

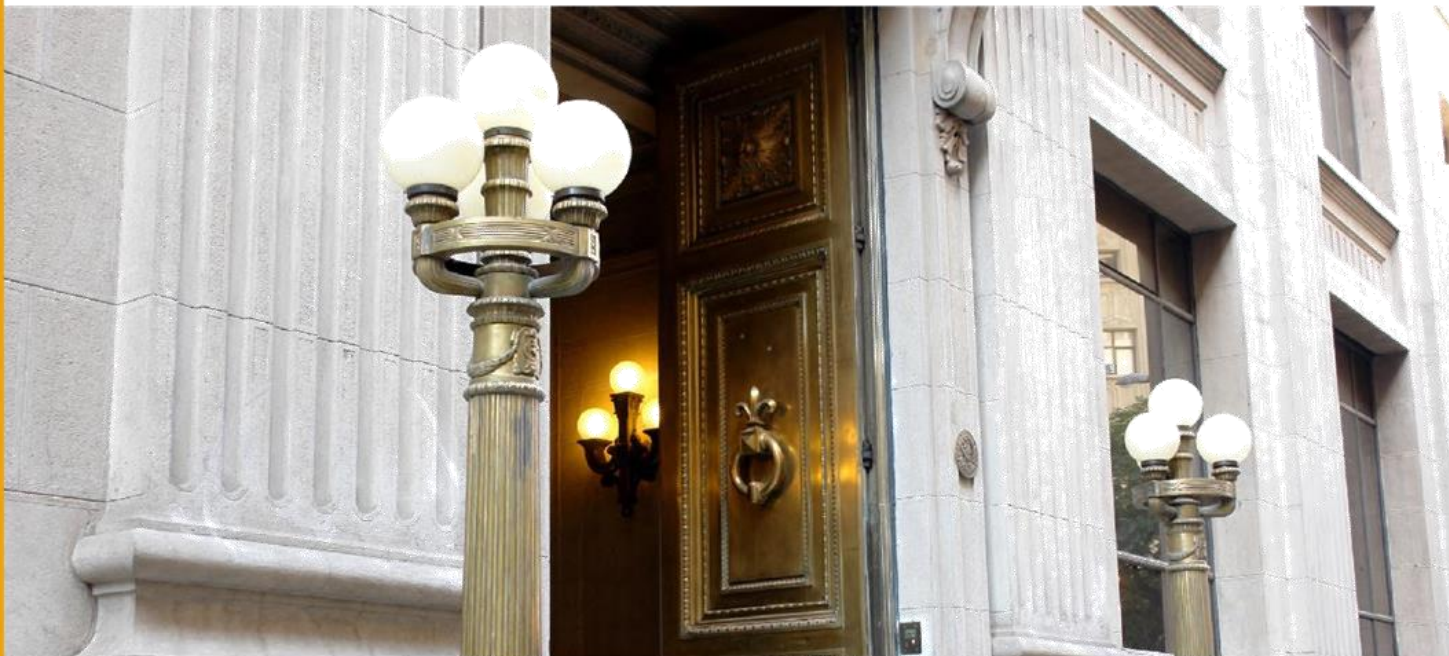
DOCUMENTOS DE TRABAJO

Utilización de noticias de prensa como indicador de confianza económica en tiempo real

María del Pilar Cruz N.
Hugo Peralta V.
Juan Pablo Cova M.

N° 938 Enero 2022

BANCO CENTRAL DE CHILE





La serie Documentos de Trabajo es una publicación del Banco Central de Chile que divulga los trabajos de investigación económica realizados por profesionales de esta institución o encargados por ella a terceros. El objetivo de la serie es aportar al debate temas relevantes y presentar nuevos enfoques en el análisis de los mismos. La difusión de los Documentos de Trabajo sólo intenta facilitar el intercambio de ideas y dar a conocer investigaciones, con carácter preliminar, para su discusión y comentarios.

La publicación de los Documentos de Trabajo no está sujeta a la aprobación previa de los miembros del Consejo del Banco Central de Chile. Tanto el contenido de los Documentos de Trabajo como también los análisis y conclusiones que de ellos se deriven, son de exclusiva responsabilidad de su o sus autores y no reflejan necesariamente la opinión del Banco Central de Chile o de sus Consejeros.

The Working Papers series of the Central Bank of Chile disseminates economic research conducted by Central Bank staff or third parties under the sponsorship of the Bank. The purpose of the series is to contribute to the discussion of relevant issues and develop new analytical or empirical approaches in their analyses. The only aim of the Working Papers is to disseminate preliminary research for its discussion and comments.

Publication of Working Papers is not subject to previous approval by the members of the Board of the Central Bank. The views and conclusions presented in the papers are exclusively those of the author(s) and do not necessarily reflect the position of the Central Bank of Chile or of the Board members.

Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Agustinas 1180, Santiago, Chile
Teléfono: (56-2) 3882475; Fax: (56-2) 38822311

USING THE PRESS AS A REAL-TIME ECONOMIC CONFIDENCE INDICATOR¹

María del Pilar Cruz N.
Banco Central de Chile

Hugo Peralta V.
Banco Central de Chile

Juan Pablo Cova M.
Banco Central de Chile

Abstract

Through the application of textual border analysis tools, this work presents the construction of a high-frequency indicator, generated in real time, based on the computerized reading of the news from the main print media from January 2015 to December 2020 in Chile. This indicator captures the emotional tone of economic and opinion news by making use of an extensive —and novel— purpose-built dictionary in Spanish. This lexicon of words was subjected to challenging statistical tests of robustness, complemented by tests of predictive precision. The latter were carried out by comparing the degree of similarity between a classification by automated means and another by manual means in a random sample. The economic application shows that the constructed indicator has a high correlation with confidence indicators based on surveys, and a high predictive capacity in the face of shock phenomena hitting the economy.

Resumen

Mediante la aplicación de herramientas de análisis textual de frontera, este trabajo presenta la construcción de un indicador de alta frecuencia, generado en tiempo real, basado en la lectura computarizada de las noticias de los principales medios de prensa entre enero 2015 y diciembre 2020 en Chile. El indicador captura el tono emocional de las noticias económicas y de opinión haciendo uso de un extenso y novedoso diccionario en español construido para el propósito. Este léxico de palabras fue sometido a exigentes pruebas estadísticas de robustez, complementadas con pruebas de precisión predictiva. Estas últimas se realizaron comparando el grado de similitud entre una clasificación por medios automatizados y otra por medios manuales en una muestra aleatoria. La aplicación económica muestra que el indicador construido tiene una alta correlación con indicadores de confianza basados en encuestas, y una alta capacidad predictiva ante fenómenos de *shock* en la economía.

