The Korean Economic Review
Volume 35, Number 2, Summer 2019, 471-511.

Deciphering Monetary Policy Board Minutes with Text Mining: The Case of South Korea*

Young Joon Lee** · Soohyon Kim*** · Ki Young Park****

We quantify the Monetary Policy Board minutes of the Bank of Korea (BOK) by using text mining. We propose a novel approach that uses a field-specific Korean dictionary and contiguous sequences of words (n-grams) to capture the subtlety of central bank communications. Our text-based indicator helps explain the current and future BOK monetary policy decisions when considering an augmented Taylor rule, suggesting that it contains additional information beyond the currently available macroeconomic variables. In explaining the current and future monetary policy decisions, our indicator remarkably outperforms English-based textual classifications, a media-based measure of economic policy uncertainty, and a data-based measure of macroeconomic uncertainty. Our empirical results also emphasize the importance of using a field-specific dictionary and the original Korean text.

JEL Classification: E43, E52, E58
Keywords: Monetary Policy, Text Mining, Taylor Rule, Machine Learning, Bank of Korea

I. Introduction

As succinctly encapsulated in the title, "Text as Data" (Gentzkow, Kelly, and Taddy, 2017), text is data in that it is a useful source of information. However, text

Received: Sept. 3, 2018. Revised: Jan. 31, 2019. Accepted: March 8, 2019.
* We thank the participants at the seminars of the Bank of Korea (BOK) and Korea Institute for International Economic Policy for their helpful comments and suggestions. The financial support of the BOK is greatly appreciated. The views expressed herein are those of the authors and do not necessarily reflect those of the BOK. The usual disclaimers apply.
** First Author, Ph.D. Candidate, School of Business, Yonsei University, Email: yj.lee@yonsei.ac.kr
*** Second Author, Economist, Bank of Korea and School of Economics, Yonsei University, Email: soohyonkim@bok.or.kr
**** Corresponding Author, Associate Professor, School of Economics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 120-749, South Korea, Email: kypark@yonsei.ac.kr

has been underutilized given its difficult quantification and interpretation compared with numerical data. As computing power increased, text-mining analysis evolved from a labor-intensive manual discipline to a sub-field of "big data" analysis. Bholat, Hansen, Santos, and Schonhardt-Bailey (2015) emphasized that a computer-enabled approach to text mining can process more texts than any person could ever read and also extract information that human readers might easily miss.

Other fields have actively applied text mining, such as marketing and political sciences. However, text-mining has not been the main tool for economic analysis, particularly for monetary and macroprudential policies. However, development and application of appropriate tools to analyze central bank communications are important, given the increased importance of transparency in conducting monetary policy. A former Governor of the Federal Reserve Board (Warsh, 2014) emphasized the importance of textual discourse as a potential source of additional information as follows: "No surprise, Fed policymakers far more often reveal their differing judgments on economic variables in their discussion around the table than in their actual votes." Warsh profoundly evaluated the usefulness of text mining approach with the following justification: "a more rigorous and constructive means of judging the effects of the Fed's new transcript policy can be found by evaluating the text of the transcripts themselves."

With all these considerations in mind, we use text mining to extract the quantitative information about monetary policy decision-making from 231,699 documents between May 2005 and December 2017.[1] By converting the qualitative contents of the Bank of Korea's (BOK) Monetary Policy Board (MPB) minutes into quantitative indicators, we measure the sentiment and tone of the minutes and assess whether the MPB communication conveys any additional information that is excluded in the available macroeconomic data. Accordingly, our text-based indicators help explain the current and future monetary policies when considering an augmented Taylor rule. In addition, in terms of explaining the current and future monetary policy decisions, our indicators significantly outperform English text-based indicators, a media-based measure of economic policy uncertainty of Baker, Bloom, and Davis (2016), and a measure of macroeconomic uncertainty developed by Jurado, Ludvigson, and Ng (2015). By comparing these various measures, we provide guidance concerning the direction of future research in this field. Our study also shows the importance of using a field-specific dictionary and the original Korean text, not a translated text.

We make unique contributions in several aspects. First, to our knowledge, this study is the first to apply sentiment analysis to monetary policy decision-making of

[1] We use 151 sets of minutes from the Monetary Policy Board (MPB), 206,223 news articles related to interest rates, and 25,325 bond analyst reports. Section 3 presents a highly detailed explanation regarding our data.

the BOK. This study demonstrates that lexicon-based indicators have ample information about monetary policy beyond information contained in macroeconomic variables. Our results suggest several venues for future research. For example, one may interpret our indicators as a latent variable of the BOK policy rate and feed these indicators into a standard VAR or DSGE models that analyze the effect of monetary policy. Alternately, our indicators can be used to evaluate the effectiveness of the BOK's communication concerning the future direction of its monetary policy. A necessity arises in analyzing what kind of information the BOK documents convey and by how much, considering that central bank communication has emerged as an important tool for central banks to manage public expectations for inflation and economic activities. In this regard, our study demonstrates the usefulness of text mining in tasks related to central banking.

Second, in terms of methodology, we adopt several advanced techniques in this field. We use n-grams because of the difficulty in determining the tone or sentiment of a single word (a uni-gram) or a bi-gram phrase that combines positive and negative words, such as "lower unemployment" or "sluggish recovery." With n-grams (from 1- to 5-grams) as features, we consider the context and can capture the subtlety of central bank communications. We develop two kinds of sentiment indicators based on two contrasting but complementing methods to determine the polarity—hawkish, neutral, or dovish, in our case—of these n-grams. These methods are the market and lexical approaches. One advantage of the market approach is that it does not rely on the researchers' subjective selection of seed words and only uses the market information. However, indicators based on this approach may naturally produce statistically significant outcomes when we examine the market impact of the indicators because this approach decides the polarity of a word (n-gram in our case) based on its statistical association with market information (e.g., stock returns). By contrast, lexical approach decides the polarity based on the proximity to the pre-determined seed words. Performance depends on the specific choice of the seed words. We use a state-of-the-art sentiment induction algorithm called the SentProp framework by Hamilton, Clark, Leskovec, and Jurafsky (2016) to reduce this problem of researchers' discretion over the seed words. Furthermore, we use documents that are not used in building our lexicons to evaluate the accuracy of our lexicon classifications. We manually label 2,341 sentences from the introductory statements of the BOK Governor's news conferences and perform an out-of-sample test. We confirm that the accuracy is quite high.

Third, we use our own natural language processing (NLP) tool called Korean NLP Python Library for Economic Analysis (eKoNLPy) to address the difficulties associated with the Korean language, such as field-specific non-Korean loan words and irregular conjugation of verbs and adjectives. The eKoNLPy tool was developed by Lee (2018), one of our coauthors, which is specifically designed for text mining research in the field of economics and finance. For example, it can recognize words, such as "일드커브 (yield curve)" and "스티프닝 (steepening)," whereas the previous tools cannot. The eKoNLPy is made public at GitHub (https://github.com/ entelecheia/eKoNLPy) to encourage additional research in this area and to enhance comparability. Lastly, in terms of future applications, our automated approach can be easily extended to measure other information, such as macroeconomic uncertainty and stock market sentiments.

The rest of the paper is structured as follows. Section 2 reviews the related literature. Section 3 explains our data and methodology to develop text-based indicators that capture the sentiment of monetary policy. Section 4 empirically evaluates the performance of our text-based indicators and compares it with other measures. Section 5 summarizes our finding and discusses future research avenues.

II. Literature Review

We limit our discussion relevant to central banking rather than attempting an extensive literature review on text mining applied to economics.[2] Early studies that use text mining approach rely more on the frequency of specific words rather than on more sophisticated methods that measure tones or sentiments. For example, Choi and Varian (2012) show that the number of search keywords, such as "jobs" and "welfare & unemployment," is a good predictor of leading indicators of the US labor market as well as economic cycles. McLaren and Shanbhogue (2011) show that the volume of the related Google searches can predict changes in unemployment as well as house prices. A recent study by Baker et al. (2016) constructs economic policy uncertainty (EPU) measure by counting the number of news articles containing specific words, such as "uncertainty" and "economy;" this measure is shown to be associated with investment and employment in policy-sensitive sectors, such as defense, health care, finance, and infrastructure construction.

Concerning monetary policy, several studies attempted to extract additional information from social media or central bank communication. Meinusch and Tillmann (2017) quantify beliefs about the timing of the exit from quantitative easing ("tapering") of market participants by using data from Twitter. Related to central bank communication, Lucca and Trebbi (2011) build semantic scores by using the discussions of FOMC statements from news on the days of announcements and find that long-term Treasury yields react to their semantic

[2] See Gentzkow et al. (2017) for a survey of text mining approach to economic research and Bholat et al. (2015) for a survey specific to central banking.

scores. Hansen and McMahon (2016) cluster sentences in FOMC communication with Latent Dirichlet Analysis (LDA); within a factor augmented vector autoregression (FAVAR) framework, they find that FOMC communication on forward guidance has a strong impact on the financial market. In line with our research to measure the sentiment of monetary policy, Picault and Renault (2017) manually classify all sentences in European Central Bank (ECB) press conferences and build a field-specific dictionary. Their measure of ECB monetary policy sentiment helps explain current and future ECB monetary decisions, and markets are highly volatile when the Governing Council's views on the economic outlook are negative. Nyman et al. (2018) attempt to extract high-frequency sentiment from Bank of England's daily market commentary and show that the sentiment series track down the commonly used measure of volatility very well, such as VIX; it even moves as a leading indicator ahead of such measure.

Text mining is also applied to the area of financial stability. Based on over 1,000 releases of financial stability reports (FSRs) and speeches/interviews by central bank governors from 37 central banks over the past 14 years, Born, Ehrmann, and Frtzscher (2014) find that FSRs with optimistic tones lead to significant and lasting positive stock market returns, whereas no such effect exists for pessimistic ones. Nopp and Hanbury (2015) analyze CEO letters and annual management reports of 27 banks under ECB supervision and subsequently find that sentiment scores of the textual data accurately reflect major economic events as well as the aggregate Tier 1 capital ratio evolution. Bholat et al. (2017) analyze confidential letters sent by Prudential Regulation Authority to banks and financial firms under supervision by using a machine learning method; the letters are found to vary, depending on the riskiness of the firm in terms of negative words and direct languages.

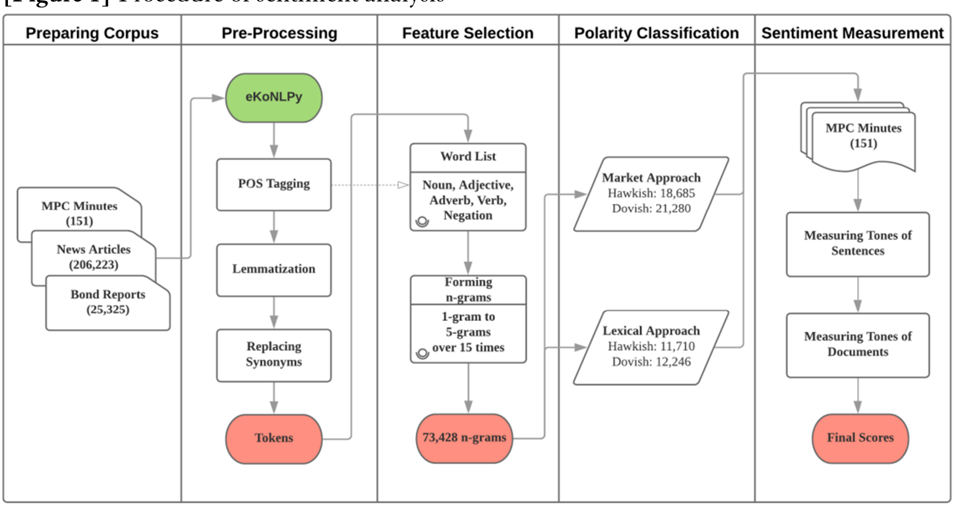
We only find two existing studies while predicting that the text mining approach using the Korean language in economic analysis is at the very early stage. Won et al. (2017) provide a useful guidance on future research direction by comparing the methods of bag-of-words and word2vec and demonstrating the outperformance of word2vec. However, as Won et al. (2017) indicate, word2vec is often problematically classifying antonyms as similar words. We address this problem by using n-gram embedding. Notably, the research of Pyo and Kim (2017) is the first to attempt sentiment analysis on economic analysis by using the Korean language. Pyo and Kim (2017) construct the sentiment index of financial markets by using news articles and show that these investor sentiment measures are statistically associated with asset prices, such as government bond yields and exchange rate. While Pyo and Kim (2017) and our study use sentiment analysis, several points of departure emerge. First, we use n-grams to capture the subtlety of texts by considering contexts. Second, we use a field-specific dictionary specifically designed for text mining in economics and finance. Third, we also construct the sentiment index based on market approach in addition to the lexical approach.[3]

II. Data and Methodology

According to Liu (2009), sentiment analysis is a series of methods, techniques, and tools about detecting and extracting subjective information from language, such as opinion and attitudes. The roots of sentiment analysis are in the studies on public opinion analysis at the beginning of the 20th century and in the text subjectivity analysis performed by the computational linguistics community in the 1990s. The current form of sentiment analysis became popular with the advancement of computer technology and the availability of texts on the web.[4] Although sentiment analysis has a long history in linguistics and political science, it has only been recently utilized in economics. Specifically, empirical studies in economics scarcely use the Korean language.

