While short-term predictability is narrowly defined as the ability of the public to anticipate monetary policy decisions correctly over short horizons, the broader, ultimately more meaningful concept of longer-term predictability also encompasses the ability of the private sector to understand the monetary policy framework of a central bank, i.e., its objectives and systematic behavior in reacting to different circumstances.
see Rosa & Verga 2008; Heinemann & Ullrich, 2007; and Hayo & Neuenkirch, 2010). However, there are three fundamental problems with these indicators: they are based on idiosyncratic identification of the communication, i.e., the researcher prints bias by focusing exclusively on the words and phrases she considers relevant in the tone of the CB; their construction becomes particularly complicated when large numbers of documents are analyzed; and they are inconsistent when there are major changes in narratives.

In this regard, a second generation of empirical studies has emerged that address this problem and make use of Natural Language Processing (NLP) techniques. This body of research aims to decipher the semantic content of CB communication and to interpret both the quantity and quality of messages conveyed through official documents and other central banking-related texts. This body of research mainly exploded two data analysis methodologies: Latent Dirichlet Allocation (LDA) and sentiment analysis. The key idea behind this is to uncover the semantic structure of CB publications and extract signals (or sentiment) from large amounts of unstructured data, such as the words contained in the documents.

In the context of central banking, the LDA algorithm in combination with sentiment analysis has been used since the work of Hansen & McMahon (2016). They explore how multidimensional aspects of Fed communication have effects on both the market and real economic variables. Similarly, this method has been previously applied to study different aspects of CB communication. For example, the consequences of transparency in monetary policymakers’ deliberations reflected in monetary policy meeting transcripts (Hansen et al 2018); the role of information from inflation reports on market interest rates (Hansen et al 2019); the power of CB talk as a predictor of future financial market behavior (Petropoulos & Siakoulis, 2021); the impact of communications on the returns of financial variables (Gu et al., 2022; Möller & Reichmann, 2021); the effects of communication on consumer inflation expectations (Szyszko et al., 2022); the influence on macroeconomic, monetary and financial variables, and the idiosyncratic effects of central bankers’ discourse or persistence of sentiment in communication (Hayo & Zahnner, 2023); and the impact of press conference speech on financial markets (Gorodnichenko et al., 2023).

4 Section 4 briefly describes both methodologies.
In parallel, using tone indicators derived from different validated dictionaries for CB sentiment analysis (see Loughran & Mcdonald, 2011, 2014; Apel & Blix-Grimaldi, 2012, 2014; Bennani & Neuenkirch, 2017; Gonzalez & Tadle, 2021; and Apel et al., 2022), several studies have constructed tone indicators to capture bias in CB communications. The dictionaries categorize the words used by the CBs as either hawkish or dovish with the goal of capturing the policy bias of central bankers, thus, they allow to assign a numerical value and create a measurement that accurately reflects this bias. Using this type of indicators, there is growing evidence for different CBs that support the idea that communication can help anticipate future interest rate decisions (Tobback et al., 2017; Baranowski et al., 2021; Astuti et al., 2022; and Priola et al., 2022).

In the particular case of Banco de México, not much research has focused on communication. To the best of my knowledge, only Cermeño & Navarrete (2011), López Marmolejo (2013), García-Herrero et al. (2019) and Solís (2023) have previously addressed this topic. The former constructs a bias index for Banxico’s monetary policy statement and integrates it into an ordinal regression using changes in the next period’s monetary stance as the dependent variable, revealing the short-term predictive effectiveness. López Marmolejo develops a communication index based on relevant sentences from the statements and incorporates it to a Taylor rule OLS estimation, suggesting that Banxico’s communication signals help to predict short-term interest rates and align market expectations. García-Herrero et al. (2019) propose an index to evaluate Banxico’s communication, relying on their own interpretation of the policy stance derived from official documents and speeches by Banxico officials. Their findings highlight the impact of this communication on both the volatility and volume of money market rates and suggest a discernible relationship between a hawkish message and an increase in money market rates. Finally, using intraday data on asset prices around monetary policy announcements in Mexico, Solís (2023) identifies the presence of two factors in monetary policy announcements: surprises about the current policy rate, conveyed through the interest rate decisions themselves, and surprises concerning its future trajectory, communicated through policy statements. Solís links the path factor to communication surprises (or “a subtle form of forward guidance”), emphasizing Banxico’s ability to manage expectations about the future path of interest rates through statements.

This research highlights the potential influence of Banxico’s communication strategy in shaping expectations about future macroeconomic outcomes, helping
to reduce uncertainty, noise, and volatility in financial markets linked to central bank operations.

This paper contributes to this growing literature. The main difference with the aforementioned studies is the type of identification of central bank language. Cermeño & Navarrete (2011), López Marmolejo (2013) and García-Herrero et al. (2019) use a subjective method of identification, i.e., they assign values to Banxico’s communication depending on their idiosyncratic evaluation of the phrases and words they consider relevant. In contrast, the methodology presented in this paper focuses on analyzing the predictive power of Banxico’s communication through the lens of LDA models and sentiment analysis, therefore, the identification relies on statistical processing of the information, subtracting subjectivity from the interpretation of CB communication.

3. Banco de México’s Communication Strategy

Like in many other CBs around the world, Banxico’s communication strategy has evolved.5 Since the early 2000s, the CB began publishing the monetary policy statement (MPS) reporting its monetary policy decisions, as well as a quarterly inflation report (QIR) aimed at analyzing economic developments and its consistent monetary policy to meet the objective of maintaining low and stable inflation. This was followed in 2001 by the formal adoption of an Inflation Targeting Regime (ITR). Since then, the strategy and communication design have undergone some changes. The first one culminated in 2011, when, after adopting an operational target for the overnight interbank interest rate in 2008 and gradually decreasing the number of monetary policy meetings (MPM) per year, the central bank determined a calendar with 8 pre-set dates for its monetary policy decisions.

At the MPM, the Governing Board can raise, leave unchanged, or cut the policy rate to influence monetary and financial conditions in line with its mandate to maintain low and stable inflation. In addition, in the event of extreme economic and financial developments that require Banxico’s intervention, the Governing Board may adjust the monetary policy stance at dates other than those established in advance. After each meeting, Banxico releases the MPS, informing the modifications on the monetary stance.

5 A similar analysis to the one outlined in this section can also be found in Banco de México (2023) and Heath and Acosta Margain (2023).
In addition, Banxico has made other major developments toward greater transparency and effective communication. For example, since 2011, the institution publishes Minutes of the mpm, and in 2018 announced that transcripts will be issued three years after the date of each meeting, making available to the public transcripts of 2018, 2019, and part of 2020 so far. In 2018, Banxico updated its communication strategy and shifted to an Inflation Forecast Targeting (IFT) scheme. In 2020, the General Communication Criteria of the Governing Board and Banxico’s staff were updated and made public for the first time. Since August 2021, the bank conveyed to publish and update the inflation forecast in every monetary policy statement and identify the direction of the vote of each of the members who participated in said meeting, indicating the members who adopted the decision taken and, if applicable, those who voted for an alternative decision and what it consisted of. Finally, in May 2022, Banxico took a major step by formally delivering verbal forward guidance in the monetary policy statements. Table 1 summarizes the evolution of Banxico’s communication strategy.
### Table 1: Evolution of Banco de México’s communication strategy