¹ The authors wish to thank Professor Stephen Hansen for his valuable comments during the internal seminar organized by the Central Bank of Chile “Unstructured Data Analysis for Macroeconomics and Monetary Policy”, held on 13 September 2021. We also thank Pablo Filippi and Carlos Medel for their valuable insights to this research, Consuelo Edwards for her editing services, *Observatorio Tecnológico* for their support in the application of text mining tools, Cristián López for his support in hardware services, Valentina Vásquez for her research assistance, as well as Alejandra Rozas, Carlos Madeira, Carlos Medel, Consuelo Edwards, Daniel Pérez, Erika Arraño, Felipe Lozano, Joaquín Pérez, José Manuel Salinas, Josué Pérez, Juan Pablo Ugarte, León Sanz, Luis Felipe Villanueva, Marcos Correa, Mario González, Miguel Musa, Natalia Oyanedel, Pablo Filippi, Sebastián Becerra, Sebastián Silva, whose valuable contribution made it possible to generate a set of manually tagged news articles. The opinions and conclusions expressed in this document are the sole responsibility of the authors and do not necessarily represent the views of the Central Bank of Chile. Emails: mcruz@bcentral.cl, pilarcruz@yahoo.com, hperalta@bcentral.cl, jcova@bcentral.cl.

1. INTRODUCTION

Since the invention of the printing press, the written media have been an essential mechanism of information, thanks to their ability as a credible source for keeping up to date judgments and visions about the future. The first newspapers appeared in the early 17th century in France, Spain and England ([Sánchez Sánchez \(2012\)](#)) and in the 18th century in the United States. In the 20th century the press and local journalism took the form of a news market, containing narratives with the capacity to influence the decision matrix of individuals and, therefore, affect the economic and social development of the country ([Bernabeu \(2002\)](#)).

The explosive development of these narratives has accelerated the need to identify their contents in order to integrate them effectively with economic phenomena. In response to these challenges, computational text analysis has emerged with great force in recent years, through which relevant information can be extracted from unstructured texts and be transformed into structured information. The increase in computer processing capacity has made it possible to create algorithms applicable to language and to considerably reduce processing time and costs.

This technical progress, has in turn, led to important innovations in nowcasting models. Official activity statistics take time to generate and compile, so their availability has lags. The literature in this field ([Banbura, Giannone, & Reichlin \(2011\)](#)) has so far been oriented towards predicting the trajectory of economic variables through the combined use of activity indicators with business and consumer survey indicators. However, more recently, studies such as [Kalamara, Turrell, Redl, Kapetanios, & Kapadia \(2020\)](#) and [Thorsrud \(2016\)](#) have proven that the performance of these models can be improved by considering the news in the press, due to its real-time availability and its predictive value in periods of economic stress.

The main contribution of this work focuses precisely on generating a high-frequency indicator built in real time based on the computerized reading of the news in Chile, with predictive capacity about the evolution of economic conditions, the business cycle and the trajectory of confidence levels in the economy. It uses a database of about 935 thousand pieces of news contained in six major newspapers in the country for the period 2015-2020.

Another contribution is the generation of a Spanish dictionary built from the same news database, which is extensive in the number of tagged words, comparable to its English similes, complex in its variety of grammatical forms, tested on a sample of manually tagged news articles and robust to the statistical testing on news.

The paper is organized as follows. Section 2 contains a synthesis of the related research; section 3 presents a characterization of the IS-News and its transformation into a database suitable for text mining; section 4 presents the construction of the dictionary for reading the news; section 5 presents the construction of a manually tagged set of news; section 6 describes the construction of the IS-News indicator; section 7 shows the results of the IS-News and their relationship with economic indicators, to close with the main conclusions in section 8.

2. SYNTHESIS OF RELATED RESEARCH

Sentiment Analysis is a text mining tool that allows to obtain qualitative data in real time, without resorting to population surveys. It is a field of research that performs a computational treatment of opinions and feelings contained in texts. Its specific application on printed news is relatively recent and its accelerated diffusion is explained by its contribution to economic analysis and modeling, its ability to anticipate changing conditions and its low production costs compared to surveys ([Shapiro, Sudhof, & Wilson \(2020\)](#)).

From these text analysis tools, two areas of research can be identified that address different objectives: one that detects “intensity” and one that detects “tone”. Intensity measures are based on the number of times certain words appear in the text, and has been extensively disseminated by [Baker, Bloom, & Davis \(2016\)](#) through the Economic Policy Uncertainty (EPU) index. This index counts the number of times words such as uncertainty or recession appear in the analyzed text, similar to what has been done by [Altig et al. \(2020\)](#), [Cerdeira, Silva, & Valente \(2016\)](#) and [Becerra & Sagner \(2020\)](#). The tone measures, on the other hand, called “Sentiment Analysis” capture the underlying sentiment of optimism or pessimism in texts, using broad tagged lexicons and also automated or machine learning tools. The research area analyzing this paper detects the tone of the texts in the news.

Due to the relevance of using databases with a marked tonal polarity, an important part of sentiment analysis research uses documents with a high subjective and judgmental content, as is the case with blogs, social networks, or product reviews. The press, meanwhile, does not contain this same density of tone as it tries to give the impression of objectivity ([Balahur et al. \(2013\)](#)). In these types of texts, the recount of judgments and polarity is often found in the form of third-party opinions or else in news stories that call upon the opinion of third parties. The editorial line may choose to emphasize or moderate the final texture of the message, but ultimately results in a discourse structure that is more complex to analyze. Therefore, success in capturing the polarity of a news article depends on correctly identifying the sentiment implicit in it (“opinion mining”), isolating the sign of the news piece itself ([Saberri & Saad \(2017\)](#)).