Sentiment analysis generally takes the following processes: (i) preparation of the corpus of interests, (ii) pre-processing of texts, (iii) feature selection, (iv) polarity or sentiment classification of features, and (v) measurement of sentiments of sentences and documents. We briefly explain our actions in each step. Table 1 and Figure 1 summarize our discussion.[5]

[Figure 1] Procedure of sentiment analysis



Notes: This figure, along with Table 1, summarizes our discussion.

[3] We explain our application of text mining approach, including market and lexical approaches.

[4] See Liu (2009) for a review on the evolution of sentiment analysis.

[5] Additional examples and details of text mining for central banks can be found in Bholat et al. (2015).

[Table 1] Process of sentiment analysis

1. Preparing the corpus	• 231,699 documents – 151 MPB minutes – 206,223 news articles related to interest rates – 25,325 bond analyst reports
2. Pre-processing texts	• Tokenization • Normalization – removing stop words – stemming and lemmatization • Morphological analysis of the Korean language → eKoNLPy – spacing – terminology and foreign words
3. Feature selection	• N-grams as a feature – 73,428 n-grams
4. Polarity classification	• Market approach – Naive-Bayes classifier • Lexical approach – ngram2vec – SentProp of Hamilton et al. (2016) • Evaluation – 2,341 manually classified sentences – out-of-sample test
5. Sentiment measurement	• Measuring tones of sentences from the features of n-grams • Measuring tones of documents from tones of sentences

Notes: This table summarizes our actions in each step of sentimental analysis. See Section 3 for additional details.

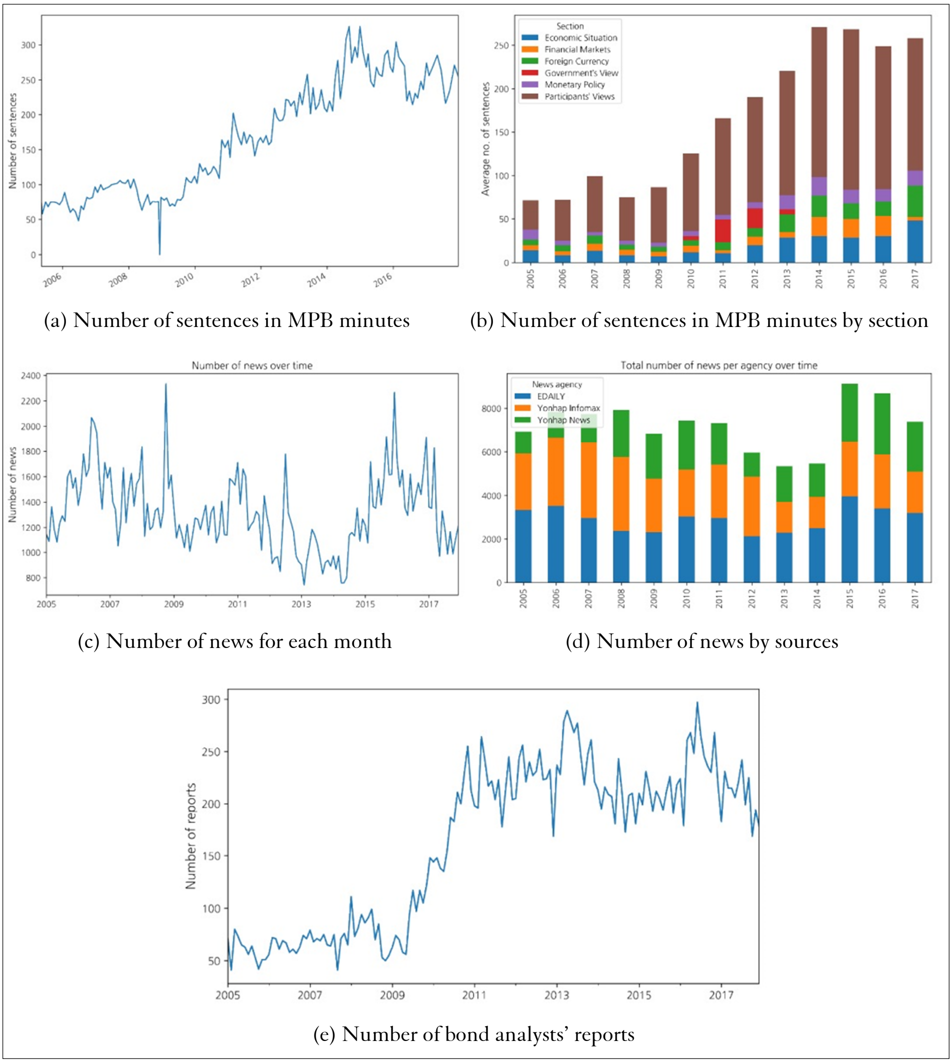
3.1. Preparing the Corpus

We collect 231,699 documents for the period of May 2005 to December 2017, which include 151 minutes of MPB meetings; 206,223 news articles; and 25,325 bond analyst reports. Table 2 shows the types and numbers of documents and the average and maximum number of sentences. Although our target texts are the MPB minutes, we use a large amount of other documents to build field-specific lexicons.

[Table 2] Statistics of the corpus

Document type	No. of docs	Average no. of sentences	Max no. of sentences
MPB minutes	151	165	326
News articles	206,223	15	340
Bond reports	25,325	49	2,515
Total	231,699	19	2,515

[Figure 2] Text data



(a) Number of sentences in MPB minutes (b) Number of sentences in MPB minutes by section

(c) Number of news for each month (d) Number of news by sources

(e) Number of bond analysts' reports

MPB Minutes. The MPB minutes are released at 4 pm on the first Tuesday two weeks after each meeting since September 2012.[6] The minutes consist of several sections:

- Outline, which provides the information about the administrative details.

[6] Given this convention of disclosing the minutes two weeks after the market closes, performance of event studies that attempting to gauge the market impact of monetary policy is difficult. The minutes were released six weeks after each meeting during April 2005 to September 2012.

- Summary of discussion on the current economic situation, which contains the MPB members' discussion on economic situation, FX and international finance, financial markets, and monetary policy.
- A discussion concerning monetary policy decision records the views of individual members.
- Result of deliberation on monetary policy.

We download the files of MPB minutes from May 2005 to December 2017 (151 min) from the BOK website.[7] We use only the second and third sections. Panels (a) and (b) in Figure 2 display the number of sentences in MPB minutes for each section over time. The length of the minutes has increased after the global financial crisis.

News Articles. We collect news articles that include the word "interest rates (금리)" from Naver and Infomax from January 2005 to December 2017.[8] These news articles contain information on the general economy, monetary policy, financial market, and public perception on the BOK's future monetary policy stances. We use only the articles from the top three news agencies (in terms of number of articles produced) because of the numerous duplicate articles from the originators. The number of news articles for our final use is 206,223. Among them, 42% (86,538) are from Yonhab Infomax, 33% (68,728) from EDAILY, and 25% (50,957) from Yonhab News. We remove the header and footer from the articles. Panels (c) and (d) in Figure 2 show the number of news articles over time.

Bond Analysts' Reports. We also use bond analysts' reports for the following two reasons: first, bond analyst reports show the experts' views on the monetary policy and the bond market; second, we incorporate the informal styles of writing into our lexicons. Generally, bond analysts write in a more informal manner compared with journalists. We obtain the reports from WIEfn, a financial information service provider in Korea.[9] Panel (e) of Figure 2 shows the number of reports from January 2005 to December 2017.

Our corpus is large in size and covers various topics. Figure 3 shows the various topics of our corpus, which we extract using Latent Dirichlet Allocation method, a topic modeling method. Table 3 shows the relative frequencies of the topics.

---

[7] During the sample period, 152 meetings were conducted, with 151 minutes. An emergency meeting during the global financial crisis had no minutes. One can download the minutes from the following link: http://www.bok.or.kr/portal/singl/crncyPolicyDrcMtg/listYear.do?mtgSe=A&menuNo=200755

[8] https://news.naver.com, http://news.einfomax.co.kr

[9] https://www.wisereport.co.kr

[Figure 3] Topic word clouds of the corpus

[Table 3] Average weights of the topics

No.	Topic name	Total	Minutes	News	Report
1	Foreign Currency	5.24	11.20	5.94	3.75
2	Financial Policy	2.69	2.24	3.15	1.99
3	Bond Issue Market 1	3.35	0.73	1.29	6.79
4	Monetary Policy	3.81	12.56	4.47	2.20
5	Bond Issue Market 2	2.79	1.32	2.67	3.08
6	Financial Crisis	1.79	1.03	2.09	1.36
7	Swap Market	4.30	3.05	4.02	4.82
8	Inflation	3.32	10.56	2.68	3.89
9	Credit Ratings	1.42	0.38	0.93	2.26
10	Real Estate	1.20	0.15	1.71	0.46
11	Global Government Bond	2.23	0.40	1.34	3.78
12	Macro Stability	1.26	2.74	1.16	1.32
13	Real Estate Policy	3.01	2.44	3.99	1.46
14	Eurozone	3.94	1.27	3.33	5.07
15	Economic Growth	5.02	13.60	4.58	5.19
16	Money Market	0.88	0.79	1.09	0.55
17	Global Stock Market	2.11	0.37	2.74	1.20
18	Global Monetary Policy	5.52	4.87	4.36	7.40
19	Financial Instruments	2.34	0.22	3.42	0.74
20	Corporate Valuation	1.64	0.27	2.13	0.95
21	Capital Requirements	0.77	0.42	0.48	1.27
22	Government Bond Futures	4.69	0.46	4.44	5.34
23	Domestic Stock Market	0.81	0.17	1.21	0.23
24	Soho Money Market	1.34	0.65	2.06	0.22
25	Bond Market Price	1.01	0.33	0.52	1.82
26	Carry Trade	1.41	0.66	1.87	0.71
27	Government Regulation	1.43	0.87	1.75	0.95
28	Fund Market	4.53	2.19	5.40	3.30
29	Corporate Restructuring	1.43	1.66	1.59	1.16
30	Consumption/Income	1.93	5.68	1.72	2.04
31	Liquidity Provision	0.61	0.21	0.46	0.88
32	Politics	2.13	0.65	1.51	3.21
33	Raw Material Price	1.61	0.90	1.54	1.78
34	Bond Investment Strategy	1.80	0.13	1.44	2.46
35	Housing Debt	5.99	8.03	7.81	2.95
36	Corporate Credit Ratings	3.61	0.46	1.86	6.60

3.2. Pre-processing Texts

3.2.1. Typical Steps of Pre-Processing

Pre-processing texts include tokenization and normalization. Tokenization is a step to split long strings of texts into small pieces or tokens, which are generally words. Tokenization can incorporate part of speech (POS) tagging, which assigns a part of word, such as nouns, verbs, adjectives, and so on. Normalization is the process of transforming a text into a single canonical form. Normalization includes the following: punctuation removal, stop word removal, conversion of numbers to their word equivalents, stemming, lemmatization, and case folding.[10]

A typical text pre-processing procedure for English includes (i) converting all words to lower case, (ii) removing numbers and punctuation by using a Porter (1980) stemming algorithm to reduce inflected words to their word roots (e.g., "increasing" to "increase," "unemployment" to "unemploy") or lemmatization (e.g., "better" to "good"), and (iii) removing stop words (e.g., a, the, an, of, to, etc.).

3.2.2. eKoNLPy

Several issues exist in converting the Korean text into numerical expressions (e.g., bag of words and word embedding). The first issue is related to spacing. Unlike English, postpositions are not space-delimited, and spacing rules are not strictly observed. Second, numerous foreign words exist that do not follow the foreign language notation standards; many of these words are field-specific. Third, various notations exist for synonyms (e.g., inflation for "인플레이션," "인플레," and "물가"). This issue can be important when one uses n-grams. Various notations of synonyms increase the number of word combinations and dilute the frequency of n-grams. Fourth, numerous verbs and adjectives conjugate irregularly. Irregular conjugation also aggravates the explosion of dimension in n-gram models, which hinders polarity classification. The first issue of spacing is relatively well taken care of by currently available Korean morpheme analyzers. For example, one can use KoNLPy.[11] However, other issues are not. We use eKoNLPy developed by Lee (2018), one of our coauthors, for this reason. eKoNLPy constructs a dictionary specific to economics and finance and uses its own morphological analyzer.

Related to the second issue, eKoNLPy is equipped with pre-supplied 4,202 field-specific terms acquired from readily available economic term dictionaries on the Internet to fully support economics and finance domain-specific terms (i.e., jargon and foreign words).[12] eKoNLPy has the functionality to easily add custom terms and foreign words to the dictionary for POS tagging. For the third issue, eKoNLPy pre-defines 1,325 pairs of synonyms in the dictionary and supports the function of replacing synonyms to deal with various notations of synonyms. The last issue, that is, conjugation of adjectives and verbs, can be handled by stemming or

[10] Stop word removal is to drop stop words, such as "it," "the," and "etc." Stemming is to simply count stems (e.g., using "bank" for "banking" and "banks"). Lemmatization is to group the inflected forms of words so that they can be analyzed as a single item. POS tagging often helps lemmatization. For example, "saw" can be the past tense of a verb "see" or a noun.

[11] KoNLPy is a Python package for NLP of the Korean language (http://konlpy.org/en/v0.5.1/references).

[12] Several online economic term dictionaries are available at Naver, Maekyung, Hankyung, etc.

lemmatization. Normalizing irregular conjugation of Korean words can be addressed by lemmatization rather than by stemming because lemmatization considers the morphological analysis of words. eKoNLPy supports the lemmatization of 1,291 adjectives and verbs, which are frequently used in the economics and finance domain to deal with this problem.

