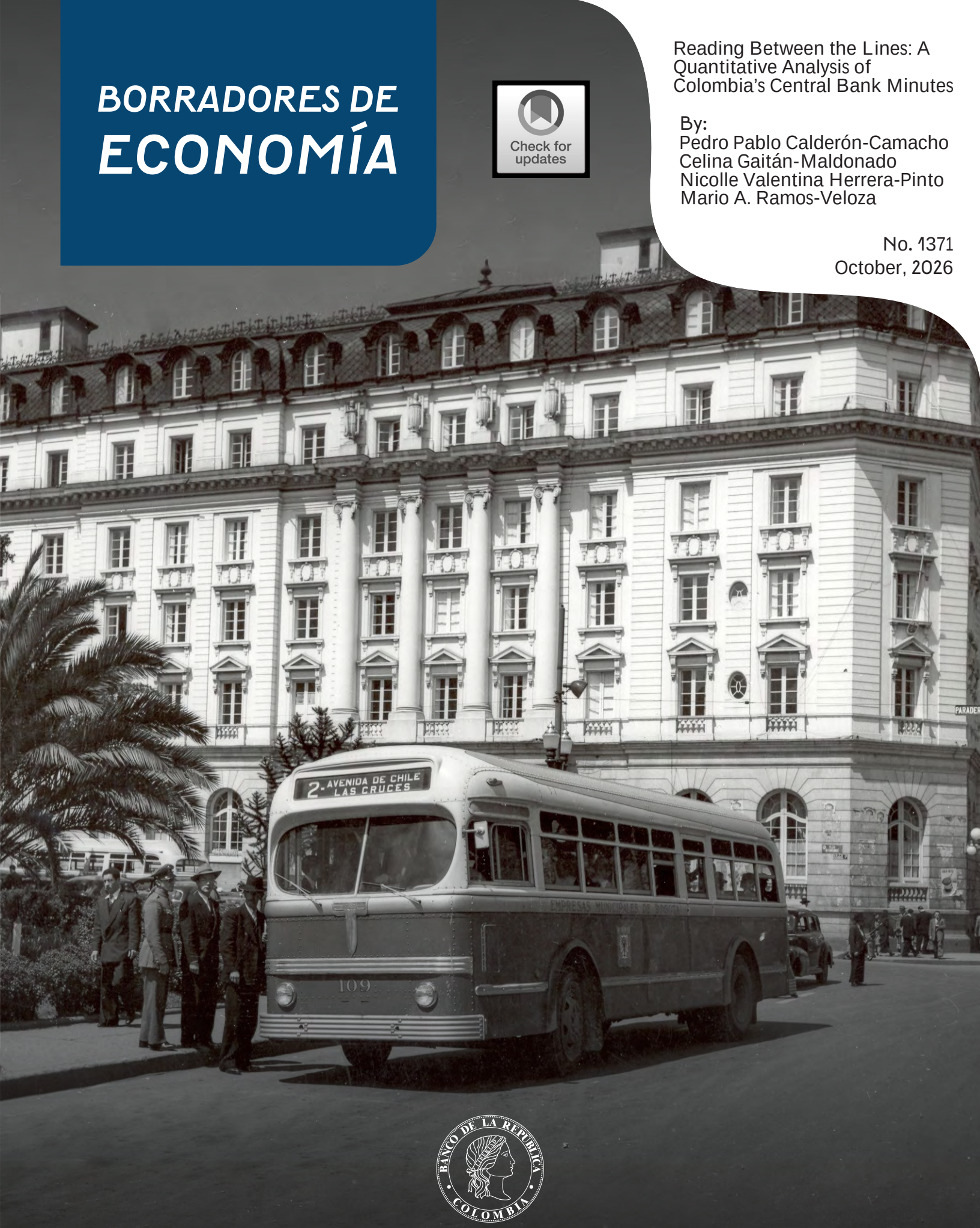
BORRADORES DE ECONOMÍA

Check for updates

# Reading Between the Lines: A Quantitative Analysis of Colombia's Central Bank Minutes

By:
Pedro Pablo Calderón-Camacho
Celina Gaitán-Maldonado
Nicolle Valentina Herrera-Pinto
Mario A. Ramos-Veloza

No. 1371
October, 2026




BANCO DE LA REPÚBLICA
COLOMBIA

# Reading Between the Lines: A Quantitative Analysis of Colombia's Central Bank Minutes*

Pedro Pablo Calderón-Camacho  Celina Gaitán-Maldonado

Nicolle Valentina Herrera-Pinto  Mario A. Ramos-Veloza†

The opinions contained in this document are the sole responsibility of the authors, and do not commit Banco de la República nor its Board of Directors.

**Abstract**

Central bank communication is a critical policy tool for inflation-targeting regimes. We analyze the tonality of monetary policy minutes to uncover inflationary pressures in Colombia. Three key findings emerge: (1) tonality reflects easing and tightening of monetary policy, rising preemptively before rate hikes (e.g., the pre-2015 oil shock) and falling during policy rate reductions; (2) during crises such as COVID-19, tonality fell during the acute COVID-19 shock but surged in its aftermath, signaling latent inflationary risks; and (3) the tonality of central bank minutes is associated with subsequent adjustments in inflation expectations, with a more hawkish tone followed by a downward revision in expected future inflation relative to the current inflation. This association is stronger when policy decisions are unanimous. We emphasize the role of central bank communication in conveying information relevant to the formation of inflation expectations.

**JEL Classification:** E52, E58, D84, C55, C53

***Keywords:*** *Minutes, Natural language processing, Inflation expectations, Tonality.*

*The authors gratefully acknowledge the comments of Julián Parra, Carlos Huertas and the audience of Banco de la República on an earlier version of this manuscript. Any remaining errors are the sole responsibility of the authors.

†Corresponding author: mramosve@banrep.gov.co.

# Leyendo entre líneas: un análisis cuantitativo de las minutas del banco central de Colombia

Pedro Pablo Calderón-Camacho Celina Gaitán-Maldonado

Nicolle Valentina Herrera-Pinto Mario A. Ramos-Veloza

Las opiniones contenidas en el presente documento son responsabilidad exclusiva de los autores y no comprometen al Banco de la República ni a su Junta Directiva.

**Resumen**

La comunicación de los bancos centrales es un instrumento de política fundamental en los regímenes de meta de inflación. Analizamos la tonalidad de las minutas de política monetaria para identificar las presiones inflacionarias en Colombia. Surgen tres hallazgos principales: (1) la tonalidad refleja las fases de endurecimiento y relajamiento de la política monetaria, aumentando de manera preventiva antes de los incrementos en la tasa de interés (por ejemplo, previo al choque petrolero de 2015) y disminuyendo durante las reducciones de la tasa de política; (2) durante crisis como la del COVID-19, la tonalidad cayó durante el choque agudo de la pandemia, pero repuntó en su fase posterior, señalando riesgos inflacionarios latentes; y (3) la tonalidad de las minutas del banco central está asociada con ajustes posteriores en las expectativas de inflación, de modo que un tono más hawkish es seguido por una revisión a la baja de la inflación esperada con respecto a la inflación actual. Esta asociación es más fuerte cuando las decisiones de política son unánimes. Destacamos el papel de la comunicación del banco central en la transmisión de información relevante para la formación de expectativas de inflación.

**Clasificación JEL**: E52, E58, D84, C55, C53

***Palabras clave**: Minutas, Procesamiento de lenguaje natural, Expectativas de inflación, Tonalidad.*

# 1 Introduction

What information do central bank meeting minutes convey beyond the interest rate decision itself, and does it matter for how economic agents form inflation expectations? This question is central to monetary policy and specifically inflation-targeting regimes, where communication has become a central instrument for anchoring expectations, reducing uncertainty, and enhancing policy credibility Blinder et al. (2008); Bholat et al. (2019). We address this question by constructing a sentence-level tonality index for the meeting minutes of Colombia's Banco de la República (BanRep) between 2007 and 2025. Using three Natural Language Processing (NLP) methods, we classify each sentence as hawkish, dovish, or neutral, and then compute the tonality as the Board's assessment of inflationary pressures. We then examine whether the tonality is informative for the formation of inflation expectations.

Meeting minutes play a key role in central bank communication. They are quickly available after each decision and provide a detailed record of the Board's deliberations, the macroeconomic context that informed the decision, document majority and minority viewpoints, and articulate the rationale underlying the policy choice. Thus, minutes provide a more detailed and deliberative account than press releases, which are often brief and formulaic announcements of the rate decision and are more timely than quarterly inflation reports. As Blinder et al. (2008) note, the move toward greater transparency has made the qualitative content of deliberations increasingly relevant to market participants and professional forecasters. Therefore, minutes are a natural complement to studies of press releases, offering a fuller window into how the Board's risk assessment evolves.

Central bank communications can be assessed according to the inflationary risks they convey along a hawkish-to-dovish spectrum, thereby reflecting the monetary policy orientation underlying interest rate decisions. A hawkish tone signals heightened concerns about inflationary pressures and a preference for a more restrictive policy stance, whereas a dovish tone indicates more subdued inflation concerns and greater support for accommodation. This classification is informative for both policymakers and market participants, as it facilitates the interpretation of the central bank's economic outlook and helps anticipate the likely direction of future monetary policy decisions, with potential implications for financial markets and economic expectations Hansen and McMahon (2016). Therefore, tone indicators complement the policy rate, which captures the contemporaneous policy decision, by revealing the *intensity* of the inflationary concerns underlying that decision Carotta et al. (2023), potentially providing signals regarding the direction and magnitude of subsequent policy adjustments.

Similar to the assessments conducted by the central bank, economic agents form expectations based on their own evaluations of macroeconomic risks and perceptions of the credibility of monetary policy decisions González et al. (2011). Therefore, the ability of monetary policy to anchor expectations depends, in part, on whether central bank communications convey information about the likely course of policy clearly and credibly. When central bank deliberations communicate a hawkish assessment—indicating either elevated inflationary pressures or the Board's willingness to respond forcefully—agents who perceive this information as credible and previously unanticipated may revise

their expected inflation paths. The direction of this revision depends on the dominant channel: newly revealed information about inflationary pressures may lead expectations upward, whereas a stronger perceived commitment to containing inflation may lead them downward.

In Colombia, previous studies have analyzed the role of BanRep communications. Guío-Martínez et al. (2020) analyzed the main topics discussed in minutes and monetary policy reports between 2007 and 2019, providing a descriptive map of content without constructing a tonality measure. Melo-Velandia et al. (2020) show that the narrative information in the minutes affects the yield curve but do not study inflation expectations directly. More closely related to our study, Arango et al. (2023) construct tone measures for BanRep *press releases* using latent semantic analysis and find evidence that they are related to inflation expectations. Similarly, Vega and Lahura (2020) construct dictionary-based tone indicators for the press releases of Colombia's, Chile's, and Peru's central banks. Our study extends this body of work in three key respects. First, we study the minutes rather than press releases: longer, more deliberative documents that present a fuller picture of the Board's risk assessment, internal disagreement, and forward-looking guidance. Second, we apply and systematically compare three NLP methods, including a fine-tuned Spanish-language transformer and a large language model, approaches not previously used for BanRep communications. Third, we extend the analysis to sector-level heterogeneity in expectations, dimensions that are absent from prior work on Colombia.

This study also contributes to the growing international literature on the application of NLP techniques to central bank communication. Hansen and McMahon (2016) use topic modeling of Federal Open Market Committee (FOMC) statements to identify the content of Federal Reserve deliberations and study their macroeconomic effects. Recently, transformer-based models have been applied to ECB minutes Niţoi et al. (2023) and FOMC statements Hansen and Kazinnik (2024), and large language models have been used to classify the tone of central bank communications in financial contexts. Carotta et al. (2023) study the tone and readability of Uruguay's central bank communications and their relation to firms' inflation expectations. Our study applies these frontier methods to a Spanish-language emerging market context, contributing evidence of an institutional setting and linguistic environment distinct from those studied in the existing literature.

Relative to existing work applying NLP methods to central bank communication, including studies specifically focused on Colombia, our contribution is threefold. First, we construct and directly compare three tonality indices for the same corpus — a context-sensitive dictionary, a fine-tuned transformer, and a large language model — and show that they yield a consistent picture of inflationary pressures across monetary policy phases, rising during tightening periods and falling when inflationary pressures and interest rates decrease. The indices peaked following the COVID-19 pandemic, coinciding with global supply chain disruptions, the Russia-Ukraine war, and domestic price pressures, and exhibited anticipatory behavior ahead of shifts toward a tighter monetary policy stance. Second, we extend the tonality index along a forward-looking dimension, distinguishing prospective from backward-looking statements in the minutes. Third, we document that tonality is associated with subsequent adjustments in inflation expectations, that this association is systematically stronger when

policy decisions are unanimous, and that it differs across sectors. We interpret these expectations results as evidence of the predictive and informational content of central bank communication rather than as a causal effect. We discuss the identification limitations of this exercise in Section 5. Taken together, tonality emerges as a valuable and systematically constructed complement to existing tools for monetary policy analysis.

The remainder of this paper is organized as follows. Section 2 presents an analysis of the minutes of the Colombian Central Bank and their evolution. Section 3 outlines the three methodologies employed to construct the tonality indicators. Section 4 describes the dynamics of the tonality indices and whether minutes include forward-looking information. Section 5 investigates whether and to what extent tonality functions as an independent channel of communication associated with inflation expectations and how unanimity relates to this association. Finally, Section 6 concludes.