Addressing the ambivalence that often exists in these texts, i.e., between news content and the feeling about it, [Balahur et al. \(2013\)](#) investigate how to gain effectiveness in their reading. Based on 1,592 quotes from English newspaper articles, they show that annotators’ agreement when tagging rises from 50% to 81% when some specific guidelines are agreed upon, such as identifying the tone of the news, irrespective of its content, isolating the internal judgment to tag with greater impartiality, avoiding interpretations of what the text means to convey, to minimize subjectivity, and proceeding to tag only those quotes with a markedly visible tone.

Notwithstanding the above, identifying the tone of the news becomes particularly complex in the face of the specially designed strategies used by the media to attract the readers’ attention. [Reis et al. \(2015\)](#) find that there is a close relationship between the news’ polarity and popularity, and that headlines with a negative sentiment correlate directly with the popularity of the news and the dynamics of the comments posted by readers.

Despite these calibration difficulties, sentiment analysis saw its first publications in the early 2000s and experienced very accelerated growth a few years later. Using Google Scholar and Scopus citation counts, [Mäntylä, Graziotin, & Kuutila \(2018\)](#) note that in 2000 there were only 37 publications in this field, rising to around 7,000 in 2016. The authors conclude that what enabled this dizzying boom has been the possibility of analyzing huge volumes of texts with computational text mining tools. Some of the research that seems most relevant to us is reviewed next.

Work published by the Federal Reserve of San Francisco ([Shapiro, Sudhof, & Wilson \(2020\)](#)) develops a time series that captures sentiment derived from news stories drawn from economic newspapers between 1980 and 2015. By generating a model that combines various internationally recognized labeled lexicon, they show that daily news sentiment is a good predictor of survey-based confidence indices.

On the other hand, recent research by [Aguilar, Ghirelli, Pacce, & Urtasun \(2021\)](#) constructs a daily frequency sentiment index based on Spain’s press news (DENSI). This index manages to outperform the Economic Sentiment Indicator (ESI), which is one of the most popular survey-based indicators in the European Union. It concludes that the DENSI index is significantly better at predicting GDP in the short term and has been even more useful in predicting the Covid-19 crisis. Unlike other studies, the sentiment index calculated by these authors does not use a dictionary or combination of

dictionaries but a pool of words pointing in a positive direction and others in a negative one, resembling the EPU methodology of [Baker, Bloom, & Davis \(2016\)](#).

The [Buckman, Shapiro, Sudhof, & Wilson \(2020\)](#) publication concludes that the use of a daily-news-based sentiment indicator reveals a sharp drop caused by the rising tide of negative news related to Covid-19. What is important about this finding is that this drop occurs two months ahead of the Conference Board's Confidence Index and the University of Michigan Index of Consumer Sentiment surveys.

[Song & Kyung-Shik \(2019\)](#) demonstrate that, by using simple sentiment analysis based on tagged lexicons, the economic news is an effective source of information to generate an indicator for the Korean economy. However, they point out that international comparisons of this type of indexes are quite difficult to establish, due to the particulars of each language, which demands more complex linguistic processing.

The study by [Kalamara, Turrell, Redl, Kapetanios, & Kapadia \(2020\)](#) from the Bank of England shows that, with information extracted from three popular U.K. newspapers not specialized in financial markets, economic forecasts can be improved for output, inflation, and unemployment. They find that said improvements depend on the methodologies used, and that the most outstanding results are obtained during periods of economic stress and combining word counting techniques with supervised machine learning.

Research by [Larsen & Thorsrud \(2015\)](#) confirms the widespread belief that changes in expectations caused by news is an important autonomous driver of economic fluctuations. Using the main Norwegian business newspapers, the authors identify the topics with the highest predictive power, from which they construct an aggregate index and show that unexpected changes in the index cause significant and persistent fluctuations in markets, especially in credit ones.

[Cruz, Peralta, & Ávila \(2020\)](#) used computational linguistics to analyze the Business Perceptions Report of the Central Bank of Chile (IPN) and generated an index with high correlations with business and economic confidence indicators.

Finally, central banks have also incorporated written communication increasingly as an important way to disseminate their policy guidance. Using computational linguistics, the IMF ([International Monetary Fund \(2018\)](#)) shows that the tone of written messages impacts the short end of the yield curve and also medium- and long-term interest rates, and that the effective use of this communication can give monetary policy more room to maneuver in the face of transitory shocks to inflation. [Hansen, McMahon, & Prat \(2018\)](#) show that the increased transparency applied in the external communication of the Federal Open Market Committee (FOMC) changes the way its messages are conveyed, but not its monetary policy formulation.

3. CHARACTERIZING THE NEWS TO CREATE A SENTIMENT INDEX

The formulation of the IS-News is based on the reading and processing of the main printed daily newspapers in the country, of national and international coverage, namely: *El Mercurio*, *La Tercera*, *Pulso*, *La Segunda*, *Diario Financiero*, and *Estrategia*.² Regional print media³ were excluded because their coverage is not nationwide, as well as those with lower periodicity (weekly or bimonthly), because the index is based on high-frequency data. The database for the construction of the index comes from information generated by *NexNews*.

Despite the instantaneous and immediate nature of on-line media, we chose to work with print media because of the advantages they offer for the purpose of adequately calibrating the tone of the news.

² In late 2016, this paper began publishing only on digital media, so from then on it is deleted from the database.

³ Relevant media to construct databases for regional sentiment indexes.