Given that eKoNLPy is developed aimed at text mining for economic analysis from the beginning, we expect its outperformance in economic analysis compared with KoNLPy. For example, we consider the following sentence:

> "한국은행이 12일 금융통화위원회(금통위) 회의를 열고 기준금리를 현행 연 1.50로 동결했다."

We find that eKoNLPy successfully recognizes "금융통화위원회 (Monetary Policy Board)" and "금통위 (MPB);" whereas KoNLPy, which uses a general-purpose dictionary, cannot.[13] We consider the following phrase on bond market:

> "금리 박스권 상단 상향과 일드 커브 완만한 스티프닝 전망."

eKoNLPy recognizes "일드커브 (yield curve)" and "스티프닝 (steepening)," whereas KoNLPy cannot.

3.3. Feature Selection

Not all words are used to express opinions, thereby necessitating the conduct of feature selection to restrict words or phrases to a targeted list of words that express opinions. Restricting words also facilitates the speed of processing by reducing the dimension of term (word) vectors. In addition, single words often lose the context. For example, whereas the word "recovery" in isolation appears to carry a positive message, the phrase "sluggish recovery" does not.[14] When positive and negative

---

[13] The result from KoNLPy is as follows:
"한국은행/NNP," "이/JKS'" "12/SN," "일/NNBC," "금융/NNG," "통화/NNG," "위원회/NNG," "(/SSO," "금/NNG," "통/NNG," "위/NNG," ")/SSC," "회의/NNG," "를/JKO," "열/VV," "고/EC" "기준/NNG," "금리/NNG," "를/JKO," "현행/NNG," "연/NNG," "1/SN," "./SY," "50/SN," "%/SY," "로/JKB," "동결/NNG," "했/XSV," "다/EF," "./SF"
The result from eKoNLPy is as follows:
"한국은행/NNP," "이/JKS," "12/SN," "일/NNG," "금융통화위원회/NNG," "금통위/NNG," "회의/NNG," "를/JKO," "열/VV," "고/EC," "기준금리/NNG," "를/JKO," "현행/NNG," "연/NNG," "1/SN," "./SY," "50/SN," "%/SY," "로/JKB," "동결/NNG," "했/XSV," "다/EC." The abbreviations are for POS tagging, such as NNP, JKS, and SN. For example, NNG, JKS, and SN imply general noun, nominative case postposition, and number, respectively. See the table in Appendix B for additional details.

[14] Apel and Grimaldi (2014) use two-word combinations (bi-grams) of a noun and an adjective, such as "higher inflation" or "slower growth," to make the hawkish-dovish classification.

words are combined, such as a bi-gram phrase, “lower unemployment,” the sentiment is not easy to measure. Thus, we use n-grams to address this problem.[15] However, increasing the length of n-grams has a trade-off. With extremely long n-grams (say, 10-grams), we might fall into the problem of over-fitting to the sample; as the lexicons are highly specific to the target documents, applying those lexicons to other types of documents is difficult, such as news articles or experts’ writings. Moreover, a curse of dimensionality arises with n-grams.[16] An exponential growth exists in the number of features, and the probability of seeing the n-grams with the same features becomes small. This explosion of dimension also causes computational problems regarding memory size and speed of processing.

We set the n of n-gram to five, with additional rules, to address this tradeoff.[17] To avoid the explosion of dimension, we use the limited word set in forming n-grams by limiting the part-of-speech tag of words to nouns (NNG), adjectives (VA, VAX), adverbs (MAG), verbs (VA), and negations.[18] To improve the accuracy of the classification and to avoid multiple counting, we consider only the highest n-gram when multiple overlapping n-grams are found in each sentence. We also drop n-grams that occur less frequently than 15 times.[19]

The final word set comprised 2,712 words, and we obtain the resulting 73,428 n-grams. Notably, our n-grams naturally include single words (1-grams) because we cover from 1- to 5-gram.[20] The next step is to classify the polarity of these n-grams for measuring the sentiments of sentences or documents.

---

[15] Picault and Renault (2017) define the field-specific lexicon by considering n-grams (from 1-gram to 10-grams) appearing at least twice in their sample. They classify the polarity of n-grams by calculating the probability of belonging to which category of sentences (dovish, neutral, hawkish or positive, negative, neutral), after manually classifying all sentences pronounced during ECB introductory statements. By limiting n-grams to those having a probability over 0.5 in each class, their final field-specific lexicon is composed of 34,052 n-grams.

[16] As text is represented as very high-dimensional but sparse vectors, reducing dimensionality while preserving the important variation across documents is challenging. With the introduction of n-gram, which is a contiguous sequence of n words in a text, this problem is aggravated. With a 1,000 unique word corpus, a bigram model needs $1{,}000^2$ values; a trigram model will need $1{,}000^3$; and so on.

[17] Including Hutto and Gilbert (2014), numerous studies report that n-gram approach improves the performance of sentiment analysis. In general, they use bigram to 5-gram. While Dey, Jenamani, and Thakkar (2018) propose the first fully automatic score calculation algorithm to create a domain-independent n-gram sentiment dictionary, they use up to tri-grams because of computational burden.

[18] Appendix B shows the eKoNLPy tagset for POS tagging.

[19] From the perspective of n-grams being effective opinion-bearing features for sentiment analysis, ideal n-grams should contain the following factors: the agent or doer (i.e., the person or entity who maintains or causes that affective state), the target (i.e., the entity about which the affect is felt), how (i.e., the direction or the degree of affective state), and negation, if any.

[20] We also check the sensitivity of n in n-gram and find that our result is insensitive to the choice of n (= 2, 3, 4, and 5). We also find that a high n increases in-sample performance and lowers out-of-sample performance in terms of the accuracy of polarity classification. This result suggests that a high n renders n-gram highly document-specific.

3.4. Polarity Classification

If no well-known lists of polarity words exist, such as Harvard-IV or LM dictionary, we must classify the polarity of our selected features (n-grams in our case) on our own.[21] Several categories of polarity classification may exist. First refers to supervised vs. unsupervised (automated) approaches depending on whether it needs human intervention or not. Google Cloud Sentiment Analysis API is an example of supervised classification in which the classifier is trained on the massive amount of documents. An example of unsupervised approach is semantic orientation by using pointwise mutual information (PMI) to measure the similarity between words and polarity prototypes.[22]

Second refers to machine-learning- vs. lexical-based methods. The former uses training corpora annotated with polarity information, and the latter uses polarity lexicons. In the lexical-based approach, three methods exist to obtain the polarity lexicons, namely, manual, dictionary-based, and corpus-based. The manual method is extremely time-consuming and prone to human errors.[23] The dictionary-based approach searches the dictionary, starting from the seed words, to determine their synonyms and antonyms. This approach requires a well-constructed lexical database, such as WordNet or thesaurus.[24] The disadvantage of this approach is its

---

[21] For unigrams (single words), the oldest is the General Inquirer (Stone, Dunphy, and Smith, 1966) also known as Harvard IV-4. The latter has numerous categories of word lists, including 1,915 words of positive outlook and 2,291 words of negative outlook. In the financial context, negative words are used for sentiment analysis (Tetlock, 2007). A widely used word list in finance literature is that of Loughran and McDonald (2011), which has lists of single words by category (Negative, Positive, Uncertainty, Litigious, Modal, and Constraining). Their research indicates that the LM dictionary has a good correlation with financial metrics. The LM dictionary is available at https://sraf.nd.edu/textual-analysis/resources/.

[22] Semantic orientation is a concept from computational linguistics and defines the position of a word or string of words between two opposite concepts. PMI is a technique for quantifying the similarity between two random variables based on probability theory. Using PMI, the similarity between two lexicons is measured as

$$PMI(w_1, w_2) = \log \frac{P(w_1, w_2)}{P(w_1)P(w_2)},$$

where w_1 and w_2 are the two lexicons under consideration. Lucca and Trebbi (2011) and Tobback, Nardelli, and Martens (2016) employ this Semantic Orientation PMI (SO-PMI) method to measure the sentiments of the FOMC statements and the ECB's press statements. They measure their sentiment indicators by counting the co-occurrences of words or strings in their documents with words that are normally associated with the sentiments of monetary policy (dovish or hawkish) by using Google Search.

[23] Hence, it is used in the specific tasks with only small-focused list of words, such as the cases of Apel and Grimaldi (2014) and Bennani and Neuenkirch (2016).

[24] WordNet® is a large English lexical database. Nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. https:

inability to determine field-specific polarity words. The corpus-based method finds polarity words by searching patterns that occur together, along with seed words in a large corpus. This approach has a major advantage in that it can find field- and context-specific sentiment words and their polarities by using a corpus of that domain. Thus, the corpus-based approach becomes the most suitable for economics or finance area, where jargon or words with different connotations are prevalent.

We classify polarity of n-grams in two ways. One is the market approach that classifies polarity from market information by using machine learning. The other is the corpus-based approach that classifies polarity by using word (n-gram in our case) embedding and seed words, which we call lexical approach.[25]

3.4.1. Market Approach

Numerous attempts have been made to extract information about market expectations from asset prices, as presented in the survey of Söderlind and Svensson (1997). One can extract information to the extent that financial market is efficient and asset prices reflect information of financial market. Then, why cannot one extract information from text? With this consideration in mind, we classify the polarity of features using market information and call it "market approach."

To determine the relative weights of words, Jegadeesh and Wu (2013) use words as the explanatory variable and stock returns as the dependent variable. If the coefficient of a word is positive and large, then, the word has high weight. That is, market reactions were estimated and used as relative weights. One advantage of Jegadeesh and Wu (2013) is that it does not rely on subjective judgment. Their work distinguishes itself with our market approach because the former starts from the known word lists, such as Harvard IV-4 or LM dictionary and uses regression-based approach to determine the relative weights (the sign and the magnitude) of the terms; it also uses uni-gram, not n-grams. Our market approach also uses market information to determine the polarity of features. However, to extract n-grams from a large corpus and classify their polarity, we use the machine learning method.

For our market approach, we use the Naïve Bayes Classifier (NBC), a simple probabilistic classifier. NBC is very simple but still competitive with highly advanced methods, including support vector machines.[26] NBC is called naive, because it assumes that all features are independent given the class (hawkish or

//wordnet.princeton.edu.

[25] Word embedding is the collective name for a set of language modeling and feature learning techniques in NLP, where words or phrases from the vocabulary are mapped to vectors of real numbers. Conceptually, it involves a mathematical embedding from a space with one dimension per word to a continuous vector space with a low dimension (https://en.wikipedia.org/wiki/Word_embedding).

[26] Refer to the following link for additional details on NBC and support vector machine: https://www.ocf.berkeley.edu/~janastas/supervised-learning-with-text-II-03-16-Lecture.html#1.

dovish in our case). Although NBC is not a feature selection tool, this independence assumption makes us use the conditional probability of each feature as a polarity score.

Machine learning methods, including NBC, are mostly supervised ones, which depend on the existence of labeled training documents. Training documents are generally obtained from the review by the public when it is available. Otherwise, several experts must manually label training documents. The former is unavailable for monetary policy. The latter is labor- and cost-intensive and is subject to experts' judgments. To circumvent these problems and exploit the information of financial market, we label news articles and reports in our corpus as hawkish (dovish) if the one-month change in call rates is positive (negative) on the day they are released.[27]

We randomly divide our labeled sentences (more than 4 million sentences) into a training set and a test set by 9:1 ratio.[28] Using 5-grams (from 1- to 5-grams) as features for each sentence, we train the classifier and check its accuracy. The trained NBC yields the conditional probability of each feature given the class (hawkish/dovish), which we use as a polarity score of the feature:

$$polarity\ score = \frac{p(feature \mid hawkish)}{p(feature \mid dovish)} = \frac{p(feature\ \&\ hawkish)/p(hawkish)}{p(feature\ \&\ dovish)/p(dovish)}. \quad (1)$$

An *n-gram* is roughly labeled as "hawkish" if it presents itself more often in "hawkish" *documents* compared with "dovish" ones.[29]

Given that we use random sampling and a probabilistic classifier, every training yields different posterior probability of each class. To obtain good predictive performance, we repeat this procedure 30 times and use the average of the polarity scores as a final one, which is called bagging in machine learning.[30] Although not the direct performance measure of our lexicon, the average accuracy of NBC is 86% (positive precision: 90%, positive recall: 84%, negative precision: 82%, negative recall: 88%).

[27] The actual threshold is $\pm 3dp$ to exclude meaningless movements.

[28] One may suggest using three kinds of sets, namely, training, test, and validation. Given that we evaluate our polarity classification with out-of-sample documents in Section 3.4.3, we need no validation set, and our out-of-sample evaluation is highly rigorous.

[29] Alternatively, one may consider the following:

$$polarity\ score = \frac{p(hawkish \mid feature)}{p(dovish \mid feature)} = \frac{p(feature\ \&\ hawkish)}{p(feature\ \&\ dovish)}.$$

In our case, because $p(hawkish) = 0.53$ and $p(dovish) = 0.47$, these two formulas produce similar results.

[30] Bagging is an abbreviation of bootstrap aggregating, which involves having each model in the ensemble endowed with equal weight.

We classify the polarity of our lexicon as hawkish (dovish) if the polarity score is greater (less) than 1, excluding lexicon in the gray area by using intensity of 1.3 as a threshold.[31] The final number of lexicon is 18,685 for hawkish and 21,280 for dovish. Table 4 presents a sample of polarity lexicons.