<table>
<thead>
<tr>
<th>Year</th>
<th>Advances in the communication strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>Start of the publication of monetary policy decision bulletins. The publication of Quarterly reports is established, accompanied by a presentation and a press conference.</td>
</tr>
<tr>
<td>2001</td>
<td>Formal adoption of an Inflation Targeting Regime.</td>
</tr>
<tr>
<td>2003</td>
<td>The bank sets specific dates for monetary policy decisions. At least one bulletin is published at the end of the month and another if the policy changes from one month to the next.</td>
</tr>
<tr>
<td>2006</td>
<td>Monetary policy decisions decrease from 23 to 12 a year.</td>
</tr>
<tr>
<td>2008</td>
<td>Monetary policy decisions decrease from 12 to 11 a year. The central bank begins to implement its monetary policy through an operational target for the overnight interbank interest rate.</td>
</tr>
<tr>
<td>2010</td>
<td>Inflation reports include fan charts for inflation and growth forecasts.</td>
</tr>
<tr>
<td>2011</td>
<td>Start of the publication of the minutes of Governing Board’s meetings. Monetary policy decisions decrease from 11 to 8 a year. Video broadcast of the presentation of the Quarterly Inflation Report.</td>
</tr>
<tr>
<td>2016</td>
<td>The launch of the “Banxico Educa” website is announced. This site aims to inform and educate the public at large about the objectives, goals, and duties of Banco de México. It is also intended to be a means of dissemination and education on economic and financial culture for the country.</td>
</tr>
<tr>
<td>2017</td>
<td>The central projection path for inflation and economic activity is included in the fan charts.</td>
</tr>
<tr>
<td>2018</td>
<td>The central bank shifts its monetary policy strategy to Inflation Forecast Targeting. Inflation reports include for the first time point forecasts of annual quarterly inflation. The minutes begin to disclose the identity of the voters and, in case of dissent, also incorporate a section explaining the reasons behind the dissenters’ vote. The minutes and monetary policy statements are published simultaneously in Spanish and English on the corresponding date. Announcement that transcripts of the meetings will be made available to the public three years after the meeting. The Governing Board makes available transcripts of speeches and public presentations by the members.</td>
</tr>
<tr>
<td>2020</td>
<td>The General Communication Criteria for the Governing Board and Banxico’s staff are updated and made public for the first time. It is established that the press releases, monetary policy statements and minutes would be made clearer and more concise in its extension.</td>
</tr>
<tr>
<td>2021</td>
<td>First disclosure of transcripts. An update of headline and core inflation forecasts for the following eight quarters is published after each monetary policy decision. The monetary policy statements identify the direction of the vote of each of the members of the Governing Board who participated in said meeting, indicating the members who adopted the decision taken and, if applicable, those who voted for an alternative decision and what it consisted of.</td>
</tr>
<tr>
<td>2022</td>
<td>Forward guidance is included in monetary policy statements.</td>
</tr>
</tbody>
</table>

Source: Banxico’s official documents and press releases from the Governing Board.
3.1 Monetary Policy Statements

Given the major changes in Banxico’s communication strategy, the QIR, MPS, and Minutes have positioned as the main publications in which the CB conveys information about its view on the macroeconomic outlook and explains the rationale behind the monetary policy decisions. MPS are scrutinized by CB watchers because these documents are recognized as a potential source of “monetary policy news” since they are published right after the MPM (Bomfim, 2003). Therefore, this publication may contain relevant information and potentially influence financial adjustments and market expectations about the future path of short-term interest rates (Blattner et al., 2008; Milani & Treadwell, 2012).

Our focus in this paper is on the MPS as a way of facilitating CB communication. Particularly, we decided to study the publications between 2008 and 2022 to have a full sample in the period in which Banxico has been conducting its monetary policy strategy through an operational target for the overnight interbank interest rate. The documents consist of a few pages (no more than 3 at their longest over the period under study) containing mainly the following information: i) international developments since the last meeting; ii) a review of economic and financial outlooks for the domestic economy; iii) Banxico’s view on both observed and expected inflation; iv) the balance of risks to the inflation forecast; v) a brief explanation of the rationale behind the monetary policy decision; vi) since May 2021 an update of the inflation forecast.

As highlighted by Hansen et al. (2019), the CB qualitative communication contains information that most of the time is challenging to summarize into quantitative data, so it represents an additional source to traditional macroeconomic time series. In that sense, our database is constructed with the text of the documents leaving aside the forecast tables. In the period under study occurred a total of 131 policy decisions. We collected the statements from Banxico’s official website and stored the information into plain text documents. Before applying any text analysis technique, it is necessary to pre-process the plain text documents and perform a text mining methodology to remove all unnecessary information and obtain a Communication Corpus and a Document-Term Matrix (DTM). All these steps can be consulted in the Appendix.
4. The Predictive Power of Central Bank Communication

We estimated the predictive power of CB communication following a similar methodology as in Tobback et al. (2017), Baranowski et al. (2021), Astuti et al. (2022), and Priola et al. (2022). First, using a Latent Dirichlet Allocation (LDA) model, we derived a series of latent topics that were the focus of Banxico’s language between 2008 and 2022. Then, since topics alone do not provide relevant measurable information, we decided to apply dictionary-based sentiment analysis to obtain our Hawkish-Dovish Tone (HDT) index of Banxico’s monetary policy statements.

With the information gathered, we focused on determining whether the tone of the documents can help to predict upcoming monetary policy decisions. Specifically, we examined whether the qualitative information provided by Banxico contains relevant “monetary policy news” by including the HDT index in an ordinal logit regression using changes in the monetary policy rate at the next decision as the dependent variable. We also incorporated gaps and the federal funds rate in the analysis so that the logit estimation can be considered as an ordinal representation of a Taylor rule-like framework augmented by communication.

4.1 Topic Modeling

Topic Modeling is a statistical technique for revealing the underlying semantic structure in a large collection of documents (Kherwa & Bansal, 2019). In this field of computer science, the most notable contribution in the past two decades is the LDA algorithm, first proposed by Blei et al. (2003). The intuition behind the model is quite simple: every document can be thought of as a mixture of topics, and every topic as a mixture of words. In this regard, LDA is a generative probabilistic model that employs an unsupervised learning process: given the number of documents, $N$, the number of unique terms (words) in the corpus, $V$, and the number of topics, $k$, it aims to identify the underlying distribution by estimating the mixture of words that are associated with each topic, while also determining the mixture of topics that describes each document. At the highest level, LDA represents a three-level hierarchical probabilistic assignment model, where each document is modeled as a weighted mixture of the underlying set of topics, and each topic is modeled, in turn, as a weighted mixture of the underlying set of terms in the corpus. The algorithmic details of LDA can be found in Blei et al. (2003) and Hansen et al. (2019).
In the context of CB communication, let’s think of topics as overarching themes that emerge from the diverse array of messages the CB conveys. Just as CBs address a wide range of economic and monetary issues in their communications, LDA identifies these underlying themes by analyzing the frequency and co-occurrence patterns of words across documents. For instance, a CB’s messages might cover topics like inflation, interest rates, economic growth, and financial stability. LDA helps uncover these latent themes and their proportions in the communication corpus, offering a holistic view of the key subjects being addressed. In our LDA model, documents can be thought of as individual MPs. Each of these documents represents a mixture of different topics covering various aspects of monetary policy and the economy. LDA extracts the underlying topics and their contributions to each document, shedding light on the multifaceted nature of central bank messaging. This process allows discerning the nuanced policy directions and priorities embedded within these communications. Just as certain keywords or phrases signal specific themes in CB communication, LDA identifies words that serve as indicators of topics. For instance, words like “inflation,” “rate,” “growth,” and “stability” might be strong indicators of distinct topics related to monetary policy. LDA’s probabilistic approach captures the likelihood of words appearing within each topic, which aligns with how CBs convey different emphases through specific terminology. By uncovering these word-topic relationships, LDA provides a quantitative foundation for understanding the semantic content of CB communication, revealing the underlying narrative and policy signals.