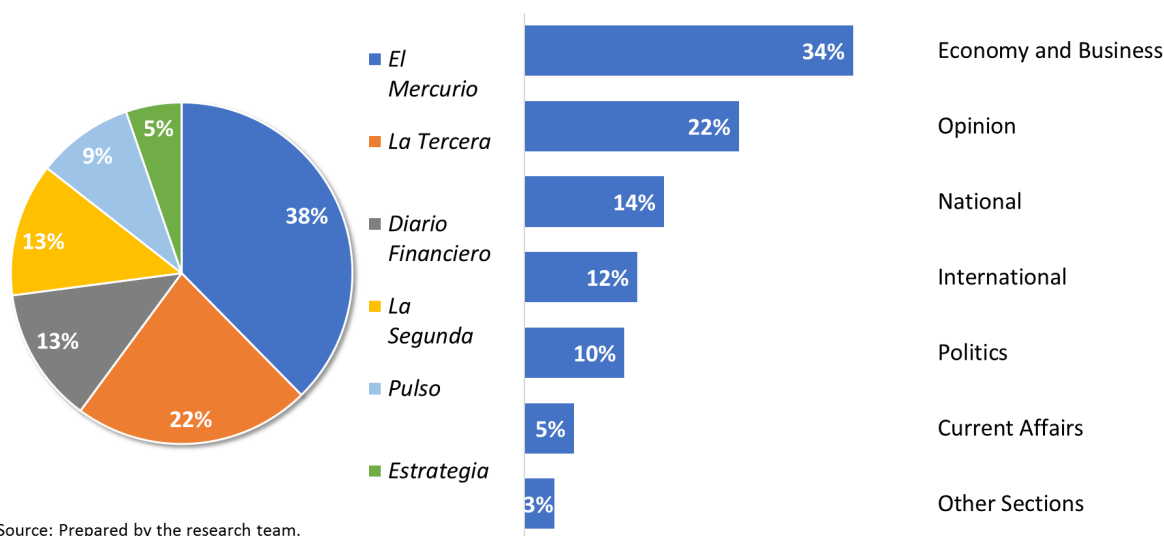
Most relevant is that the printed format must go through stricter fact-checking and editing filters, which helps to increase the degree of reliability of the publications.

In general, these media also include deeper and more diverse analysis of news events, as well as op-eds with a high emotional content. The readers of these newspapers, meanwhile, face fewer distractions when reading (“finishability”⁴), a fact that has contributed to maintaining their relevance as an informative medium.

Figure 1

NEWS BY MEDIA AND SECTIONS, 2015-2020

(filtered base, % of total news by media)



Source: Prepared by the research team.

The number of news articles in the selected media totals 935 thousand for the period under study (between 2015 and 2020). Two events of great impact on the Chilean economy are included in this period: the social crisis of October 2019 and the arrival of the pandemic in March 2020. This base was purged (see section 3.1), reducing its size to 417 thousand news articles, which include the sections Economy and Business, National, International, Politics, Current Affairs and Opinion (Figure 1).

In this new base, the “Economy and Business” and “Opinion” sections are the most relevant, as they account for an average of 56% of the total news, while the “National”, “International” and “Politics” sections each represents around 12%. By media, news in *El Mercurio* and *La Tercera* account for 60% of the total base, while *Diario Financiero*, *La Segunda*, *Pulso*, and *Estrategia* make up the remaining 40% (figure 1).

3.1 CREATING THE DATABASE

Each piece of news in the database contains title, subtitle, date, press medium, section, body, and, in some cases, boxes or additional information. In order to have a database suitable for use in developing

⁴ This term was coined in 2015 by publishing groups’ executives to indicate the quality of the print media to retain the readers’ concentration, avoiding interleaved ads or access to additional distracting information.

the IS-News index, it was necessary to apply the following processes of grouping, cleansing, and filtering:

- Elimination of too short or too long articles,⁵ to remove the noisier ones, because they are either repetitions or so extended that they distort the measurement of the index on the corresponding day.
- Grouping of sections, which allows to discard the ones not related to the economy or businesses (i.e., sports, sciences, special supplements, entertainment, and miscellaneous).⁶
- Application of various filters to the database sections. The version that yields the best results for generating the IS-News index is the one that combines the op-ed with the economic sections (241 thousand news).⁷
- Finally, use of the *Python spaCy* library to be able to apply various *Natural Language Processing* (NLP) techniques such as lemmatization, part-of-speech (POS) tagging, or entity recognition. This text processing contributes to facilitate the subsequent construction of the dictionary, because it reduces the size of the universe of terms to be tagged and facilitates identifying contexts where the words are used. As will become clear when analyzing the construction of the dictionary, this is especially useful when there is disagreement regarding the tone orientation of the lemmas and the selection of the tags.⁸

4. CONSTRUCTING A DICTIONARY TO READ THE NEWS

The most used techniques to perform Sentiment Analysis (SA) are, on the one hand, Machine Learning techniques, which encompass a wide range of statistical models capable of learning from massive databases and generate text tone predicting models. This technique's predictive potential makes it particularly advantageous, but its quality depends on the training set, which needs to be voluminous and complex to learn about the unit lexicons of the language, as well as about both simple and complex sentences.

A second group of commonly used techniques to perform SA contains those that use previously tagged lexicons. This approach is based on a selection of terms that reflect a clear sentiment orientation and can be tagged. Unlike Machine Learning, it is not based on algorithms, but on semantic dictionaries containing terms classified with valences. Words with positive valences are used to collect desired states, and those with negative valences are used to collect undesired ones. In this way words that are contained in the tagged lexicon are detected and added together according to the corresponding polarity. These are the techniques used in this research.

In English, there is abundance of tagged lexicons available, and with specific designs according to the texts to be read, as can be financial and economic, press, marketing, psychological, entertainment, or general texts. This even allows to originate other dictionaries that result from the combination of tagged lexicons with specific designs, as presented by [Shapiro, Sudhof, & Wilson \(2020\)](#) for the reading of economic news in the United States.

In Spanish, these resources with assigned valences are still scarce, which makes it difficult to obtain the necessary syntony between the tagged lexicon and the texts making up the data. In view of this

⁵ Containing fewer than 100 words (synopsis, briefs) or more than 2,500 words (stories or interviews).

⁶ The original database contains over 27,000 sections, from which unwanted ones are deleted by searching the keywords that are related to the section (e.g., soccer, football, tennis, or the like, to detect sport news).

⁷ One alternative form of classification, less intensive in human labor, is the application of topic modeling such as *Latent Dirichlet Allocation*. However, given its high cost in terms of processing time, it will benefit from future improvements for this project.

⁸ This application of the *spaCy* library is not cost-free; the size of the database expands around 100-fold, so it demands significant reinforcements of the computational capacity.

limitation, this research has generated a voluminous and customized tagged lexicon for reading the printed news in the country.

There are fundamentally two approaches to solve the dilemma of how to create a lexicon that is suitable for performing Sentiment Analysis (Medhat, Hassan, & Korasky (2014)):

- Creating a corpus lexicon, that solves very effectively the problem of finding words of opinion according to the context's specificity but requires contents with a strong tone orientation. It can be applied by using an initial list of seed words, which is then completed sequentially with other words with polarity through semantic methods. It is best applied in very specific topics that shape the corpus of specific characteristics or dimensions, for example, corpus associated with security, reliability, texts in social networks (Cruz Cornejo (2017)).
- Creating a purpose dictionary lexicon, which does not require a certain volume of data, but may present the difficulty of associating the tone of the words with the specific context of analysis (e.g., economic and financial news), and which is more demanding of the use of human calibrations. This way of constructing the lexicon was the one used in this research to create the IS-News dictionary, and which was perfected with the use of algorithms to recognize contexts (subsection 6.2).