[Table 4] Sample of polarity lexicons by market approach

Hawkish	Dovish
총액/NNG;대출/NNG;한도/NNG;축소/NNG	금리/NNG;지준율/NNG;인하/NNG
재할인율/NNG;인상/NNG	하락/NNG;거래/NNG;마치/VV;내리/VV
인플레이션/NNG;심리/NNG;확산/NNG	금리/NNG;인하/NNG;가계/NNG;부채/NNG;증가/NNG
자본/NNG;유출입/NNG;변동/NNG;완화/NNG	금리/NNG;인하/NNG;소식/NNG;상승/NNG
자본/NNG;유출입/NNG;규제/NNG;우려/NNG	정책/NNG;공조/NNG;금리/NNG;인하/NNG
물가/NNG;안정/NNG;견조/NNG;성장/NNG	실물/NNG;경기/NNG;침체/NNG;우려/NNG
금리/NNG;인상/NNG;인플레이션/NNG;우려/NNG	유동성/NNG;경색/NNG;해소/NNG
소비자/NNG;물가/NNG;상승률/NNG;금리/NNG;인상/NNG	경기주체/NNG;심리/NNG;부진/NNG
물가/NNG;불안/NNG;확산/NNG	금융위기/NNG;세계/NNG;확산/NNG
잠재/NNG;성장률/NNG;경제/NNG;성장/NNG	구조조정/NNG;자본/NNG;확충/NNG
완만/NNG;속도/NNG;확장/NNG	비둘기/NNG;금리/NNG;인하/NNG
콜/NNG;금리/NNG;인상/NNG;금리/NNG;인상/NNG	위기설/NNG;불안/NNG
총액/NNG;한도/NNG;대출/NNG;금리/NNG;인상/NNG	인하/NNG;파급효과/NNG
금리/NNG;인상/NNG;물가/NNG;상승/NNG;압력/NNG	유로존/NNG;경제/NNG;지표/NNG;부진/NNG
경제/NNG;예상/NNG;회복/NNG	경기/NNG;후퇴/NNG;우려/NNG;완화/NNG
금리/NNG;인상/NNG;콜/NNG;금리/NNG;인상/NNG	신용스프레드/NNG;부담/NNG
발행/NNG;압력/NNG;약화/NNG	국고채/NNG;하락/NNG;금리/NNG;내리/VV
자본/NNG;규제/NNG;우려/NNG	금리/NNG;인하/NNG;지준율/NNG;인하/NNG
금리/NNG;콜/NNG;인상/NNG	금리/NNG;인하/NNG;경제/NNG
금리/NNG;인상/NNG;물가/NNG;상승/NNG	정책/NNG;공조/NNG;차원/NNG;금리/NNG;인하/NNG
자금/NNG;해외/NNG;이탈/NNG	금리/NNG;인하/NNG;효과/NNG;없/VA
경기/NNG;위축/NNG;속도/NNG;둔화/NNG	금리/NNG;인하/NNG;실망/NNG
물가/NNG;상승/NNG;압력/NNG;점차/MAG;크/VV	미국발/NNG;금융/NNG;불안/NNG
자산시장/NNG;불안/NNG	cp/NNG; 금리/NNG; 급락/NNG
세제개편안/NNG;불확실성/NNG	cd/NNG; 금리/NNG; 인하/NNG
금리/NNG;인상/NNG;물가/NNG;불안/NNG	대출/NNG;예금/NNG;금리/NNG;인하/NNG
금리/NNG;인상/NNG;소식/NNG;상승/NNG	금융위기/NNG;세계/NNG;경제/NNG;침체/NNG
원화/NNG;절상/NNG;금리/NNG;인상/NNG	경제/NNG;구조적/VAX;취약/NNG
고유가/NNG;높/VA	금리/NNG;예금/NNG;금리/NNG;내리/VV
금리/NNG;인상/NNG;경기/NNG;위축/NNG	경제/NNG;수출/NNG;감소/NNG

3.4.2. Lexical Approach

While our market approach that associates the release dates of documents with changes in call rates, is easy and also effective in maximizing market information, one may suggest different methods of incorporating market information. For example, one can use the stock market index instead of call rates. One can also measure the change in call rates over one week, not one month. Rather than addressing all these possibilities, we construct another indicator that takes an opposite stance in that it does not use market information at all. Instead, it uses the seed words, which we call lexical approach.

Lexical approach is based on an intuitive observation. If two words appear

[31] Intensity measures the relative strength of the polarity, which is a ratio of the conditional probability of the feature given the class (hawkish/dovish) with the greater of the two in numerator.

together frequently in the same context, they tend to have the same polarity. Then, the polarity of an unknown word can be determined by calculating the relative frequency of co-occurrence with another word. This process is possible using the concept of PMI. One can use Semantic Orientation-PMI (SO-PMI) proposed by Turney (2002) for polarity classification.

Two problems exist, despite the quite intuitive quality of this approach. First, this approach sometimes fails to recognize antonyms because it judges the polarity based on co-occurrence. Second, the outcome is affected by choices of seed words. To address the first problem, we use ngram2vec by Zhao, Liu, Li, Li, and Du (2017) instead of word embedding. They show that n-gram embedding is effective in finding antonyms. For the second problem, we adopt the SentProp framework by Hamilton et al. (2016), a state-of-the-art domain-specific sentiment induction algorithm. The SentProp framework addresses this issue by bootstrapping seed words.

We place the seed set of words (n-grams in our case) and our n-grams in a vector space (lexical graph) and measure the proximity of our n-grams to this seed. The polarity of an n-gram is proportional to the probability of a random walk from the seed set hitting that n-gram. Each feature will have two probabilities, one for hawkish and the other for dovish. A final polarity score is the relative ratio of the two as in Equation (1).

We train ngram2vec by using the entire 231,699 documents of our corpus. The parameters we use for training are 5-grams for center words, 5-grams for context words, window size of 5, negative sampling size of 5, and 300 dimensions for vector representation.[32] Our corpus has 344,293 unique n-grams, with a minimum frequency limit of 25, which yield 410,902,512 pairs of n-grams (21.7 GB in size). With this resulting n-gram vector, we bootstrap by running our propagation 50 times over 10 random equally sized subsets of the hawkish and dovish seed sets. Table 5 shows the seed sets.

Similar to the market approach, we classify the polarity of our lexicons as hawkish (dovish) if the polarity score is greater (less) than 1, excluding lexicon in the gray area by using intensity of 1.1 as a threshold. The final number of lexicon is 11,710 for hawkish and 12,246 for dovish. Table 6 presents a sample of polarity lexicons.

We count the number of common n-grams to see if the market and lexical approach provide a similar result. Given 39,965 n-grams from the market approach and 23,956 n-grams from the lexical approach, a total of 14,154 common n-grams exist. Among them, 9,791 n-grams (69% of common n-grams) have the same

[32] Every word can be a center or context word. For example, in the sentence, “We train ngram2vec using the entire…,” if “ngram2vec” is a center word with window size of 1, then, “train” and “using” are context words. For additional information on these and negative sampling, refer to https://cs224d.stanford.edu/lecture_notes/notes1.pdf.

polarity.[33]

[Table 5] Seed words for polarity induction

Hawkish		Dovish	
높/VA	팽창/NNG	낮/VA	축소/NNG
인상/NNG	매파/NNG	인하/NNG	비둘기/NNG
성장/NNG	투기/NNG;억제/NNG	둔화/NNG	악화/NNG
상승/NNG	인플레이션/NNG;압력/NNG	하락/NNG	회복/NNG;못하/VX
증가/NNG	위험/NNG;선호/NNG	감소/NNG	위험/NNG;회피/NNG
상회/NNG	물가/NNG;상승/NNG	하회/NNG	물가/NNG;하락/NNG
과열/NNG	금리/NNG;상승/NNG	위축/NNG	금리/NNG;하락/NNG
확장/NNG	상방/NNG;압력/NNG	침체/NNG	하방/NNG;압력/NNG
긴축/NNG	변동성/NNG;감소/NNG	완화/NNG	변동성/NNG;확대/NNG
흑자/NNG	채권/NNG;가격/NNG;하락/NNG	적자/NNG	채권/NNG;가격/NNG;상승/NNG
견조/NNG	요금/NNG;인상/NNG	부진/NNG	요금/NNG;인하/NNG
낙관/NNG	부동산/NNG;가격/NNG;상승/NNG	비관/NNG	부동산/NNG;가격/NNG;하락/NNG
상향/NNG	(Total 25 seeds)	하향/NNG	(Total 25 seeds)

[Table 6] Sample of polarity lexicon by lexical approach

Hawkish	Dovish
인상/NNG	인하/NNG
확장/NNG	하향/NNG
상향/NNG	부진/NNG
투기/NNG;억제/NNG	회복/NNG;못하/VX
금리/NNG;상승/NNG	금리/NNG;하락/NNG
상회/NNG	악화/NNG
채권/NNG;가격/NNG;하락/NNG	침체/NNG
인플레이션/NNG;압력/NNG	하락/NNG
과열/NNG	변동성/NNG;확대/NNG
견조/NNG	위축/NNG
팽창/NNG	하회/NNG
물가/NNG;상승/NNG	둔화/NNG
부동산/NNG;가격/NNG;상승/NNG	완화/NNG
성장/NNG	채권/NNG;가격/NNG;상승/NNG
긴축/NNG	물가/NNG;하락/NNG
흑자/NNG	위험/NNG;회피/NNG
요금/NNG;인상/NNG	하방/NNG;압력/NNG
상방/NNG;압력/NNG	부동산/NNG;가격/NNG;하락/NNG

33 Ribeiro, Araújo, Gonçalves, Gonçalves, and Benevenuto (2016) conduct an apple-to-apple comparison of existing sentiment analysis methods and find that the accuracy widely varies depending on the dataset used. In general, results on Twitter datasets are good. The average accuracy of 24 methods is 84.2%, ranging from 52.0% to 95.3%. When comments from BBC forums and Digg.com are used, the performance is not as good as in the case of Twitter datasets. The averages are 67.4% and 62.4%, respectively, ranging from 13.3% to 93.9%.

낙관/NNG	비관/NNG
변동성/NNG;감소/NNG	요금/NNG;인하/NNG
위험/NNG;선호/NNG	적자/NNG
매파/NNG	비둘기/NNG
부동산/NNG;과열/NNG;억제/NNG	둔화/NNG;경기/NNG;침체/NNG
부동산/NNG;과열/NNG	경기/NNG;침체/NNG;빠지/VV
과열/NNG;우려/NNG	악화/NNG;경기/NNG;침체/NNG
과열/NNG;억제/NNG	경기/NNG;침체/NNG
과열/NNG;막/VV	침체/NNG;빠지/VV
경기/NNG;과열/NNG	침체/NNG;가능성/NNG;높/VA
부동산/NNG;과열/NNG;우려/NNG	경기/NNG;침체국면/NNG;빠지/VV
경기/NNG;과열/NNG;우려/NNG	침체/NNG;경기/NNG;침체/NNG
가격/NNG;억제/NNG	둔화/NNG;침체/NNG
투자/NNG;과열/NNG	경기/NNG;침체/NNG;빠지/VV;않/VX
부동산/NNG;가격/NNG;억제/NNG	이미/MAG;침체/NNG
경기/NNG;과열/NNG;억제/NNG	길/VA;침체/NNG
과열/NNG;조짐/NNG	침체/NNG;빠지/VV;우려/NNG
인플레이션/NNG;긴축/NNG	침체국면/NNG;빠지/VV
경기/NNG;과열/NNG;막/VV	이미/MAG;경기/NNG;침체/NNG
경제/NNG;과열/NNG	침체/NNG;최악/NNG
긴축/NNG;압력/NNG	경제/NNG;침체/NNG;빠지/VV
과열/NNG;방지/NNG	침체/NNG;늪/NNG

3.4.3. Evaluation

We evaluate the accuracy of our lexicon classification in several ways. In principle, the criteria of judging the accuracy is how well the classification of sentiment agrees with human judgments.[34] In addition, we use the typical set of metrics to measure the performance: Accuracy, Recall, Precision, and F1 score, which we will explain. Finally, we compare the performance of our lexicons with Korean Sentiment Analysis Corpus (KOSAC).[35]

First, we compare the performance of our indicators by using the documents not used in building our lexicons. Documents for evaluation are introductory statements from the BOK Governor's news conference about monetary policy decisions. With the documents from May 2009 to January 2018, we manually label 2,341 sentences as hawkish, neutral, and dovish. To check the consistency of our classification, we train an NBC with randomly selected 60% of hawkish and dovish sentences and test the remaining sentences. With 30 times of iteration, the average accuracy of classifiers is approximately 86%, which we think is above par accuracy.

[34] According to research by Amazon's Mechanical Turk, human raters typically only agree approximately 79% of the time (https://mashable.com/2010/04/19/sentiment-analysis/#skdJb2rbx5qg). In this regard, its reference to this 79% as a benchmark would be good when evaluating the raw accuracy of sentiment analysis.