Our LDA communication model takes as inputs the data in our DTM (see Appendix A): \( N = 131; \ V = 1,649; \) and the number of topics, \( k \). Despite the many advantages of using LDA models, they have two limitations. First, determining the optimal number of topics can be challenging due to the unknown value of \( k \). Secondly, interpreting the results can be ambiguous since the model only reveals the structure of topic-per-word and document-per-topic probabilities, meaning the results are a weighted bag of words associated with the topics of each document and the words of each topic. However, it lacks the ability to assign labels to each of the topics, thus requiring some prior knowledge to interpret the content and infer an appropriate structure for each of them.

We have effectively addressed the first issue by training a total of 99 models for \( k \) values ranging from 2 to 100. We then ran the algorithms developed by Griffiths & Steyvers (2004), Cao et al. (2009), Arun et al. (2010), and Deveaud et al. (2014). This procedure yields four validated metrics to determine
the number of topics to consider, which suggest that the optimal value of \( k \) lies between 13 and 61 (see Figure 1).

Figure 1: Optimal Number of Topics

![Figure 1: Optimal Number of Topics](image)

Source: Author’s calculations. Notes: Standardized values of each metric. The optimal number of topics is found where the metrics are minimized (top panel) or maximized (bottom panel).

When estimating the LDA model, it is crucial to consider the trade-off between the number of topics and accuracy in assigning unique terms.\(^6\) That is, using a large value for \( k \) may improve the accuracy in assigning unique terms to each topic, but it may also lead to a more complicated interpretation. Conversely, using a small value for \( k \) may simplify the interpretation, but it comes with a risk of assigning a small number of topics that are too general. After considering

\(^6\) In LDA model estimation, it is essential to balance the number of topics with the accuracy in assigning unique terms, which involves a dilemma between interpretability and goodness of fit (Chang et al., 2009).
the recommended $k$-range of the four metrics, as well as analyzing previous literature on CB communication (see Section 2), which typically uses between 5 and 30 topics, we evaluated different LDA models for $k$ ranging from 5 to 35. Finally, we selected a rather ad hoc 11 topics as the optimal value for our communication corpus.

With the information derived from the LDA model, we identified 7 potential categories into which Banxico directed its language over the period 2008-2022: Monetary Policy, Economic Activity, International Affairs, Exchange Rate, Inflation Risk, Balance of Risks, and COVID-19 Pandemic. As depicted in Table 2, each of the inferred topics can be linked to specific subjects. For example, topics 1 and 11 encompass words closely related to monetary policy (“interest rate”, “decision”, “overnight interbank interest rate”, “monetary policy”, “balance”), making it possible to label these topics as identifiable components of Banxico’s monetary policy language. Topics 2, 3, and 10 are tied to economic activity, as words such as “economy”, “activity”, “gap”, “growth”, to name a few, exhibit a high likelihood of being part of them. Topics 4 and 7 represent international affairs, whereas topics 5, 6, 8, and 9 are associated with the balance of risks, exchange rate, the COVID-19 pandemic, and inflation risk, respectively.

In addition, the document-per-topic probabilities allow us to describe the thematic distribution over time within all MPS (see Figure 2).

---

7 This type of selection is supported by the recommendations of Blei (2012), who states that interpretability is a legitimate reason to choose a value of different from the one that yields the most efficient model. He highlights the disconnection between how topic models are evaluated and the reasons why we expect them to be useful (Blei, 2012, as cited in Hansen et al., 2018, p. 18). However, it is important to acknowledge a limitation of this ad hoc selection method. As highlighted by Ballester & Penner (2022), while the methodology employed in this paper can be valuable for constructing a topic model that properly reflects reality, the results may not be robust to changes in the data (the words contained in each document) or the dimensions (the number of topics).
Table 2: Terms with the highest Probability per Topics

<table>
<thead>
<tr>
<th>Topics Structure</th>
<th>LDA Topics</th>
<th>Terms within Topics (Top 10 with the highest Probability)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monetary Policy</td>
<td>Topic 1</td>
<td>continuous Banxico interest rate México day decision overnight interest rate</td>
</tr>
<tr>
<td></td>
<td>Topic 11</td>
<td>monetary policy expectations low decrease good regard monetary balance affect average</td>
</tr>
<tr>
<td>Economic Activity</td>
<td>Topic 2</td>
<td>economy activity contraction narrow Governing Board month purpose emergent last expect</td>
</tr>
<tr>
<td></td>
<td>Topic 3</td>
<td>country product equal gap growth rate development great near behavior</td>
</tr>
<tr>
<td></td>
<td>Topic 10</td>
<td>exhibit shock might domestic particular record tendency process hold special</td>
</tr>
<tr>
<td>International Affairs</td>
<td>Topic 4</td>
<td>recovery year United States although hold previous toward demand annual forecast</td>
</tr>
<tr>
<td></td>
<td>Topic 7</td>
<td>international part anticipate economic growth change minor zone previous prevail still</td>
</tr>
<tr>
<td>Balance of risks</td>
<td>Topic 5</td>
<td>risk factor environment inform uncertainty large median amount prudent track</td>
</tr>
<tr>
<td>Exchange Rate</td>
<td>Topic 6</td>
<td>price depreciate currency commodity related spoil general service decision possible</td>
</tr>
<tr>
<td>covid-19 Pandemic</td>
<td>Topic 8</td>
<td>core headline pandemic term effect financial market global trajectory action necessary</td>
</tr>
<tr>
<td>Inflation Risk</td>
<td>Topic 9</td>
<td>increase inflation pressure target major level base quarter point adjust</td>
</tr>
</tbody>
</table>

Source: Author’s calculations. Notes: The table displays the top 10 terms with the highest probability of belonging to each topic. Each term is represented in its root form (see Appendix A). The LDA model is estimated using the original Spanish documents, and the terms shown in the table are translations by the author.
Throughout the period, topics related to Monetary Policy show a high prevalence among policy statements. International Affairs and Economic Activity also capture most of the language in specific periods, notably standing out in the post-Global Financial Crisis era. It is also worth mentioning that the Exchange Rate topic shows a high prevalence between 2014 and 2016, when Banxico acknowledged that external shocks could pose some challenges to the achievement of its objectives. Another important highlight is that we identify themes related to the covid-19 pandemic and Inflation Risk, which explain most of the language since 2020, in line with the macroeconomic narrative of this period.