4.1 METHODOLOGY FOR CREATING A DICTIONARY

The methodology for creating a dictionary in Spanish consisted of several sequential stages, all of them referenced to news databases in Spanish. No other words from other Spanish dictionaries or from translated English dictionaries were included. In other words, the tagged terms were obtained entirely from the printed news themselves, with the purpose of having a domain-specific dictionary.

Briefly, the stages were the following:

- Use of the *Python spaCy* library, which helps identify the grammatical class⁹ and the lemma of each word, as well as recording the publication where the news appears. With this processing, the size of the original database is multiplied roughly by 100,¹⁰ because each word contains a new set of linguistic information.
- With the purpose of avoiding the difficulty of processing this densified data base in one take, we opt to subdivide it randomly and work on the parts one by one until the entire database is covered. In other words, to obtain the terms that originate the dictionary, the database is handled by segments, thus facilitating the computational processing that the complete database demands.
- The base was subdivided by randomly selecting¹¹ news articles equivalent to 10% of the total base (47 thousand news), in the form of a sampling without replacement.¹² The processing of the first random sample of news pieces identified a total of 9,596 terms and the second random sample, when compared with the first, added barely 325 new terms (3.4%). For this reason, a third sampling was ruled out due to its low probability of adding any new terms. Thus, with 90 thousand randomly collected news articles, the universe of terms originating the tagged dictionary was covered.
- Identification of the terms (verbs, adjectives, and adverbs) in the random samples is done with the functions of lemmatization and computational POS tagging by *spaCy*. The complete universe of unique words found in the two samples was 9,921, of which 7,616 were selected for tagging, as

⁹ For example, verbs, adjectives, adverbs, nouns, and determiners, plus names of entities or locations.

¹⁰ The complete news base, including functionalities added by *spaCy*, is around 100 GB in size.

¹¹ For the sample to be representative, the random choice includes the dimension of the media, sections, and dates, and it is made in a stratified way, that is, preserving their proportions in the full base.

¹² Sampling without replacement is that which is carried out without taking back to the base those news articles that were chosen to build the sample.

2,305 were removed because of their low frequency.¹³

- Upon completing the process described above, the next step is to tag the set of unique words with positive, negative, or neutral tone by each member of the research team. Following international recommendations (Balahur et al. (2013)), the tagging criteria are standardized, in the sense that only those words with a clear tonality are appraised, and that those terms with two or more diverging meanings or with a diffuse connotation are tagged as neutral.
- Upon completing of the manual tagging process, it was decided to maintain the tags of the words that posted 100% coincidence while, in the cases that presented partial coincidences, it was decided to maintain them only if the discrepant valuation was not the opposite.
- For the rest, where the divergence was bigger, the team proceeded to reevaluate word by word, analyzing the contexts in which they were used.

The outcome of this methodology was the generation of a tagged dictionary¹⁴ with valences different from zero, based on printed news, with a total of 5,419 lemmas with their inflections,¹⁵ out of which 374 are unique lemmas (257 verbs, 67 adjectives and 50 adverbs) (figure 2; Appendix).

Figure 2

MAIN TAGGED COMPLEX DICTIONARIES

(with positive or negative valence)

DICCIONARY	FEATURE SPACE	LANGUAGE	LEXICON SIZE (TAGGED WORDS)
Hu-Lui (HL)	Movie Reviews	English	6,789
IS News	News Articles	Spanish	5,419
Harvard General Inquirer (GI)*	General English	English	4,206
Loughran-McDonald (LM)*	Financial Reports	English	2,683
Business Perception Report Central Bank of Chile (BPR CBCh)	Economic and Business Reports	Spanish	774
Financial Stability Report Bank of Spain (FSR BoS)	Economic and Business Reports	Spanish	565
Financial Stability Report Central Bank of Chile (FSR CBCh)	Economic and Business Reports	Spanish	361

* These dictionaries contain additional categories to positive and negative, which means that words can belong to several other different categories such as "degree of uncertainty", "power", "strength", among others. In the LM there are about 1.553 words in other categories which gives rise to a total of 4.236 words tagged in all categories. The whole dictionary contains 26 categories and a total of 11.788 words tagged in all categories. Note that a same word can be counted several times; the word "about" is counted seven times because it is tagged in seven different categories.

The extension of this dictionary compares favorably with other English-language dictionaries, popular in SA, such as the Harvard General Inquirer (GI), which has a length of 4,206 words and is a general-purpose dictionary developed by the Harvard University. It also compares positively with the Loughran-McDonald (LM), which is a little smaller than the GI, with a total of 2,683 words, and

¹³ Words with very low frequency, present in fewer than six pieces of news in the sample, may represent an error or be irrelevant in the indicator's calculation.

¹⁴ The dictionary can be found on [here](#).

¹⁵ Inflections are conjugations, plurals, verb tenses, and others. Each verb in the dictionary has associated an average of 20 inflections, for example: *augmentar*, vs *augmentado*, *augmentando*, *augmentó*, *augmentará*, *augmentaría*, and so on.

which has the particularity of being for the specific domain of economics and finance. The Hu-Lui (HL) dictionary, on the other hand, has a total of 6,789 terms, but because it is created from magazines with movie reviews, it limits its use to the economic-financial area.¹⁶

In Spanish, tagged lexicon resources for sentiment analysis are scarcer. However, among the most popular is the translated Spanish version of *SentiStrength*, designed specifically for the context of social networks; also, the dictionary created by [Díaz Rangel, Sidorov, & Suárez Guerra \(2014\)](#), which presents a proprietary repertoire with more than 2,000 terms, classified into six basic emotions: joy, anger, fear, sadness, surprise, and repulsion, associated with a probability of being used with an emotional content ([Pérez Albertos \(2018\)](#)).

