



Dictionary-based sentiment analysis of monetary policy communication: on the applicability of lexicons

Aleksandra Rutkowska¹ · Magdalena Szyszko²

Accepted: 21 April 2024
© The Author(s) 2024

Abstract

In this study, we discuss central bank communication sentiments assessed using different dictionaries. Policy communication by monetary authorities is considered an essential additional policy tool. Various text-mining techniques could evaluate its sentiment. Dictionary methods are a prominent example here because they offer simplicity and are a starting point for applying more elaborate methods. We decided to test four lexicons: generic, economic and financial and two domain-specific lexicons to conclude on their applicability to assess the sentiments of monetary releases and provide recommendations for their further application. This study surveys four lexicons qualitatively and conducts a few exercises: lexicon content comparison, performance tests for highly positive and negative messages, and statistical tests of dictionary alignment and correlation. The sample covers small open economies implementing inflation targeting. The study concludes that all investigated dictionaries could be used for detecting central bank intentions: they do not return contradictory results. The choice among them for a particular study depends on the study's setup and goals—as presented in our recommendations

Keywords Monetary policy communication · Central bank releases · Dictionary methods

1 Introduction

In this study, we compare lexicons by Minqing and Bing (2004), Loughran and McDonald (2011), Apel and Grimaldi (2014), and Bennani and Neuenkirch (2017) as tools to assess the sentiment of monetary policy post-decision releases. This examination is primarily methodological because we evaluate the applicability of lexicons for central bank (CB) communication analysis. No benchmark facilitates this exercise, and thus, we test dictionaries from various perspectives, starting from their qualitative comparison. Our procedure consists of two main parts: (1) Dictionary-based comparative exercises performed

✉ Aleksandra Rutkowska
aleksandra.rutkowska@ue.poznan.pl

Magdalena Szyszko
magdalena.szyszko@poznan.merito.pl

¹ Poznan University of Economics and Business, al.Niepodleglosci, Poznan, Poland

² WSB Merito University in Poznan, Powstancow Wlkp 5, Poznan, Poland

qualitatively and quantitatively without reference to actual central bank releases, and (2) empirical comparison of associations between dictionaries provided for our sample of 15 small open economies implementing inflation forecast targeting. Various techniques of text mining could be applied for CB communication, as presented by Bholat et al. (2015). We decided to examine dictionary methods in detail as lexicons are the input to most other methods, even the hybrid one.

Policy communication, especially central bank communication with markets and the general public, has recently been gaining importance as an extraordinary tool to drive market expectations. This issue is worth examining due to the recent proliferation of literature that derives sentiments from monetary policy releases and uses them as a variable in econometric modelling. Central bank communication sentiment is found to be an explanatory factor for the expectations of future policy actions (Hubert and Labondance 2021), inflation expectations (Baranowski et al. 2021), market and economic indicators (Hansen and McMahon 2016), asset prices (Jegadeesh and Wu 2016; Schmeling and Wagner 2019) or simply as a useful indicator of the monetary policy stance (Picault and Renault 2017). The most general message from previous studies on sentiments is that the words provide information beneath the rough information expressed in monetary policy decisions on interest rates or unconventional measures.

For the empirical part of the study, we collected data from 15 small open economies classified as European (even if geographically they are Central Asian or Caucasian countries having a minority of their territory in Europe or being situated on both continents). Implementing inflation targeting (IT) by their CBs creates natural room for clear communication with markets. For this study, we chose small open economies at different stages of economic development. We set aside European Central Bank and Federal Open Market Committee announcements. The majority of studies consider these leading central banks. We contribute to the literature by extending the sample to less prominent economies.¹

The novelty of the study is twofold. First, providing empirical results for these less studied economies is a value-added as we discuss 15 central banks and about 421 million citizens affected by central banks' policies. All central bank covered apply different structure of the policy releases and different wording which enlarges the perspective of this study. Still, the most important study's contribution to the literature is the multicriterial comparison of dictionaries used to derive monetary policy tone. We compare the lexicons qualitatively, run a comparative exercise on the counterfactual model announcements, and provide an association analysis across different lexicons. For the latter, we avoided standard and disputable correlation analysis. We chose to present the sign alignment, the direction of change alignment and finally, association analysis based on entropy and mutual information. The paper differs methodologically from the existing works.

This examination has become increasingly important as central bank communications have evolved. Modern CBs' communication complements or even replaces standard policy tools. However, as announcements consist of words, communication is a sophisticated policy tool. Textual documents (corpora) incorporate some ambiguity. The message of CBs' releases is implicit concerning CB dovishness or hawkishness, even if CBs have recently moved towards more precise expression of policy inclination. The transformation of qualitative information into numbers is a challenging task that requires a few assumptions; the first concerns the dictionary applied. Once the choice is made, natural language processing (NLP) techniques can approximate sentiment.

¹ The sample covers Albania, Czechia, Georgia, Hungary, Iceland, Kazakhstan, Moldavia, Norway, Poland, Romania, Russia, Serbia, Sweden, Turkey, and the UK.

When one seeks to perform sentiment assessment, classifying words or expressions as positive or negative, or hawkish or dovish in the case of monetary policy, is necessary. Most authors use lexicons already presented in the literature and are available as open-access input to computational software. These dictionaries differ substantially in terms of their coverage and the algorithm applied for classification. They return different measures of sentiments. The following question arises: which dictionaries better analyse the sentiment of monetary policy releases? We address this research question as previous works presented in the literature do not prioritise any available dictionaries. Multiple authors apply one lexicon and use others as a robustness check. Moreover, previous works are not supported by any formal analysis of the lexicons' applicability to a diversified sample of CBs. The exemplary justification provided by Baranowski et al. (2021) acknowledges, "We did not consider the Loughran and McDonald (2011) dictionary as it was developed to capture the tone within a relatively broad financial context. In other words, we expect the tailor-made monetary policy tone dictionaries to better extract the timbre of the texts analysed". This is the standard reference to dictionary choices made by many authors, including our previous works. Thus, we consider providing a formal comparison of various lexicons to be a novel contribution to the economic literature.

The sentiment embedded in corpora could be defined differently, as for example, *the disposition of an entity toward an entity, expressed via a specific medium* (Algaba et al., 2021). In this study, the *entity* that sends the message about its policy settings is a central bank, and the *entity* that receives the message is the general public. *Disposition*, namely the sentiment, could be positive or negative if one examines general texts. In the monetary policy context, negative or dovish words or phrases reflect an economic situation that could lead to policy easing; positive or hawkish phrases indicate possible tightening of monetary policy. The *medium* (means of communication) that we examine is a textual document that a central bank publishes just after a monetary policy decision. It could be abridged and anonymised minutes, a governor's statement or a press release. With this study's perspective in mind, we choose a document presented in written form to the general public that explains the rationale for a policy decision together with a reference to the economic stance and its outlook. We realise that CB communication goes beyond a single measure. Nonetheless, due to the volume and diversity of CB communication means and following the standard approach presented by other authors, we chose only one release to identify the timbre of CBs' message.

The remainder of this paper is organized as follows: The next section briefly qualitatively reviews the existing lexicons, and the third section describes the study's data and methods. The fourth section presents our results and takeaways regarding the lexicons' applicability in further research. Finally, we present a discussion and the conclusion of this study.

2 Qualitative comparison of dictionaries

A lexicon elaborated for the purpose of textual analysis is a collection of words (pairs of words or a sequence of words) and associated sentiment scores (Algaba et al. 2021). A variety of dictionaries exist for evaluating the sentiments in corpora. Their application, instead of the narrative assessment of the text, increases the objectivity and repeatability of examinations. However, objectifying the procedure does not avoid the problems linked to dictionary elaboration and composition. Thus, the first step of our procedure involves

the qualitative assessment of dictionaries and their comparison. For this study, we have primarily chosen four English-rooted dictionaries by Minqing and Bing (2004) (hereinafter referred to as BG), Loughran and McDonald (2011) (LM), Apel and Grimaldi (2014) (ABG), and Bennani and Neuenkirch (2017) (BN). They are the most commonly used for deriving sentiments. Prior to this study, these lexicons were applied in a few examinations, and examples are given in the presentation of each lexicon (see Subsections above describing each of the lexicons). They are also available under different licenses for standard computational software. This availability is an important feature when large dictionaries are discussed. The first characteristic that differentiates dictionaries applied is their coverage (and source). Our sample has general sentiment lexicons (such as BG) and domain-specific dictionaries. The latter are based on textual data from a specific field and are meant to be applied in that field. Dictionary specialization can be quite broad: we examine the LM lexicon, a general economic and financial lexicon, and two dictionaries tailored to monetary policy (ABG, BN). The expression *domain-specific dictionary* could be understood differently. The LM lexicon was elaborated due to the limited applicability of generic lexicons to economic and financial texts. Nonetheless, the LM lexicon does not fully capture monetary policy language, which motivated the creation of more specific dictionaries.