---

8 For instance, on February 17, 2016, Banxico made an unscheduled decision to raise the policy rate by 50 basis points. This move responded to increased volatility in international financial markets and deteriorating external conditions for the Mexican economy, including the continued fall in oil prices. These factors were recognized by the CB as negatively affecting public finances, the current account, and the exchange rate, thus increasing the risk of misalignment of inflation expectations with the target.
4.2 Sentiment Analysis

Since the LDA model alone does not provide relevant measurable information on the semantic content of CB communication, we calculated a measure of the policy bias (hawkish/dovish) for each MPS. Using dictionary-based sentiment analysis, we constructed a Hawkish-Dovish Tone (HDT) index. To this end, we rely on the Loughran-McDonald [LM] (Loughran & McDonald, 2011, 2014) and González-Tadle [GT] (Gonzalez & Tadle, 2021) dictionaries. We decided to use an intercept between the bag of words of both dictionaries because the former has been widely validated by the literature on CB communication, while the latter is a novel contribution to the analysis of CB documents written in Spanish in Latin American economies.9

The LM and GT dictionaries are made up of two types of words: positives/hawkish and negatives/dovish. The positives/hawkish category contains terms that CBs use when implementing a restrictive policy stance, such as “increase,” “inflationary,” “high,” “accelerate,” and others. On the other hand, the negatives/dovish category includes terms that CBs emphasize when taking an accommodative policy, such as “adverse,” “cut,” “decrease,” “low,” “disinflation,” and others. We use these categories to create an average tone index that considers the CB’s policy tone. To calculate the index, we use the following formula:

$$HDT_i = \frac{\sum_{i=1}^{N} positive_i - negative_i}{\sum_{i=1}^{N} positive_i + negative_i}$$

where \(\sum_{i=1}^{N} positive_i - negative_i\) is the total number of positive words minus the total number of negative words in document \(i\), normalized by the total coincidences between the dictionaries and document \(i\), \(\sum_{i=1}^{N} positive_i + negative_i\). This formula results in an index of the average sentiment of each MPS, on a scale of 1 to -1, where 1 represents the exclusive use of hawkish language and -1 the total use of dovish terms.

Following Astuti et al. (2022), we built our HDT weighted index utilizing the subset of documents belonging to \(k\) topics identified by the LDA model. Thus, we performed sentiment analysis considering the topic-per-word and document-per-topic

---

9 The LM dictionary is a collection of English words represented in their root form. To match our analysis, we translated the dictionary using the DeepL software. Then, 5 synonyms are considered for each word, seeking congruence between the dictionary language and the specific wording of Banco de México.
The Predictive Power of Central Bank Communication: Evidence from Mexico

distributions. First, we computed the individual sentiment for each topic \((k)\) in each document \((i)\): \(HDT_i^k, k = 1,\ldots, K\). As a final step, we calculated the weighted average:

\[
HDT_w = \frac{\sum_{k=1}^{K} w_{k,i} HDT_i^k}{\sum_{k=1}^{K} w_{k,i}} \tag{2}
\]

Where \(w_{k,i}\) denotes the document-per-topic distribution, defined as the proportion at which the topic \((k)\) appears within the document \((i)\). Since the sum of the weights \(\left(\sum_{k=1}^{K} w_{k,i}\right)\) is equal to 1, equation (2) can be simplified as:

\[
HDT_w = \sum_{k=1}^{K} w_{k,i} HDT_i^k \tag{3}
\]

Figure 3 shows our tone index and changes in Banxico’s policy rate between 2008 and 2022. There are several periods in which joint movements of the variables are observed, such as 2008, 2013-2014, 2016-2018, 2019-2020 and 2021-2022. In addition, it is worth noting that the Pearson’s correlation coefficient emphasizes a moderately positive linear relation between the changes in the policy rate and the tone index, with a value of 0.53.

Figure 3: Changes in the policy rate and Banco de México’s communication tone

Source: Author’s calculations.
4.3 Econometric Estimation

Our objective is to determine whether CB communication can accurately predict interest rate decisions. Specifically, we should expect the use of restrictive language to be an indicator of an upcoming policy rate hike. Conversely, accommodative language should suggest an imminent interest rate cut. If the language used is neutral or uncertain, it would be more likely to expect no changes in monetary policy in the short term.

Following the aforementioned criteria, we decided to include our HDT weighted index into an ordinal logit model. Since policy interest rates typically move by a discrete amount (e.g., by 25 basis points), we selected this type of econometric estimation because it allows capturing the discrete nature of variables. Therefore, to investigate the research question we used the changes in the monetary policy rate \( \Delta r_t \) as the dependent variable. \( \Delta r_t \) take values of: 

-3 → 75 basis points (bp) cut; 
-2 → 50bp cut; 
-1 → 25bp cut; 
0 → no change in the monetary policy stance; 
1 → 25bp hike; 
2 → 50bp hike; 
3 → 75bp hike.10

Moreover, in our analysis we also incorporated other relevant macroeconomic variables such as the inflation and output gaps and the federal funds rate. The inflation gap is measured by the difference between the annual percent changes on the Consumer Price Index and Banxico’s inflation target (3%), and the output gap by the cyclical component of the overall indicator of economic activity (IGAE), measured through the Hodrick-Prescott filter. Since data on interest rate decisions are not collected on a regular monthly frequency, the macroeconomic variables represent the information available at the time of the decision (e.g., the October 2020 decision considers the September 2020 gaps).11 Finally, the federal funds rate is represented by the target established by the Federal Open Market Committee in the decision preceding that of Banxico.12

---

10 Cermeño & Navarrete (2011), Téllez-León & Venegas-Martínez (2013), Picault & Renault (2017), Baranowski et al. (2021), and Priola et al. (2022) apply a similar methodology, choosing only three values for the discrete change in the policy rate (cuts, holds and hikes). Following Astuti et al. (2022), we employ various categories at different levels of analysis to accurately assess the influence of communication on specific monetary policy changes.

11 Gaps are taken in first differences, which is consistent with similar studies. See Bennani & Neuenkirch (2017), Astuti et al. (2022) and Priola et al. (2022).

12 Starting from December 2008, the Federal Reserve has implemented a target range for the federal funds rate. Throughout this period, the upper limit of this range
Let's start with a discrete representation of a Taylor-type rule for policy rates, considering it as a function of inflation and output gaps:

$$\Delta r_{t+1} = \rho \Delta r_t + \alpha \Delta (\pi_t - \pi^T) + \beta \Delta (y_t - y^*_t) + \epsilon_{t+1}$$

(4)

Where $\Delta r_{t+1}$ is the discrete change in the policy rate at $t+1$, i.e., at the next monetary policy meeting, $\Delta r_t$ is the change in the policy rate at $t$, which plays the role of a smoothing term of monetary policy, $\pi_t - \pi^T$ is the inflation gap, $y_t - y^*_t$ is the output gap, $\alpha$, $\beta$, $\rho$ are parameters, and $\epsilon_t$ is an error term.