[35] http://word.snu.ac.kr/kosac

Second, we check the accuracy of our lexicons by using labeled sentences that are completely out-of-sample. We use the popular metrics for comparison. Accuracy is the most intuitive performance measure, which is simply a ratio of correctly predicted observation to the total observations:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{False Positives} + \text{True Negatives} + \text{False Negatives}}, \quad (2)$$

where the predicted “Positive” and “Negative” refer to a model’s prediction, and the terms “True” and “False” refer to whether the prediction corresponds to the actual value. Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. Precision is an informative measure when the cost related to False Positive is high:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}. \quad (3)$$

Recall is the ratio of correctly predicted positive observations to all observations of actual Positives:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}. \quad (4)$$

F1 score is the weighted average of Precision and Recall. Therefore, this score considers False Positives and False Negatives.

$$\text{F1 score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}. \quad (5)$$

Although F1 does not seem to be intuitively straightforward, it is more useful than Accuracy in case of an uneven class distribution.

For the lexicon generated by market approach, the accuracy is 68% (positive precision: 63%, positive recall: 75%, positive F1: 68%, negative precision: 74%, negative recall: 62%, negative F1: 67%). For the lexicons generated by lexical approach, the accuracy is 67% (positive precision: 69%, positive recall: 71%, positive F1: 70%, negative precision: 65%, negative recall: 62%, negative F1: 63%).

Finally, to contextualize the numbers, we compare the performance of our lexicons with KOSAC. Notably, KOSAC is a general-purpose Korean sentiment dictionary, whereas we use the field-specific dictionary.[36] In comparison, the result of KOSAC shows relatively poor performance: the accuracy is only 53% (positive precision: 71%, positive recall: 57%, positive F1: 63%, negative precision: 29%, negative recall: 43%, negative F1: 35%), which is lower than 68% and 67% of ours.

3.5. Measuring Sentiments

With the lexicons in hand, the last step is to measure the tone of our target documents. We adopt a two-step approach to measure the tone of documents. First, we calculate the tone of a sentence based on the number of hawkish and dovish features (n-grams) in each sentence. Specifically, the tone of a sentence s is defined by the following formula:

$$tone_s = \frac{No.\ of\ hawkish\ features - No.\ of\ dovish\ features}{No.\ of\ hawkish\ features + No.\ of\ dovish\ features}. \tag{6}$$

Then, we calculate the tone of a document i using the following formula:

$$tone_i = \frac{No.\ of\ hawkish\ tone_{s,i} - No.\ of\ dovish\ tones_{s,i}}{No.\ of\ hawkish\ tone_{s,i} + No.\ of\ dovish\ tone_{s,i}}. \tag{7}$$

It creates a continuous variable $tone_i$ for each document, which is bound between −1 (dovish) and +1 (hawkish).[37] This index is our indicator of quantifying the sentiment of monetary policy. We denote the indicator based on market and lexical approaches by $tone^{mkt}$ and $tone^{lex}$, respectively. We examine their statistical properties and explanatory power of current and future monetary policy decisions.

IV. Empirical Analysis

We attempt to answer the following questions:

1. Can our lexicon-based indicators ($tone^{mkt}$ and $tone^{lex}$) explain the BOK's

[36] Given the lack of research conducted in Korean language in this field, it is the only one we can use for comparison.

[37] This way of measuring tones is similar to that of diffusion index in that it only considers direction. Kennedy (1994) finds that the diffusion indexes for industrial production and employment have predictive power in a twenty-five-year sample beginning in 1967. From the perspective of signal-extraction, by ignoring the magnitudes, this way of calculation insulates it from idiosyncratic shocks and offers a clean view of the persistent component.

current and future monetary policy decisions? In particular, do these indicators have additional information that are unavailable in the existing macroeconomic data?
2. Is it important to use a field-specific dictionary?
3. Is it important to use the original Korean text, not Korean-to-English text?

For the first question, our answer is an astounding "yes." We consider an augmented Taylor rule to compare the explanatory powers of our indicators with other macroeconomic variables. Our indicators have additional explanatory power. For the second and third questions, we recommend sticking to the original Korean text and using a dictionary specific to economics and finance terminologies.

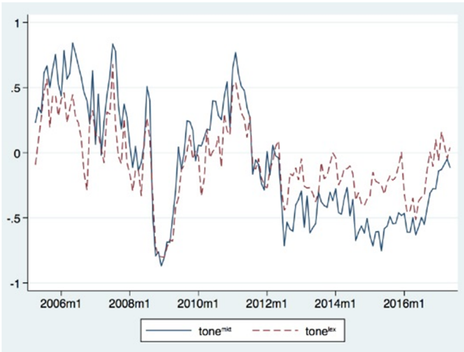
4.1. Measures of MP Sentiment

On the basis of our methodology, we develop lexicon-based indicators that capture the sentiment (or tone) of the BOK MPB's minutes: $tone^{mkt}$ and $tone^{lex}$. The former uses the market approach, whereas the latter uses lexical approach.

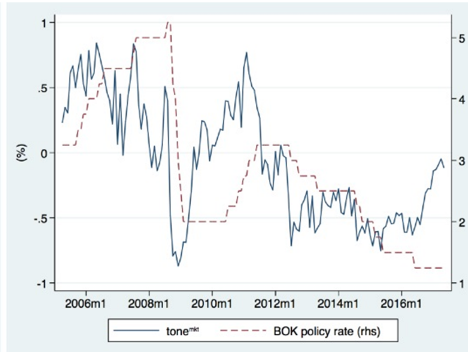
Figure 4 shows the time-series of $tone_t^{mkt}$ and $tone_t^{lex}$ with those of the BOK policy rate and other measures of economic uncertainty. Panel (a) in Figure 4 shows the time-series of $tone_t^{mkt}$ and $tone_t^{lex}$. They move closely with each other. The correlation coefficient between the two indicators is 0.85. This co-movement is interesting, given that these two indicators are constructed differently. Panel (b) shows that our indicator $tone^{mkt}$ captures the movements of the BOK policy rate. Panels (c) and (d) compare our indicator with other measures, such as economic policy uncertainty index by Baker et al. (2016) and uncertainty index by Jurado et al. (2015).[38] Given that uncertainty tends to rise in economic downturns, during which the central bank becomes dovish, we expect the negative correlation between $tone^{mkt}$ and measures of uncertainty. Panel (c) shows that, except for the period of the recent financial crisis, our indicator and economic policy uncertainty index do not seem to move in opposite directions. The correlation coefficient between the two is only −0.06. By contrast, Panel (d) shows that our indicator is negatively associated with uncertainty index. The correlation coefficient is −0.54.

[38] These uncertainty measures are known to be highly correlated with real economic activities as shown in Jurado et al. (2015) and Baker et al. (2016). For example, in the case of Jurado et al. (2015), hundreds of macroeconomic and financial indicators were considered to extract their uncertainty measure. Given that most of those variables are necessarily included in the information set of the central bank, expecting that their measure based on those indicators would be related to monetary policy decisions is natural. For measures of economic policy uncertainty and uncertainty, we use the Korean versions. As of August 2018, one can use economic policy uncertainty indexes of 25 countries at http://www.policyuncertainty.com. For the Korean version of Jurado et al. (2015), refer to Shin, Zhang, Zhong, and Lee (2018) and the journal website.

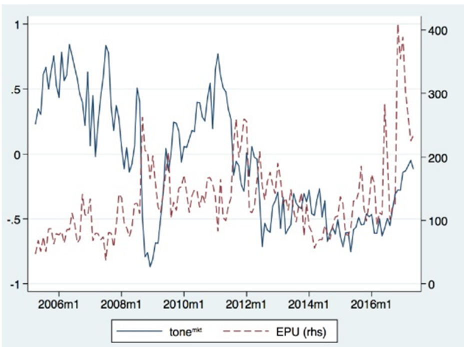
[Figure 4] MP sentiments, BOK policy rate, and other measures of uncertainty



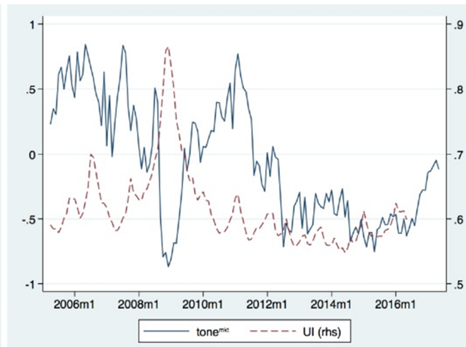
(a) $tone^{mkt}$ and $tone^{lex}$



(b) $tone^{mkt}$ and BOK policy rate



(c) $tone^{mkt}$ and EPU



(d) $tone^{mkt}$ and UI

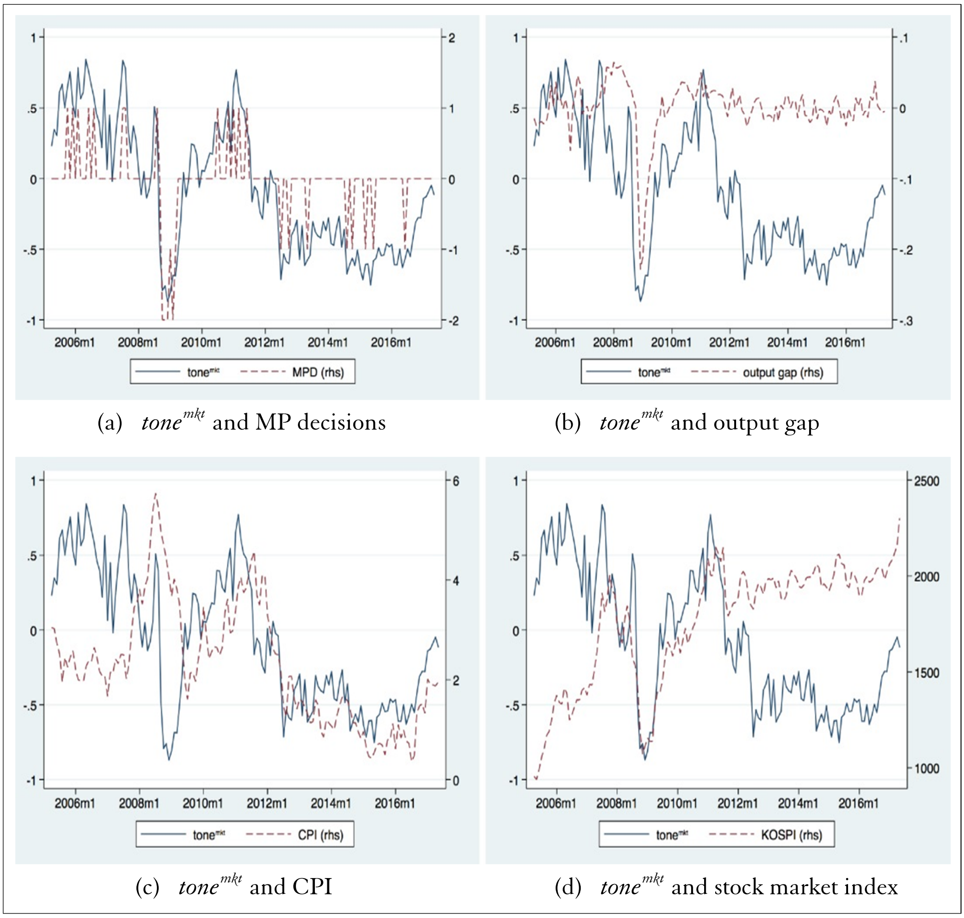
Notes: This figure shows the time series of our text-based indicators $tone^{mkt}$ and $tone^{lex}$ with the BOK policy rate, the Korean version of economic policy uncertainty measure by Baker et al. (2016), and the Korean version of macroeconomic uncertainty measure based on Jurado et al. (2015) and Shin et al. (2018).

Figure 5 compares our indicator $tone^{mkt}$ with monetary policy decisions (MPD_t), output gap ($y_t - y_t^*$), CPI (π_t), and stock market index. Following Picault and Renault (2017), to account for the BOK's non-standard monetary policy, MPD_t takes a value of zero when no change in monetary policy stance exists, +1 for a hawkish monetary policy decision (an increase of policy rate by 25 basis points), −1 for a dovish monetary policy decision, either through a policy rate cut or a non-standard policy, and −2 for a very dovish decision with a policy rate cut and a non-standard policy.[39] Output gap ($y_t - y^*$) is defined as the difference between the industrial production and its trend from Hodrick-Prescott filter. Panel (a) clearly shows that our indicator tracks the changes in monetary policy stance that considers BOK's non-standard policy. Panels (b) to (d) shows the time-series of output gap,

[39] Appendix A shows the chronological list of the BOK's non-standard monetary policy.

CPI, and stock market index (KOSPI). Correlation coefficient of $tone^{mkt}$ with monetary policy decision, output gap, CPI, and KOSPI are 0.58, 0.40, 0.40, and −0.36, respectively.