The superiority of domain-specific dictionaries seems obvious. They do not miscategorise words that could be neutral in the specific context, such as *tax*, *risk* or *inflation*. They also omit emotionally charged words, which is important if one is examining the formalized messages of policy-makers. However, their specialization could be a constraint: if a lexicon is based on a narrow sample of documents, it might not be efficient in capturing sentiments expressed in similar documents published by other entities. We mention this because one of the lexicons used in this study was derived from a single CB's announcements, namely ABG for Sveriges Riksbank. Our objection to overly specialized dictionaries also concerns the source for elaboration and validation. A single central bank uses a similar structure of releases (and other means of communication). This means that when preparing the given type of announcement, the CB does not start from a blank page (Ehrmann and Talmi 2020). Announcements released by one central bank are similar and could be (relatively) easily processed to create a lexicon. However, substantial cross-country differences between contents might exist, especially when the central banks under discussion adopt a new monetary regime, and their communication practices evolve. This is the case in our sample.

The specialization of a dictionary could also be problematic if the audience for a document does not consist of specialists. Non-specialists might need non-jargon words and expressions to formulate a subjective assessment of the document. It is worth noticing that CBs increasingly seek to address the general public as studied by Binder (2017). The specialization of dictionaries does not reflect the understanding capabilities of this target group.

The next important characteristic of lexicons concerns their composition. Dictionaries could consist of single words (unigrams) or strings of words (bigrams for two words, n-grams for more). This issue is relevant once a lexicon is applied to evaluate CB messages: in economic reality *more* does not often mean *positive*. A non-positive relationship between the direction of change in a variable and the expected policy response impedes classification. The clearest example here is *unemployment*. Contrary to the majority of expressions, such as *high growth rate*, *increasing inflation*, *strong capital inflow*, *rapid consumption growth*, the expression *increasing unemployment* does not denote a positive economic situation or an incentive for hawkish policy actions. A similar situation occurs with the expression *strengthening of national currency* whose effect subdues inflationary

pressure. *Deflation* also creates ambiguity regarding the interpretation of positive or negative sentiment: *accelerating deflation* promises accommodative policy. In this study, the cases when an increase in the intensity of a variable indicates weaker economic conditions and a more dovish policy perspective are called *inverse expressions*. Their proper interpretation as positive or negative is constrained for lexicons that only apply unigrams. When bigrams or n-grams (strings of words, expressions) are used, the ambiguity regarding the direction of the variable changes and its meaning for economic development is reduced. In our sample, we include both types of dictionaries: unigram and n-gram-based.

The general sentiment of a document is based on scoring. That constitutes the next feature of lexicons that could differentiate them. The scoring of a single unigram or n-gram is mostly done on an ordered scale (binary or classification into three classes, such as negative, positive, and neutral). Three dictionaries from our sample, BG, ABG and BN, apply binary classification as positive/negative or hawkish/dovish for monetary policy-specific lexicons. LM lexicon by Loughran and McDonald (2011) divided words into six categories: positive, negative, litigious, uncertain, constraining, and superfluous. For standard sentiment assessment, only two binary categories are applied (positive and negative), and some studies use the category of uncertain expressions to discuss the risk incorporated in descriptive messages.

When discussing the applicability of lexicons for monetary purposes, three caveats should be made. The first one involves the highly sophisticated language of CBs. The description of economic development and policy-making is never a simple story. A central bank's perspective is that optimal policy results occur when its ultimate goals are a targeted level for inflation and a closed output gap for production. The word *inflation* is generally neutral, especially when inflation is held near the CB's targeted level. The expression *increasing inflation rate* is attributed a positive score because in most cases it triggers hawkish monetary policy. That is unless this increase does not start at levels substantially below the inflation target. During one monetary policy committee meeting, policy-makers discuss and weigh many factors with different effects on the general price level regarding their strength and direction. The decision balances the perspectives. The same applies to dictionary analysis of CB releases: it is not perfect, but it returns the overall semantic orientation of the document under analysis. This specific feature of monetary policy language creates a strong preference for lexicons that use bigrams or n-grams. Nevertheless, there is no option to distinguish nuances such as *inflation below* or *inflation above the target* while keeping the lexicon simple.

The second caveat that constraints lexicon applicability concerns their inability to assess the context around the expression (Algaba et al. 2021). This is the case when negations are used (such as *not increased*, *not improved*) or downtoners (such as *barely* or *hardly*). It applies to dictionaries that use unigrams or bigrams. A more complex approach that uses n-grams or entire sentences, as presented by Picault and Renault (2017) for the ECB introductory statements allows for reducing this ambiguity. We appreciate this approach; however, it is far from the simplicity that we seek in lexicons applicable to a diversified sample. As mentioned above, a central bank uses a standardized form of communication in terms of expressions and the structure of a document. The ECB is the exemplary case here (Berger et al. 2011). As a consequence, this kind of classification that presents the probability of an n-gram assignment to some categories needs to be tailored individually to each CB. The applicability of the Picault and Renault (2017) procedure for central banks whose communication evolves over years, being quite scant after IT introduction, is reduced. Moreover, as suggested by Windsor (2021), communication evolved over time not only when strategy changes: since

the Great Recession central bank efforts and communication have focused more on less aggregated view rather than macro modelling.

The third caveat refers to the differences between national languages and English. The majority of studies applying lexicons for the English language are provided for monetary areas where English serves as the primary language of communication—the USA, the UK, the euro area, and Canada. However, as most resources and dictionaries are provided in English, using translations from national languages is a common practice (Algaba et al. 2021). The works of Baranowski et al. (2021) or Montes et al. (2016) are examples using translations of CB announcements. In our sample, only the UK and Iceland use English as the primary communication language. We acknowledge that for the remaining CBs, we are not able to capture possible nuances in their communication sentiment because we use releases translated into English. Dictionaries elaborated for national languages are a rare exception. Apel and Grimaldi (2014) elaborated their lexicon for the Swedish version of Riksbak minutes; however, they provided its translation into English. An English version of the word list provided in ABG is applied in other studies. Ghirelli et al. (2021) is another example that uses corpora and a lexicon in the national language (Spanish).

Usually, English translations of CBs' releases are available online. It could be assumed that these translations, certified by CB staff, are as close as possible to genuine announcements. It is more convenient to use them than to translate or create a dictionary. Moreover, English is characterized by rather simple conjugation and inflexion relative to many national languages. Multiple variations in inflexion and conjugation could constrain a dictionary's usefulness for natural language processing techniques if it is presented in the national language.

The features of the four dictionaries chosen for this study are discussed below: a generic lexicon, a general economic and financial dictionary and two domain-specific lexicons for monetary policy announcements.

2.1 BM generic lexicon

Minqing and Bing (2004) presented an opinion-based lexicon that consists of 6786 English words. Text mining techniques were used to assess the sentiment of online customer reviews. The dictionary is classified as generic because customers freely express their opinions on the chosen features of products. Unigrams are classified via a binary scheme into positive or negative. If unigrams are applied there is no option to properly classify inverse expressions occurring in monetary policy messages.

Even if generic lexicons, especially opinion-based ones, do not appear to be the obvious choice to assess the sentiment of economic and financial announcements, there are a few examples of their application in the literature. They could be successfully used to check the robustness of results obtained with other lexicons. They were also used by Petropoulos and Siakoulis (2021) to discuss whether central bank speeches help to predict stock market distortions. Finally, the BG lexicon, together with other dictionaries, was used by Szyszko et al. (2022) to check the effect of the sentiment of monetary releases on consumer expectations. We see the potential for applying general lexicons to discuss monetary policy issues, especially when consumers are expected recipients of a CB message.

2.2 LM economic and financial lexicon

The most commonly applied dictionary based on financial reports is by Loughran and McDonald (2011). The authors used 10-K mandatory reports filed annually by publicly traded companies in the US to create their word list. As the extent of information considered goes far beyond standard financial statements, these reports are an effective source for elaborating an economic and financial dictionary. The authors were motivated to create this dictionary due to miscategorization of financial and economic words—almost three-fourths of contextually neutral expressions were considered negative by commonly used generic lexicons (Loughran and McDonald 2015).