Traditionally, this specification is widely accepted as good estimation of the CB reaction function (Orphanides, 2010), so we should expect the parameters to be significant in the regression. Following Apel & Blix-Grimaldi (2014), Picault & Renault (2017) and Priola et al. (2022), we augmented this model with communication by including our HDT weighted index in the specification:

$$\Delta r_{t+1} = \rho \Delta r_t + \alpha \Delta (\pi_t - \pi^T) + \beta \Delta (y_t - y^*_t) + \phi HDTw_t + \epsilon_{t+1}$$

(5)

If CB communication adds valuable information to anticipate interest rate decisions, beyond that stemming from the other regressors, the tone in should help predict the change in the monetary policy stance in $t + 1$. Therefore, we should expect $\phi$ to be positive and statistically significant. Subsequently, we decided to run the logit regression considering only the smoothing term and the HDT weighted index, to investigate the predictive power of communication alone:

$$\Delta r_{t+1} = \rho \Delta r_t + \phi HDTw_t + \epsilon_{t+1}$$

(6)

Finally, as highlighted by recent studies (see Figueroa & Padilla, 2022), there appears to be evidence of monetary policy coordination between Banxico and the Federal Reserve. This suggests that, due to the proximity in the decision-making of both CBs, a pattern emerges wherein Banxico’s decisions typically align in direction and magnitude with those of the Federal Reserve. To account for this potential coordination, we have incorporated the change in the federal funds rate at $t$ ($\Delta r_{t,fedfunds}$) into equations (5) and (6):

is the focal point for our statistical analysis. This variable is also transformed into first differences to maintain consistency with the econometric model.
Table 3 shows the results of logit estimations. As can be noted, the gaps do not exhibit statistical significance across all specifications. On the other hand, the communication augmented models display a better performance with higher McFadden $Pseudo - R^2$ and lower (Akaike) information criteria. The tone index parameter is statistically significant at the 5% level across the different specifications, suggesting that it is a relevant variable in anticipating the next policy decision. Moreover, this performance holds when we only consider the smoothing term and the HDT weighted index.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta r_{t+1}$</td>
<td>8.823***</td>
<td>8.174***</td>
<td>8.157***</td>
<td>7.695***</td>
<td>7.432***</td>
<td>8.855***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.976)</td>
<td>(1.006)</td>
<td>(1.004)</td>
<td>(1.049)</td>
<td>(1.152)</td>
<td>(0.972)</td>
<td></td>
</tr>
<tr>
<td>$\Delta (\pi_t - \pi^T)$</td>
<td>0.400</td>
<td>0.323</td>
<td>0.464</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.389)</td>
<td>(0.396)</td>
<td>(0.403)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta (y_t - y^*_t)$</td>
<td>4.516</td>
<td>3.215</td>
<td>4.318</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.510)</td>
<td>(6.297)</td>
<td>(6.745)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$HDTw_t$</td>
<td>2.440**</td>
<td>2.597**</td>
<td>2.820**</td>
<td>2.973**</td>
<td></td>
<td>5.578***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.159)</td>
<td>(1.152)</td>
<td>(1.199)</td>
<td>(1.119)</td>
<td></td>
<td>(0.954)</td>
<td></td>
</tr>
<tr>
<td>$\Delta r_{t+1}^{fedsfunds}$</td>
<td>4.056***</td>
<td>3.904***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.961)</td>
<td>(0.927)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.339</td>
<td>0.352</td>
<td>0.349</td>
<td>0.406</td>
<td>0.401</td>
<td>0.335</td>
<td>0.106</td>
</tr>
<tr>
<td>Parallel Assumption</td>
<td>holds</td>
<td>holds</td>
<td>holds</td>
<td>holds</td>
<td>doesn't holds</td>
<td>holds</td>
<td>holds</td>
</tr>
<tr>
<td>LLR</td>
<td>-117.512</td>
<td>-115.244</td>
<td>-115.738</td>
<td>-105.794</td>
<td>-106.644</td>
<td>-118.340</td>
<td>-158.981</td>
</tr>
<tr>
<td>AIC</td>
<td>253.024</td>
<td>250.488</td>
<td>247.477</td>
<td>233.588</td>
<td>231.288</td>
<td>250.68</td>
<td>331.962</td>
</tr>
</tbody>
</table>

Source: Author’s calculations. Notes: Ordinal logit models estimated by maximum likelihood. Asymptotic standard errors are given in parentheses. Significant at the (*** ) 1 percent level; (**) 5 percent level; (*) 10 percent level. Brant test was performed to validate the parallel lines assumption. In model D, $\Delta r_t$ and $HDTw_t$ parameters violate the parallel lines assumption. LLR stands for Log-Likelihood Ratio.

13 These findings are consistent with similar studies that also do not find statistical significance for macroeconomic variables in the cases of the Swedish Riksbank (Apel & Blix-Grimaldi, 2014), the ecb (Picault & Renault, 2017; Priola et al., 2022), the Bank of England and the Federal Reserve (Priola et al., 2022).
Furthermore, models augmented with the change in the federal funds rate exhibit higher McFadden $Pseudo-R^2$ values and lower (Akaike) information criteria. Despite the significance of the estimated coefficients for the change in the federal funds rate in each regression, violations of the parallel lines assumption are observed in model D. Conversely, model E, an econometric specification incorporating the smoothing term, the tone index, and the change in the federal funds rate as explanatory variables, outperforms all other models. These findings suggest a potential correlation between Banxico’s future decisions and its communication, as well as the interest rate announcements from the Federal Reserve.

Simultaneously, when applying a ‘from general to specific’ approach, taking model D as the starting point, looking for individual significance in the coefficients of the independent variables and the non-rejection of the null hypothesis in the Brant test to verify the fulfillment of the parallel lines assumption, the estimates lead us to consider models C and E as the best specifications.

While the results suggest a potential coordination in Banxico’s monetary policy with the Federal Reserve’s interest rate decisions, there are notable correlations between the smoothing term and the tone index with the information contained in the lagged federal funds rate. As a result, incorporating this variable into our analysis may give rise to multicollinearity issues and does not enhance the predictive capability of our model.\textsuperscript{14} Given the complexity of exploring detailed policy coordination effects, beyond the scope of this paper, we opt for the simplicity of analysis and consider model C (hereafter benchmark model) as the most suitable econometric specification.

In summary, the results suggest that Banxico’s monetary policy statements may provide insights into future short-term monetary policy decisions. Moreover, as the tone index remains informative even after accounting for macroeconomic variables, the analysis indicates that communication captures information not reflected in inflation and output gaps.

However, as shown in Table 3, model F, which includes only the smoothing term, outperforms model G, which considers the tone index alone. Therefore, while the tone index is statistically significant across all specifications and communication appears to be a relevant explanatory variable, it is possible that

\textsuperscript{14} Certainly, model C demonstrates a success rate of 72.3%, slightly surpassing model E, which exhibits a success rate of 70.8%. Detailed results are available upon request.
the primary source of predictive power in the model stems from the significant autocorrelation observed in the dependent variable.

The estimated coefficients of the logit regression do not directly reveal information about the contributions of independent variables on the dependent variable, unlike the linear models. Therefore, to yield an economic interpretation, we calculated the average marginal effects (AME) related to the benchmark model, as well as models F and G. To facilitate the interpretation of the results, the tone index is standardized and denotes variations in terms of standard deviations. The AME provide information of the estimated change in the predicted probabilities of moving up or down the ordinal response categories for a one-unit change in the independent variable, while holding other variables constant. In our analysis, the latter means we report the effects on the probability of changes in the policy rate (±75bp, ±50bp, ±25bp, or no changes) for a one standard deviation increase in the tone index, or for a unit change in the discrete policy rate. Table 4 shows the results.

A positive AME indicates that an increase in the independent variable is associated with higher odds of moving to a higher category of the dependent variable. A negative AME indicates the opposite, that is, a higher odd of moving to a lower category. For instance, according to the benchmark model, when Banxico tightens its monetary policy, the probability of observing a 25bp, 50bp, or 75bp rate hike in the next decision increases by 70%, 17%, and 1% (not significant), respectively. On the other hand, when Banxico eases its monetary policy, this probability decreases by 35%, 12%, and 1% (not significant).