# 2 Data - Minutes

Since the adoption of inflation targeting, regular meetings of the *Banco de la República* board of directors (BDBR) have been critical in aligning monetary policy decisions with the central bank's objectives of transparency and predictability. These meetings enable the BDBR to analyze prevailing macroeconomic trends, assess short- and medium-term economic outlooks, and evaluate the risks to price stability. Policymakers then debate alternative measures to align inflation with the inflation target, ensuring that decisions are based on both current data and forward-looking scenarios. Although the BDBR convenes monthly, adjustments to the monetary policy rate occur exclusively during eight scheduled meetings per year (January, March, April, June, July, September, October, and December).[1]

A fundamental aspect of the BDBR's communication strategy is the dissemination of meeting minutes discussing the decision about the interest rate, which are available on the bank's website following each meeting.[2] The minutes summarize key arguments from majority and minority viewpoints, clarify the rationale for policy changes, and include a statistical annex with critical indicators such as disaggregated inflation data, GDP demand components, and labor market dynamics, as well as a technical document summarizing the key determinants of capital market dynamics.[3] By illustrating internal deliberations and providing granular economic data, minutes can anchor public and market expectations, reduce uncertainty, and enhance policy credibility González et al. (2019). This transparency mechanism not only reinforces accountability but also empowers households, firms, and investors to anticipate future monetary actions within a structured framework.

[1]Before 2018, the monetary policy rate could be adjusted every month; the reduction to eight annual *decision* meetings aligned the BDBR with international standards and allowed for more thorough analysis of economic data. See `https://www.banrep.gov.co/es/comunicado-24-noviembre-2017-frecuencia-reuniones`.

[2]In recent years, the interval between the meeting date and the release of minutes has decreased, from two weeks in 2007 to approximately three to six days after 2017.

[3]Minutes are available in Spanish since 2007 (`https://www.banrep.gov.co/es/minutas`) and English since 2008 (`https://www.banrep.gov.co/en/minutes`). In this study, we use the Spanish version of the minutes.

The structure of the minutes has undergone important changes since its initial publication in June 2007, and four distinct types of minutes can be identified.

**Type 1**: The first type of minutes, published between June 2007 and May 2014, was divided into three main sections: background, monetary policy discussion and options, and monetary policy decisions. The Background section was further subdivided to describe the most important aspects of recent inflationary developments, financial variables, and the external context. Additionally, a statistical annex was included to provide a detailed description.

**Type 2**: In 2014, the BDBR implemented significant modifications to its minutes to enhance the transparency of its monetary policy vision and stance. Thus, the second type of minutes, spanning from June 2014 to May 2019, placed greater emphasis on the discussion and monetary policy options sections, which were expanded. Conversely, the Background section, subsequently referred to as the Macroeconomic Context, adopts a more concise approach to the analysis of the aforementioned variables. The other two sections remained, with increased emphasis on discussion. Furthermore, the statistical annex was intended to be supplemented by working papers, which present the behavior of various internally monitored markets.[4]

**Type 3**: The third type of minutes was used between July 2019 and November 2020, a period that largely coincided with the COVID-19 pandemic. During this period, the Macroeconomic Context section was shortened to give greater prominence to the arguments supporting the Board's monetary policy decisions. These minutes also included a statistical annex summarizing the main developments in inflation, economic activity, the labor market, financial conditions, and the international economic environment, all of which were key inputs into the decision-making process. Although the narrative discussion of the macroeconomic context was reduced, the annexes provide important information on the economic conditions under which monetary policy decisions were made.

**Type 4**: Finally, from December 2020 onward, the minutes include a brief descriptive summary of the main macroeconomic developments, to facilitate the understanding of the Board's decision-making process. Nevertheless, the main focus of the minutes remained unchanged: the monetary policy decision and the discussion surrounding its adoption.

Notable differences exist in the length of minutes for each of these types. As illustrated in Figure 1, the average word count of the minutes decreased by nearly half between 2014 and 2018 compared to that from 2007 to 2013. Additionally, between 2019 and 2021, the minutes had an average length of approximately 650 words. This relatively shorter format did not result from the uncertainty associated with the COVID-19 pandemic but rather from a deliberate decision to give greater prominence to the discussion and rationale underlying the Board's monetary policy decisions. Nevertheless, the minutes continued to highlight the multiple risks facing the Colombian economy and describe the extraordinary measures adopted to provide liquidity to financial markets during the pandemic. Since 2022, the minutes have averaged 1,200 words, reflecting a trend similar to that observed before the pandemic.

[4] See https://www.banrep.gov.co/es/noticias/junta-directiva-modifica-minutas.

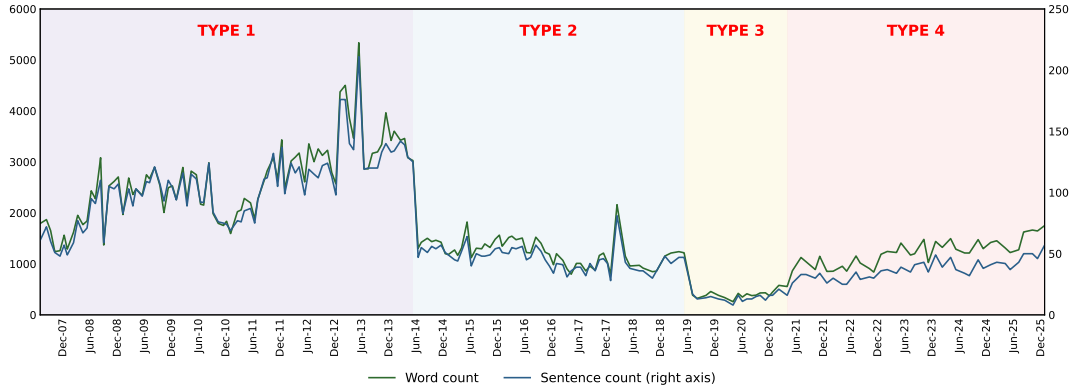



Figure 1: Average number of words and sentences per minute

Another change in the content of the minutes concerns the reporting of consensus in the policy decisions of the Banco de la República's board. Before 2016, the minutes explicitly indicated whether the decision was reached by consensus (i.e., unanimity). Starting in 2016, the minutes provide more detailed information about the voting breakdown, specifying the number of members voting in favor of the final decision and those who support alternative policy options. We find that a consensus among all members occurred in 38.3% of the meeting minutes, a value that is not affected by the change in the reporting format.

In conclusion, despite structural changes over time and the resulting textual variations, the minutes have consistently maintained three key focuses over the years: the macroeconomic context in which the policy discussion takes place, deliberations among board members reflecting different viewpoints, and the decisions adopted. These sections enable the analysis of the minutes and the comparison of the BDBR's perspectives over time.

# 3 Corpus and Classification of Tonality

In this section, we describe the corpus of minutes and outline the NLP methods used to construct the tonality indicators. Based on the previous literature, we categorized each sentence into three groups: hawkish, dovish, and neutral, according to the degree of inflationary pressure they describe or signal about the economy. For instance, Hawkish Sentences reflect pressures that could be driven by strong demand, tightening labor market conditions, or adverse supply shocks. In contrast, dovish sentences reflect low inflationary pressures due to deteriorating labor conditions, weakened gross capital formation, or declines in related indices (e.g., cement dispatches and industrial production). Finally, the neutral sentences included those with mixed or ambiguous signals. Thus, tonality reflects the board's expressed assessment of inflationary pressures. Throughout this study, we use "tonality," "tone," and "sentiment" interchangeably to refer to this construct.

Subsections 3.2 and 3.3 detail our Word2Vec (W2V) (a semi-supervised approach) and fine-tuning

of BETO (a supervised NLP model) for document classification in central bank minutes. In Subsection 3.4, we discuss how our corpus is classified based on the reasoning skills of AI developments. In Section 4 we present the tonality indices, their formulas, and the aggregation procedure to derive quarterly indicators comparable to economic indicators. We also discuss their evolution and how they reflect the monetary policy stance.

## 3.1 Corpus cleaning and preparation

Our corpus consists of 194 minutes of data published between June 2007 and December 2025. For our analysis, we divided each minute into individual sentences and treated each sentence as an independent document. To focus specifically on the tone of the macroeconomic context discussed in the minutes, we removed the titles and sections pertaining to monetary policy decisions. In this way, we eliminated duplicate information and did not include sentences that described the final decision of the BDBR. Leaving them will bias the estimated tonality towards the decision, highlighting the focus on the monetary policy rate instead of the tonality. After this process, the corpus comprised 12,468 documents (i.e., sentences).

## 3.2 Context sensitive dictionary

Dictionary-based analysis is a widely used methodology for sentiment and tone extraction from text. This approach relies on lists of predefined words associated with specific sentiments and counts the occurrences of these terms to classify the tone. Researchers often apply existing dictionaries, such as those employed by Bennani and Neuenkirch (2017), Jegadeesh and Wu (2017), and Shapiro et al. (2022), to analyze textual data. As Loughran and McDonald (2011) pointed out, contextual ambiguity is a significant concern: words can have different interpretations depending on their context, which requires careful consideration in dictionary design to ensure alignment with the study's specific goals and context. Another limitation is linguistic specificity, as most tools and dictionaries are tailored for English, complicating their application in other languages.

To address these issues, we propose a methodology similar to that of Rice and Zorn (2021), which constructs context-sensitive dictionaries. Starting with a set of *seed words*, the method iteratively incorporates nearest neighbors, which are semantically related terms that share contextual or conceptual relevance, ensuring that the dictionary adapts to both the target language and the analytical framework. Specifically, we use as seed words the most common terms in the hawkish, dovish, and neutral dictionaries developed for Colombia by Vega and Lahura (2020) and to find the neighbors, we use the Word2Vec (W2V) skip-gram methodology developed by Mikolov et al. (2013); more details of this methodology are provided in Appendix C.[5]

[5]Vega and Lahura (2020) constructed a dictionary of hawkish, dovish and neutral phrases based on central bank press releases, between February 2003 and December 2018. Each phrase was manually classified by the authors, resulting in a dictionary of 1,021 hawkish, 929 dovish, and 185 neutral phrases. The dictionary was directly provided by the authors.

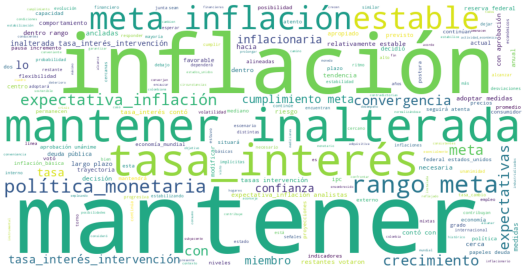
Figure 2 shows the most frequently used terms in each classification. As the size of the term increases, they are presented more frequently. For dovish tone, the most frequent terms are 'crecimiento', 'inflación', 'desaceleración', 'inflación anual', 'caída', 'disminuir', 'menor', 'inferior', 'baja', 'debajo'. For hawkish tone the most frequent terms are 'crecimiento', 'inflación', 'precio', 'aumento', 'mayor', 'incremento', 'superior', 'alto', and for the neutral tone: 'inflación', 'mantener', 'mantener inalterada', 'tasa de interés', 'meta inflación', 'estable' and 'rango meta'. In addition, words such as "inflación básica", "crecimiento económico", and "tasa interés intervención" are considered as a single term because of their importance.




(a) Dovish




(b) Hawkish




(c) Neutral

Figure 2: Dictionary word cloud, based on Vega and Lahura (2020)

From each of these seed or center terms, the five nearest neighbors were found according to W2V. This means that we train three W2V models, one for each classification.[6] As an example, one of the most common terms of the hawkish tone is "aumento", then the W2V methodology looks for the five closest neighbors, such as "aumentar","alto", "incremento", and "inflación básica", with a cosine similarity of (0.991, 0.989, 0.984 and 0.967) respectively.[7] As a result, the proposed context-sensitive dictionary of hawkish tone will have the words "aumento", 'aumentar", "alto", "incremento", and "aumento inflación básica".

Therefore, the proposed dictionary incorporates words that maintain semantic and contextual meanings based on the most important words in each tone. The list of terms obtained by applying this

[6]For more details on the W2V methodology, see Appendix C.

[7]A higher cosine similarity means that the words are closer, contextually and semantically.

methodology is presented in Table 1. This list constitutes the dictionary used to calculate the W2V indices presented in Section 4. We found that dovish, hawkish, and neutral terms generally aligned with the expected economic policy stances. In the dovish category, terms such as *bajar, reducción, menor*, and *desaceleración* logically reflect a focus on economic easing, reduced intervention, and concerns about weakening growth. Hawkish terms include *aumento, incremento,* and *creciente, mayor, aceleración*, which appropriately emphasize tightening measures, such as rate hikes or inflationary pressures. *Tasa* and *alcista* also align with proactive policy actions. The neutral category features terms such as *estable, mantener, inalterado, cumplimiento*, and *convergencia*, which convey the idea that stability is associated with neutrality.

| **Hawkish** | **Dovish** | **Neutral** |
|---|---|---|
| mayor, aumentar, aumento, superior, alto, incremento, crecer, alto dinamismo, mayor dinamismo, aceleración, aceleración crecimiento económico, mayor crecimiento económico, crecimiento económico superior, presión, alcista, mayor inflación básica, aumento inflación básica, aumento tasa interés intervención, contractiva. | menor, reducir, bajo, debajo, desaceleración, inferior, débil, caída, reducción, disminución, disminución inflación básica, inflación básica inferior, descenso inflación básica, baja inflación anual, inflación anual baja, menor inflación anual, crecimiento económico débil, crecimiento económico menor, bajar tasa interés intervención, reducir tasa interés intervención, bajo dinamismo, dinamismo bajo, expansiva. | mantener, cumplimiento meta, inalterado, inalterada, estable, rango, convergencia meta, cumplimiento, confianza, mantener tasa interés intervención, ratificar tasa interés intervención, ancla, anclada. |

Table 1: Context Sensitive Dictionary

## 3.3 Classification using BETO-Transformers

The second methodology for constructing our tonality indices relies on the use of transformers, specifically BETO, which has been shown to outperform Multilingual BERT in Spanish-language classification tasks according to Cañete et al. (2020).[8] Although BETO can classify the corpus directly into our three categories, we enhanced the classification precision by adopting the transfer learning framework (Pan and Yang, 2020). This involves fine-tuning the pre-trained model on a labeled subset of minutes, enabling it to learn the domain-specific linguistic patterns and contextual nuances inherent in central bank communication.