The LM lexicon consists of 4150 unigrams. Each is attributed to one of the following six categories: positive, negative, litigious, uncertain, constraining, and superfluous. As mentioned above, this extended categorization is not a standard approach. Only two binary categories (positive and negative) are applied for detecting the sentiments of economic announcements. Unigrams with different classifications are omitted. The dictionary avoids standard, emotionally affected words from everyday language. It does not attribute negative ranks to economically neutral words, such as *tax* or *cost*. As it is unigram based, it does not allow a user to accurately classify inverse expressions.

The LM lexicon is widely used to recognize the timbre of monetary releases as the primary dictionary or for robustness checks; for a survey see Baranowski et al. (2021), Jegadeesh and Wu (2016), Hansen and McMahon (2016), Hansen et al. (2018), and Schmeling and Wagner (2019). It is also part of today's most popular hybrid approaches combining word-of-bag with machine learning (see Danowski et al. 2021; Jegadeesh and Wu 2016; and Hubert and Labondance 2021) also provided an interesting lexicon, as they used unigrams classified as uncertain to develop the measure of uncertainty in monetary releases.

2.3 BN domain-specific lexicon based on unigrams

The domain-specific lexicon to classify monetary policy announcements was presented by Bannani and Neuenkirch (2017). The lexicon was based on 1618 speeches delivered by members of the ECB Governing Council between 1999 and 2014. The speeches were not directly linked to policy decision announcements. They presented the individual opinions of policy-makers and had not been edited before they were presented. The lexicon is unigram based.²

The lexicon is modest in terms of the number of unigrams covered. However, the keywords all fall within monetary policy jargon. Moreover, they were not derived from official ECB statements; thus, we can expect broader coverage than that within the structured and edited releases, especially when the speakers can be assumed to be of different nationalities. The authors did not incorporate nouns specific to monetary policy, such as *inflation*, *prices*, and *output gap*. Thus, inverse expression bias is unavoidable. However,

² The authors enumerated keywords that indicated a more hawkish monetary policy: *accelerat**, *better*, *boom**, *emerg**, *expansion*, *fast**, *favo(u)rabl**, *firm**, *great**, *high**, *improv**, *increas**, *larger*, *positive*, *rais**, *ris**, *stabili**, *stable*, *strengthen**, *strong**, *subdued*, *unsustainable*, *upside*, *upswing*, *upturn*, *upward** They also included a list of dovish keywords: *collaps**, *contraction*, *dampen**, *decelerat**, *declin**, *decreas**, *delay**, *depression*, *destabili**, *deteriorat**, *difficul**, *diminish**, *disappear**, *downside*, *downswing*, *downturn*, *downward**, *fall**, *fragil**, *low**, *negative*, *poor*, *recession**, *slow**, *sluggish*, *small**, *struggling*, *sustainable*, *unfavo(u)rabl**, *unstable*, *weak**, *worse*.

this approach excludes other noted ambiguities: variables expressed by nouns are generally neutral. The monetary policy context requires reference not only to the direction of variable changes but also to their relationship towards targets. Excluding nouns from the lexicon is one option here.

The BN lexicon is commonly used in sentiment assessment due to its simplicity and universal entries. Examples can be found in Baranowski et al. (2021), Baranowski et al. (2021), Dybowski and Kempa (2020), Parle (2022), and Szyszko and Rutkowska (2022).

2.4 ABG domain-specific lexicon based on bigrams

Apel and Grimaldi (2014) provided a dictionary based on Sveriges Riksbank statements. Notably, the primary language of this lexicon was Swedish. These authors appreciated the idea of using bigrams to classify monetary announcements as dovish or hawkish. They distinguished eleven nouns (or expressions)³ as those that most related to ultimate policy goals and frequently occurred in CB's communication. Additionally, dovish adjectives were indicated,⁴ together with the basic forms of each adjective. This dictionary avoids the missed classification of inverse expressions: the only inverse noun included is *unemployment*, with the accurate classification as dovish (*strong/increasing unemployment*) or hawkish (*weak/decreasing unemployment*). The word *unemployment* seems to be very important when the stance and outlook of the economy are discussed; however, *unemployment* is also important in other contexts. Nonetheless, when dictionaries are elaborated, there is a compromise between simplicity and domain adequacy.

The ABG lexicon was used in empirical studies by Baranowski et al. (2021), Dossani (2021), and Szyszko and Rutkowska (2022). This lexicon has also been recently extended by (Apel et al. 2021); an example of the application of this extended lexicon is presented by Parle (2022).

3 Data and methods

Our sample covers small open economies located in Europe,⁵ operating with a national currency and implementing IT (not necessarily fully fledged IT). We purposely avoided world-leading central banks. The economies under consideration adopted IT at different moments, and this is why the sample starting points differ. We use the widest range of data possible.⁶ The sample ends in June 2019. An empirical comparison of dictionaries is done for Czechia (CZ), Hungary (HU), Iceland (IS), Norway (NO), Poland (PL), Sweden (SE), the United Kingdom (UK) (long CB experience in IT implementation); Albania (AL), Georgia (GE), Romania (RO), Serbia (RS), Turkey (TR) (moderate experience in IT implementation); Kazakhstan (KZ), Moldavia (MD), Russia (RU) (late adopters of IT). This scale of study of non-world-leading central banks is novel in the economic literature.

³ Namely, *inflation, price, wage, oil price, cyclical position, growth, development, employment, unemployment, recovery and cost*.

⁴ The dovish adjectives were *decreas**, *slow**, *weak**, and *low** while the hawkish adjectives were *increas**, *fast**, *stron*r*, and *high**.

⁵ Economies that we discuss are classified as European even if geographically they are Central Asian or Caucasian countries with a minority of their territory in Europe or situated on both continents.

⁶ In some cases, we collected data for a shorter period than that since IT implementation. The example here is Iceland – there are no minutes dated prior to 2009 available.

Our procedure consists of two main parts.

- Dictionary-based comparative exercises: (1) Monetary policy expressions cross-check; (2) Purely dovish/hawkish cases exercise.
- Empirical part: (1) Sentiment estimations for the sample; (2) Consistency of signs; (3) Consistency of the direction of change; (4) Mutual information examination.

The first part of the procedure does not rely on empirical data. Instead, we perform two exercises to compare the word list classifications presented in the chosen lexicons and the sentiments they return for a purely hawkish or purely dovish corpus. Extreme differences in classification on a single unigram/bigram level could disqualify a dictionary. The same applies to biased returns from the purely hawkish/dovish case.

The empirical part of the study approximates sentiments for our sample and verifies the alignments of sentiments derived from different dictionaries.

We collected corpora of approximately 2000 monetary releases published along with policy decisions (statements, announcements, abridged minutes, press releases). The corpora were preprocessed before being input into the algorithm (tokenization, lemmatization or stemming process). These automated search and word-counting processes were performed via ‘tidytext’ packages (Silge and Robinson 2016).

3.1 Content comparison

As one might expect that monetary policy lexicons are advantageous in approximating policy announcements, we decide to verify whether their content and classification is reflected in other dictionaries. Due to their elaboration based on monetary policy messages, they offer a good starting point for the content cross-check. If monetary policy expressions are reflected with similar scores obtained by the other dictionaries, they have potential for application to monetary announcements. Moreover, due to the differed sources of elaboration, we do not expect the same coverage of lexicons. What we would like to verify with this exercise is the degree of opposite classification that might occur.

3.2 Purely dovish/hawkish exercise

The most demanding part of our study is to find a relevant benchmark for the sentiment assessment. We decide to create purely positive (hawkish) and purely negative (dovish) descriptions of the economic stance using the language of central banks. They can be found in the Supplementary Material. This part of the study is designed to eliminate dictionaries that return counterintuitive sentiments from a benchmark text.

3.3 Sentiment variable

The sentiment variable that is pivotal for this study is derived from corpora after words have been identified as negative or positive according to the four lexicons. In accordance with Uang et al. (2006) the algorithm counts the words and returns the simple index of communication sentiment calculated as presented by Eq. (1):

$$sentiment_{i,t} = \frac{PositiveWords_{i,t} - NegativeWords_{i,t}}{PositiveWords_{i,t} + NegativeWords_{i,t}} \quad (1)$$

where $sentiment_{i,t}$ is the sentiment of CB i 's release from the t -th month; $PositiveWords_{i,t}$ is the number of expressions indicating strong economic conditions, and $NegativeWords_{i,t}$ is the number of expressions indicating weak economic conditions. The procedure returns a continuous variable $sentiment_{i,t}$ for each minute or press release, the value of which varies from -1 (all words are dovish) to 1 (all words are hawkish).