Table 4: Average marginal effects

<table>
<thead>
<tr>
<th></th>
<th>Pr(-3)</th>
<th>Pr(-2)</th>
<th>Pr(-1)</th>
<th>Pr(-0)</th>
<th>Pr(0)</th>
<th>Pr(1)</th>
<th>Pr(2)</th>
<th>Pr(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta r_t$</td>
<td>-0.013</td>
<td>-0.121**</td>
<td>-0.347***</td>
<td>-0.404*</td>
<td>0.700***</td>
<td>0.174**</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>$HDT w_t$</td>
<td>-0.004</td>
<td>-0.039</td>
<td>-0.111**</td>
<td>-0.129</td>
<td>0.223***</td>
<td>0.055*</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>$\Delta r_t$</td>
<td>-0.012</td>
<td>-0.134**</td>
<td>-0.410***</td>
<td>-0.485*</td>
<td>0.784***</td>
<td>0.239**</td>
<td>0.018</td>
<td></td>
</tr>
<tr>
<td>$HDT w_t$</td>
<td>-0.080*</td>
<td>-0.243***</td>
<td>-0.292***</td>
<td>-0.274</td>
<td>0.508***</td>
<td>0.297***</td>
<td>0.084*</td>
<td></td>
</tr>
</tbody>
</table>

Source: Author's calculations.
Notes: Average marginal effects of models C, F and G. Statistical significance at 1% (***), 5% (**) and 10% (*).
As for the tone index, we found that a one-standard-deviation increase (decrease) change in the tone of communication increases on average the probability of a rate hike (rate cut). Accordingly, the probability of observing a rate hike of 25bp, 50bp, or 75bp in the next decision increases by 22%, 6% and 0.4% (not significant) when Banxico delivers a hawkish message in the monetary policy statement. Conversely, when the message is dovish, this probability decreases by 11%, 4% (not significant), and 0.4% (not significant). These results hold when we consider the ame derived from models F and G.

During the period under study, Banco de México’s Governing Board met on 130 occasions; on 76 of these meetings, it was decided to leave the policy rate unchanged, on 23 it was cut by at least 25bp, and it was raised by at least 25bp on 31 occasions. Our benchmark model is effective in anticipating, on average, 7 out of 10 changes in Banco de México’s monetary policy stance. Specifically, it is highly efficient when the monetary stance remains unchanged, successfully anticipating 93% of the time. When large changes occurred (±75bp) the model has a hit success rate of 66% and 75% for negative and positive shifts, respectively. On the other hand, the model anticipates 5 out of 10 times that the policy rate increased by 25bp and 50bp, respectively. It shows similar results in the case of 50bp cuts (4 out of 9). Finally, the model presents its worst performance for changes of 75bp, as it failed to predict even one of the 11 times the CB cut its rate by this amount. Similar predictive ability is observed when examining the results from model F, confirming that the tone index contains limited predictive ability. Table 5 summarizes the outcomes.

Table 5: Predictive power of Banco de México’s communication

<table>
<thead>
<tr>
<th>Model</th>
<th>Observed</th>
<th>−75 pb</th>
<th>−50 pb</th>
<th>−25 pb</th>
<th>hold</th>
<th>+25 pb</th>
<th>+50 pb</th>
<th>+75 pb</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmark</td>
<td>Prediction</td>
<td>3</td>
<td>9</td>
<td>11</td>
<td>76</td>
<td>16</td>
<td>11</td>
<td>4</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>Success Rate (%)</td>
<td></td>
<td>66.67</td>
<td>44.45</td>
<td>0.00</td>
<td>93.40</td>
<td>50.00</td>
<td>54.60</td>
<td>75.00</td>
</tr>
<tr>
<td>F</td>
<td>Prediction</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>71</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Success Rate (%)</td>
<td></td>
<td>66.67</td>
<td>44.45</td>
<td>0.00</td>
<td>94.70</td>
<td>0.00</td>
<td>54.60</td>
<td>75.00</td>
</tr>
<tr>
<td>G</td>
<td>Prediction</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>75</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Success Rate (%)</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>98.7</td>
<td>0.00</td>
<td>27.30</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Source: Author’s calculations.
These findings suggest that, while communication is indeed a significant variable in our analysis, the ability to successfully anticipate the next policy decision with a high degree of accuracy may not be plausible without acknowledging the influence of the smoothing parameter in the monetary policy. On the contrary, the predictive effects of communication in our model align well with Solís (2023) definition, emphasizing the presence of two factors in monetary policy announcements. The first factor is associated with surprises about the current policy rate (the change in the policy rate), while the other factor is linked to surprises about its future path communicated through policy statements (the tone index).

Nevertheless, the limited predictive capability of the tone index could be related to the fact that the message conveyed through the policy statement contains both past and new information regarding monetary policy. Consequently, while our index effectively captures the monetary stance, it may not differentiate whether that stance pertains to the current or future state of affairs. Instead, it gauges the overall tone and content of CB communication, which may include a mix of current and future-oriented information. Therefore, one limitation of the index may be attributed to the fact that it predominantly reflects the current rather than future monetary policy stance. Moreover, while statements can be an important source of guidance, the lag between monetary policy meetings, typically six weeks, could dilute the impact of the message. As a result, other communication channels, such as minutes and officials’ speeches, may become more relevant for anticipating Banxico’s actions.

At the same time, during most of the period under study, Banxico held the policy rate steady, or was in a tightening cycle, which explains why the model is more effective in anticipating this type of decisions; interest rate cuts were not systemic; there were occasions in which the CB changed its monetary stance without modifying the communication strategy. For example, unanticipated changes were observed, such as in February 2016 (oil price shock) and March-April 2020 (covid-19 pandemic), when the CB modified its monetary stance in extemporaneous decisions.

Overall, our analysis suggests that a rate hike (cut) and a hawkish (dovish) monetary policy statement are associated with higher odds of observing a restrictive (accommodative) decision at the next monetary policy meeting. Furthermore, the analysis reveals that communication captures information not reflected in traditional macroeconomic variables. However, when it comes to effectively
anticipating the next monetary policy decisions, communication conveyed through policy statements contains limited predictive ability.

5. Conclusions

Following the world’s leading CBs and the observed trends in monetary policy making, Banxico has made significant efforts to improve its communication and transparency strategy over the last 20 years. In this paper, we have measured the role of CB qualitative communication, based on the language contained in the monetary policy statements of this institution.

We have leveraged Natural Language Processing techniques to uncover aspects of CB communication. In particular, using a Latent Dirichlet Allocation model, an unsupervised machine-learning technique, and dictionary-based sentiment analysis, we have developed our Hawkish-Dovish Tone weighted index, which gauges the bias in the tone of communication. The LDA model was useful to identify a series of latent topics on which Banxico’s language focused between 2008 and 2022. Using this information, we could capture CB’s talk on 7 categories: Monetary Policy, Economic Activity, International Affairs, Exchange Rate, Inflation Risk, Balance of Risks, and COVID-19 Pandemic. Notably, the topics of Monetary Policy emerged as the primary driver of Banxico’s message. The model also captured interesting patterns related to recent macroeconomic developments, such as the prevalence of the Exchange Rate topic during volatility in international financial markets and the fall in oil prices in 2016, the rise of the COVID-19 Pandemic topic since 2020, and discussions on Inflation Risk in recent years.