Using the corpus of minutes, we drew a stratified sample of 2,734 sentences (700 hawkish, 785 dovish, and 1,249 neutral) while preserving the original temporal and length distributions (Figure 1).

[8]BETO is a 110-million-parameter BERT variant pre-trained on a 3-billion-token Spanish corpus. It was trained using whole-word masking over more than two million steps. Details on its training corpus are available at `https://github.com/josecannete/spanish-corpora`

To address class imbalance and reduce potential bias in our classification towards the neutral category, we re-sampled the dovish and neutral categories to ensure 700 sentences per class while preserving the original temporal and length strata. Such undersampling has been shown to increase recall and F1 scores without materially reducing accuracy (Cedeño-Valarezo et al., 2022).

To estimate and evaluate our BETO classification, we use five-fold cross-validation to obtain a reliable estimate of its out-of-sample performance (Chen and Du, 2025; Berrar, 2018). Training and testing on the whole sample would give an overly optimistic assessment because the model has already "seen" the data on which it is being evaluated. Cross-validation addresses this concern by repeatedly splitting the labeled sample into training and validation folds, such that each observation is used for validation exactly once. The average performance across the five folds therefore provides a less split-dependent assessment of how the model generalizes to unseen sentences and allows us to examine the stability of its classification performance across alternative sample partitions. Given the small size of our labeled sample, the main fine-tuning hyperparameters, particularly the learning rate and batch size, were specified ex ante following Niţoi et al. (2023), whose application uses a labeled corpus of comparable size and similar sentence-level textual units. We subsequently evaluated this fixed specification using stratified five-fold cross-validation. In each iteration, the model was trained on four folds and validated on the remaining fold, while maintaining the same hyperparameter configuration across all five iterations.

### 3.3.1 Training and Validation Loss

Table 2 reports the loss values of the categorical cross entropy, or negative log likelihood, computed over the three sentiment classes. Therefore, lower values indicate that the model assigns a higher probability to the correct labels. Across the five folds, the training loss declines monotonically, whereas the validation loss generally reaches its minimum around epoch 3. The average validation loss decreases from 1.05 in epoch 1 to 0.81 in epoch 2 and reaches its minimum value of 0.73 in epoch 3. In the subsequent epochs, the average validation loss increases to 0.78 and 0.97, while the training loss continues to decline. This divergence is a classic sign of overfitting, as additional training improves the fit to the training data without improving performance on unseen observations. Four of the five folds reach their minimum validation loss at epoch 3, while Fold 3 reaches a slightly lower value at epoch 4. Overall, the results indicate that epoch 3 provides the best average trade-off between fit and generalization across the five folds. This pattern is consistent with the findings of Niţoi et al. (2023), who fine-tuned FinBERT using a similarly sized corpus.

| Epoch | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Mean |
|---|---|---|---|---|---|---|
| | | | Training loss | | | |
| 1 | 1.06 | 1.08 | 1.03 | 1.05 | 1.05 | 1.05 |
| 2 | 0.87 | 0.82 | 0.79 | 0.85 | 0.86 | 0.84 |
| 3 | 0.70 | 0.55 | 0.63 | 0.70 | 0.62 | 0.64 |
| 4 | 0.41 | 0.42 | 0.30 | 0.31 | 0.47 | 0.38 |
| 5 | 0.29 | 0.34 | 0.29 | 0.22 | 0.31 | 0.29 |
| | | | Validation loss | | | |
| 1 | 1.05 | 1.04 | 1.05 | 1.04 | 1.06 | 1.05 |
| 2 | 0.76 | 0.82 | 0.80 | 0.82 | 0.85 | 0.81 |
| 3 | 0.68 | 0.74 | 0.80 | 0.76 | 0.69 | 0.73 |
| 4 | 0.74 | 0.75 | 0.78 | 0.92 | 0.73 | 0.78 |
| 5 | 0.93 | 0.95 | 0.94 | 0.99 | 1.06 | 0.97 |

Table 2: Loss in Training and Validation by fold

### 3.3.2 Classification Performance

Table 3 summarizes the classification results of the five-fold cross-validation. We report the accuracy, precision, recall, and F1-score for each sentiment category: hawkish, dovish, and neutral (see Appendix A for a detailed description of these metrics). Overall, the classification performance is stable across folds, with macro-F1 scores consistently around 0.70 and limited variability. Both the hawkish and dovish classes achieve balanced performance (F1 between 0.69 and 0.77), with occasional trade-offs between precision and recall, depending on fold-specific data splits. The neutral category is notably more challenging, exhibiting asymmetric precision and recall across folds, which is an expected outcome, given the semantic overlap with the other two sentiment classes. Because it is defined as a residual category for mixed or ambiguous signals, neutral classification shows greater instability, as borderline sentences frequently shift their classifications based on subtle lexical differences. Most classification errors involve confusion between neutral and hawkish or dovish statements rather than confusion between polar opposites.

| **Class** | Stat | **Fold 1** | **Fold 2** | **Fold 3** | **Fold 4** | **Fold 5** |
|---|---|---|---|---|---|---|
| Hawkish | Precision | 0.76 | 0.71 | 0.80 | 0.62 | 0.81 |
| | Recall | 0.64 | 0.79 | 0.62 | 0.79 | 0.65 |
| | F1-Score | 0.69 | 0.75 | 0.70 | 0.69 | 0.72 |
| | Support | 140 | 140 | 140 | 140 | 140 |
| Dovish | Precision | 0.77 | 0.67 | 0.77 | 0.68 | 0.67 |
| | Recall | 0.70 | 0.84 | 0.77 | 0.86 | 0.91 |
| | F1-Score | 0.73 | 0.74 | 0.77 | 0.76 | 0.77 |
| | Support | 140 | 140 | 140 | 140 | 140 |
| Neutral | Precision | 0.63 | 0.75 | 0.62 | 0.83 | 0.77 |
| | Recall | 0.79 | 0.48 | 0.75 | 0.39 | 0.64 |
| | F1-Score | 0.70 | 0.59 | 0.68 | 0.53 | 0.70 |
| | Support | 140 | 140 | 140 | 140 | 140 |
| Accuracy | | 0.71 | 0.70 | 0.71 | 0.68 | 0.74 |
| Macro and weighted avg | Precision | 0.72 | 0.71 | 0.73 | 0.71 | 0.75 |
| | Recall | 0.71 | 0.70 | 0.71 | 0.68 | 0.74 |
| | F1-Score | 0.71 | 0.69 | 0.71 | 0.66 | 0.73 |

Table 3: Classification Statistics across Folds

## 3.4 ChatGPT Prompt

For the third methodology, we address the limitation of the BETO-based approach in classifying neutral sentences, which can be challenging because of the overlap in vector representations between this class and the two extremes (hawkish and dovish), as neutral sentences often share features with both. To address this issue, we propose a classification using a generative AI model capable of capturing these nuanced distinctions through a targeted prompt (see Figure 3). Sentences were classified as hawkish if they indicated inflationary pressure, higher inflation risk, contractionary monetary policy, elevated demand, or expectations that surpassed inflation targets. Sentences were classified as dovish if they indicated a more accommodative monetary policy, lower interest rates, subdued inflation expectations, or below-target economic growth rates. Finally, neutral sentences were defined as those that mentioned mixed signals, compensatory effects, or relatively stable conditions.

Eres un experto en política monetaria del Banco Central de Colombia. Tu tarea es clasificar oraciones extraídas de minutas de política monetaria en una de tres categorías: Halcón, Paloma o neutral. Se considera Halcón cuando la oración refleja presiones inflacionarias en aumento o señales de sobrecalentamiento económico. Esto incluye:

- Mayor actividad económica, crecimiento en la inversión, aumento del consumo, reducción del desempleo o mayor uso de la capacidad productiva.
- Desbalances fiscales o comerciales, mayor liquidez de cualquier tipo o incrementos en los agregados monetarios que puedan generar presiones inflacionarias.
- Expectativas de inflación al alza, inflación total (IPC) o sus componentes por encima del rango meta sin señales claras de convergencia, depreciación de la moneda vinculada a la inflación.
- Incrementos en el Índice de Precios al Productor (IPP).

- Endurecimiento de la política monetaria, como aumentos en la tasa de interés.
- Sorpresas al alza en indicadores económicos como ventas al por menor, producción industrial, crecimiento del PIB o confianza de los hogares.
- Revisiones al alza en pronósticos de inflación o crecimiento.
- Deterioro de los términos de intercambio.
- Factores de incertidumbre que reflejan mayores costos internacionales, volatilidad en mercados financieros o factores geopolíticos que presionan los precios.
- Señales mixtas: Si hay una presión inflacionaria al alza junto con estabilidad en otro rubro, se clasifica como Halcón.

Se clasifica como Paloma cuando la oración indica presiones desinflacionarias y desaceleración de la actividad económica. Esto incluye:

- Caída en la inversión y confianza, reducción del consumo o deterioro del mercado laboral (menores presiones salariales y aumento del desempleo).
- Expectativas de inflación a la baja, inflación total (IPC) o sus componentes por debajo del rango meta sin presiones al alza.
- Menores costos de producción y menores variaciones del Índice de Precios al Productor (IPP).
- Un enfoque más expansivo de la política monetaria, como reducciones en la tasa de interés o medidas de estímulo.
- Revisiones a la baja en pronósticos de inflación o crecimiento.
- Mejoras en los términos de intercambio.
- Factores de incertidumbre que reflejan riesgos de menor crecimiento.
- Señales mixtas: Si hay una presión inflacionaria a la baja junto con estabilidad en otro rubro, se clasifica como Paloma.

La oración es neutral cuando no hay una tendencia clara ni hacia presiones inflacionarias crecientes ni decrecientes. Esto ocurre cuando:

- Se presentan señales mixtas o efectos compensados en la inflación (aumentos en ciertos rubros contrarrestados por disminuciones en otros).
- Se mencionan choques de oferta o demanda explícitamente señalados como temporales.
- Las expectativas inflacionarias permanecen dentro del rango meta.
- La inflación total (IPC) se encuentra alineada con la meta sin señales de desviación.
- El entorno macroeconómico es estable y los indicadores están alineados con su comportamiento esperado.
- Si una sola oración presenta múltiples señales que podrían apuntar en direcciones diferentes, se debe evaluar el balance. Si las señales inflacionarias predominan, Halcón; si las desinflacionarias predominan, Paloma; y si se compensan, neutral.

**Instrucciones:**

- Para oraciones que mencionen tasas de interés reales específicas, si se encuentran por debajo de sus promedios históricos o apuntan a estimular el crédito, Paloma; si se ubican por encima o restringen el crédito, Halcón.
- Explicaciones en el cambio de la inflación por cestas: aumento, Halcón; disminución, Paloma.
- Decisión de política monetaria: subir tasa, Halcón; bajar, Paloma; mantener, neutral (salvo aclaración).
- En caso de términos como “ligeramente" o “subió/bajó un poco", clasifícalas como neutral si no hay señales claras de presiones inflacionarias.

Responde únicamente con la etiqueta correspondiente.

Figure 3: Prompt to classify Inflationary Pressure of Sentences

As our generative AI model, we selected GPT-4o[9], which has gained significant popularity since its release in May 2024. This choice was motivated by the extensive use of GPT models for sentiment analysis, particularly in the economic and monetary policy contexts. A notable example is the work by Hansen and Kazinnik (2024), who used GPT-3.5 and GPT-4 to identify monetary policy shocks in FOMC statements. Their study demonstrated that by employing a detailed prompt, the models can replicate predefined classifications, thus validating their usefulness in complex economic interpretation tasks. For our analysis, we accessed GPT-4o through OpenAI's API, which allowed us to perform automated sentence classification. We set the temperature parameter to zero to reduce sampling variability and make the classification as deterministic as possible.

Given that the model weights may change over time and that the API does not use a fixed model version, exact reproducibility cannot be assured. To assess the stability of our classification, we re-ran the exercise on the 2100 labeled sentences approximately one year after the initial implementation, using the same prompt.[10] The two classifications coincide on 96% of the sentences. The hawkish and dovish classes, which determine the tonality index dynamics, are stable across runs, as approximately 97.5% of these sentences keep the same label. Most of the differences occur in the Neutral category, which is the most difficult to identify because it groups mixed or ambiguous sentences. As the polar classes barely change and the aggregate performance is the same in both runs (accuracy of 0.83 and macro-F1 of 0.83), with no change in the balance between classes, we argue that the tonality index is stable and its dynamics are reproducible regardless of when the classification is run. [11]

## 3.5 Overall classification performance

To compare the classification performance of our three methodologies, we analyze the precision, recall F1-score and Support statistics for how they classify the three categories (2100 sentences manually labeled for BETO classification) and the general accuracy. Table 4 presents the aggregate classification metrics.

The results revealed substantial variations in the performance of the three models. The Word2Vec (W2V) model exhibits the weakest overall performance, with a macro-averaged F1-score of 0.63 and an accuracy of 0.62. Notably, this model struggles with the "Neutral" class, achieving a precision of only 0.50, indicating a high rate of false positives when identifying neutral statements. Although its recall for the neutral class is relatively high (0.71), this comes at the cost of precision, suggesting that the model defaults toward neutrality at the expense of accurately classifying hawkish and dovish tones.