3.4 Sign and change alignment

Dictionaries with different numbers of words and expressions return different sentiment values. Therefore, we check whether sentiments derived from a single document, based on different dictionaries, are aligned in terms of signs. The same procedure is applied to verify the consistency of signs in the case of changes in sentiment. Thus, we can conclude whether different lexicons are aligned in terms of the direction of change they identify.

For each pair x, y of dictionaries:

```

1: for  $t = 1$  to  $N$  do
2:   if  $\text{sgn}(sentiment_{i,t}^x) == \text{sgn}(sentiment_{i,t}^y)$  then
3:     same sign+ = 1
4:   end if
5: end for
6:  $SC_{x,y} = \frac{\text{same sign}}{N}$ 

```

3.5 Mutual information measure

Mutual information, based on the concept of entropy,⁷ measures the information of a random variable contained in another random variable (Dionisio et al. 2004). Mutual information measures the reduction in uncertainty about variable X from observing variable Y . Important for the purpose of our work, unlike the Pearson correlation, it captures both linear and nonlinear dependence between X and Y . Note that mutual information does not imply causality.

Let us denote by X and Y two random variables and assume that each of them can be described by their probability distributions (P_X and P_Y , respectively). The self-information of measuring X as outcome x is defined as (Behrendt et al. 2019):

$$I_X(x) = -\log_2(P_X(X = x)). \quad (2)$$

According to Shannon (1948), for a discrete random variable X with probability distribution P_X , the average number of bits required to optimally encode independent draws can be calculated as:

⁷ Shannon or information entropy is the measure of the uncertainty associated with a random variable (Shannon 1948).

$$H_X(X) = - \sum_x P_X(X = x) \log_2 P_X(X = x) = E[I_X(x)] \tag{3}$$

or – in the case of continuous variables—as:

$$H_X(x) = - \int p_X(x) \log_2 p_X(x) dx, \tag{4}$$

where $p_X(x)$ denotes the probability density function.

If we denote the joint distribution of X and Y by P_{XY} , then we can define the joint entropy by:

$$H(X, Y) = - \sum_x \sum_y P_{X,Y}(X = x, Y = y) \log_2(P_{X,Y}(X = x, Y = y)) \tag{5}$$

Based on the two measures, one can define conditional entropy as:

$$H(Y/X) = H(X, Y) - H(X). \tag{6}$$

Based on the concept of entropy and self-information, one can define the *mutual information* as:

$$I(X, Y) = H(X) - H(X/Y) = H(Y) - H(Y/X) = H(X) + H(Y) - H(X, Y). \tag{7}$$

To normalize mutual information to take values from 0 to 1, Dionisio et al. (2004) we transform it to the so-called *global correlation coefficient* λ :

$$\lambda(X, Y) = \sqrt{1 - \exp(-2I(X, Y))} \tag{8}$$

4 Results

In this section, we discuss the dictionary-based exercises and empirical results. The conclusion on the dictionaries’ applicability will be provided Section Interpretation.

4.1 Dictionary-based exercises

The first exercise consists of a simple comparison of specific monetary policy expressions presented in the BN and ABG lexicons with the coverage and classification of a generic lexicon and one tailored to discuss economic and financial content. By performing this exercise, we seek to verify the degree of contradictions on monetary policy language, which we believe can be derived from monetary policy lexicons, and other dictionaries. The results can be found in the Supplementary Material. The comparison of BN vs. BG and LM yields a similar conclusion—10 monetary policy expressions are not classified by these two dictionaries (3 hawkish and 7 dovish). It applies for highly specialized words such as *strengthen*, *upturn*, *contraction*, *downside* and for a few words that seem to be commonly used such as *fall*, *fast*, *small*. Unigrams *sustainable*, *unsustainable* and *subdued* are classified inversely by BG. The only opposite classification in LM occurs for *sustainable*. Non-coverage or inverse expressions between BN and two lexicons occurs in approximately one-third of cases.

Table 1 Sentiments derived for purely dovish and purely hawkish releases

Dictionary	No. of positive n-grams	No. of negative n-grams	Sentiment
<i>Dovish release</i>			
BG	3	7	- 0.40
LM	2	5	- 0.43
BN	8	10	- 0.11
ABG	4	9	- 0.38
<i>Hawkish release</i>			
BG	4	3	0.14
LM	5	3	0.25
BN	12	1	0.85
ABG	5	1	0.67

The content comparison becomes complicated when ABG is considered as the reference dictionary. Contrary to the other lexicons, it applies bigrams. We generated a set of all possible bigrams and verify the scores of each adjective and noun that constitute a bigram.

Only two out of eleven nouns included in ABG were classified in the other lexicons: *recovery* as positive by BG and *unemployment* as negative by LM. This is not surprising because BG is based on more general content, and LM avoids the classification of words that are neutral in a financial context. A similar situation occurs when adjectives are discussed: *strong* (positive), *weak*, *slow* (negative) are classified by BG and LM and *fast* (positive) by BG only. While ABG distinguishes 45 bigrams, 22 of them are not classified in BG (even if we assume that only one word needs to be scored), and 30 of them are not classified in LM. The low coverage of ABG vs. BG and LM is not surprising. However, average content compatibility is preferable to substantial opposite scoring, which is not happening in these cases.

The second dictionary-based exercise that we performed involves applying linguistic analysis of tone to purely dovish and purely hawkish policy releases. The releases can be found in Supplementary Material. They were prepared by us and cross-checked by a monetary policy expert and professional English language editor before running the exercise. The sentiment was estimated according to Eq. (1). The results of the estimations are presented in Table 1.

First, note that no misclassification occurs: dovish announcements are assessed as negative messages by all lexicons applied (negative values of the sentiment variable). The same applies for hawkish messages. The conclusion is that all lexicons could be considered useful for the analysis of monetary policy releases.

In our pursuit of creating corpora that return extreme sentiments, we were confronted with the lexicons' features. The number of unigrams or bigrams given positive scores (a hawkish message) is lower than those with negative scores. This confirms that the number of negative expressions included in the dictionaries is higher.

Interestingly, a larger distance between positive and negative sentiment value is reported for ABG, and the lowest is reported for the BG dictionary. In this exercise, we were searching for differences in sentiment classification, and both monetary policy lexicons perform better than the other two. Note that BN returns the least negative value of the dovish message and the most positive value for the hawkish message. It also identified the most words as positive/negative in our exemplary releases.

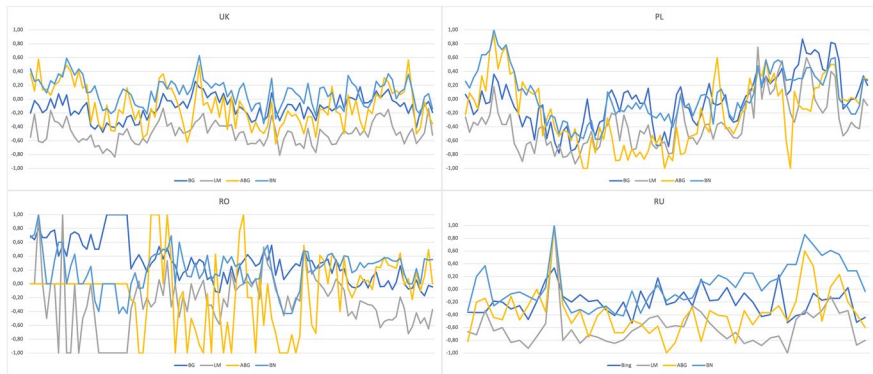


Fig. 1 Sentiment estimations—examples. Notes: Due to the large sample, only selected economies are presented: the UK, one of the most IT-experienced countries worldwide; PL a post-transition economy with long experience in IT; RO, a moderately experienced inflation targeter; and RU, a new adopter of the IT framework. The length of the series presented depends on the length of IT implementation

The conclusion that could be drawn from this exercise is that even if the lexicons considered do not capture the same sentiment and do not cover every detail of policy announcements, they all capture what is valid in CB communication. This is in line with the recommendation presented for dictionaries by Muddiman et al. (2019).