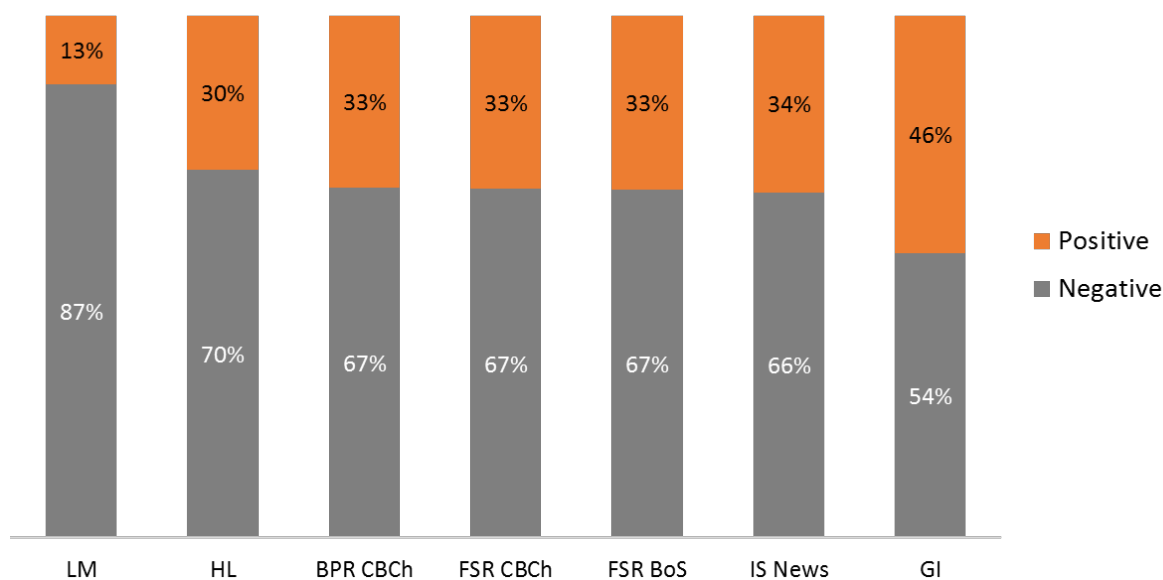
Finally, there are three Spanish dictionaries created for the specific domain of economic and financial reports. These are the dictionary for the CBCh's Financial Stability Report ([Becerra & Cruces \(2021\)](#)), of its peer at the Bank of Spain ([Moreno Bernal & González Pedraz \(2020\)](#)) and of the CBCh's Business Perceptions Report ([Cruz, Peralta, & Ávila \(2020\)](#)). These highly specific dictionaries are characterized by having words with unique connotations that differ even from other dictionaries for financial contexts. For example, the word delinquency may have a negative connotation, but in a financial stability report, it may be neutral.

The analysis of the result of negative and positive valences in each of the aforementioned dictionaries reveals that negative tags in IS-News represent 66% of the total number of words with non-neutral valences, similar to what is found in the other two Spanish dictionaries (IPN and IEF), and in the HL (figure 3). The prevalence of negative over positive tags is consistent with the findings of [Reis et al. \(2015\)](#) (section 2).

Figure 3

TAGGED WORDS IN SELECTED DICTIONARIES

(with positive or negative valence)



¹⁶ This comparison must take into consideration that the derivations or inflected forms in English are less numerous than in Spanish.

4.2. LIMITATIONS IN THE USE OF DICTIONARIES FOR TEXT CLASSIFICATION

Valued dictionaries are usually created manually, which, while granting reliability, transforms them into a highly time-intensive task. Even so, it is necessary to bear in mind some limitations associated with its effectiveness to perform Sentiment Analysis (Mechulam Burstin & Salvia Varela (2018)), as presented below:

- Correctly identifying and tagging the tone of the words, paying attention to the contexts in which they are used. Indeed, a given word may have a neutral sentiment orientation in an economic context, but negative in a scientific or legal one, or vice versa. Using a sample of corporate financial reports (10-K¹⁷) in the United States between 1994 and 2008, Loughran & McDonald (2011) show that nearly three-quarters of the words identified as negative by the GI dictionary are typically not considered negative in financial contexts, such as liability, depreciate, tax, or cost.
- The texts, sentences, or phrases may have an affective polarity, even if they contain no word expressing by itself a negative feeling (e.g., my clothes got all wet in the rain).
- Conversely, there may be sentences or phrases that do not contain an affective polarity, but do include words with a polarity (how bad were my clothes soaked in the rain?)
- Casual writing, with irony or sarcasm, which often makes for ambiguous texts, can also lead to a misreading (you won't believe how my clothes got in the rain).
- The contraposition in polarity, where, in the same sentence or text, a single polarity is not distinguishable, but on the contrary, two opposing polarities are superimposed (the rain ruined my clothes, but I'm OK because they were too worn out).

Notwithstanding these limitations that affect dictionaries in general, robustness tests can be performed to ensure that the effectiveness of the dictionary is not tied to any particular term or group of terms, which is discussed in subsection 7.1 below.

5. CONSTRUCTING A SET OF MANUALLY TAGGED NEWS

With the objective of evaluating the predictive capacity of the IS-News dictionary, a manual tagging procedure is being performed on a set of news articles, like the one by Shapiro, Sudhof, & Wilson (2020). This procedure is carried out by surveying a group of 23 CBCh researchers, who manually classify 840 news articles according to the polarity of sentiment they detect from their reading. The 840 news articles are randomly selected from the database used by the research, based on some specific criteria:

- News no shorter than 200 words—which could lack a clear polarity—and no longer than 400—which could make it more complex for the researcher to define a tag.
- News containing keywords such as: mentioned, explained, said, wrote, pointed out, or similar, because, often, this type of text contains a certain polarity of sentiment, synthesized in the form of a third party's opinion.
- Randomly selected news, stratified by year, in order not to alter the original arrangement of the news in the database.

The survey is run on the electronic platform,¹⁸ which indicates to each researcher the instructions for tagging news articles and the progress status of the evaluation process. After reading each text, researchers are asked to identify its polarity on a five-level scale, from the very negative minimum,

¹⁷ 10-K is a yearly financial report of companies listed in the U.S. stock exchange, containing more details than the one that is delivered every year to stockholders, and is a requirement of the U.S. SEC.