The BETO model, a BERT-based architecture pretrained on Spanish corpora, demonstrated marked improvement over W2V across all metrics. With an overall accuracy of 0.71 and a macro-averaged F1-score of 0.70, BETO achieved a more balanced performance across the three classes. The model

[9]For technical specifications, see: `https://platform.openai.com/docs/models/gpt-4o`

[10]The initial classification was performed on May 7, 2025, and the second one on March 30, 2026.

[11]The comparison of this classification is available upon request.

performs particularly well on the dovish class (F1-score of 0.75) and shows improved precision for the neutral class (0.72) compared with W2V. However, BETO's recall for the neutral class (0.61) remained the lowest among its class-specific metrics, suggesting some difficulty in comprehensively identifying all neutral instances.

The ChatGPT model unambiguously outperformed both alternatives, achieving the highest metrics across all classes and aggregation levels. With an overall accuracy of 0.83 and macro-averaged F1-score of 0.83, ChatGPT demonstrated superior and balanced classification capabilities. Notably, it achieves high precision and recall for both the hawkish (0.81/0.89) and dovish (0.86/0.90) classes, indicating reliable identification of tonality when the policy signal is clear. The model performed respectably on the neutral class (F1-score of 0.75), although this remained the most challenging category, consistent with the other models. Given its better performance across evaluation dimensions—accuracy, precision, recall, and F1-score—ChatGPT was the preferred model for constructing the tonality index used in our main analysis. Its balanced performance minimizes the risk of measurement errors in the independent variable, thereby enhancing the reliability of our subsequent econometric estimates.

| **Model** | **Class** | **Precision** | **Recall** | **F1-score** | **Support** |
|---|---|---|---|---|---|
| W2V | Hawkish | 0.70 | 0.60 | 0.64 | 700 |
| | Dovish | 0.79 | 0.56 | 0.65 | 700 |
| | Neutral | 0.50 | 0.71 | 0.58 | 700 |
| | Accuracy | 0.62 | | | |
| | Macro and weighted avg | 0.66 | 0.62 | 0.63 | 2100 |
| BETO | Hawkish | 0.74 | 0.70 | 0.71 | 700 |
| | Dovish | 0.71 | 0.82 | 0.75 | 700 |
| | Neutral | 0.72 | 0.61 | 0.64 | 700 |
| | Accuracy | 0.71 | | | |
| | Macro and weighted avg | 0.72 | 0.71 | 0.70 | 2100 |
| ChatGPT | Hawkish | 0.81 | 0.89 | 0.85 | 700 |
| | Dovish | 0.86 | 0.90 | 0.88 | 700 |
| | Neutral | 0.82 | 0.70 | 0.75 | 700 |
| | Accuracy | 0.83 | | | |
| | Macro and weighted avg | 0.83 | 0.83 | 0.83 | 2100 |

Table 4: Overall Classification Metrics

# 4 Dynamics and Characteristics of Tonality Indices

After classifying the corpus along the hawkish-to-dovish spectrum, we construct several quarterly indicators to summarize the tone of the minutes. We adopt a quarterly frequency to ensure simplicity and

comparability with economic indicators. For each period $t$, we calculate the proportion of sentences belonging to each category: $H_t$ (hawkish), $D_t$ (dovish), and $N_t$ (neutral). For example, $H_t$ is derived as the ratio of hawkish sentences to the total number of sentences in the period. $D_t$ and $N_t$ are computed in a similar way. Given that the categories are mutually exclusive, we have $H_t + D_t + N_t = 1$. Finally, we define the sentiment measure $S_t$ as the difference between the proportions of hawkish and dovish sentences: [12]

$$S_t = H_t - D_t \tag{1}$$

## 4.1 Tonality and Monetary Policy Tightening

Figure 4 shows the sentiment indices derived from (i) the fine-tuned BETO model ($S_t^B$), (ii) the Word2Vec approach ($S_t^W$), and (iii) the ChatGPT prompt-based method ($S_t^G$). We compare our indicators with the tightening and easing phases of Monetary Policy. Following Adrian and Estrella (2008) we define tightening and easing phases based on the direction of change in the nominal policy rate; a phase change occurs only when the direction of the policy rate's movement reverses. This is a distinct concept from the restrictiveness of the policy stance, which depends on the real rate gap relative to the neutral rate. The figure uses shaded yellow areas to represent tightening monetary policy phases (characterized by increasing policy rates), whereas green areas indicate easing periods (characterized by decreasing policy rates). In our classification, when the rate is held constant, the phase remains the same as in the previous periods. Consequently, we consider a new phase to begin only when the direction of the BDBR's monetary policy rate changes.

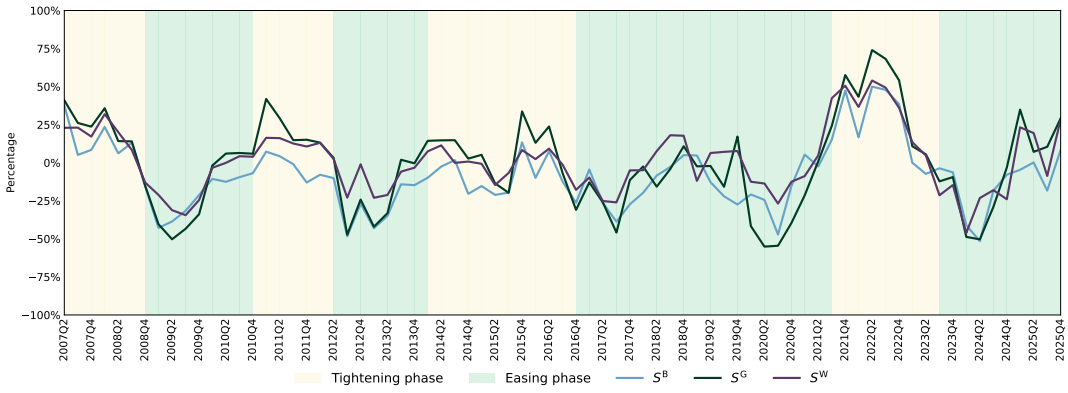



Figure 4: Tonality indicators and Monetary Policy Stance

Between 2007 and the first quarter of 2009, price increases were mainly driven by food prices, especially perishable goods such as potatoes, vegetables, and eggs, due to the climatic effects of *El*

[12] We use this unweighted formula because it is easy to compare across the methodologies proposed in this paper. As a robustness check, we inspect alternative formulas used in previous research using transformer models to compute the indices based on probability or tokens. See appendix B.

*Niño* (early 2007) and *La Niña* (late 2007). There was additional pressure from rising meat prices associated with strong domestic demand. Furthermore, the growing demand for biofuels and higher oil prices have increased the costs of transportation and agricultural inputs, such as fertilizers and pesticides. The increased demand from Venezuela for Colombian food products, along with robust domestic consumption, reinforced these pressures. Simultaneously, inflation in services also rose, most notably in rent prices. From the beginning of 2008, regulated inflation added pressure, with increases in gas tariffs and fuel prices. Consequently, both general inflation and inflation expectations remained above the official targets for most of the period leading up to the Global Financial Crisis of 2008. During this period, the tonality indices started at 40% and remained high until the second quarter of 2008.

The global financial crisis, which originated in the subprime mortgage market, led to reduced liquidity for emerging economies and diminished investor interest in assets such as Colombian Treasury securities (TES), thereby impacting foreign investment. This crisis caused a significant decline in both domestic and international economic conditions, which was marked by high uncertainty. Consequently, these conditions contributed to a domestic negative output gap, with inflation nearing its long-term target of 3%, prompting an easing monetary policy. During this phase, the tonality indices started at one of their lowest levels for the period and then increased steadily until the next policy shift. Between 2011 and 2014, the economy exhibited positive signs, with GDP recovering to an annual growth rate of 3.5%–4%. This recovery was mainly driven by improved terms of trade, largely due to rising oil prices and low global interest rates, which also resulted in a real appreciation of the currency. This currency appreciation, along with favorable external inflation conditions and an appropriate monetary policy stance, helped keep the headline inflation close to its target. During this period, monetary policy was in a tightening phase until the third quarter of 2012, followed by an easing phase until the second quarter of 2014.

From 2014 to 2016, Colombia's economy faced several adverse effects. Initially, the drop in oil prices led to currency depreciation and an increase in the prices of imported inputs. Additionally, a severe *El Niño* phenomenon, lasting eight months, caused a decrease in crop yields and hydroelectric power supply, leading to higher food and regulated prices. Subsequently, with a rise in headline inflation and expectations, monetary policy shifted to a tightening phase. In this phase, the tonality indices also showed a reduction, but it was smaller than in any other tightening phase.

Between 2017 and 2021, a prolonged easing phase occurred. Economic growth was slow, and there were few inflationary pressures until 2019. However, this easing phase continued throughout the pandemic. The coronavirus pandemic between the second and fourth quarters of 2020 generated significant contractions in global and local economic activity. Isolation measures and sectoral restrictions negatively affected employment, household income, and business revenue, resulting in excess productive capacity. In response, the BDBR reduced the monetary policy interest rate to provide liquidity in both domestic and foreign currencies, seeking to ease the financial burden on households and firms, improve cash flow, and facilitate new credit and servicing of existing debt. The government implemented

measures to support credit granted to companies and workers in an environment of high uncertainty, allowing for a greater volume of credit in these sectors. However, concerns about a possible resurgence of the pandemic, especially in Europe, increased in September 2020. From the fourth quarter of 2020 to the fourth quarter of 2021, the monetary policy rate remained at 1.75%, while the tonality reached its minimum value in the second quarter of 2020.

After the pandemic, inflation began to rise steadily in 2021, driven by several major shocks that affected price dynamics and expectations. Two major international shocks disrupted global supply chains: freight transport costs experienced a historic, sharp surge starting in late 2020 and continuing through 2021-2022 due to pandemic-induced disruptions and the Russian invasion of Ukraine, which began in February 2022. In addition, local shocks contributed to inflationary pressures, including social unrest and truck strikes in 2021, a series of gasoline price increases between mid-2022 and mid-2023 aimed at reducing the public budget deficit, and nominal wage shocks. These combined shocks intensified global inflationary pressures on food prices, imported goods, and regulated items, affecting inflation expectations and leading to an increase in headline inflation to 13.12% by 2022. In response, the BDBR initiated a tightening phase between the fourth quarter of 2021 and the fourth quarter of 2023.

In 2023, although inflation remained above its target, it began to decrease, as did inflation expectations, which gradually declined from 7.7% to 6.4% by the end of the year. The fall in the international prices of food and other imported goods helped alleviate inflationary pressures. In addition, there was a slowdown in consumption, a drop in credit growth, and appreciation of the exchange rate. These factors allowed for a gradual reduction in the policy rate without abandoning the objective of consolidating the return of inflation to its target range in the medium term. Nevertheless, relevant risks persisted, such as the *El Niño* phenomenon, adjustments in domestic fuel prices, and international uncertainty.

In summary, we can draw some key characteristics of the relationship between tonality and the tightening phases of monetary policy. First, at the beginning of the tightening phases, the tonality indices start at high values, with the highest increase occurring during the pandemic. In contrast, during the easing phases, the indices start at low levels, typically negative. Second, the tonality tends to shift during each phase, moving from hawkish to dovish in tightening phases. Finally, the volatility of the indices is higher during easing phases.

## 4.2 Forward looking component

Central bank minutes describe the Board's assessment of past, current, and future conditions that may affect the evolution of the economy and the expected policy path, which is relevant for the formation and anchoring of inflation expectations Coibion (2019). Building on the sentence tonality using ChatGPT, we introduced an additional layer that categorizes each sentence based on its temporal orientation, differentiating forward-looking statements from contemporaneous or backward-looking ones. Subsequently, we constructed a prompt from an economic perspective (see Figure 5), following Silva

et al. (2025).

Eres un experto en comunicación de política monetaria del Banco de la República. Tu tarea es clasificar oraciones extraídas de las minutas en una de tres categorías: Forward-Looking, Backward-Looking o Neutral. La clasificación debe basarse en el contenido económico sustantivo del mensaje, no únicamente en el tiempo verbal. Se considera Forward-Looking cuando la oración contiene señales, expectativas, proyecciones u orientación sobre desarrollos futuros. Esto incluye:

- Proyecciones de inflación, crecimiento u otras variables.
- Trayectoria futura esperada de la tasa de interés.
- Forward guidance explícita o implícita.
- Evaluación de riesgos futuros.
- Intención futura de la Junta.
- Necesidad de acciones futuras.
- Escenarios prospectivos o condicionales.

Ejemplos de Forward-Looking: "Se espera que la inflación continúe convergiendo hacia la meta." "Podrían requerirse ajustes adicionales en la tasa de interés." "Los hallazgos del año pasado enfatizaron la necesidad de continuar los esfuerzos para abordar las disparidades en el acceso a los servicios bancarios." Se considera Backward-Looking cuando la oración evalúa, describe o racionaliza condiciones pasadas o actuales sin implicar orientación futura. Esto incluye:

- Descripción de datos observados.
- Evaluación de inflación reciente.
- Análisis de crecimiento ya ocurrido.
- Explicación de choques pasados.
- Justificación ex-post de decisiones de política.
- Evaluación de condiciones actuales sin proyección.

Ejemplos de Backward-Looking: "La inflación anual se ubicó en 7,4% en diciembre." "La Junta decidió aumentar la tasa en 25 puntos básicos." "En las mismas fechas para los créditos de vivienda las observaciones fueron 14,14%, 13,84% y 13,66%,"
Se asigna Neutral cuando la oración:

- No expresa claramente orientación prospectiva.
- No constituye una evaluación retrospectiva sustantiva.
- No cumple claramente criterios de Forward-Looking ni Backward-Looking.