[Figure 5] MP sentiment and macroeconomic variables



(a) $tone^{mkt}$ and MP decisions

(b) $tone^{mkt}$ and output gap

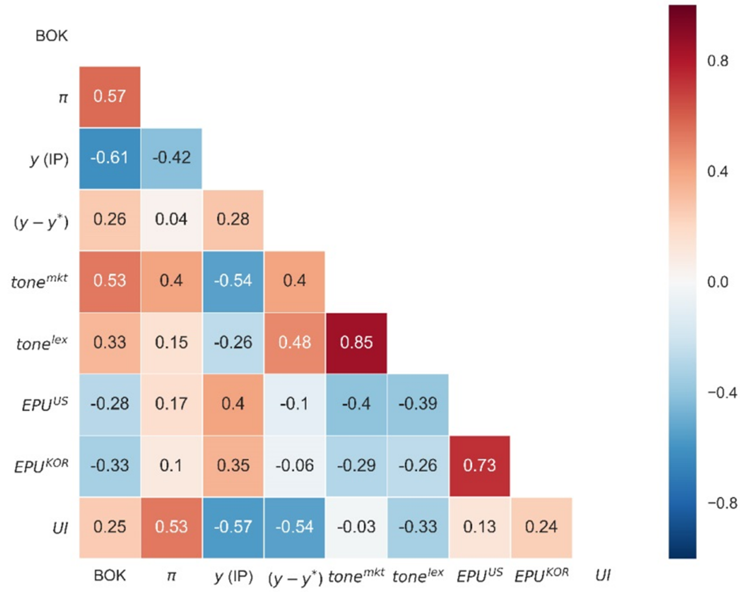
(c) $tone^{mkt}$ and CPI

(d) $tone^{mkt}$ and stock market index

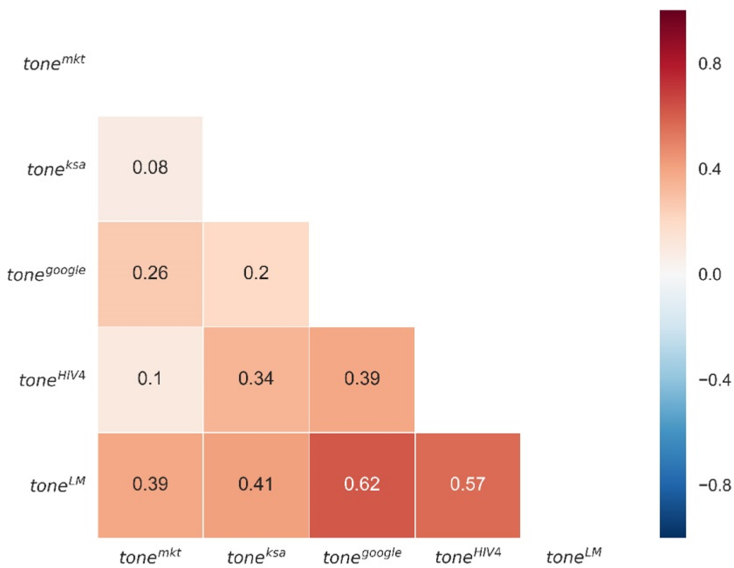
Notes: Panels (a) to (d) show the time series of $tone^{mkt}$, monetary policy decision MPD_t, output gap, CPI, and stock market index (KOSPI).

Panel (a) in Figure 6 shows the correlation coefficients among the selected variables. The BOK policy rate and industrial production have a strong negative correlation. BOK policy rate is positively correlated with inflation (π) and $tone^{mkt}$, whereas it is negatively correlated with the economic policy uncertainty measure of the US and South Korea. We formally test the explanatory power of various indicators in an augmented Taylor rule specification.

[Figure 6] Correlation coefficients



(a) With macroeconomic variables



(b) With other text-based indicators

Notes: Panels (a) and (b) show the correlation coefficients among key variables. Among the correlation coefficients with $tone_t^{mkt}$, $Corr(tone_t^{mkt}, UI_t)$ $Corr(tone_t^{mkt}, tone_t^{ksa})$, and $Corr(tone_t^{mkt}, tone_t^{HIV4})$ are not statistically significant at 10% level.

4.2. Explaining BOK's Monetary Policy Decisions

To assess the relation between the BOK MPB's policy rate decisions and the information content of MPB minutes, we test the explanatory power of our lexicon-based indicators ($tone_t^{mkt}$ and $tone_t^{lex}$) on contemporaneous and future decisions. We basically consider three kinds of specifications: a typical specification expressed in level and two kinds of differenced forms used in Picault and Renault (2017) and Apel and Grimaldi (2014).

First, we consider the following specification:

$$MP_t = \alpha + \rho MP_{t-1} + \gamma_1(\pi_t - \pi^*) + \gamma_2(y_t - y_t^*) + \gamma_3 tone_t + \gamma_4 X_t + \varepsilon_t\,, \tag{8}$$

where MP_t is the BOK's monetary policy stance and $(\pi_t - \pi^*)$ is the inflation gap defined as the difference between CPI and the inflation target (π^*), which is set to 3% before 2016 and 2% thereafter. $(y_t - y^*)$ is the output gap, where y_t is the monthly industrial production index and y^* is the trend component obtained from Hodrik-Precott filter. $tone_t$ is our lexicon-based indicator ($tone_t^{mkt}$ and $tone_t^{lex}$). Considering the possibility that the BOK may refer to the movement of exchange rate for its monetary policy decision, X_t is a variable related to exchange rates. We consider "exchange rate gap $((e_t - e^*)$ or $(e_t - e^*)/e_t)$," where e_t is the KRW/USD monthly exchange and e^* is the trend component obtained from Hodrik-Precott filter, and a proxy of exchange rate volatility $(vol(e_t))$, with which we use the rolling standard deviation for the past twelve months.

Table 7 shows the result of OLS estimation of (8) with robust standard errors when we use the BOK base rate (BOK) for monetary policy stance (MP). Our indicators provide ample information even after accounting for other macroeconomic variables.[40] Although Column (1) shows that inflation and output gaps are statistically significant, the explanatory power of inflation gap vanishes as we include the lagged dependent variable in Column (2). Throughout Columns (2) to (7), inflation gap is statistically insignificant as long as MP_{t-1} is present as an additional regressor. Notably, despite the presence of MP_{t-1}, our lexicon-based indicators of $tone_t^{mkt}$ and $tone_t^{lex}$ are highly significant with p-values less than 0.001 in all specifications. When we include a variable related to exchange rate in Columns (6) and (7), output gap is no longer statistically significant. Although exchange rate gap or its volatility are not highly significant, it seems that they take away some of the explanatory power of output gap. Moreover, our estimation result is notably consistent with those of previous studies, such as those of Shin (2015) and

[40] Given that uncertainty measures, such as EPU_t and UI_t necessarily contain information on the macroeconomy, they can be related to monetary policy decisions. We also estimate Equation (8) by adding these variables and find that they are statistically insignificant.

Kim and Kwark (2016): The BOK is highly concerned about output gap than with inflation gap. The policy rate is highly persistent; thus, including the lagged dependent variable noticeably lowers the magnitudes and statistical significance of macro variables.

[Table 7] OLS Estimation of Taylor rule

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Dependent variable: BOK_t			
BOK_{t-1}		0.983***	0.960***	0.973***	0.944***	0.957***	0.944***
		(0.013)	(0.015)	(0.013)	(0.026)	(0.018)	(0.027)
$(\pi_t - \pi^*)$	0.458***	−0.005	−0.019	−0.006	−0.005	−0.014	−0.006
	(0.054)	(0.012)	(0.013)	(0.012)	(0.012)	(0.011)	(0.012)
$(y_t - y^*)$	6.579***	2.114***	1.432**	1.354**	1.059	1.208	1.146
	(1.640)	(0.469)	(0.436)	(0.476)	(0.595)	(0.697)	(0.693)
$tone_t^{mkt}$			0.186***		0.173***	0.188***	0.170***
			(0.047)		(0.035)	(0.049)	(0.037)
$tone_t^{lex}$				0.241***			
				(0.069)			
$\frac{(e_t - e^*)}{e_t}$					−0.005		−0.005
					(0.004)		(0.004)
$vol(e_t)$						−0.000	0.000
						(0.001)	(0.001)
Constant	3.144***	0.033	0.103**	0.078*	0.152*	0.131	0.142
	(0.097)	(0.032)	(0.039)	(0.037)	(0.072)	(0.076)	(0.078)
N	145	144	143	143	143	143	143
R^2	0.28	0.99	0.99	0.99	0.99	0.99	0.99

Notes: This table shows the result of estimating Equation (8) with OLS. BOK is the policy rate set by the BOK. $(\pi_t - \pi^*)$ and $(y_t - y^*)$ are inflation and output gaps, respectively. $tone^{mkt}$ and $tone^{lex}$ are our lexicon-based indicators. $(e_t - e^*)/e_t$ is exchange rate gap, and $vol(e_t)$ is the rolling standard deviation of KRW/USD exchange rate with the window of $[t-12, t]$ and used as a proxy for exchange rate volatility. The numbers in parentheses are robust standard errors. *, **, and *** denote p-value < 0.05, p-value < 0.01, and p-value < 0.001, respectively.

Second, we consider the same specification used in Picault and Renault (2017). They use an ordered probit model to estimate the coefficients of the forward-looking Taylor rule and compare the explanatory powers of their own lexicon-based indicators with those of macroeconomic variables and other types of indicators. For estimation, to ensure stationarity, they estimate the differenced version of Equation (8)[41]:

[41] Using the augmented Dickey-Fuller test, we find that policy rate (BOK_t), inflation gap $(\pi_t - \pi^*)$, output gap $(y_t - y_t^*)$, expected inflation (π_t^e), output leading indicator (y_t^e), and uncertainty index (UI_t) have a unit root; whereas $tone_t^{mkt}$, $tone_t^{lex}$ and EPU_t turn out to be

$$\Delta MP_t = \rho \Delta MP_{t-1} + \gamma_1 \Delta(\pi_t - \pi^*) + \gamma_2 \Delta(y_y - y_t^*)$$
$$+ \gamma_3 \Delta \pi_t^e + \gamma_4 \Delta y_t^e + \gamma_5 tone_t + u_t \,, \tag{9}$$

where π_t^e is the expected inflation of Consumer Survey Index and y_t^e is the leading Economic Composite Index.[42] For X_t, we consider our lexicon-based indicators ($tone_t^{mkt}$ and $tone_t^{lex}$). For changes in monetary policy stance (ΔMP_t), we use two variables following Picault and Renault (2017). First, we focus on changes in policy rate (ΔBOK_t). During the sample period, the BOK MPB increases its policy rate by 25 basis points on thirteen occasions and decreases it on thirteen occasions (nine times by 25 basis points, twice by 50 basis points, and twice by 100 basis points).[43] Thus ΔBOK_t takes values of −1.0, −0.5, −0.25, 0, and +0.25 in our sample. The other is a variable of monetary policy decision (MPD_t) to account for the BOK's non-standard monetary policy. In a forward-looking approach, Equation (9) is rewritten as

$$\Delta MP_{t+k} = \rho \Delta MP_t + \gamma_1 \Delta(\pi_t - \pi^*) + \gamma_2 \Delta(y_t - y_t^*)$$
$$+ \gamma_3 \Delta \pi_t^e + \gamma_4 \Delta y_t^e + \beta tone_t + u_t \,. \tag{10}$$

We use $k = 1$ and $k = 2$ and report the case of $k = 2$ for brevity.

If the MPB minutes do provide any additional information to previously released macroeconomic data, then, our indicators of $tone^{mkt}$ and $tone^{lex}$ should be significant in Equations (9) and (10). We also expect a positive coefficient: highly hawkish (dovish) sentiment should be associated with tight (loose) monetary policy. Table 8 shows the estimation result when we measure the change in monetary policy stance (ΔMP_t) as the change in policy rate (ΔBOK_t). Our indicators $tone^{mkt}$ and $tone^{lex}$ are highly significant. Moreover, adding $tone^{mkt}$ or $tone^{lex}$ noticeably raises the value of R2. Comparing Columns (2) and (3), adding $tone^{mkt}$ raises R^2 from 0.095 to 0.446. When the dependent variable is ΔBOK_{t+2}, we observe similar results. The result from Table 8 strongly suggests that our lexicon-based indicators contain additional information that are not captured by

---

stationary.

[42] Statistics Korea (http://kostat.go.kr) reports three kinds of Economic Composite Index, namely, leading, coincident, and lagging. The leading index is based on nine indicators that are closely related to the following future economic activities: construction orders received, opening-to-application ratio, inventory circulation indicator, consumer expectation index, machinery orders received, import of capital goods, stock price index, total liquidity, interest rate spread, three-year treasury bonds less call rate, and net barter terms of trade.

[43] Given that we use the monthly data for our empirical analysis, we measure the changes in policy rate on a monthly basis, not on each meeting. For example, a regular MPB meeting on October 9, 2008 lowered the rate by 25 basis points and an emergency MPB meeting on October 27, 2008 lowered the rate by 75 basis points. We record $\Delta BOK_t = -100$ basis points in October, 2008.

macroeconomic data.[44]

Table 9 shows the result when the dependent variable is MPD_t that considers BOK's non-standard policy. We obtain similar results. $tone^{mkt}$ and $tone^{lex}$ are highly significant, considerably raising R^2, whereas other measures are not.[45]

Third, we also estimate the different specification used in Apel and Grimaldi (2014) by using ordered probit:

$$\Delta MP_{t+1} = \alpha_1 \Delta MP_t + \alpha_2 X_t + \alpha_3 GDPgrowth_t + \alpha_4 Inflation_t + \varepsilon_{t+1}.$$

If macroeconomic variables contain all the relevant information for the next policy rate, then, the estimate of α_2 would be largely insignificant. We estimate the same regression equation with or without $tone_t^{mkt}$ after replacing GDP growth rate with IP growth rate for monthly estimation. We obtain the following result. The numbers in parentheses are standard errors.

$$\widehat{\Delta MP}_{t+1} = \underset{(0.65)}{1.90}\Delta MP_t + \underset{(4.64)}{7.28} IPgrowth_t + \underset{(0.10)}{0.12} CPI_t, \text{ pseudo } R^2 = 0.08.$$

With $tone_t^{mkt}$, we obtain

$$\widehat{\Delta MP}_{t+1} = \underset{(0.90)}{-1.67}\Delta MP_t + \underset{(0.86)}{4.20} tone_t^{mkt} + \underset{(5.33)}{9.73} IPgrowth_t - \underset{(0.14)}{0.28} CPI_t,$$

pseudo $R^2 = 0.37$.