4.2 Empirical results

The empirical part of this study begins with sentiment estimations according to our four dictionaries. Figure 1 presents examples of these estimations. The case of Romania is worth commenting on here: at the beginning of the research period, this country CB's releases were short and not informative. Therefore, few expressions included in the lexicons were detected and scored. The sentiment variable changes rapidly between extreme values, and the sentiments derived from different lexicons are not aligned. The situation stabilizes when monetary policy releases become longer. We observe similar situations for some other CBs (such as Georgia) as long as they do not elaborate on a more informative communication pattern. Generally, Fig. 1 suggests that there is potential for alignment and correlation analysis. Our estimations of sentiment for the sample and figures presenting the results are available online (see XX, 2022).

In the empirical part of the study, we compare the sign and direction of change alignment, the sentiments derived from different lexicons, and the correlation assessed with the application of the mutual information measure. The summary of our statistical analysis is shown in Table 2. The table presents a pair of dictionaries with the highest (H) and the lowest (L) alignment in terms of sign, direction of change, and the mutual information coefficient (Table 3). Additionally, full results of the calculations are provided in Tables 5, 6, and 7.

The LM and BN lexicons are the least consistent in terms of sign alignment, which occurs on average only in 37 percent of cases. This is also confirmed by country-level analysis, as in the majority of cases, the lowest consistency is reported for this pair of dictionaries. In contrast, the general economic and financial dictionary aligns more with the generic lexicon—this pair of lexicons displays the highest alignment. However, the differences

Table 2 Pairs of the best and the worst performing lexicons—summary

	SA - H	SA - L	DC - H	DC - L	CI - H	CI - L
AL	LM-ABG	BG-ABG	LM-BN	BG-ABG	BG-LM	BG-ABG
CZ	BG-LM	LM-BN	BG-LM	LM-ABG	ABG-BN	BG-ABG
GE	ABG-BN	BG-ABG	BG-LM	LM-ABG	LM-BN	LM-ABG
IS	BG-LM	LM-BN	BG-LM	BG-ABG	BG-LM	BG-BN
HU	BG-LM	LM-BN	BG-LM	BG-ABG	BG-LM	BG-ABG
KZ	BG-LM	LM-BN	ABG-BN	BG-ABG	ABG-BN	BG-ABG
MD	BG-LM	LM-BN	LM-BN	BG-ABG	BG-LM	LM-ABG
NO	BG-LM	LM-BN	ABG-BN	LM-ABG	LM-BN	BG-ABG
PL	ABG-BN	LM-ABG	BG-LM	BG-ABG	ABG-BN	BG-ABG
RO	BG-BN	BG-ABG	BG-LM	LM-ABG	BG-LM	ABG-BN
RU	LM-ABG	LM-BN	ABG-BN	LM-ABG	LM-BN	BG-ABG
RS	ABG-BN	LM-BN	ABG-BN	BG-ABG	ABG-BN	LM-ABG
SE	BG-LM	LM-BN	ABG-BN	BG-BN	ABG-BN	BG-ABG
TR	BG-BN	BG-LM	BG-LM	LM-ABG	ABG-BN	BG-ABG
UK	BG-LM	LM-BN	BG-LM	LM-ABG	ABG-BN	BG-ABG

SA Sign alignment; DC Direction of change alignment; CI Mutual information coefficient. H denotes the pair of dictionaries with the highest coefficient according to a given criterion; L denotes the pair of dictionaries with the lowest coefficient. Numerical results from the calculations are available in the Tables 5, 6, and 7

Table 3 Average values of coefficients for the sample

Average of	BG-LM	BG-ABG	BG-BN	LM-ABG	LM-BN	ABG-BN
Direction of signs	0.57	0.57	0.51	0.50	0.34	0.55
Direction of change	0.60	0.45	0.54	0.46	0.57	0.58
Mutual information	0.24	0.11	0.15	0.11	0.22	0.23

between the remaining pairs are not striking, being slightly more visible when the direction of change is compared. In the latter case, we still report the highest alignment for BG-LM and the lowest for BG-BN.

Finally, we present the global correlation coefficient based on mutual information and note that the generic lexicon (BG) and general economic and financial lexicon (LM) are the most correlated, on average, and mostly aligned regarding the sentiment signs and direction of change. We obtained slightly lower coefficients for the monetary policy-specific lexicons, but they could also be considered more consistent. The lowest correlation and alignment are diversified across the lexicons.

4.3 Robustness of the tone estimates

As our tone estimates based on the Eq. (1) do not consider the frequency of a term occurrence in the corpora, we decided to present additional estimates of tone with the application of the Term Frequency-Inverse Document Frequency (TF-IDF) indices proposed in

Table 4 Mutual information between tone estimates and TF-IDF measures

	BN	LM	BG
AL	0.246	0.529	0.402
CZ	0.312	0.714	0.685
GE	0.399	0.836	0.565
IS	0.291	0.501	0.409
HU	0.306	0.665	0.539
KZ	0.574	0.458	0.745
MD	0.370	0.671	0.445
NO	0.459	0.538	0.694
PL	0.387	0.782	0.683
RO	0.374	0.787	0.864
RU	0.520	0.389	0.243
RS	0.329	0.705	0.758
SE	0.386	0.761	0.444
TU	0.327	0.513	0.388
UK	0.189	0.724	0.491

Salton and McGill (1986). The TF-IDF in a document d , in a corpus of documents D , for a t term is formulated as presented by Eq. (9):

$$tfidf(t, d, D) = tf(t, d) * idf(t, D) \quad (9)$$

Term frequency $tf(t, d)$ is the number of times term t appears in document d . Inverse document frequency $idf(t, D)$ for a term t in a corpus of documents D is:

$$idf(t, D) = \log\left(\frac{N}{n}\right) \quad (10)$$

where N is the total number of documents in the corpus D and n is the number of documents that contain that term t (Mee et al. 2021). Hence, the word found in only a few documents has a very high IDF value, while an IDF value of 0 indicates that the term is present in all documents. A list containing terms ordered by TF-IDF score presents terms frequently used in the document (but uncommon in the corpus) at the top of the list and terms common in all documents at the bottom of the list. The sentiment score is also calculated as the difference between the positive and negative words in the documents, but the TF-IDF measure weights each word.

We calculated mutual information coefficients between these two estimates to assess the difference between tone estimations derived from the standard coefficients and TF-IDF measure (see Table 4) for all dictionaries except the bigram-based ABG. The highest mutual information occurs for the LM lexicon (average-0.68)—the frequency of a word occurrence matters the least for this lexicon. The standard correlation between estimates is on average 0.9 for all lexicons.

We do not observe substantial differences when the tone is measured with TF-IDF scores as compared to standard measures. For the graphical presentation for chosen economies and lexicons see Figs. 2, 3 and 4.⁸

⁸ The graphical presentation for the remaining countries and dictionaries available upon request.

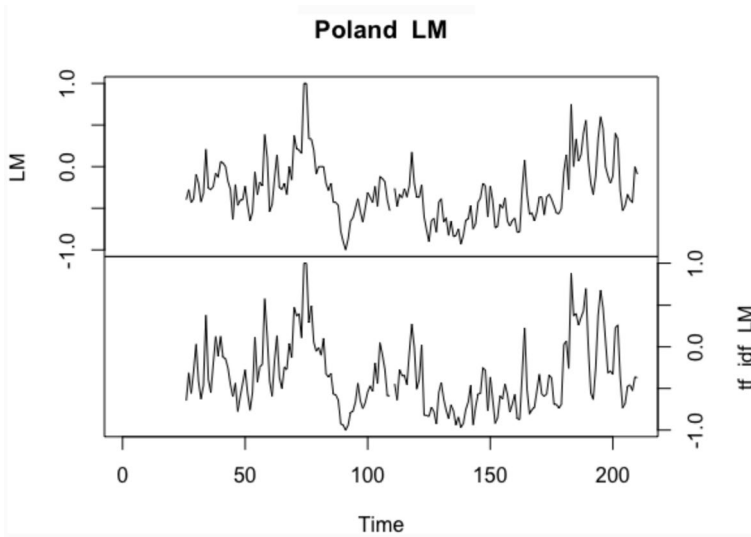


Fig. 2 Standard measure vs TF-IDF indicators—Poland

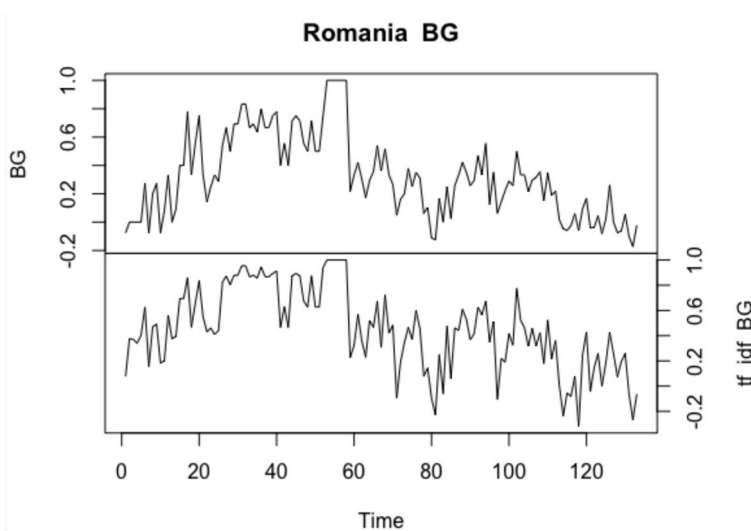


Fig. 3 Standard measure vs TF-IDF indicators—Romania

Moreover, we find frequency indicators less relevant to discussing monetary policy issues as the tone estimates usually open the discussion on the causality between the tone and financial and economic variables. Beyond the strand of the literature that uses tone estimates, the idea is to check whether one release affects a variable. Recipients of a central bank release refer to the single document, not the corpora, when they react to the release.