**Importante:** La clasificación no depende exclusivamente del tiempo verbal. Una oración en pasado puede ser Forward-Looking si su mensaje principal es prospectivo.
Si el núcleo del mensaje es la necesidad de una acción futura, debe clasificarse como Forward-Looking aunque mencione hechos pasados.
**Regla decisiva:** Identifica el núcleo económico del mensaje. Si la oración señala expectativas, proyecciones o acciones futuras, clasifícala como Forward-Looking. Si la oración describe datos observados o decisiones ya tomadas, clasifícala como Backward-Looking. Si no hay señal temporal sustantiva clara, clasifícala como Neutral.
Responde únicamente con la etiqueta correspondiente: Forward-Looking, Backward-Looking o Neutral.

Figure 5: Prompt to classify Forward-looking Sentences

After classification, we found that, on average, 28% of the sentences in our sample referred to forward-looking phenomena. This proportion has risen over time and is approximately 40% over the past five years. To analyze whether the forward-looking component of minutes offers a distinct perspective compared to the overall tonality, Figure 6 presents a comparison between the index based

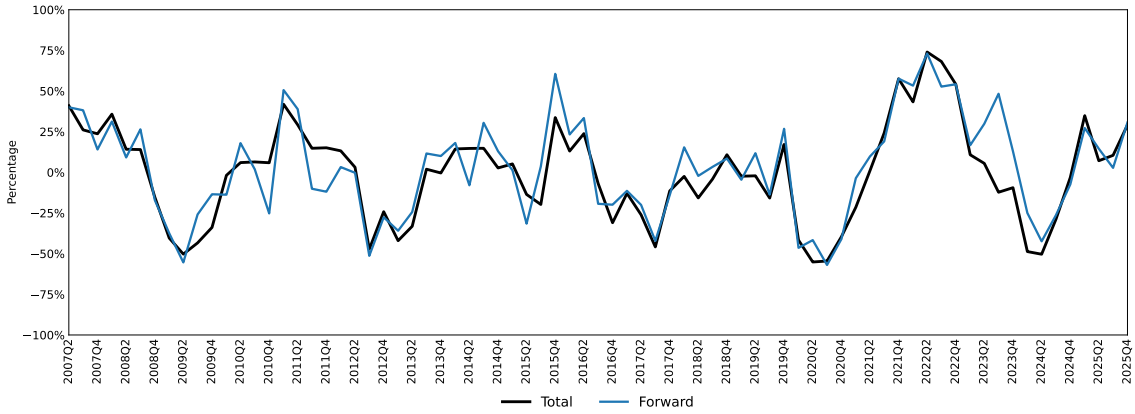



Figure 6: Sentiment vs Forward looking

exclusively on forward-looking components and the aggregate tonality presented in Section 4.1. In general, we observe the same dynamics in both indicators, with a divergence occurring between the first quarter of 2023 and the first quarter of 2024, where the forward-looking component suggests increased inflationary pressures. Another key difference is the higher volatility of the forward-looking tonality. Thus, minutes clearly convey information about the future dynamics and might be useful to form expectations, as we discuss in the next section.

# 5 Does tonality predict adjustments in Inflation Expectations?

Analyzing the transmission from minutes tonality to inflation expectations is crucial, as these expectations not only predict future inflation but also actively shape it through the macroeconomic channels. According to the New Keynesian framework, inflation expectations propagate into observed inflation through three distinct mechanisms: first, via the direct channel of the New Keynesian Phillips curve, where firms embed their revised outlook on future prices into current pricing strategies; second, through the wage-setting channel, where workers and unions incorporate these expectations into nominal wage negotiations, increasing marginal costs, and ultimately feeding them into consumer prices; and third, through the intertemporal substitution channel, where shifts in expected inflation modify the real interest rate, thereby influencing output gaps and subsequent demand-driven price pressures. While the first two channels may imply an indexation process, the last one represents the primary monetary policy channel and implies a reduction in the inflation rate when the monetary policy rate increases.

## 5.1 Baseline results

In this subsection, we examine whether the tonality of Colombian Central Bank minutes conveys information on monetary policy and is associated with the formation of inflation expectations. We define the dependent variable as the deviation of $h$-quarter-ahead inflation expectations from current inflation, $\mathbb{E}_t[\pi_{t+h}] - \pi_t$, which we refer to as the expected inflation gap. This variable captures how agents anticipate inflation to move from its current level over the following $h$ quarters rather than the level of expected inflation. The object of interest is thus the anticipated adjustment of expectations relative to current conditions — whether agents expect inflation to continue rising, to hold steady, or to revert toward more moderate levels — rather than expectations in isolation. This is precisely the margin along which tonality and the policy rate could plausibly matter: both are informative not about the current state of the economy per se, but about how the Board's assessment and its policy response are expected to shape the path inflation takes from here. The underlying hypothesis is that economic agents interpret the tone of central bank communications as a signal of future policy actions and the state of the economy. A positive value of tonality $S_t$ indicates that the central bank is concerned about inflationary pressures.

We include three additional variables to account for potential confounding macroeconomic factors. First, we include the lagged dependent variable to capture the inherent inertia in expectation formation. Second, we include contemporaneous GDP growth, $\Delta y_t$, and changes in the real exchange rate, $\Delta q_t$, to account for domestic and external sources of inflationary pressure. Third, we include the change in the monetary policy rate, $\Delta i_t$ to account for the direct effect of policy actions. The estimated equation is as follows:

$$\mathbb{E}_t[\pi_{t+h}] - \pi_t = \gamma_0 + \gamma_1(\mathbb{E}_{t-1}[\pi_{t+h-1}] - \pi_{t-1}) + \gamma_2\Delta y_t + \gamma_3\Delta q_t + \gamma_4\Delta i_t + \gamma_5 S_t + u_t \tag{2}$$

We estimate the model using quarterly data from Colombia at four horizons ($h = 1, 2, 3, 4$). Inflation expectations are obtained from the Quarterly Survey of Economic Expectations published by BanRep.[13] This survey summarizes the perspectives of economic agents in various economic sectors. According to Sanchez-Jabba et al. (2023), these measures have a greater influence on inflation than those from financial markets or economic models. Headline inflation (CPI) and real GDP are obtained from Colombia's national statistics agency (DANE), and the real exchange rate and monetary policy rate are published by BanRep.[14] The tonality index $S_t$ is constructed using the ChatGPT methodology described in the previous section. All variables were tested for stationarity prior to estimation.[15]

[13]For further information on the survey, see the methodological guide published by Banco de la República https://www.banrep.gov.co/sites/default/files/encuesta-expectativas-trimestral-guia-metodologica.pdf.

[14]The real exchange rate is the bilateral index computed with respect to the United States.

[15]We prefer these variables as they are available in news and used as reference by economic agents. However, we also check specifications with the real exchange rate for main commercial partners, or replacing quarterly GDP growth and the real exchange rate by their gaps, but the estimated coefficients of the monetary policy rate and tonality did not change with respect to what we present here.

A potential concern regarding the validity of our estimates is reverse causality: Could inflation expectations influence the tonality of the minutes? We argue that this is unlikely, given the temporal sequencing of information. The tonality index is constructed exclusively from the minutes published prior to the expectation-formation period. In Colombia, minutes are released within three working days following each monetary policy meeting, whereas the Quarterly Survey of Economic Expectations collects responses over a window of several weeks, beginning in the first month of each quarter. Consequently, the tonality embedded in the most recently published minutes is known to survey respondents before they submit their expectations; future expectations cannot retroactively affect the content of previously published minutes. This timing rules out reverse causality from expectations to minutes, but it does not establish that minutes are the first source of tone-relevant information available to survey respondents: the BDBR also issues a policy statement and holds a press conference the same day as the decision, several days before the minutes are released. However, our variables and the timing argument do not rule out the possibility that our dependent variable may mechanically revert during episodes in which the Board communicates in a more hawkish tone, a concern that the lagged dependent variable only partially addresses.

Before presenting the results, it is important to be precise about what this exercise establishes. The tonality index is not an exogenous policy instrument; by construction, it reflects the Board's own assessment of inflationary pressures, external conditions, and the rationale for its decision — the same set of fundamentals that also shape private-sector expectations. Therefore, a regression of expectations on tonality cannot be interpreted as isolating a causal effect of "communication" in itself; the estimated coefficient may instead combine the direct informational content of the minutes, the central bank's private assessment of the outlook, and omitted macroeconomic fundamentals observed by both the Board and survey respondents.

A related interpretive question concerns the sign of $\gamma_5$ that we should expect. As discussed in the literature, two channels can plausibly link central bank communication to expectations, with opposing sign predictions. Under the standard information channel of monetary policy (Nakamura and Steinsson, 2018; Jarociński and Karadi, 2020), if a hawkish tone reveals genuinely new information about inflationary pressures — information not already anticipated by forecasters — the natural prediction is that expectations should be revised upward, not downward, as agents learn that the central bank sees more inflation risk than they had assumed. Under a second, commitment/credibility channel, a hawkish tone instead signals a clearer or more decisive future policy response, which can lower expected inflation even holding fixed the information about current conditions, because agents revise downward their expectation of how much inflation the Board will tolerate going forward. These two channels are observationally difficult to fully separate using survey data alone, and we do not claim to do so definitively. We estimate the sign empirically rather than imposing it.

Table 5 presents the estimates. Several findings stand out. First, there is a high persistence in the dependent variables, as the coefficient on the lagged dependent variable, $\hat{\gamma}_1$, is close to 0.9 and is statistically significant across all horizons. A coefficient below unity also indicates that the expected

inflation gap is mean-reverting, consistent with a gradual return toward the target over time. Second, the effect of GDP growth is non-significant, while there is a negative effect of the real exchange rate at all horizons. Third, the change in the policy rate, $\hat{\gamma}_4$, has a negative and significant effect on the expected inflation gap. The magnitude of this effect increases with the horizon (from -0.050 at $h = 1$ to -0.237 at $h = 4$), which is consistent with the lagged transmission of monetary policy to the inflation rate. Fourth, tonality is associated with the formation of inflation expectations. The estimated coefficient $\hat{\gamma}_5$ is negative across all horizons, and its magnitude grows substantially from -0.141 at $h = 1$ to -0.506 at $h = 4$. This suggests that the association between tone and expectations strengthens over longer horizons. This negative relationship is not the prediction that the information-revelation channel would generate; we view our results as more consistent with the commitment/credibility channel, which also operates more powerfully as agents update their long-run expectations.

| | $\mathbb{E}_t[\pi_{t+1} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+2} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+3} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+4} - \pi_t]$ |
|---|---|---|---|---|
| $\gamma_0$ | 0.003 | -0.004 | -0.019 | -0.040 |
| | (0.019) | (0.046) | (0.067) | (0.072) |
| $\gamma_1$ | $0.908^{***}$ | $0.904^{***}$ | $0.891^{***}$ | $0.900^{***}$ |
| | (0.042) | (0.047) | (0.045) | (0.038) |
| $\gamma_2$ | 1.009 | 1.017 | 0.454 | 1.336 |
| | (1.813) | (4.078) | (5.671) | (6.065) |
| $\gamma_3$ | $-2.027^{***}$ | $-2.564^{*}$ | $-3.163^{+}$ | $-3.082^{+}$ |
| | (0.503) | (1.183) | (1.678) | (1.823) |
| $\gamma_4$ | $-0.050^{*}$ | $-0.165^{**}$ | $-0.246^{**}$ | $-0.237^{**}$ |
| | (0.022) | (0.051) | (0.073) | (0.080) |
| $\gamma_5$ | $-0.141^{*}$ | −0.099 | −0.186 | $-0.506^{*}$ |
| | (0.066) | (0.150) | (0.210) | (0.228) |

**Note:** Standard errors in parentheses.
Significance levels: $^{+}$; $p < 0.10$, $^{*}$ $p < 0.05$; $^{**}$ $p < 0.01$, $^{***}p < 0.001$

Table 5: Expected inflation gap and Tonality.

However, statistical significance is present only at the one- and four-quarter horizons, with weaker estimates at the intermediate horizons. This finding does not map cleanly onto either channel in isolation, and we flag this as a limitation of the current evidence rather than forcing an interpretation that the data do not clearly support. Moreover, the absence of statistical significance at all horizons may mask state-dependent effects. Therefore, we examine whether the estimated coefficient of tonality depends on the degree of consensus among monetary policy board members. The underlying idea is that unanimity among members may signal a clearer macroeconomic outlook in the face of various shocks or adverse scenarios that may arise. We hypothesize that unanimity is associated with a stronger relationship between tonality and expectations. Using the minutes, we construct a dummy variable $U_t$ that takes the value of one when the BDBR decision on the policy rate is unanimous and zero, otherwise. Thus, we estimate the following equation to check for the state-dependent effects of unanimity.

$$\mathbb{E}_t[\pi_{t+h}] - \pi_t = \gamma_0 + \gamma_1(\mathbb{E}_{t-1}[\pi_{t+h-1}] - \pi_{t-1}) + \gamma_2 \Delta y_t + \gamma_3 \Delta q_t + \gamma_4 \Delta i_t + \gamma_6 S_t \times \mathbf{1}_{\{U_t=0\}} + \gamma_7 S_t \times \mathbf{1}_{\{U_t=1\}} + u_t \quad (3)$$

Where $\mathbf{1}_{\{U_t=0\}}$ is an indicator function that takes the value of 1 when there is no unanimity and $\mathbf{1}_{\{U_t=1\}}$ is another indicator function that is one when there is unanimity. In this case, $\gamma_6$ and $\gamma_7$ capture the effects of tonality without and with unanimity, respectively.