Again, our indicator $tone_t^{mkt}$ is highly significant and raises R^2 from 0.08 to 0.37.

---

44 We also run a horse race by adding all three variables of $tone^{mkt}$, EPU_t, and UI_t. The estimation result is as follows (the numbers in parentheses are standard errors):

$$\widehat{\Delta MP}_{t+2} = \underset{(0.88)}{-0.06}\Delta MP_t + \underset{(0.39)}{0.20}\Delta(\pi_t - \pi^*) + \underset{(5.24)}{5.15}\Delta(y_t - y_t^*) + \underset{(0.97)}{0.85}\Delta \pi_t^e$$
$$+\underset{(0.49)}{0.11}\Delta y_t^e + \underset{(0.41)}{1.58} tone_t^{mkt} + \underset{(0.003)}{0.001} EPU_t - \underset{(2.36)}{4.720} UI_t, \text{ pseudo } R^2 = 0.22.$$

Notably, $tone^{mkt}$ remains highly significant, and R^2 does not increase subtantially with the addition of EPU_t and UI_t.
When we use $tone^{lex}$ instead of $tone^{mkt}$, we obtain a similar result:

$$\widehat{\Delta MP}_{t+2} = \underset{(0.84)}{-0.59}\Delta MP_t + \underset{(0.38)}{0.16}\Delta(\pi_t - \pi^*) + \underset{(5.18)}{6.47}\Delta(y_t - y_t^*) + \underset{(0.96)}{0.58}\Delta \pi_t^e$$
$$+\underset{(0.47)}{0.01}\Delta y_t^e + \underset{(0.57)}{1.92} tone_t^{mkt} + \underset{(0.00)}{0.003} EPU_t - \underset{(2.23)}{0.950} UI_t, \text{ pseudo } R^2 = 0.19.$$

45 We also run the same specification by OLS and obtain similar results.

This result strongly suggests that our indicator $tone_t^{mkt}$ contains relevant information on the future monetary policy stance beyond information in macroeconomic variables.

[Table 8] Ordered probit, changes in BOK policy rate

	(1)	(2)	(3)	(4)
	Dependent variable: ΔBOK_t			
ΔBOK_{t-1}	1.893**	1.790**	−0.209	−0.839
	(0.622)	(0.632)	(0.736)	(0.797)
$\Delta(\pi_t - \pi^*)$	0.142	0.0163	−0.364	−0.490
	(0.331)	(0.341)	(0.517)	(0.431)
$\Delta(y_t - y^*)$	7.068	5.614	6.025	8.351
	(4.362)	(4.634)	(5.298)	(5.160)
$\Delta\pi_t^e$		1.734	1.553	1.635
		(0.910)	(1.262)	(1.107)
Δy_t^e		0.322	0.153	0.0661
		(0.450)	(0.637)	(0.536)
$tone_t^{mkt}$			5.327***	
			(1.114)	
$tone_t^{lex}$				4.515***
				(0.797)
N	143	143	143	143
pseudo R^2	0.076	0.095	0.446	0.364
	Dependent variable: ΔBOK_{t+2}			
ΔBOK_{t-1}	2.406***	2.339***	0.773	0.692
	(0.671)	(0.678)	(0.777)	(0.797)
$\Delta(\pi_t - \pi^*)$	0.359	0.326	0.290	0.174
	(0.338)	(0.343)	(0.373)	(0.367)
$\Delta(y_t - y^*)$	5.521	5.190	6.167	6.659
	(4.721)	(4.945)	(5.127)	(5.073)
$\Delta\pi_t^e$		0.629	0.607	0.536
		(0.901)	(0.963)	(0.946)
Δy_t^e		0.0952	0.160	0.00654
		(0.449)	(0.482)	(0.468)
$tone_t^{mkt}$			1.406***	
			(0.383)	
$tone_t^{lex}$				1.970***
				(0.542)
N	142	142	142	142
pseudo R^2	0.110	0.113	0.203	0.189

Notes: This table displays the ordered probit estimation result of Equations (9) and (10) when changes in policy rate ΔBOK are used as changes in monetary policy stance ΔMP. *, **, and *** denote p-value < 0.05, p-value < 0.01, and p-value < 0.001, respectively.

[Table 9] Ordered probit, changes in MP stance

	(1)	(2)	(3)	(4)
		Dependent variable: MPD_t		
MPD_{t-1}	0.759***	0.748***	−0.0467	−0.275
	(0.215)	(0.216)	(0.267)	(0.292)
$\Delta(\pi_t - \pi^*)$	0.109	0.0166	−0.336	−0.513
	(0.331)	(0.340)	(0.512)	(0.433)
$\Delta(y_t - y^*)$	7.627	6.101	8.136	11.55*
	(4.634)	(4.917)	(5.603)	(5.705)
$\Delta\pi_t^e$		1.393	0.959	1.220
		(0.894)	(1.218)	(1.086)
Δy_t^e		0.355	−0.0585	−0.183
		(0.441)	(0.621)	(0.531)
$tone_t^{mkt}$			5.464***	
			(1.122)	
$tone_t^{lex}$				4.900***
				(0.853)
N	143	143	143	143
pseudo R^2	0.095	0.109	0.461	0.397
		Dependent variable: MPD_{t+2}		
MPD_{t-1}	0.800***	0.780***	0.131	0.0958
	(0.215)	(0.217)	(0.272)	(0.278)
$\Delta(\pi_t - \pi^*)$	0.468	0.444	0.409	0.261
	(0.338)	(0.343)	(0.375)	(0.369)
$\Delta(y_t - y^*)$	4.871	4.301	6.145	6.981
	(4.746)	(4.982)	(5.243)	(5.220)
$\Delta\pi_t^e$		0.512	0.510	0.413
		(0.902)	(0.970)	(0.951)
Δy_t^e		0.169	0.235	0.0406
		(0.445)	(0.479)	(0.463)
$tone_t^{mkt}$			1.585***	
			(0.418)	
$tone_t^{lex}$				2.258***
				(0.581)
N	142	142	142	142
pseudo R^2	0.114	0.116	0.215	0.205

Notes: This table displays the ordered probit estimation result of Equations (9) and (10) when the variable of monetary policy decision MPD is used as changes in monetary policy stance ΔMP. *, **, and *** denote p-value < 0.05, p-value < 0.01, and p-value < 0.001, respectively.

4.3. Comparison with Other Text-Based Indicators

One may wonder if the original Korean text should be used. What if one translates Korean text into English and applies the standard procedures for English text? It is a legitimate question given the availability of more advanced and diverse text mining techniques for English text. Moreover, is it really important to use a field-specific dictionary? To answer these questions, we compare our indicators with four other indicators: an indicator that specializes in Korean texts and uses a general-purpose dictionary ($tone^{ksa}$), and three English-based indicators ($tone^{google}$, $tone^{HIV4}$, and $tone^{LM}$). $tone^{ksa}$ is based on Kind Korean Morpheme Analyzer (KKMA) project developed by Seoul National University Intelligent Data Systems Laboratory; KKMA is one of the most popular tools for analyzing Korean texts.[46] However, a general-purpose dictionary was used. For English-based text analysis, we translate all the MPB's minutes into English by using Google Cloud Translation.[47] $tone^{google}$ measures the tone of minutes by using the service of sentiment analysis provided by Google Cloud Natural Language.[48] $tone^{HIV4}$ is based on the general-purpose Harvard IV-4 dictionary, and $tone^{LM}$ is based on the field-specific dictionary of Loughran and McDonald (2011).

Figure 7 shows the time-series of $tone^{ksa}$, $tone^{google}$, $tone^{HIV4}$, and $tone^{LM}$ with that of $tone^{mkt}$. Panel (a) shows that $tone^{ksa}$ has considerably less variations than with $tone^{mkt}$. $tone^{ksa}$ does not fluctuate substantially even during the recent financial crisis. Although English-based indicators have more variations than with $tone^{ksa}$, the degree of co-movement with $tone^{mkt}$ seems rather weak. Panel (b) in Table 6 shows the correlation coefficients. The correlation coefficients of $tone^{mkt}$ with $tone^{ksa}$, $tone^{google}$, $tone^{HIV4}$, and $tone^{LM}$ are 0.08, 0.26, 0.10, and 0.39, respectively. Although $tone^{mkt}$ and $tone^{ksa}$ are based on the original Korean texts, the statistical association is extremely low.

We compare the performance of these indicators based on Equations (9) and (10) when performance is measured by the statistical significance of an individual coefficient and R^2. Our conjecture is

(i) $tone^{mkt}$ outperforms $tone^{ksa}$.

(ii) $tone^{LM}$ outperforms $tone^{google}$ and $tone^{HIV4}$.

(iii) $tone^{mkt}$ outperforms $tone^{LM}$, $tone^{google}$, and $tone^{HIV4}$.

Roughly, (i) and (ii) are aimed at testing the benefit of using a field-specific dictionary, and (iii) aims to compare the results from Korean and Korean-to-English texts.

Table 10 shows the estimation result. When the dependent variable is the current change in the BOK policy rate (ΔBOK_t), Columns (1) and (2) show that $tone^{ksa}$ fails to capture the change in BOK policy rate although it is based on the most

---

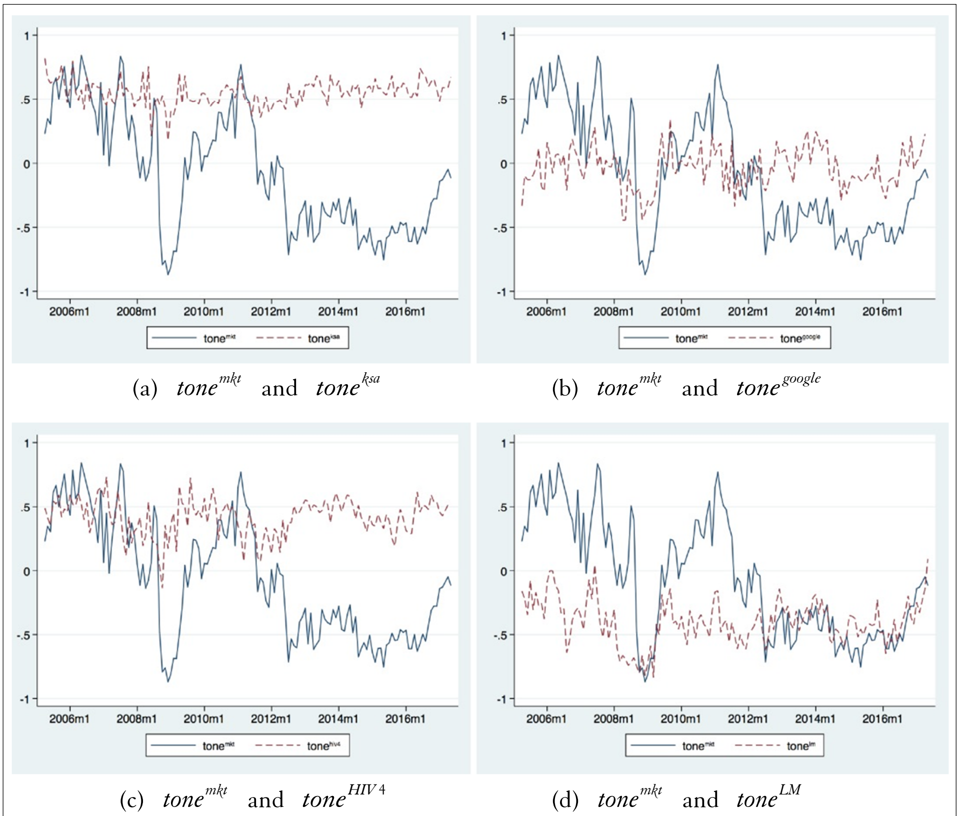
[46] See Lee, Yeon, Hwang, and Lee (2010) for additional detail.

[47] https://cloud.google.com/translate/.

[48] https://cloud.google.com/natural-language/. Google Cloud Translation and Google Cloud Natural Language are part of Google's AI and Machine Learning products.

popular tool for the Korean text analysis. Columns (3) to (5) show that, whereas $tone^{google}$ and $tone^{HIV4}$ are statistically insignificant, $tone^{LM}$ is statistically significant at 10% level. These results suggest that its use of a field-specific dictionary is important. When we compare $tone^{mkt}$ (Column [1]) with $tone^{LM}$ (Column [5]), R^2 in Column (1) is far higher than R^2 in Column (5), attesting the importance of the original Korean text. When the dependent variable is the future change in BOK policy rate ΔBOK_{t+2}, $tone^{mkt}$ outperforms other indicators in terms of its statistical significance and R^2. Considering our discussion based on Figure 7 and Table 10, using the original Korean text with a field-specific dictionary is desirable. The validity of our approach to quantify the information in the MPB minutes is confirmed.

[Figure 7] MP sentiment and other text-based indicators



(a) $tone^{mkt}$ and $tone^{ksa}$ (b) $tone^{mkt}$ and $tone^{google}$

(c) $tone^{mkt}$ and $tone^{HIV4}$ (d) $tone^{mkt}$ and $tone^{LM}$

Notes: Panels (a) to (d) show the time-series of text-based indicators. Correlation coefficient of $tone^{mkt}$ with $tone^{ksa}$, $tone^{google}$, $tone^{HIV4}$, and $tone^{LM}$ are 0.08, 0.26, 0.10, and 0.40, respectively.