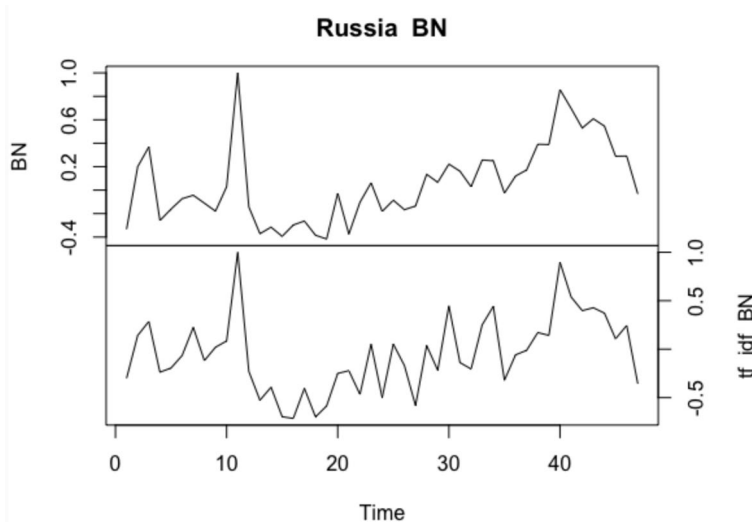


Fig. 4 Standard measure vs TF-IDF indicators—Russia

4.4 Interpretation

Having performed qualitative and quantitative exercises, we are ready to present a few conclusions on the applicability of the four lexicons for assessing monetary policy releases. First, sentiments derived from different lexicons differ in levels but are consistent. The dovish–hawkish exercise returned aligned results in terms of the expected signs of sentiment. We also report sign and direction of sign alignment between all pairs of dictionaries. Some degree of consistency between lexicons occurs regardless of the different content of the non-domain-specific lexicons compared to the dictionaries tailored to monetary policy.

As expected, the domain-specific dictionaries are more consistent. However, and less expected, the generic lexicon and financial and economic dictionary are the most aligned. This result suggests that monetary policy jargon differs substantially not only from general language but also from the language of economics and finance.

The sentiments derived from different lexicons are correlated as measured by the mutual information measure. The reduction in uncertainty on sentiment measured by one dictionary knowing that of another reaches approximately 40% in some cases (BG-LM for AL, PL, UK; ABG-BN for PL, SE, UK), but we expected a more stable result. For some countries, the correlation based on common information is very low (RU, GE, RS). The results vary significantly in the study sample depending on the country, which is more indicative of the differences in the quality of the documents released by CBs.

We also report more consistent results for CBs that are more experienced in inflation targeting implementation. Generally, we register higher coefficients for these CB releases. As mentioned above, the length and consistency of announcements allow the algorithm to classify more words. However, the important conclusion that can be drawn from this study is that applying dictionary methods works for small open economies in addition to leading CBs. Nonetheless, their application could be constrained mainly by the quality (length) of monetary announcements.

This study does not provide unambiguous results regarding lexicon superiority, however, we did not reject the applicability of any of these lexicons to discuss monetary policy content. Thus, we present the following recommendations for further studies.

First, it is a standard research procedure to use more than one lexicon. We suggest choosing the least correlated, on average or more specifically, for a given country, if one lexicon is used to test the robustness of the results. If the results from the dictionaries with relatively low associations are juxtaposed, the validity of the robustness check increases. If one seeks to identify the association between sentiment and CB decisions, and the latter is expressed qualitatively,⁹ the lowest direction of change alignment between dictionaries could be considered the most important criterion when a main dictionary and another for robustness are chosen.

Alternatively, the averaging of sentiments derived from different lexicons could be applied. If this option is chosen, there is no need to justify, often in an arbitrary way, the lexicon choice. The averaging eliminates the drawbacks of a specific lexicon applied for this specific purpose. The empirical results of this study indicate that averaging has potential, as lexicons return different but not contradictory results.

The choice of a lexicon should be also based on the goal of the study. If one aims to discuss the communication effect on consumers (their expectations, private consumption), the application of a general dictionary appears reasonable. Even if a lexicon is based on content simpler than that in monetary policy releases, households might easily identify everyday language. Domain-specific lexicons could be applied if professionals are considered.

If a study sample covers a CB with well-established communication practices over the considered period of time, we recommend applying the ABG lexicon or its recent extension (see Apel et al., 2021). In the dovish–hawkish exercise, the ABG lexicon returned the highest distance between negative and positive sentiments, and it seems to outperform the other lexicons. Moreover, we prefer lexicons that apply bigrams due to the inverse expressions that monetary policy releases refer to. Generally, we regard the idea of bigram-based lexicons as more suitable for detecting monetary policy sentiments. Adjectives, used as dovish or hawkish modifiers, allow users to obtain a more accurate classification of bigrams. In our opinion, applying this lexicon follows the recommendation to create (and use) dictionaries from content to account for context as suggested by Muddiman et al. (2019). Nonetheless, our preference for the ABG lexicon must be considered in conjunction with the study's goal and the need to check the robustness of the results.

5 Discussion and conclusion

Assessing the tone of monetary policy communication has two dimensions. First, it allows us to assess the clarity and consistency of the policy message and actions, especially when directed to the general public. Second, the sentiment variable is the input for econometric models, which could be useful for finding associations between policy communication and its effects. In this study, we discussed the applicability of four commonly used lexicons, and their features and presented recommendations for their application. We built on existing literature. However, as there is no perfect advice on lexicon choice, our findings should be discussed in light of what we did not provide in this study.

⁹ This could be a useful solution during a period of relative stability of interest rates and the application of extraordinary policy measures. CBs' decisions could be encoded from - 1 (restrictive measures) through 0 (no changes) to 1 (accommodative policy measures). Such a variable could feed logit or probit models.

First, we acknowledge that the initial choice of lexicons for this study was a compromise: there exist other dictionaries, such as that by Nielsen (2011) or a psychosociological lexicon, the Harvard IV Dictionary (Stone et al., 1962), that we omitted in this work. We excluded the former because it scores unigrams into five negative and five positive categories. The Harvard lists are sentiment classifications derived from applications in psychology and sociology; they are criticised for their degree of misclassification of specialist expressions. As suggested by Loughran and McDonald (2011), up to three-fourths of unigrams are classified inaccurately.

The dictionaries applied here are well-established in the literature that uses natural language processing to identify sentiments. This is why we did not analyse the recent lexicon by Tadde (2022). This domain-based lexicon using bigrams appears promising, and we will consider further tests. Moreover, the emergence of a new lexicon suggests that the strand of the literature that we are exploring is still evolving.

Second, we know that sentiment analysis techniques are categorised into machine learning-based techniques (supervised, unsupervised, and semi-supervised), lexicon-based techniques (corpus-based and dictionary-based), and hybrid. However, we must remember that the chosen approach is always domain and data-specific. The machine learning techniques (and hybrid techniques with them) required large-scale data to train the model. Sentiment analysis of exchanged information in social networks or news that can be web-scraped guarantees an adequate amount of data, although it should be subjected to an appropriate preprocessing process to guarantee data quality.

In this study, we focused on applying tone analysis to releases published by central banks. Unfortunately, this sample of data is much smaller. As a result, we did not apply machine learning methods. Moreover, Cochrane et al. (2022) examined tools from three broad classes: dictionaries, supervised machine learning, and a method of dictionary induction based on word embeddings. They found that supervised learners were less accurate than leading dictionaries. The abovementioned embedding approach seems also to be a promising method in the textual analysis of central banks (see Baumgärtner and Zahner 2021) but also includes machine learning steps, which is why it was not addressed in this study.