Similar to tonality, unanimity is not randomly assigned; it plausibly coincides with periods in which the economic outlook is clear to board members, which could independently make agents more responsive regardless of the effect of consensus itself. Therefore, we interpret the estimates that follow as evidence of the predictive and informational content of tonality for subsequent adjustments in expectations, conditional on observable macroeconomic controls, rather than as an estimate of the causal effect. The next subsection further probes the extent to which the unanimity interaction reflects a distinct consensus/clarity channel, as opposed to observable correlates of unanimity.

The results reported in Table 6 preserve the main findings from the previous specification regarding the autoregressive component and the effect of the monetary policy rate. We find that both interaction terms, $\hat{\gamma}_6$ and $\hat{\gamma}_7$, are negative across all the horizons. A key feature of these results is the horizon profile of each coefficient. The non-unanimous coefficient is small and statistically insignificant at short and medium horizons, only becoming meaningful at h=4 (-0.418). In contrast, the unanimous coefficient is statistically significant at every horizon and grows monotonically from -0.322 at h=1 to -0.837 at h=4, suggesting that agents update their expectations systematically across all time horizons, and not only in the near term. These findings are consistent with unanimity helping to anchor short-run expectations, while at longer horizons, the tone becomes informative even in the absence of full consensus.

Notably, the effect of tonality is consistently larger in magnitude when the decision is unanimous (i.e., $|\hat{\gamma}_7| > |\hat{\gamma}_6|$). For instance, at the four-quarter horizon, a one-unit increase in tonality under unanimity reduces the expected inflation gap by -0.837, compared to -0.418 under non-unanimity. This is consistent with a stronger association between tonality and expectations when the decision is unanimous. Finally, the estimates are statistically different across regimes: the bottom row of the table reports the p-value for the null hypothesis $H_0 : \gamma_6 = \gamma_7$. We reject equality at conventional significance levels for horizons 1 (p=0.013), 2 (p=0.014), and 3 (p=0.063). Only at the four-quarter horizon (p=0.201) do we fail to reject the effects as the same, possibly due to greater uncertainty surrounding very long-term expectations. We also investigate whether the 2016 change in the reporting format, from a binary unanimity indicator to a full vote breakdown, changes our results by including a format-change dummy in an alternative specification. This dummy was not statistically significant and left the remaining parameter estimates unchanged, confirming the robustness of our main findings.

A similar pattern emerges when this specification is separately estimated across economic sectors using the disaggregated information of the Quarterly Survey of Economic Expectations (see Table D.1). The estimated coefficients of tonality and unanimity remain negative for most sectors across nearly all

| | $\mathbb{E}_t[\pi_{t+1} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+2} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+3} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+4} - \pi_t]$ |
|---|---|---|---|---|
| $\gamma_0$ | 0.000 | −0.011 | −0.025 | −0.044 |
| | (0.019) | (0.045) | (0.065) | (0.072) |
| $\gamma_1$ | $0.918^{***}$ | $0.913^{***}$ | $0.900^{***}$ | $0.906^{***}$ |
| | (0.041) | (0.046) | (0.045) | (0.038) |
| $\gamma_2$ | 0.803 | 0.643 | −0.006 | 1.008 |
| | (1.747) | (3.929) | (5.572) | (6.042) |
| $\gamma_3$ | $-2.013^{***}$ | $-2.547^{*}$ | $-3.179^{+}$ | $-3.103^{+}$ |
| | (0.484) | (1.139) | (1.647) | (1.815) |
| $\gamma_4$ | $-0.047^{*}$ | $-0.158^{**}$ | $-0.237^{**}$ | $-0.229^{**}$ |
| | (0.021) | (0.049) | (0.071) | (0.080) |
| $\gamma_6$ | −0.091 | 0.015 | −0.065 | $-0.418^{+}$ |
| | (0.066) | (0.152) | (0.216) | (0.237) |
| $\gamma_7$ | $-0.322^{**}$ | $-0.512^{*}$ | $-0.630^{*}$ | $-0.837^{*}$ |
| | (0.095) | (0.219) | (0.312) | (0.344) |
| $p$-value $H_0 : \gamma_6 = \gamma_7$ | 0.013 | 0.014 | 0.063 | 0.201 |

**Note:** Standard errors in parentheses.
Significance levels: $^{+}$; $p < 0.10$, $^{*}$ $p < 0.05$; $^{**}$ $p < 0.01$, $^{***}p < 0.001$

Table 6: Expected inflation gap, Tonality and Unanimity.

the time horizons considered. Moreover, statistical significance tends to increase at longer horizons, particularly at three and four quarters, while the magnitude of the effect is larger when monetary policy decisions are unanimous. One important difference is that the estimates for the financial sector are smaller across all the horizons analyzed. One possible explanation is that a substantial part of the information subsequently disclosed in the minutes is already known by these institutions, reducing the additional informational content provided by their publication.

## 5.2 Unanimity or Clarity?

The finding that the association between tonality and inflation expectations is stronger under unanimity raises a natural question: Does this pattern reflect a genuine consensus channel, or is it merely a byproduct of policy environment characteristics that correlate with unanimity?

In this section, we consider three specific concerns. First, large rate adjustments may be both more likely to attract unanimous support and more likely to shift expectations sharply, introducing a magnitude confound. Second, the phase of monetary policy — tightening versus easing — may independently shape both the likelihood of consensus and the responsiveness of agents to central bank signals. Third, episodes of low uncertainty may simultaneously make the appropriate policy response obvious to all board members and render agents more attentive to central bank guidance. Rather than modeling these channels in a separate framework, we address all three within the regression structure of Equation 4 by introducing interactions between unanimity and, respectively, the magnitude of the rate adjustment, monetary policy phase dummy, and our uncertainty index. The central question is whether the differential effect of unanimity remains statistically and economically meaningful once

these alternative explanations are absorbed.

First, we examine its relationship with the magnitude of monetary policy rate adjustments. If unanimity is related to clarity, a larger adjustment must be associated with a clear outlook on economic momentum. However, it is also possible that board members prefer smaller adjustments; therefore, there is a negative association between the size of the adjustment and unanimity. Thus, we include the terms $\gamma_8(|\Delta i_t| \times \mathbf{1}_{U_t=0}) + \gamma_9(|\Delta i_t| \times \mathbf{1}_{U_t=1})$, where $|\Delta i_t|$ denotes the absolute value of the quarterly change in the monetary policy rate.

$$\begin{aligned}\mathbb{E}_t[\pi_{t+h}] - \pi_t = \gamma_0 + \gamma_1(\mathbb{E}_{t-1}[\pi_{t+h-1}] - \pi_{t-1}) + \gamma_2 \Delta y_t + \gamma_3 \Delta q_t + \gamma_4 \Delta i_t + \gamma_6 S_t \mathbf{1}_{\{U_t=0\}} \\ + \gamma_7 S_t \mathbf{1}_{\{U_t=1\}} + \gamma_8 |\Delta i_t| \mathbf{1}_{\{U_t=0\}} + \gamma_9 |\Delta i_t| \mathbf{1}_{\{U_t=1\}} + u_t.\end{aligned} \tag{4}$$

Table 7 presents the estimates of Equation 4. The estimates of $\gamma_1 - \gamma_4$ remain similar to those obtained in the baseline specifications. Estimates of $\gamma_8$ and $\gamma_9$ are not statistically different from zero and from each other. Moreover, the estimates of the effect of tonality under unanimity remain statistically different for two and three quarters, as in the baseline results. However, for the first quarter, the statistical significance is lost, despite the estimated effect of tonality being similar.

| | $\mathbb{E}_t[\pi_{t+1} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+2} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+3} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+4} - \pi_t]$ |
|---|---|---|---|---|
| $\gamma_0$ | -0.007 | -0.018 | -0.015 | -0.033 |
| | 0.02 | 0.048 | 0.069 | 0.075 |
| $\gamma_1$ | 0.941*** | 0.915*** | 0.891*** | 0.896*** |
| | 0.044 | 0.048 | 0.047 | 0.041 |
| $\gamma_2$ | 0.197 | 0.725 | 0.752 | 1.876 |
| | 1.791 | 4.031 | 5.719 | 6.214 |
| $\gamma_3$ | -1.943*** | -2.610* | -3.280+ | -3.177+ |
| | 0.487 | 1.162 | 1.677 | 1.85 |
| $\gamma_4$ | -0.044 | -0.182** | -0.248** | -0.241* |
| | 0.027 | 0.062 | 0.088 | 0.098 |
| $\gamma_6$ | -0.1 | 0.048 | -0.039 | -0.39 |
| | 0.07 | 0.162 | 0.231 | 0.253 |
| $\gamma_7$ | -0.244* | -0.551* | -0.760* | -0.973* |
| | 0.111 | 0.256 | 0.363 | 0.398 |
| $\gamma_8$ | 0.022 | 0.027 | -0.03 | -0.037 |
| | 0.023 | 0.054 | 0.078 | 0.088 |
| $\gamma_9$ | 0.049 | -0.042 | -0.088 | -0.093 |
| | 0.043 | 0.096 | 0.136 | 0.15 |
| $p$-value $H_0 : \gamma_6 = \gamma_7$ | 0.223 | 0.031 | 0.066 | 0.172 |

**Note:** Standard errors in parentheses.
Significance levels: $^{+}$; $p < 0.10$, $^{*}$ $p < 0.05$; $^{**}$ $p < 0.01$, $^{***}p < 0.001$

Table 7: Expected inflation gap, Tonality, Unanimity and Size of the adjustment.

Second, we investigate whether unanimity is related to the tightening and easing phases of monetary policy. During tightening phases, the macroeconomic outlook is often polarized. Board members may hold differing opinions regarding the persistence of inflationary pressures, the appropriate pace

of monetary tightening, or the risks associated with excessive tightening. In such contexts, achieving consensus may be more challenging because of the high costs associated with policy errors and divergent views on optimal trade-offs. Conversely, during easing phases, the direction of action may be more apparent (e.g., supporting growth), and disagreements may be less pronounced. Consequently, the likelihood of unanimity may not be random but is correlated with the prevailing policy regime.

Thus, we add to the estimation $\gamma_8(\mathbf{1}_{T_t=0} \times \mathbf{1}_{U_t=0}) + \gamma_9(\mathbf{1}_{T_t=1} \times \mathbf{1}_{U_t=0}) + \gamma_{10}(\mathbf{1}_{T_t=0} \times \mathbf{1}_{U_t=1})$, where $\mathbf{1}_{T_t=1}$ is a dummy variable that captures the tightening monetary policy phase.

$$\begin{aligned}\mathbb{E}_t[\pi_{t+h}] - \pi_t = \gamma_0 + \gamma_1(\mathbb{E}_{t-1}[\pi_{t+h-1}] - \pi_{t-1}) + \gamma_2 \Delta y_t + \gamma_3 \Delta q_t + \gamma_4 \Delta i_t + \gamma_6 S_t \mathbf{1}_{\{U_t=0\}} \\ + \gamma_7 S_t \mathbf{1}_{\{U_t=1\}} + \gamma_8 \mathbf{1}_{\{T_t=0\}} \mathbf{1}_{\{U_t=0\}} + \gamma_9 \mathbf{1}_{\{T_t=1\}} \mathbf{1}_{\{U_t=0\}} + \gamma_{10} \mathbf{1}_{\{T_t=0\}} \mathbf{1}_{\{U_t=1\}} + u_t.\end{aligned} \tag{5}$$

The estimates in Table 8 show that there is no differential effect of the phase on the inflation expectation gap, as $\gamma_8, \gamma_9, \gamma_{10}$ are close to zero. More importantly, the effect of unanimity also holds in this specification, at $\gamma_6$ and $\gamma_7$ are statistically different for all horizons. This suggests that the stronger association under unanimity is not simply explained by systematic differences in the policy environment across phases, although we cannot rule out other unobserved correlates of unanimity.

| | $\mathbb{E}_t[\pi_{t+1} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+2} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+3} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+4} - \pi_t]$ |
|---|---|---|---|---|
| $\gamma_0$ | -0.042 | -0.090 | -0.067 | 0.028 |
| | 0.038 | 0.088 | 0.126 | 0.137 |
| $\gamma_1$ | 0.887*** | 0.897*** | 0.890*** | 0.910*** |
| | 0.047 | 0.050 | 0.048 | 0.040 |
| $\gamma_2$ | 1.607 | 1.805 | 1.521 | 2.014 |
| | 1.836 | 4.028 | 5.691 | 6.118 |
| $\gamma_3$ | -1.950*** | -2.320+ | -3.133+ | -3.188+ |
| | 0.505 | 1.175 | 1.704 | 1.869 |
| $\gamma_4$ | -0.031 | -0.128* | -0.188* | -0.185* |
| | 0.025 | 0.057 | 0.081 | 0.090 |
| $\gamma_6$ | -0.071 | 0.070 | 0.018 | -0.332 |
| | 0.067 | 0.154 | 0.220 | 0.240 |
| $\gamma_7$ | -0.350** | -0.607* | -0.803* | -1.116** |
| | 0.102 | 0.235 | 0.338 | 0.371 |
| $\gamma_8$ | 0.071+ | 0.151 | 0.139 | 0.026 |
| | 0.042 | 0.094 | 0.133 | 0.144 |
| $\gamma_9$ | 0.001 | 0.001 | -0.088 | -0.179 |
| | 0.047 | 0.106 | 0.150 | 0.164 |
| $\gamma_{10}$ | 0.034 | 0.037 | -0.017 | -0.180 |
| | 0.045 | 0.103 | 0.146 | 0.158 |
| $p$-value $H_0 : \gamma_6 = \gamma_7$ | 0.009 | 0.006 | 0.019 | 0.040 |

**Note:** Standard errors in parentheses.
Significance levels: $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

Table 8: Expected inflation gap, Tonality, Unanimity and Monetary Policy Tightening.

Finally, we investigate whether unanimity and clarity regarding the economic outlook are related.