Likewise, we found a moderately positive linear correlation between the changes in Banxico’s policy rate and our tone index. Moreover, the analysis revealed that a rate hike (cut) and a hawkish (dovish) monetary policy statement are associated with higher odds of observing a restrictive (accommodative) decision at Banxico’s next meeting. These results are consistent with recent research empirically testing the same hypothesis in other CBs (see Astuti et al., 2022; and Priola et al., 2022).

Furthermore, the analysis reveals that communication captures information not reflected in traditional macroeconomic variables. The latter was an interesting upshot of our study. It is worth mentioning that CB communication can capture prospective and retrospective qualitative information about the CB’s view.
Therefore, we interpret these results as follows: CB communications are informative and influential and have the ability to encompass information on important macroeconomic variables in the determination of monetary policy, i.e., by having more and better information on upcoming economic outcomes, and on its own behavior, the CB shares information that by itself is relevant in the short run. Accordingly, our model accurately predicted most decisions when the monetary stance remained steady and was less effective predicting changes of 75bp, 50bp and 25bp. Moreover, it failed to anticipate a single one of the 11 decisions in which the central bank cuts its rate by 25bp. In other words, the model provides information to anticipate with 70% effectiveness what Banxico will do in its next decision.

Despite communication is indeed a significant variable in our analysis, it contains limited forward-looking information. Certainly, the benchmark model achieved a success rate of 70% in anticipating the next monetary policy decision. However, successfully predicting the change in the policy rate in future decisions was not achievable without considering the influence of the smoothing parameter in monetary policy. The primary source of predictive power in the model stems from the significant autocorrelation observed in the dependent variable. Nonetheless, caution is warranted in interpreting this approach to the predictive power of CB communication. It’s important to note that the constructed index, employed to assess the CB forward-looking perspective within the statement, encompasses more than this particular aspect. CBs disseminate both past and new information regarding monetary policy, especially through statements. Consequently, while our index gauges the overall verbal monetary policy stance and content of CB communication, it may not differentiate whether that stance pertains to the current or future state of affairs. Therefore, the limited predictive ability of the index may be attributed to the fact that it predominantly reflects the current rather than future monetary policy stance. Additionally, although statements can provide valuable guidance, the lag between monetary policy meetings could diminish their impact. Hence, other communication channels, such as minutes and speeches by officials, could be more pertinent for predicting Banxico’s actions.

Going forward, there are various implications of our analysis. First, if communication conveyed through policy statements does not exclusively focus on forward guidance but mostly encompasses the current monetary policy stance, it could be interesting to explore its relationship with contemporaneous variables.
For instance, one possibility is to assess whether there is consistency between the published statement and the subsequent action taken, such as the policy rate target level. The positive linear correlation between changes in Banxico’s policy rate and our tone index could shed light on the coherence between the central bank’s actions and words.

Secondly, the text-processing techniques outlined in this article could be useful to develop further investigation on the predictability of monetary policy. On the one hand, we could leverage the results of the LDA model to explore how different aspects of communication, such as monetary policy, economic activity, international affairs, to mention a few, may be useful to forecast policy and other market interest rates. In addition, it might be interesting to explore the possibility of building more precise dictionaries capable of capturing the unique features of Banxico’s wording. On the other hand, we could apply the same methodology to other text sources, such as quarterly reports, minutes, or speeches by officials, to verify if our conclusions hold across these documents. This would help determine if the forward-looking information contained in these documents further improves the predictability of Banxico’s monetary policy decisions. At the same time, it could be interesting to study the news about central banking on the day of the decision, and how analysts and financial markets react to the communications in the very short term.

Another potential investigation pertains to exploring if CB qualitative communication shapes expectations of professional forecasters. Hence, a feasible approach could encompass integrating indicators akin to our tone index within varied econometric models to examine communication’s impact on private agents’ expectations. It’s reasonable to anticipate that effective CB communication could affect these variables, even if temporarily.

Overall, our analysis indicates that Banxico’s monetary policy statements contain valuable information about monetary policy decisions in the short term. The tone index, even after considering macroeconomic variables, remains informative, suggesting that it captures information not reflected in traditional economic indicators. However, the limited forward-looking information in the tone index indicates that other factors, such as changes in the current policy rate, play a more significant role in predicting future monetary policy decisions. These results highlight the importance of considering a range of factors, including both qualitative and quantitative information, in predicting CB policy decisions.
As stated by Ben Bernanke (2015), monetary policy today is 98 percent talk and only two percent action, so the ability to shape market expectations about future policy through public statements is one of the most powerful tools CBs have. Thus, following Gabriel Markhlouf (2020), we should expect the days of ‘mumbling with great incoherence’ to be over.

References


**Appendix A: Text mining and document preprocessing**

To apply the techniques implemented in this research to the publications of the central bank, a process of filtering and cleansing the documents must be carried out. This is commonly known as text mining, and it involves extracting useful and relevant information from a set of unstructured text documents. It is used in the field of data mining and artificial intelligence to analyze large amounts of text and discover hidden patterns, trends, and insights.

The objective of text mining is to transform textual data into structured and comprehensible information, enabling deeper analysis and informed decision-making. Various techniques and algorithms are applied during the text mining process, such as natural language processing (NLP), text classification, entity extraction, sentiment analysis, among others. Some common tasks in text mining include identifying keywords, detecting topics, document classification, grouping similar documents, and extracting specific information such as names of individuals, locations, or dates. In summary, text mining is a technique that allows converting large volumes of unstructured data into valuable information, facilitating the understanding and analysis of texts in different contexts and applications.

For our specific case, the first step is to download the publications. This includes all Banxico’s monetary policy statements from 2008 to 2022. All information was obtained in PDF format from Banco de México’s official website. I chose to analyze the publications in the original Spanish language for two reasons: first, the availability of the documents, and second, to eliminate any bias that might be introduced in the translation. Secondly, before applying any text mining techniques, it is necessary to preprocess the documents to obtain a set of plain text data. The filtering process involves removing all non-text-related objects, such as headers, page numbers, covers, indexes, table of contents, footnotes,
appendices, etc. Despite the many efforts made by academics and researchers to develop techniques for including graphics, tables, figures, charts, and other visual elements in linguistic analysis, no such method is available. Therefore, even though such material can add valuable content to the analysis, it is necessary to eliminate them.

With this information, a corpus of central bank documents can be constructed. Specifically, a corpus is a structured and systematic collection of documents or other linguistic data that are gathered and organized for the purpose of analysis and study. The central bank communication corpus provides a valuable data source for the empirical analysis in this research. Our corpus consists of 131 plain text documents, for each monetary policy statement. A second filtering step is performed, which involves converting each letter to lowercase, removing all numbers, punctuation marks, and stop words. Stop words are common words that are excluded or filtered in natural language processing and text analysis. These words are very frequent in language but often do not contribute substantial meaning to the context or understanding of the text. Common examples of Spanish stop words include “de,” “el,” “y,” “es,” “en,” “a,” “una,” “para,” among others. These words are often adverbs, conjunctions, prepositions, and pronouns that are frequently used in sentence construction but do not provide distinctive or relevant information for text analysis. By eliminating these words, noise or interference in the analysis can be reduced, and the most important and distinctive words (or terms) in the text can be identified more accurately.