The work by Dun et al. (2021) that compares dictionaries and supervised learning in terms of their effectiveness in assessing the tone of media comments on government spending found them comparable, suggesting that merging these methods increases the accuracy of a media policy signal. This variable could be compared with sentiment.

The benefit of dictionary-based methods is their ease of understanding and evaluation through their straightforward and transparent quantification of an underlying corpus. Moreover, machine learning procedures involve the application of dictionaries or human assistance in the training phase of the exercise. In future research, if a larger data sample becomes available, it will involve the incorporation of machine learning models and topic modelling techniques. These methodologies could significantly aid in identifying the policy topics pertaining to central banks.

Third, regardless of the lexicon applied, the sentiment assessment depends on the coefficients employed. In this study, we presented a standard version of the sentiment variable as presented by Eq. (1), compared to the TI-IDF measures. There are variations of this simple coefficient. Moreover, as presented by Parle (2022), Picault and Renault (2017) and Tadde (2022), the recent strand of the literature addresses not only improved scoring but also more accurate sentiment representation. The definition of the sentiment variable could substantially affect the results. In the robustness check conducted for the dictionary selection problem, it is evident that there are no significant differences between choosing a standard sentiment index and TF-IDF. TF-IDF effectively identifies the terms most relevant to a particular document, making it potentially more valuable for accurately assessing sentiment.

The most important conclusion from this study concerns the applicability of dictionary methods to assess policy release sentiments in small open economies and not only in the leading global economies. Moreover, we reported that the four lexicons commonly applied in such studies generally return different but consistent results. The takeaway for further studies is that the choice of a lexicon for the core part of the study and robustness checks could be based on low correlations or alignments, as presented in this work. We also recommend choosing the most relevant lexicon for the study's goal. If one investigates the communication effects on consumers - the least qualified group of economic agents—the general dictionary as the BG should be more consistent with the way consumers communicate (simple language, broader coverage). Domain-based lexicons are suggested when specialists are studied.

Eventually, the reported results could be discussed regarding their applicability to the central bank. The lexicons applied are based on different sources, so they return consistent but not the same tones. The nuances of policy communications could be captured and analysed with different lexicons to assess if the message incorporates the intended meaning. Manipulating the information sets to direct them towards a chosen group of economic agents (separately—consumers or professionals) would be a challenging task. Still, some central banks have already started to be directed more towards the general public as they use microblogs (Twitter/X) or create more simplified, educational content. Central banks with longer experience in IT implementation manage to create releases that tones are more aligned no matter which dictionary is used to assess the sentiment.

Ultimately, one must adopt a pragmatic approach in choosing a lexicon. Different scores and tone variables could result in different properties of the sentiment variable. This is why testing more than one lexicon is a must.

Appendix

See Tables [A1](#), [A2](#), [A3](#).

Table A1 Sign alignment

Country	BG-LM	BG-ABG	BG-BN	LM-ABG	LM-BN	ABG-BN
AL	0.38	0.33	0.69	0.74	0.62	0.59
CZ	0.76	0.54	0.42	0.58	0.34	0.61
GE	0.47	0.38	0.51	0.45	0.54	0.59
IS	0.66	0.51	0.41	0.40	0.18	0.61
HU	0.72	0.54	0.34	0.60	0.25	0.35
KZ	0.70	0.41	0.48	0.56	0.41	0.44
MD	0.62	0.39	0.45	0.23	0.23	0.61
NO	0.47	0.37	0.29	0.32	0.29	0.44
PL	0.53	0.50	0.60	0.48	0.49	0.67
RO	0.26	0.22	0.59	0.25	0.23	0.31
RU	0.87	0.77	0.66	0.89	0.53	0.60
RS	0.38	0.42	0.66	0.56	0.31	0.42
SE	0.76	0.55	0.32	0.43	0.18	0.66
TR	0.15	0.57	0.83	0.44	0.20	0.67
UK	0.85	0.59	0.46	0.54	0.34	0.71
Average	0.57	0.47	0.51	0.50	0.34	0.55

Table A2 Sign direction of change

country	BG-LM	BG-ABG	BG-BN	LM-ABG	LM-BN	ABG-BN
AL	0.61	0.48	0.64	0.63	0.73	0.60
CZ	0.65	0.52	0.53	0.50	0.58	0.63
GE	0.56	0.51	0.44	0.41	0.45	0.51
IS	0.71	0.44	0.64	0.46	0.64	0.57
HU	0.67	0.41	0.61	0.53	0.62	0.57
KZ	0.54	0.42	0.50	0.42	0.62	0.73
MD	0.56	0.35	0.47	0.40	0.63	0.46
NO	0.44	0.39	0.40	0.32	0.42	0.49
PL	0.61	0.43	0.53	0.45	0.50	0.52
RO	0.47	0.29	0.47	0.25	0.37	0.35
RU	0.57	0.52	0.54	0.46	0.54	0.65
RS	0.54	0.35	0.56	0.39	0.56	0.56
SE	0.66	0.61	0.60	0.63	0.63	0.69
TR	0.73	0.48	0.52	0.47	0.57	0.65
UK	0.71	0.57	0.65	0.56	0.66	0.70
Average	0.60	0.45	0.54	0.46	0.57	0.58

Table A3 Global correlation coefficient

country	BG-LM	BG-ABG	BG-BN	LM-ABG	LM-BN	ABG-BN
AL	0.43	0.14	0.25	0.17	0.34	0.27
CZ	0.13	0.03	0.04	0.03	0.09	0.24
GE	0.17	0.05	0.09	0.03	0.22	0.17
IS	0.25	0.14	0.11	0.16	0.25	0.20
HU	0.29	0.08	0.09	0.10	0.15	0.14
KZ	0.15	0.07	0.20	0.07	0.21	0.32
MD	0.31	0.08	0.07	0.03	0.23	0.07
NO	0.15	0.07	0.12	0.16	0.21	0.16
PL	0.40	0.17	0.34	0.23	0.36	0.42
RO	0.27	0.24	0.15	0.19	0.14	0.13
RU	0.09	0.07	0.09	0.07	0.16	0.10
RS	0.06	0.12	0.14	0.02	0.08	0.21
SE	0.33	0.13	0.17	0.18	0.27	0.41
TR	0.23	0.05	0.11	0.05	0.24	0.26
UK	0.38	0.15	0.26	0.24	0.32	0.39
Average	0.24	0.11	0.15	0.11	0.22	0.23

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11135-024-01896-9>.

Acknowledgements The data that support the findings of this study are openly available in open data repository- Open Science Framework- as: Rutkowska and Szyszko (2024).

Author contribution All authors contributed to the study conception and design. Material preparation, and analysis were performed by Aleksandra Rutkowska. The first draft of the manuscript was written by Magdalena Szyszko and all authors commented on previous versions of the manuscript. Data collection and results interpretations were done by Aleksandra Rutkowska and Magdalena Szyszko. All authors read and approved the final manuscript.

Funding This research project is funded by the National Science Centre, Poland, no. 2018/31/B/HS4/00164.