When uncertainty is low and the macroeconomic outlook is clear, members are more likely to agree on a common course of action. Consequently, economic uncertainty may act as an important determinant of consensus, and failing to account for it could confound the relationship between them. We examine the uncertainty arising from three sources: newspapers, country economic reports, and monetary policy. For uncertainty indices, we use measures constructed using local newspapers Escayola et al. (2024). In addition, we include uncertainty indices based on country reports for Colombia published by the Economist Intelligence Unit, using a 297-word dictionary related to uncertainty (LM–Loughran – McDonald), as explained by Mora-Quiñones et al. (2026). Finally, we include the Uncertainty in Economic Policy index constructed by Fedesarrollo. Thus, we include the terms $\gamma_8(I_t \times \mathbf{1}_{U_t=0}) + \gamma_9(I_t \times \mathbf{1}_{U_t=1})$, where $I_t$ denotes the uncertainty index. Table 9 shows that none of the uncertainty indicators are significant, except for a marginal effect of Economic Policy Uncertainty at intermediate horizons in the formation of inflation expectations, and that the stronger association between tonality and expectations under unanimity persists, as in our baseline results. Overall, the results in this section suggest that the stronger association between tonality and expectations under unanimity is not explained by the phase of monetary policy, the size of the adjustments in the policy rate, or the uncertainty captured by the news, economic reports or monetary policy. These checks address specific observable confounders but do not, on their own, establish a causal consensus channel.

| | $\mathbb{E}_t[\pi_{t+1} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+2} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+3} - \pi_t]$ | $\mathbb{E}_t[\pi_{t+4} - \pi_t]$ |
|---|---|---|---|---|
| | | Newspapers | | |
| $\gamma_4$ | -0.049* | -0.164** | -0.243** | -0.236** |
| | 0.022 | 0.05 | 0.073 | 0.082 |
| $\gamma_6$ | -0.084 | 0.035 | -0.04 | -0.392 |
| | 0.067 | 0.153 | 0.218 | 0.239 |
| $\gamma_7$ | -0.339*** | -0.565* | -0.709* | -0.928* |
| | 0.098 | 0.224 | 0.321 | 0.354 |
| $\gamma_8$ | 0 | 0 | 0 | 0 |
| | 0 | 0 | 0.001 | 0.001 |
| $\gamma_9$ | 0 | 0 | -0.001 | -0.001 |
| | 0 | 0.001 | 0.001 | 0.001 |
| $p$-value $H_0 : \gamma_6 = \gamma_7$ | 0.01 | 0.008 | 0.036 | 0.123 |
| | | Country Reports | | |
| $\gamma_4$ | -0.053* | -0.173** | -0.249** | -0.238** |
| | 0.022 | 0.05 | 0.073 | 0.082 |
| $\gamma_6$ | -0.078 | 0.042 | -0.044 | -0.406+ |
| | 0.067 | 0.153 | 0.219 | 0.24 |
| $\gamma_7$ | -0.328** | -0.540* | -0.677* | -0.887* |
| | 0.097 | 0.222 | 0.318 | 0.35 |
| $\gamma_8$ | 0.003 | 0.005 | 0.001 | -0.004 |
| | 0.004 | 0.009 | 0.013 | 0.015 |
| $\gamma_9$ | 0.002 | 0.001 | -0.005 | -0.01 |
| | 0.004 | 0.01 | 0.014 | 0.016 |
| $p$-value $H_0 : \gamma_6 = \gamma_7$ | 0.01 | 0.008 | 0.043 | 0.159 |
| | | Economic Policy Uncertainty | | |
| $\gamma_4$ | -0.059** | -0.186*** | -0.267*** | -0.245** |
| | 0.022 | 0.05 | 0.073 | 0.082 |
| $\gamma_6$ | -0.075 | 0.05 | -0.029 | -0.395 |
| | 0.066 | 0.151 | 0.215 | 0.239 |
| $\gamma_7$ | -0.322** | -0.511* | -0.633* | -0.859* |
| | 0.095 | 0.219 | 0.313 | 0.35 |
| $\gamma_8$ | 0 | 0.001+ | 0.001+ | 0.001 |
| | 0 | 0 | 0.001 | 0.001 |
| $\gamma_9$ | 0 | 0 | 0.001 | 0 |
| | 0 | 0 | 0.001 | 0.001 |
| $p$-value $H_0 : \gamma_6 = \gamma_7$ | 0.01 | 0.01 | 0.05 | 0.173 |

**Note:** Standard errors in parentheses.
Significance levels: $^{+}$; $p < 0.10$, $^{*}$ $p < 0.05$; $^{**}$ $p < 0.01$, $^{***}p < 0.001$

Table 9: Expected inflation gap, Tonality, Unanimity and Uncertainty.

# 6 Conclusions

The communicative strategy of modern central banks has elevated policy statements and meeting minutes to a central point of interest in monetary policy. While the final interest rate decision provides a single data point, the accompanying text offers a rich and multidimensional view of the BDBR's internal dynamics. In this study, we construct three tonality indices using Natural Language Processing (NLP) techniques by systematically quantifying the linguistic nuances embedded in these documents.

Our tonality indices effectively capture the Board's expressed assessment of inflationary pressures across different phases of monetary policy. Sentiment indices peaked during tightening phases, such as the pre-2015 oil shock and post-COVID-19 period, signifying higher inflationary risks from supply chain disruptions, higher aggregate demand, and geopolitical tensions (among others). Conversely, these indices reach their lowest levels during the easing phases. Notably, the indices exhibit anticipatory behavior, rising preemptively before monetary policy rate hikes and declining ahead of dovish turns, underscoring their utility as early indicators of policy shifts. The indices also mirrored extreme policy stances, such as the historically low policy rate of 1.75% in 2020–2021, when the tonality metrics concurrently reached minimal values. Classification performance varied across methodologies, with the GPT-4o classification outperforming the W2V and fine-tuned BETO models, demonstrating a superior contextual understanding, particularly in resolving ambiguities in neutral sentences.

We provide empirical evidence that the tonality of central bank minutes is significantly associated with the formation of inflation expectations. We find that a more hawkish tone is followed by a revision of inflation expectations relative to the current inflation. This association is economically meaningful and holds beyond the policy rate's informational content. Furthermore, we document that this association is stronger when policy decisions are reached unanimously, a pattern consistent with, though not proof of, consensus enhancing the clarity of the central bank's signal. Our robustness checks rule out several observable correlates of unanimity — the size of the rate adjustment, the monetary policy phase, and measured uncertainty — as explanations for this pattern, although they do not allow us to fully distinguish a consensus/clarity channel from other unobserved factors correlated with unanimity. Across alternative robustness specifications, the stronger tonality-expectations relationship under unanimous decisions remains statistically and economically meaningful, suggesting that this pattern cannot be simply explained by the observable features of the policy environment we examine.

Taken together, these findings suggest that the minutes convey information beyond a reflection of current economic conditions and are systematically associated with how agents revise their inflation expectations. The internal decision-making process, particularly the degree of consensus among members, is systematically associated with the strength of this relationship and subsequent adjustments in agents' expectations. This study underscores that the NLP-driven sentiment index provides both policymakers and market participants with a rigorous, data-driven instrument for navigating the complexities of monetary policy guidance.

# References


Adrian, T. and Estrella, A. (2008). Monetary tightening cycles and the predictability of economic activity. *Economics Letters*, 99(2):260–264.

Arango, L. E., Pantoja, J., and Velásquez, C. (2023). A content analysis of the central bank's press releases in colombia. *Latin American Journal of Central Banking*, 4(3):100097.

Bennani, H. and Neuenkirch, M. (2017). The (home) bias of european central bankers: new evidence based on speeches. *Applied Economics*, 49(11):1114–1131.

Berrar, D. (2018). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology*, 1:542–545.

Bholat, D., Broughton, N., Ter Meer, J., and Walczak, E. (2019). Enhancing central bank communications using simple and relatable information. *Journal of Monetary Economics*, 108:1–15. "Central Bank Communications:From Mystery to Transparency"May 23-24, 2019Annual Research Conference ofthe National Bank of UkraineOrganized in cooperation withNarodowy Bank Polski.

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):910–45.

Carotta, G., Mello, M., and Ponce, J. (2023). Monetary policy communication and inflation expectations: New evidence about tone and readability. *Latin American Journal of Central Banking*, 4(3):100088.

Cañete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., and Pérez, J. (2020). Spanish pre-trained bert model and evaluation data. In *PML4DC at ICLR 2020*.

Cedeño-Valarezo, L., Soria, J., and García, G. (2022). Técnicas para equilibrar conjuntos de datos desbalanceados. *Medwave*, 22(S1).

Chen, X. and Du, Y. (2025). Enhancing medical text classification with gan-based data augmentation and multi-task learning in bert. *Scientific Reports, nature*, 15(1):13854.

Coibion, Gorodnichenko, W. (2019). Monetary policy communications and their effects on household inflation expectations. NBER Working Paper 25482, National Bureau of Economic Research. Revised November 2021.

Escayola, E.-A., Ghirelli, C., Molina, L., Perez, J. J., and Vidal, E. (2024). Using newspapers for textual indicators: Guidance based on spanish- and portuguese-speaking countries. *Computational Economics*, 64(2):643–692.

González, A., Huertas, C., Parra, J., and Vargas, H. (2019). Proceso de toma de decisiones de política monetaria del banco de la república y comunicación sobre política monetaria. *Documentos Técnicos o de Trabajo, Banco de la República–Colombia*.

González, E., Jalil, M., and Romero, J. V. (2011). Inflación y expectativas de inflación en colombia. In López, E. and Ramírez, M. T., editors, *Formación de precios y salarios en Colombia*, volume 2, chapter 13, pages 490–519. Banco de la República, Bogotá.

Guío-Martínez, D. V., Ospina-Tejeiro, J. J., Muñoz-Bravo, G. A., and Parra-Polanía, J. A. (2020). Descripción de las Minutas e Informes de Política Monetaria a partir de herramientas de Lingüística Computacional. Borradores de Economia 1108, Banco de la Republica de Colombia.

Hansen, A. L. and Kazinnik, S. (2024). Can ChatGPT decipher Fedspeak? SSRN Working Paper 4399406. https://doi.org/10.2139/ssrn.4399406.

Hansen, S. and McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99:S114–S133. 38th Annual NBER International Seminar on Macroeconomics.

Jarociński, M. and Karadi, P. (2020). Deconstructing monetary policy surprises—the role of information shocks. *American Economic Journal: Macroeconomics*, 12(2):1–43.

Jegadeesh, N. and Wu, D. (2017). Deciphering fedspeak: The information content of fomc meetings. *Monetary Economics: Central Banks–Policies & Impacts eJournal*.

Jurafsky, D. and Martin, J. H. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*. 3rd edition. Online manuscript released January 12, 2025.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of finance*, 66(1):35–65.

Melo-Velandia, L. F., Ospina-Tejeiro, J. J., and Parra-Polania, J. A. (2020). Effects of Banco de la Republica's Communication on the Yield Curve. Borradores de Economia 1137, Banco de la Republica de Colombia.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Mora-Quiñones, R. C. A., Orozco-Gallo, A. J., and Mora-Pérez, D. A. (2026). Sentiment and uncertainty indices from economic news in colombia. Borradores de Economia 1340, Banco de la Republica de Colombia.

Nakamura, E. and Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: The information effect. *The Quarterly Journal of Economics*, 133(3):1283–1330.

Niţoi, M., Pochea, M.-M., and Ştefan Constantin Radu (2023). Unveiling the sentiment behind central bank narratives: A novel deep learning index. *Journal of Behavioral and Experimental Finance*, 38:100809.

Pan, S. J. and Yang, Q. (2020). Transfer learning. *Learning*, 21:1–2.

Rice, D. R. and Zorn, C. (2021). Corpus-based dictionaries for sentiment analysis of specialized vocabularies. *Political Science Research and Methods*, 9(1):20–35.

Sanchez-Jabba, A., Villabon-Hinestroz, E., and Romero-Torres, B. (2023). Inflation expectations measurement and its effect on inflation dynamics in colombia. Borradores de Economia 1257, Banco de la Republica de Colombia.

Shapiro, A. H., Sudhof, M., and Wilson, D. J. (2022). Measuring news sentiment. *Journal of econometrics*, 228(2):221–243.

Silva, T. C., Moriya, K., and Veyrune, R. M. (2025). From text to quantified insights: A large- scale llm analysis of central bank communication. IMF Working Paper WP/25/109, International Monetary Fund.

Vega, M. and Lahura, E. (2020). Assessing central bank communication through monetary policy statements: Results for colombia, chile and peru. Working Papers 2020-017, Banco Central de Reserva del Perú.

# A Evaluation Metrics

To assess the performance of our tonality classification model, we employ standard evaluation metrics from the machine learning literature. These metrics are computed based on the confusion matrix, which records the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) relative to a reference classification. In our context, a "positive" instance corresponds to a sentence classified as exhibiting positive tonality regarding inflationary pressures.

- **Accuracy**: measures the overall proportion of correct classifications:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{A.1}$$

  While intuitive, accuracy can be misleading when classes are imbalanced, which motivates the use of additional metrics.

- **Precision**: quantifies the reliability of positive classifications by capturing the fraction of instances labeled as positive that are indeed positive

$$\text{Precision} = \frac{TP}{TP + FP} \tag{A.2}$$

  High precision indicates that when the model predicts positive tonality, it is likely to be correct, minimizing false alarms.

- **Recall**: (also known as sensitivity) measures the model's ability to identify all relevant positive instances:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{A.3}$$

  High recall implies that the model successfully captures most instances of positive tonality, minimizing missed detections.