¹⁸ Provided by *Boostworld* in the e-learning digital platform's survey module.

to the very positive maximum.

The instructions describe the general criterion for classifying the polarity of the text, indicating the importance of differentiating the sentiment from the actual content of the news item. This precision is important because a given story may communicate a negative fact in itself, such as, for example, “fortunately we will see a lower fall in activity than initially expected,” but still contain a positive sentiment.¹⁹

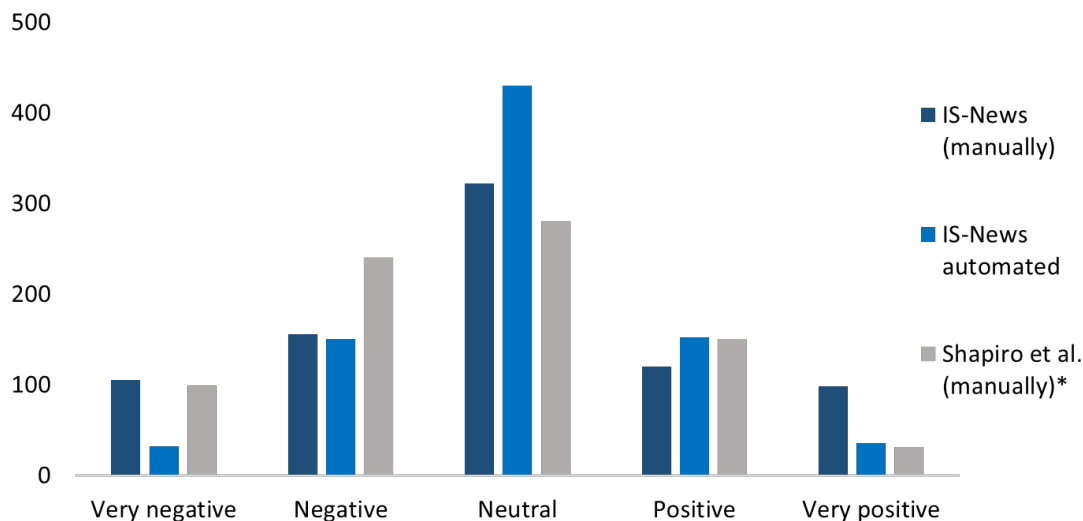
The manual tagging process was carried out in four sequential stages, to scale the burden on the evaluators. Each evaluator had to tag a total of 120 pieces of news in a period of about two months, and each piece was tagged by four different evaluators to reduce classification bias.

Figure 4 shows the histogram of manual classifications, automated news ratings and a comparison with that obtained by a similar manual tagging exercise in [Shapiro, Sudhof, & Wilson \(2020\)](#). The results show that the highest frequency of ratings occurs in neutral news (particularly in the automated rating), although with a bias towards negative categories consistent with international evidence, which shows a greater emphasis on reporting negative news. In turn, the distribution of human IS-News ratings follows a normal function with positive skewness (centered on articles considered negative) and a slight degree of kurtosis.

Figure 4

COMPARED HISTOGRAM OF NEWS SET TAGGING

(IS-News tagged manually vs automated)



* Adjusted to histogram scale in Shapiro et.al (2020)

5.1 EVALUATING THE DICTIONARY’S PREDICTABILITY IN TAGGED NEWS

The predictive capacity of the tagged dictionary is evaluated by contrasting the automated news classification with the manual one, to determine the degree of coincidence between both evaluations.

¹⁹ To convert the results of the automated news classification from continuous to discrete format, several options for assigning values in the neutral category are tested (detailed in subsection 5.1). All the options tested yield similar results.

For this purpose, Spearman's coefficient and the F1 macro score²⁰ statistics are measured as goodness-of-fit measures. To calculate the latter metric, it is necessary to transform the results of the automated indicator from continuous to discrete formulation, to enable the comparison to be made in the standard of manual tagging.²¹

Next, the results of both news classifications are contrasted considering the three above mentioned sentiment categories and identifying the percentages of coincidence between the two.

The Spearman correlation statistical test shows that it is very unlikely for manual and automated news classification to be uncorrelated (p-value close to zero; with 95% confidence). On the other hand, the F1 accuracy indicator yields a value of 0.47, indicating that 47% of the news articles have identical manual and automated classification. Figure 5 summarizes the results of the goodness-of-fit measures.

Figure 5

COINCIDENCE BETWEEN MANUAL AND AUTOMATED CLASSIFICATION OF NEWS

(goodness-of-fit measures)

TAGGED DATABASE	STATISTICAL TESTS	
	SPEARMAN RANK	MACRO-F1
IS-News discreet p80/20	-	0.473
IS-News discreet p75/25	-	0.474
IS-News discreet +-0.0050 neutral	-	0.474
IS-News continuous	0.384 (p=0.000)	

To extend the accuracy analysis provided by the previous statistics, the degree to which the manual and automated classification provides highly contrasting tags is verified. For this purpose, the results are segmented according to the following criteria: i) when the manual and automated classification are complete opposites, ii) when the differences are of one tone, and iii) when there are no differences in tone.

According to this analysis, the results coincide with indicator F1, showing that in 49% of the cases there is no difference in tone between tagged manual and automated news. It also shows that in 45% of the cases there is a change in tone towards neutrality, i.e., if the text was manually classified as negative (or positive) and the automated rating classified it as neutral, and vice versa. Thus, in 94% of the cases there is at most one degree of difference between the two sources. Finally, only 6% of the 840 news articles showed opposite results and 0.4% showed a maximum degree of difference, that is, if the article was manually classified as very negative, automatically it was classified as very positive and vice versa (figure 6).

²⁰ The F1 Macro Score represents a harmonic average of "precision" and "recall." It is used as a measure of accuracy in classification problems with unbalanced categories (for example, classification of negative, neutral, or positive texts), considering that this indicator neither underestimates nor overestimates a particular category.

²¹ This transformation is done by grouping the values obtained in three news sentiment categories: negative, neutral, and positive.