Declarations

Conflict of interests Author Aleksandra Rutkowska and Magdalena Szyszko declare they have no financial interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Algaba, A., Ardia, D., Bluteau, K., Borms, S., Boudt, K.: Econometrics meets sentiment: an overview of methodology and applications. *J. Econ. Surv.* **34**(3), 512–547 (2021). <https://doi.org/10.1111/joes.12370>
- Apel, M., Grimaldi, M.B.: How informative are central bank minutes? *Rev. Econ.* **65**(1), 53–76 (2014). <https://doi.org/10.1515/roe-2014-0104>
- Apel, M., Blix Grimaldi, M., Hull, I.: How much information do monetary policy committees disclose? Evidence from the FOMC's minutes and transcripts. *J. Money Credit Bank.* **54**(5), 1–32 (2021). <https://doi.org/10.1111/jmcb.12885>
- Baranowski, P., Bennani, H., Doryń, W.: Do the ECB's introductory statements help predict monetary policy? Evidence from a tone analysis. *Eur. J. Polit. Econ.* **66**, 101964 (2021). <https://doi.org/10.1016/j.ejpoleco.2020.101964>
- Baranowski, P., Doryń, W., Łyziak, T., Stanisławska, E.: Words and deeds in managing expectations: empirical evidence on an inflation targeting economy. *Econ. Model.* **95**, 49–67 (2021). <https://doi.org/10.1016/j.econmod.2020.12.003>
- Baumgärtner, M., Zahner, J.: Whatever it takes to understand a central banker: embedding their words using neural networks. Technical report, MAGKS Joint Discussion Paper Series in Economics (2021)
- Behrendt, S., Dimpfl, T., Peter, F.J., Zimmermann, D.J.: Rtransferentropy-quantifying information flow between different time series using effective transfer entropy. *SoftwareX* **10**, 100265 (2019)
- Bennani, H., Neuenkirch, M.: The (home) bias of European central bankers: new evidence based on speeches. *Appl. Econ.* **49**(11), 1114–1131 (2017). <https://doi.org/10.1080/00036846.2016.1210782>
- Berger, H., de Haan, J., Sturm, J.-E.: Does money matter in the ECB strategy? New evidence based on ECB communication. *Int. J. Financ. Econ.* **16**, 16–31 (2011). <https://doi.org/10.1002/ijfe.412>
- Bholat, D., Hansen, S., Santos, P., Schonhardt-Bailey, C.: Text mining for central banks. Centre for central banking studies, Bank of England (2015)
- Binder, C.: Fed speak on main street: central bank communication and household expectations. *J. Macroecon.* **52**, 238–251 (2017). <https://doi.org/10.1016/j.jmacro.2017.05.003>
- Cochrane, C., Rheault, L., Godbout, J.-F., Whyte, T., Wong, M.W.-C., Borwein, S.: The automatic analysis of emotion in political speech based on transcripts. *Polit. Commun.* **39**(1), 98–121 (2022). <https://doi.org/10.1080/10584609.2021.1952497>
- Danowski, J.A., Yan, B., Riopelle, K.: A semantic network approach to measuring sentiment. *Quality & Quantity* **55**(1), 221–255 (2021). <https://doi.org/10.1007/s11135-020-01000-x>
- Dionisio, A., Menezes, R., Mendes, D.A.: Mutual information: a measure of dependency for nonlinear time series. *Phys. A* **344**(1), 326–329 (2004). <https://doi.org/10.1016/j.physa.2004.06.144>

- Dossani, A.: Central bank tone and currency risk premia. *J. Int. Money Financ.* **117**, 102424 (2021). <https://doi.org/10.1093/rof/rfaa014>
- Dun, L., Soroka, S., Wlezien, C.: Dictionaries, supervised learning, and media coverage of public policy. *Polit. Commun.* **38**, 140–158 (2021). <https://doi.org/10.1080/10584609.2020.1763529>
- Dybowski, T.P., Kempa, B.: The European Central Bank's monetary pillar after the financial crisis. *J. Bank Financ.* **121**, 105965 (2020). <https://doi.org/10.1016/j.jbankfin.2020.105965>
- Ehrmann, M., Talmi, J.: Starting from a blank page? Semantic similarity in central bank communication and market volatility. *J. Monet. Econ.* **111**, 48–62 (2020). <https://doi.org/10.1016/J.JMONECO.2019.01.028>
- Ghirelli, C., Hurtado, S., Pérez, J.J., Urtaşun, A.: New data sources for central banks. In: *Data Science for Economics and Finance*, pp. 169–194. Springer, UK (2021)
- Hansen, S., McMahon, M.: Shocking language: understanding the macroeconomic effects of central bank communication. *J. Int. Econ.* **99**, 114–133 (2016). <https://doi.org/10.1016/j.jinteco.2015.12.008>
- Hansen, S., McMahon, M., Prat, A.: Transparency and deliberation within the FOMC: a computational linguistics approach. *Q. J. Econ.* **133**(2), 801–870 (2018). <https://doi.org/10.1093/qje/qjx045>
- Hubert, P., Labondance, F.: The signaling effects of central bank tone. *Eur. Econ. Rev.* **133**, 103684 (2021). <https://doi.org/10.1016/J.EUROCOREV.2021.103684>
- Jegadeesh, N., Wu, D.: Deciphering fedspeak: the information content of FOMC meetings. 2016 AEA Annual Meeting Working Paper, American Economic Association (2016)
- Loughran, T., McDonald, B.: When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *J. Finance* **66**(1), 35–65 (2011). <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- Loughran, T., McDonald, B.: The use of word lists in textual analysis. *J. Behav. Financ.* **16**, 1–11 (2015). <https://doi.org/10.1080/15427560.2015.1000335>
- Mee, A., Homapour, E., Chiclana, F., Engel, O.: Sentiment analysis using TF-IDF weighting of UK MPs' tweets on Brexit. *Knowledge-Based Syst.* **228**, 107238 (2021). <https://doi.org/10.1016/j.knosys.2021.107238>
- Minqing, H., Bing, L.: Mining and summarizing customer reviews. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2004)* (2004)
- Montes, G.C., Oliveira, L.V., Curi, A., Nicolay, R.T.F.: Effects of transparency, monetary policy signalling and clarity of central bank communication on disagreement about inflation expectations. *Appl. Econ.* **48**(7), 590–607 (2016). <https://doi.org/10.1080/00036846.2015.1083091>
- Muddiman, A., McGregor, S.C., Stroud, N.J.: (Re)claiming our expertise: parsing large text corpora with manually validated and organic dictionaries. *Polit. Commun.* **36**, 214–226 (2019). <https://doi.org/10.1080/10584609.2018.1517843>
- Nielsen, F.Å.: Afinn new (anew) evaluation of a word list for sentiment analysis in microblogs. In: *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big Things Come in Small Packages 718 in CEUR Workshop Proceedings*, pp. 93–98 (2011). [arxiv:1103.2903](https://arxiv.org/abs/1103.2903)
- Parle, C.: The financial market impact of ECB monetary policy press conferences—a text based approach. *Eur. J. Polit. Econ.* **108**, 10782 (2022). <https://doi.org/10.1080/00036846.2016.1210782>
- Petropoulos, A., Siakoulis, V.: Can central bank speeches predict financial market turbulence? Evidence from an adaptive NLP sentiment index analysis using XGBoost machine learning technique. *Central Bank Rev.* **21**, 145–153 (2021). <https://doi.org/10.1016/j.cbrev.2021.12.002>
- Picault, M., Renault, T.: Words are not all created equal: a new measure of ECB communication. *J. Int. Money Financ.* **79**, 136–156 (2017). <https://doi.org/10.1016/j.jimonfin.2017.09.005>
- Rutkowska, A., Szyszko, M.: Sentiment measures of selected central banks monetary releases (2024). <https://doi.org/10.17605/OSF.IO/BSPJ7>
- Salton, G., McGill, M.J.: *Introduction to Modern Information Retrieval*, pp. 375–384. McGraw-Hill Inc, USA (1986)
- Schmeling, M., Wagner, C.: Does central bank tone move asset prices? EPR. Discussion Paper **13490** (2019)
- Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423 (1948). <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Silge, J., Robinson, D.: Tidytext: Text mining and analysis using tidy data principles in R. *J. Open Source Softw.* **1**(3), 37 (2016). <https://doi.org/10.21105/joss.00037>
- Stone, P.J., Bales, R.F., Namenwirth, J.Z., Ogilvie, D.M.: The general inquirer: a computer system for content analysis and retrieval based on the sentence as a unit of information. *Behav. Sci.* **7**, 484–498 (1962). <https://doi.org/10.1002/bs.3830070412>
- Szysko, M., Rutkowska, A.: Do words transform into actions? The consistency of central banks' communications and decisions. *Equilib. Quarter. J. Econ. Econ. Policy* **13**, 31–49 (2022). <https://doi.org/10.24136/eq.2022.002>

- Szyszko, M., Kliber, A., Rutkowska, A.: Do Words affect Expectations? The effect of central banks communication on consumer inflation expectations. *Quarter. Rev. Econ. Financ.* **86**, 221–229 (2022). <https://doi.org/10.1016/j.qref.2022.07.009>
- Tadle, R.C.: FOMC minutes sentiments and their impact on financial markets. *J. Econ. Bus.* **118**, 106021 (2022). <https://doi.org/10.1016/j.jeconbus.2021.106021>
- Uang, J.-Y., Citron, D.B., Sudarsanam, S., Taffler, R.J.: Management going-concern disclosures: impact of corporate governance and auditor reputation. *Eur. Financ. Manag.* **12**(5), 789–816 (2006)
- Windsor, C.: The intellectual ideas inside central banks: what's changed (or not) since the crisis? *J. Econ. Surv.* **35**, 539–569 (2021). <https://doi.org/10.1111/joes.12413>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.