[Table 10] Comparison of text-based indicators

	(1)	(2)	(3)	(4)	(5)
			Dependent variable: ΔBOK_t		
ΔBOK_{t-1}	−0.209	1.730**	1.440*	1.804**	1.422*
	(0.736)	(0.643)	(0.664)	(0.633)	(0.654)
$\Delta(\pi_t - \pi^*)$	−0.364	0.0230	0.0251	−0.0317	−0.121
	(0.517)	(0.341)	(0.344)	(0.352)	(0.352)
$\Delta(y_t - y^*)$	6.025	5.640	5.378	5.671	6.691
	(5.298)	(4.640)	(4.650)	(4.637)	(4.711)
$\Delta\pi_t^e$	1.553	1.658	1.693	1.867*	1.964*
	(1.262)	(0.923)	(0.918)	(0.940)	(0.925)
Δy_t^e	0.153	0.263	0.0715	0.237	−0.163
	(0.637)	(0.465)	(0.474)	(0.474)	(0.497)
$tone_t^{mkt}$	5.327***				
	(1.114)				
$tone_t^{ksa}$		0.576			
		(1.184)			
$tone_t^{google}$			1.458		
			(0.843)		
$tone_t^{HIV4}$				0.492	
				(0.866)	
$tone_t^{LM}$					1.890*
					(0.783)
pseudo R^2	0.446	0.097	0.111	0.097	0.127
			Dependent variable: ΔBOK_{t+2}		
ΔBOK_{t-1}	0.773	2.502***	2.158**	2.102**	1.770*
	(0.777)	(0.700)	(0.686)	(0.705)	(0.709)
$\Delta(\pi_t - \pi^*)$	0.290	0.334	0.325	0.127	0.194
	(0.373)	(0.344)	(0.345)	(0.356)	(0.353)
$\Delta(y_t - y^*)$	6.167	5.419	4.702	5.937	6.766
	(5.127)	(4.981)	(4.964)	(5.048)	(5.064)
$\Delta\pi_t^e$	0.607	0.771	0.634	1.430	1.151
	(0.963)	(0.915)	(0.907)	(0.969)	(0.940)
Δy_t^e	0.160	0.232	−0.106	−0.212	−0.441
	(0.482)	(0.467)	(0.473)	(0.472)	(0.498)
$tone_t^{mkt}$	1.406***				
	(0.383)				
$tone_t^{ksa}$		−1.337			
		(1.197)			
$tone_t^{google}$			1.183		
			(0.815)		
$tone_t^{HIV4}$				2.256*	
				(0.935)	
$tone_t^{LM}$					2.311**
					(0.820)
pseudo R^2	0.203	0.119	0.124	0.144	0.156

Notes: This table displays the ordered probit estimation result of Equations (9) and (10). *, **, and *** denote p-value < 0.05, p-value < 0.01, and p-value < 0.001, respectively.

V. Concluding Remarks

We develop text-based indicators that quantify the sentiment of monetary policy by using a field-specific Korean dictionary and n-grams. Our indicators help explain current and future monetary policy decisions and perform better compared with other indicators. Moreover, using a field-specific dictionary and the original Korean text is important, which would benefit future research in this field.

Our empirical results suggest several future research avenues. First, examining what kind of information our indicators have (or do not have) compared with the BOK policy rate and macroeconomic variables is important. If we interpret the BOK policy rate as a threshold or latent variable of our indicator of monetary policy sentiment, feeding our measure into the standard VAR systems or DSGE models that analyze the effect of monetary policy would be interesting. Alternatively, examining the implication of the discrepancy between our indicator and policy rate would be interesting. As shown in Panel (b) of Figure 4, our indicator $tone^{mkt}$ and the BOK policy rate started diverging from 2016. Whereas $tone^{mkt}$ becomes more hawkish, the policy rate does not. One can examine the information content in $tone^{mkt}$ orthogonal to the policy rate. Another direction would be that of Hansen and McMahon (2016). They construct two separate indicators on the state of economy and forward guidance from the Fed statements and use FAVAR to examine how these two dimensions of central bank communication affect financial market and real variables. Second, our measure can be used to evaluate the effectiveness of central bank communication, including forward guidance. Regardless of whether the BOK intends or not, our indicators based on the minutes help explain the future decision of monetary policy. In this line of research, Lee, Kim, and Park (2019) propose a new method of measuring monetary policy “surprises” by measuring the differences of media tones around the dates of MPB meetings. Given that no such thing as futures exists for policy rates in South Korea, a text-based indicator would be a decent alternative. One can also examine how successfully the central bank delivers its intention by comparing the changes in tones between central bank minutes and news articles. Third, our methodology can be easily applied to construct other indicators to measure macroeconomic uncertainty, public expectation about future monetary policy stance, and stock market sentiment. One can examine how changes in these measures affect asset prices or real variables.

Despite our acknowledgement of other efforts that must be conducted in this field, we hope that our study serves as a starting point as it demonstrates the following: text mining approach can be a useful addition to the BOK and researchers’ toolbox of analyzing monetary policy and achieving its objectives.

Appendices

A. BOK non-standard policy announcements

Date	Announcements
October 2008	Currency swap with the Fed loan guarantee
November 2008	Expansion of Aggregate Credit Ceiling Loans Expansion of the range of BOK's counterparties
December 2008	Expansion of the range of BOK's counterparties Interest on reserves Emergency liquidity assistance Currency swaps with China and Japan
February 2009	Currency swap extension with the Fed
March 2009	Expansion of Aggregate Credit Ceiling Loans
June 2009	Currency swap extension with the Fed

Source: Bank of Korea website (www.bok.or.kr).

B. eKoNLPy Tagset for POS Tagging

[Table A1] Tagset used in Mecab tagger of eKoNLPy

Tag	Name	Tag	Name
NNG	General Noun	JKQ	Case Postposition (Quotation)
NNP	Proper Noun	JC	Conjunctive Postposition
NNB	General Dependent Noun	JX	Auxiliary Postposition
NNBC	Unit Word	EP	Prefinal Ending
NR	Number Word	EF	Final Ending
NP	Pronoun	EC	Conjunctive Ending
VV	Verb	ETN	Nominal Ending
VA	Adjective	ETM	Adnominal Ending
VAX	Derived Adjective	XPN	Noun Prefix
VX	Auxiliary Predicate	XSN	Noun Suffix
VCP	Positive Copula	XSV	Verbalization Suffix
VCN	Negative Copula	XSA	Adjectivization Suffix
MM	Determiner	XR	Root Word
MAG	Adverb	SF	Sentence Ending Marker
MAJ	Conjunctive Adverb	SE	Ellipsis Symbol
IC	Exclamation	SSO	Left Quotation Mark
JKS	Case Postposition (Nominative)	SSC	Right Quotation Mark
JKC	Case Postposition (Complementive)	SC	Separator Symbol
JKG	Case Postposition (Determinative)	SY	Symbol
JKO	Case Postposition (Objective)	SH	Chinese Character
JKB	Adverbial Postposition	SL	Foreign Word
JKV	Case Postposition (Vocative)	SN	Number

References

Apel, M. and M. B. Grimaldi (2014), "How Informative Are Central Bank Minutes?" *Review of Economics*, 65, 655.

Baker, S. R., N. Bloom, and S. J. Davis (2016), "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 131, 1593–1636.

Bennani, H. and M. Neuenkirch (2016), "The (Home) Bias of European Central Bankers: New Evidence Based on Speeches," *Applied Economics*, 49, 1114–1131.

Bholat, D., J. Brookes, C. Cai, K. Grundy, and J. Lund (2017), "Sending Firm Messages: Text Mining Letters from PRA Supervisors to Banks and Building Societies They Regulate," *Bank of England Working Paper* No. 688.

Bholat, D., S. Hansen, P. Santos, and C. Schonhardt-Bailey (2015), "Text Mining for Central Banks," *Centre for Central Banking Studies Handbook*, 1–19.

Born, B., M. Ehrmann, and M. Frtzscher (2014), "Central Bank Communication on Financial Stability," *Economic Journal*, 124, 701–734.

Choi, H. and H. Varian (2012), "Predicting the Present with Google Trends," *Special Issue: Selected Papers from the 40th Australian Conference of Economists*, 88, 2–9.

Dey, A., M. Jenamani, and J. J. Thakkar (2018), "Senti-N-gram : An N-Gram Lexicon for Sentiment Analysis," *Expert Systems with Applications*, 103, 92–105.

Gentzkow, M., B. T. Kelly, and M. Taddy (2017), "Text As Data," *NBER Working Paper 23276*.

Hamilton, W. L., K. Clark, J. Leskovec, and D. Jurafsky (2016), "Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora," *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, 2016, 595–605.

Hansen, S. and M. McMahon (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.

Hutto, C. and E. Gilbert (2014), "Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," in *Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*, Vol. Eighth International Conference onWeblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech.edu/papers/icwsm14.vader.hutto.pdf.

Jegadeesh, N. and D. Wu (2013), "Word Power: A New Approach for Content Analysis," *Journal of financial economics*.

Jurado, K., S. C. Ludvigson, and S. Ng (2015), "Measuring Uncertainty," *American Economic Review*, 105, 1177–1216.

Kennedy, J. E. (1994), "The Information in Diffusion Indexes for Forecasting Related Economic Aggregates," *Economics Letters*, 44, 113–117.

Kim, W. H. and N.-S. Kwark (2016), "Estimation of the Monetary Policy Reaction Function in Korea Before and After the Global Financial Crisis," *Korean Journal of Economic Studies*, 64 [in Korea].

Lee, D.-J., J.-H. Yeon, I.-B. Hwang, and S.-G. Lee (2010), "Kkma: A Tool for Utilizing Sejong Corpus Based on Relational Database," *Journal of KIISE: Computing Practices and Letters*, 16, 1046–1050 [in Korean].

Lee, Y. J. (2018), "eKoNLPy: A Korean NLP Python Library for Economic Analysis," https://github.com/entelecheia/eKoNLPy.

Lee, Y. J., S. Kim, and K. Y. Park (2019), "Measuring Monetary Policy Surprises Using Text Mining: The Case of South Korea," working paper.

Liu, B. (2009), "Sentiment Analysis and Subjectivity," in *Handbook of Natural Language Processing*, New York, NY, USA: Marcel Dekker, Inc.

Loughran, T. and B. McDonald (2011), "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance*.

Lucca, D. and F. Trebbi (2011), "Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements," *Working Paper*.

McLaren, N. and R. Shanbhogue (2011), "Using Internet Search Data as Economic Indicators," *Bank of England Quarterly Bulletin*.

Meinusch, A. and P. Tillmann (2017), "Quantitative Easing and Tapering Uncertainty: Evidence from Twitter," *International Journal of Central Banking*.

Nopp, C. and A. Hanbury (2015), "Detecting Risks in the Banking System by Sentimental Analysis," *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*.

Nyman, R., S. Kapadia, D. Tuckett, D. Gregory, P. Ormerod, and R. Smith (2018), "News and Narratives in Financial Systems: Exploiting Big Data for Systemic Risk Assessment," *Bank of England Staff Working Paper* No. 704.

Picault, M. and T. Renault (2017), "Words Are Not All Created Equal: A New Measure of ECB Communication," *Journal of International Money and Finance*, 79, 136–156.

Porter, M. F. (1980), "An Algorithm for Suffix Stripping," *Program*, 14, 130–137.

Pyo, D.-J. and J. Kim (2017), "News Media Sentiment and Asset Prices: Text-mining Approach," *KIF Working Paper*.

Ribeiro, F. N., M. Araújo, P. Gonçalves, M. A. Gonçalves, and F. Benevenuto (2016), "Sentibench-a Benchmark Comparison of State-of-the-practice Sentiment Analysis Methods," *EPJ Data Science*, 5, 1–29.

Shin, K. (2015), "Inflation Targeting in South Korea: Experience and Evaluation," conference presentation, [in Korean], http://www.kea.ne.kr/common/download?id=153&section=conf.

Shin, M., B. Zhang, M. Zhong, and D. J. Lee (2018), "Measuring International Uncertainty: The Case of Korea," *Economics Letters*, 162, 22–26.

Söderlind, P. and L. Svensson (1997), "New Techniques to Extract Market Expectations from Financial Instruments," *Journal of Monetary Economics*, 40, 383–429.

Stone, P. J., D. C. Dunphy, and M. S. Smith (1966), *The General Inquirer: A Computer Approach to Content Analysis*, MIT Press.

Tetlock, P. C. (2007), "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *The Journal of finance*, 62, 1139–1168.

Tobback, E., S. Nardelli, and D. Martens (2016), "Between Hawks and Doves: Measuring

Central Bank Communication," 1–41.

Turney, P. D. (2002), "Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 417–424.

Warsh, K. (2014), "Transparency and the Bank of England's Monetary Policy Committee," *Review by Kevin Warsh*.

Won, J.-H., W. Son, H. Moon, and H. Lee (2017), "Text Mining Techniques for Classification of Economic Sentiment," *BOK Quarterly Bulletin*.

Zhao, Z., T. Liu, S. Li, B. Li, and X. Du (2017), "Ngram2vec: Learning Improved Word Representations from Ngram Co-occurrence Statistics," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA: Association for Computational Linguistics, 244–253.