The following is an example of what a paragraph from Banco de México’s publications would look like, with and without punctuation marks, numbers and stop words. Specifically, the Monetary Policy Statement of December 12, 2022, is used:

**ORIGINAL STATEMENT**

La Junta de Gobierno del Banco de México decidió incrementar en 50 puntos base el objetivo para la Tasa de Interés Interbancaria a un día a un nivel de 10.50%, con efectos a partir del 16 de diciembre de 2022. La actividad económica mundial se recuperó moderadamente en el tercer trimestre, aunque las perspectivas para 2023 siguieron deteriorándose. La inflación global se mantiene elevada, si bien la general disminuyó en diversas economías ante menores presiones en los precios de alimentos y energéticos.
This process allows us to identify only unique and significant terms for analyzing the lexical diversity of the central bank. It can be argued that removing stop words, numbers, and punctuation makes reading the text slightly more difficult. However, the message and context of the publication remain understandable. Finally, a series of n-grams are collapsed. N-grams are continuous sequences of n elements, which can be words, symbols, or tokens in a document. In the corpus constructed so far, bi-grams and tri-grams are chains of two and three words, respectively, that often appear together in the text. For example, “crecimiento económico”, “tasa de interés”, “política monetaria”, “Estados Unidos”, “tipo de cambio”, “meta de inflación”. etc. The n-grams considered in Banxico publications are a combination of Toborda’s list (see Taborda, 2015) adapted to the central bank’s style and extended with some terms that the author considers relevant for identification:

• banco central, bancos centrales, banca central \(\rightarrow\) banco_central.
• banco central europeo \(\rightarrow\) bce.
• banco de México \(\rightarrow\) banxico.
• crecimiento económico, crecimiento de la economía \(\rightarrow\) crecimiento_económico.
• crisis financiera \(\rightarrow\) crisis_financiera.
• comercio internacional \(\rightarrow\) comercio_internacional.
• covid-19, coronavirus, sarscov2, covid \(\rightarrow\) covid.
• déficit fiscal, déficit público \(\rightarrow\) déficit_fiscal.
• europa, eurozona, europeo, europea, europeos, eurosistema, eurogrupo \(\rightarrow\) europa.
• estabilidad financiera \(\rightarrow\) estabilidad_financiera.
The Predictive Power of Central Bank Communication: Evidence from Mexico

- estados unidos, estadounidense, estadounidenses → estados_unidos.
- internacional, internacionales, mundial, mundiales, mundo → internacional.
- junta de gobierno → junta_gobierno.
- objetivo de inflación, objetivos de inflación, inflación objetivo, meta de inflación, metas de inflación, inflación meta → inflación_objetivo.
- mercados financieros, mercado financiero, sistema financiero, sector financiero → mercado_financiero.
- sistemas bancarios, sistema bancario, sector bancario, sectores bancarios → sector_bancario
- mexicana, mexicano, mexicanos, méxico → méxico.
- política fiscal, políticas fiscales → política_fiscal.
- política monetaria, políticas monetarias → política_monetaria.
- tipo de cambio, tipos de cambio, tasa de cambio, tasas de cambio → tipo_cambio.
- tipo de interés, tipos de interés, tasa de interés, tasas de interés → tasa_interés.
- tipo de interés objetivo, tipos de interés objetivo, tasa de interés objetivo, tasas de interés objetivo, tipo de interés de referencia, tipos de interés objetivo de referencia, tasa de interés objetivo de referencia, tasas de interés objetivo de referencia, tasa objetivo, tasa de referencia, tasa de fondos federales, tasa de interés fondos federales → tasa_monetaria.
- reserva federal → fed.

The following paragraph exemplifies how Banxico’s publications would look like, without punctuation, numbers, stop words, and with collapsed n-grams:

**STATEMENT WITHOUT PUNCTUATION, NUMBERS, STOP WORDS, AND WITH COLLAPSED**

junta_gobierno banxico decidió incrementar puntos base objetivo tasa_interés interbancaria día nivel efecto partir diciembre actividad económica mundial recuperó moderadamente tercer trimestre aunque perspectivas siguieron deteriorándose inflación global mantiene elevada si bien general disminuyó diversas economías menores presiones precios alimentos energéticos amplio número banco_central continuó incrementando tasa_monetaria mencionaron comenzarían moderar magnitud aumentos obstante anticipa dichas tasas permanezcan niveles altos periodo prolongado
Finally, it is necessary to stem the words. This step can be applied at the researcher’s discretion, but it is strongly recommended to reduce the remaining words in the corpus. By stemming the words, we collapse similar terms to their root form. For example, “economía”, “económico”, “economista”, “economizar” would be reduced to “econom”. This is useful as it represents a normalization of the elements in the corpus, allowing different variations of words to be counted as a single term. This reduces the amount of information to manipulate in subsequent analysis. In addition, in the lemmatization of the publications, relevant words for synonyms, such as “meta” and “objetivos”, or “incremento”, “aumento” and “elvado”, or “internacional” and “global”, among other words, are reduced to a single root term. The result of a paragraph in the corpus documents would look like as shown below:

**STATEMENT WITHOUT PUNCTUATION, NUMBERS, STOP WORDS, COLLAPSED N-GRAMS, AND REDUCED WORDS:**

junta_gobiern banxic decid increment punt bas objet tasa_interes interban-cari dia nivel efect part diciembr activ econom mundial recuper moder terc trimestr aunqu perspect sigu deterior inflacion global mantien elev si bien general disminu divers econom menor presion preci alimen energet ampli numer banco_central continu increment tasa_monetari mencion comenz moder magnitud aument obstant anticip dich tas permane ncv nivel altos period prolong decision recent fed estados_un aument rang objet tas fond federal punt bas despuݕ cuatr increment consecut punt bas asim anticip increment adicional riesg global

The latter methodology results in a corpus ready to be transformed into a Document-Term Matrix. A Document-Term Matrix (dtm) is a tabular representation used in natural language processing and text analysis. It organizes a collection of text documents into rows and terms (words or phrases) into columns. Each cell in the matrix contains a numerical value representing the frequency or importance of a specific term in a particular document. dtms are used to capture the textual content of documents in a format suitable for various analytical techniques, such as topic modeling, sentiment analysis, and machine learning algorithms. Our communication corpus yields a dtm with 131 rows, each representing a monetary policy statement, and 1,649 columns, each representing unique terms in the entire corpus. The dtm entries show the frequency of each unique term in each monetary policy statement.
Table 6: Document-Term Matrix

<table>
<thead>
<tr>
<th>Date</th>
<th>acentu</th>
<th>actual</th>
<th>acuerd</th>
<th>adopt</th>
<th>afect</th>
<th>agent</th>
<th>ahor</th>
<th>aliment</th>
<th>alivi</th>
<th>anterior</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008-01-18</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>2008-02-15</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>2008-03-14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>...</td>
</tr>
<tr>
<td>2008-05-16</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>2008-06-20</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>2008-07-18</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>2008-08-15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>2008-09-19</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>2008-10-17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>2008-11-28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>2009-01-16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>2009-02-20</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>...</td>
</tr>
<tr>
<td>2009-03-20</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>...</td>
</tr>
<tr>
<td>2009-04-17</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>2009-05-15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>2009-06-19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>2009-07-17</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>2009-08-21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>2009-09-18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Source: Author's calculations. Notes: This table presents an excerpt of the document-term matrix. Terms are presented in their root form in the original Spanish language. Columns indicate the date of publication of the monetary policy statement.

Finally, the same filtering process was applied to the dictionaries to ensure similar terms for analysis and comparison.