- **F1-score**: provides a harmonic mean of precision and recall, offering a balanced assessment when both false positives and false negatives are of concern:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{A.4}$$

  The F1-score is particularly useful for comparing model performance across different classification thresholds or against alternative specifications.

# B Alternatives indices/formulas using BETO

In this appendix, we evaluate two adjustments to the methodology used to compute BETO indicators. These adjustments were found in previous studies and offer alternatives to account for sentence length or use a probability-weighted indicator.

## B.1 Token-adjusted indices

Because sentence length varies across minutes, we complement the pure-count indices with a token-weighted alternative that assigns greater weight to longer statements. Let $\tau_i$ denote the number of tokens in sentence $i$, excluding the `[CLS]` and `[SEP]` tokens automatically added by the BETO tokenizer at the beginning and end of each sentence, respectively. Let $I_i^H$ and $I_i^D$ denote the indicators that the classifier labels sentence $i$ as hawkish and dovish, respectively, With $S_t$ sentences in period $t$ and the total tokens $T_t = \sum_{i=1}^{S_t} \tau_i$, the token-adjusted proportions are given by

$$H_t^{\text{CA}} = \frac{\sum_{i=1}^{S_t} \tau_i I_i^H}{T_t}, \tag{B.1}$$

$$D_t^{\text{CA}} = \frac{\sum_{i=1}^{S_t} \tau_i I_i^D}{T_t}, \tag{B.2}$$

The corresponding token-weighted sentiment index is defined as:

$$S_t^{\text{CA}} = H_t^{\text{CA}} - D_t^{\text{CA}} \tag{B.3}$$

Empirically, the token-weighted indices in (B.3) are highly correlated with their pure-count counterparts, and we find no major differences between them.

## B.2 Probabilities-adjusted indices

Because BETO's final layer outputs a soft-max vector, we can exploit the associated class probabilities to construct indices that weight each sentence based on model confidence. Let $p_i^H$ and $p_i^D$ be the predicted hawkish and dovish probabilities for sentence $i$, respectively. With $S_t$ sentences in period $t$, the probability-adjusted proportions $H_t^{\text{PA}}$ and $D_t^{\text{PA}}$ are defined as

$$H_t^{\text{PA}} = \frac{\sum_{i=1}^{S_t} p_i^H}{S_t}, \tag{B.4}$$

$$D_t^{\text{PA}} = \frac{\sum_{i=1}^{S_t} p_i^D}{S_t}, \tag{B.5}$$

and the corresponding probability-weighted sentiment index is defined as

$$S_t^{\text{PA}} = H_t^{\text{PA}} - D_t^{\text{PA}} \tag{B.6}$$

Because these averages place greater weight on sentences that the model classifies with high confidence, they mitigate the noise introduced by borderline cases. In practice, the index in (B.6) is highly correlated with both the pure count and token-weighted indices, reinforcing the robustness of the tone measures. Figure B.1 shows no substantial differences between the token- and probability-weighted hawkish indices and the original count-based index. For simplicity, we use the original count-based index, in which all sentences contribute equally to the calculation. Similar results are found for the dovish and sentiment indices, respectively.

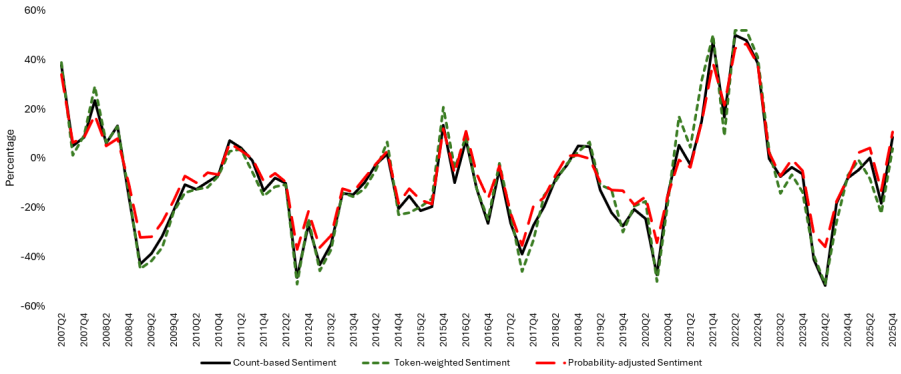



Figure B.1: BETO Sentiment Indices Comparison

# C Word2Vec methodology

The Word2Vec (W2V) methodology allows the creation of word embeddings using neural networks based on two algorithms: bag-of-words and skip-gram. The first algorithm creates vector representations by predicting each word in the corpus based on its context, and the second algorithm predicts based on each word. For small corpora, it is usually recommended to use the skip-gram algorithm because it has a larger number of examples to perform W2V training. In essence, the word embeddings of each vocabulary word consider their semantic and contextual meanings (Jurafsky and Martin (2025)).

Once the vector representations of each word have been obtained, multiple comparisons can be made. For example, to assess the semantic and contextual similarity between two words ($\mathbf{v}$ and $\mathbf{w}$), the distance between their corresponding vectors can be calculated, which is commonly referred to as the cosine similarity.

$$\cos(\mathbf{v}, \mathbf{w}) = \frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}|| \, ||\mathbf{w}||}$$

where $\mathbf{v} \cdot \mathbf{w}$ is the dot product and $||\mathbf{v}|| \, ||\mathbf{w}||$ is the norm of $\mathbf{v}$ and $\mathbf{w}$. The cosine similarity is a value between 0 and 1; the closer it is to 1, the more similar the words are, and the closer it is to 0, the less similar the words are. From cosine similarity, it is also possible to determine which words are most similar to a specific or center term; this is known as the "nearest neighbors" of a word.

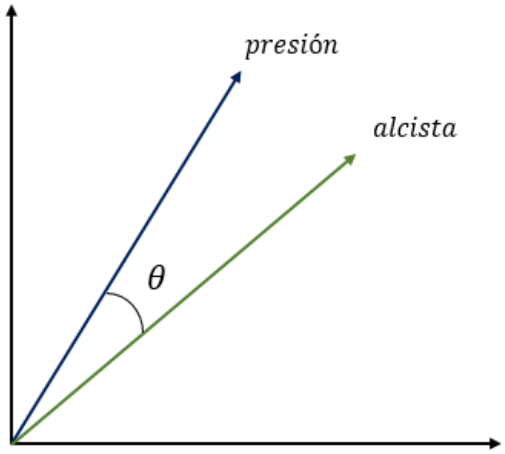



Figure C.1: Example of Cosine Similarity

# D Heterogeneity by economic sectors

| | | Manufacturing | Financial | Commerce | Transport | Academic |
|---|---|---|---|---|---|---|
| $h = 1$ | $\gamma_0$ | 0.037 | -0.017 | 0.059$^{+}$ | 0.061$^{+}$ | -0.014 |
| | | (0.024) | (0.029) | (0.033) | (0.035) | (0.021) |
| | $\gamma_1$ | 0.925$^{***}$ | 0.891$^{***}$ | 0.946$^{***}$ | 0.893$^{***}$ | 0.873$^{***}$ |
| | | (0.039) | (0.056) | (0.040) | (0.048) | (0.042) |
| | $\gamma_2$ | -1.925 | 3.005 | -2.827 | -4.111 | 1.506 |
| | | (2.323) | (2.712) | (3.078) | (3.337) | (1.900) |
| | $\gamma_3$ | -1.806$^{**}$ | 0.357 | -2.440$^{**}$ | -2.755$^{**}$ | -1.821$^{***}$ |
| | | (0.668) | (0.784) | (0.899) | (0.964) | (0.528) |
| | $\gamma_4$ | -0.082$^{*}$ | -0.093$^{**}$ | -0.088$^{*}$ | -0.068 | -0.082$^{***}$ |
| | | (0.031) | (0.035) | (0.042) | (0.044) | (0.023) |
| | $\gamma_6$ | -0.020 | -0.012 | -0.065 | 0.002 | -0.037 |
| | | (0.093) | (0.111) | (0.126) | (0.135) | (0.072) |
| | $\gamma_7$ | -0.304$^{*}$ | -0.169 | -0.276 | -0.501$^{*}$ | -0.298$^{**}$ |
| | | (0.135) | (0.159) | (0.185) | (0.195) | (0.104) |
| | $p$-value $H_0 : \gamma_6 = \gamma_7$ | 0.013 | 0.380 | 0.374 | 0.015 | 0.020 |
| $h = 2$ | $\gamma_0$ | -0.007 | -0.083 | 0.005 | 0.026 | -0.029 |
| | | (0.051) | (0.059) | (0.037) | (0.044) | (0.033) |
| | $\gamma_1$ | 0.829$^{***}$ | 0.846$^{***}$ | 0.859$^{***}$ | 0.836$^{***}$ | 0.910$^{***}$ |
| | | (0.055) | (0.057) | (0.041) | (0.050) | (0.039) |
| | $\gamma_2$ | 0.320 | 6.765 | 2.260 | -1.758 | 1.248 |
| | | (4.727) | (5.238) | (3.455) | (4.186) | (2.911) |
| | $\gamma_3$ | -3.849$^{**}$ | 0.064 | -3.837$^{***}$ | -4.072$^{***}$ | -1.812$^{*}$ |
| | | (1.369) | (1.541) | (1.001) | (1.178) | (0.841) |
| | $\gamma_4$ | -0.174$^{**}$ | -0.176$^{**}$ | -0.119$^{*}$ | -0.130$^{*}$ | -0.135$^{***}$ |
| | | (0.062) | (0.067) | (0.048) | (0.054) | (0.036) |
| | $\gamma_6$ | -0.067 | -0.128 | -0.334$^{*}$ | 0.002 | -0.036 |
| | | (0.185) | (0.205) | (0.138) | (0.167) | (0.111) |
| | $\gamma_7$ | -0.591$^{*}$ | -0.438 | -0.468$^{*}$ | -0.798$^{**}$ | -0.410$^{*}$ |
| | | (0.269) | (0.297) | (0.203) | (0.237) | (0.160) |
| | $p$-value $H_0 : \gamma_6 = \gamma_7$ | 0.066 | 0.264 | 0.546 | 0.001 | 0.010 |
| $h = 3$ | $\gamma_0$ | -0.037 | -0.094 | 0.005 | -0.020 | -0.040 |
| | | (0.044) | (0.070) | (0.042) | (0.047) | (0.044) |
| | $\gamma_1$ | 0.874$^{***}$ | 0.859$^{***}$ | 0.867$^{***}$ | 0.843$^{***}$ | 0.914$^{***}$ |
| | | (0.036) | (0.044) | (0.036) | (0.044) | (0.033) |
| | $\gamma_2$ | 1.919 | 5.235 | 0.612 | 2.219 | 0.212 |
| | | (3.903) | (5.969) | (3.890) | (4.416) | (3.699) |
| | $\gamma_3$ | -3.112$^{**}$ | -1.493 | -3.443$^{**}$ | -3.575$^{**}$ | -1.371 |
| | | (1.156) | (1.765) | (1.120) | (1.265) | (1.091) |
| | $\gamma_4$ | -0.170$^{**}$ | -0.280$^{***}$ | -0.151$^{**}$ | -0.157$^{**}$ | -0.180$^{***}$ |
| | | (0.053) | (0.078) | (0.055) | (0.057) | (0.047) |
| | $\gamma_6$ | -0.328$^{*}$ | -0.269 | -0.425$^{**}$ | -0.157 | -0.183 |
| | | (0.154) | (0.233) | (0.154) | (0.175) | (0.143) |
| | $\gamma_7$ | -0.606$^{**}$ | -0.529 | -0.597$^{**}$ | -0.896$^{***}$ | -0.508$^{*}$ |
| | | (0.223) | (0.340) | (0.227) | (0.251) | (0.206) |
| | $p$-value $H_0 : \gamma_6 = \gamma_7$ | 0.250 | 0.457 | 0.490 | 0.017 | 0.104 |
| $h = 4$ | $\gamma_0$ | -0.080 | -0.120 | -0.026 | -0.001 | -0.084 |
| | | (0.062) | (0.088) | (0.059) | (0.067) | (0.053) |
| | $\gamma_1$ | 0.862$^{***}$ | 0.860$^{***}$ | 0.854$^{***}$ | 0.825$^{***}$ | 0.909$^{***}$ |
| | | (0.040) | (0.042) | (0.041) | (0.054) | (0.030) |
| | $\gamma_2$ | 4.245 | 4.577 | 3.607 | -1.834 | 3.504 |
| | | (5.415) | (7.399) | (5.469) | (6.244) | (4.387) |
| | $\gamma_3$ | -2.148 | -1.892 | -3.535$^{*}$ | -3.136$^{+}$ | -1.233 |
| | | (1.612) | (2.200) | (1.594) | (1.776) | (1.317) |
| | $\gamma_4$ | -0.202$^{**}$ | -0.319$^{**}$ | -0.197$^{*}$ | -0.207$^{*}$ | -0.230$^{***}$ |
| | | (0.073) | (0.099) | (0.078) | (0.081) | (0.057) |
| | $\gamma_6$ | -0.464$^{*}$ | -0.389 | -0.587$^{**}$ | -0.086 | -0.296$^{+}$ |
| | | (0.212) | (0.289) | (0.217) | (0.245) | (0.171) |
| | $\gamma_7$ | -0.839$^{**}$ | -0.741$^{+}$ | -0.715$^{*}$ | -0.952$^{**}$ | -0.535$^{*}$ |
| | | (0.310) | (0.421) | (0.321) | (0.353) | (0.247) |
| | $p$-value $H_0 : \gamma_6 = \gamma_7$ | 0.225 | 0.408 | 0.771 | 0.020 | 0.175 |

*Notes:* Standard errors in parentheses. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. The coefficients correspond to the specification in Equation 3.

Table D.1: Regression results by sector.