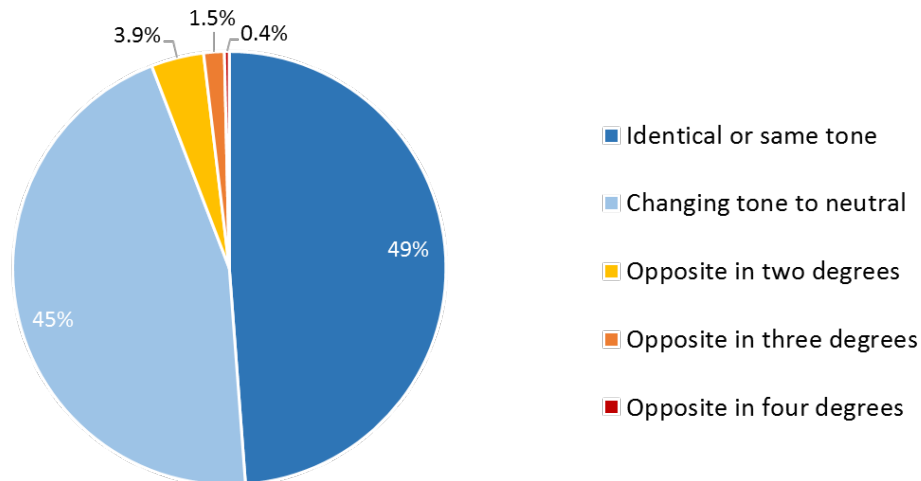
Therefore, based on the results of the goodness-of-fit statistics (in particular Spearman's correlation coefficient), as well as the comparative analysis of manual and automated tagging in terms of changes in tone, it is possible to conclude that there is statistical coincidence between the results of the two, giving goodness-of-fit to the dictionary built for the purpose.

Subsection 7.1 shows robustness tests applied to the dictionary. As we will see, even in the most demanding tests, the dictionary proved robust in its results.

Figure 6

COMPARISON BETWEEN MANUAL AND AUTOMATIC NEWS TAGGING

(results from identical to total opposite)



6. CONSTRUCTING THE HIGH-FREQUENCY IS-NEWS INDICATOR

The central objective of this research is to construct a time series, of daily frequency, based on printed news, using the text mining method called Sentiment Analysis.

Two methodologies commonly found in international research are used to calculate the series. One, which we have termed IS- News^{Dice}, is based exclusively on the use of the dictionary and manual tags assigned by the researchers. The other, which we have termed News^{PMI}, uses a corpus lexicon developed from the dictionary and introduces a measure of probability that a given word is associated with a given sentiment ("Pointwise Mutual Information, PMI"). In both methodologies, the construction of the time series is formulated according to the following sequence:

1. Every day, words with polarity are detected in the news, following a specific algorithm in each methodology (SW, formula 1).
2. The polarity of each piece of news is calculated, also following a specific algorithm (ISN, formula 1).
3. A sentiment index is calculated for each day, which is the lineal average of every one of the texts processed according to the description in point 2 (formula 2).
4. A weekly, bi-weekly, monthly, or any chosen time window index is calculated by averaging the daily indexes for the chosen window (formula 3).

- (1) $SW, ISN =$ calculation algorithm;
- (2) $ISD = \frac{\sum_{i=1}^n ISN_i}{n}$;
- (3) $\overline{IS}_t = \frac{\sum_{i=1}^t ISD_i}{t}$; where,

SW = polarity of a word w ;

ISN = sentiment index of a news article;

ISD = sentiment index for a day;

IS_t = sentiment index for period t ;

n = number of news articles per day;

t = number of days in the period analyzed.

The algorithms used to derive both series (IS-News^{Dicc} and IS-News^{PMI}) are analyzed in the next sections. The results, presented in section 6, show that the methodology using PMI has a better predictive potential than the one using the tagged dictionary.

6.1 METHODOLOGY FOR CALCULATING IS-NEWS^{Dicc} USING THE DICTIONARY

The most traditional method for generating a sentiment index is based on using a tagged dictionary, which means that the polarity of each word is determined solely by the dictionary. In order to produce the IS-News^{Dicc} series, the following procedures are required:

To create an algorithm to detect polarity for each news article, which adds words with positive and negative tags. Each text is scrolled through, identifying verbs, adjectives, or adverbs, and then looked up in the dictionary. Sentiment words can take the values 1 (positive tone) or -1 (negative tone), modifying words can take values of 1.5 (intensifies the tone) or 0.5 (attenuates the tone) and negation words, a quotient of -1, which reverses the polarity of the next three contiguous words.

To assign tags at the level of each sentence, for which those with assigned polarity are added together, modifiers and negations are applied, and the result is divided by the total number of words in the news article, to avoid longer stories having a greater weight on the sentiment index. Accordingly, the formula for the sentiment index of a news item is as follows:

$$(1) ISN = \frac{\sum_{i=1}^p SO_i}{q}$$

SO = sentiment of a sentence according to calculation algorithm;

p = number of sentences in the news article;

q = total number of words in one news article.

The sentiment index for the day or other desired time window is calculated with the linear average of every text processed, according to formulas 2 and 3 of the previous section.

6.2 IS-NEWS^{PMI} CALCULATION METHODOLOGY USING A CORPUS LEXICON

The great difference between this methodology, *Pointwise Mutual Information* (PMI²²) and the one that uses only the dictionary, lies in the way in which the polarity of each word (SW) is calculated. PMI provides a measure of the probability that a given word is associated with a given sentiment. Thus, the value assigned manually to each word according to the dictionary is adjusted based on the usage detected throughout the news corpus.

For example, if a positive word is found more frequently in positive news, its value will be amplified, whereas if it is more present in negative news, its value will be diminished. Consequently, the application of the PMI algorithm means that the sentiment associated with a news item (ISN) depends not only on the individual tag of each word, but also on its relative frequency in the rest of the text. To generate this metric, it is necessary to form the following vectors:

- A rating of positive, neutral, or negative²³ for the totality of the news items in the database, according to the tagged dictionary,
- A co-occurrence matrix between each of the dictionary words and the frequency in which they appear in positive, neutral, or negative news. This matrix forms the PMI between a word “*w*” and a sentiment category “*c*” (positive, neutral, or negative), which is arithmetically defined as follows:

$$(1) \text{ PMI}(w, c) = \log\left(\frac{p(w, c)}{p(w) * p(c)}\right), \text{ where:}$$

$p(w)$ is the word w 's share in the news as a whole;

$p(c)$ is the category c 's share in the news as a whole;

$y p(w, c)$ is the probability that the word w in a context c appears in the news as a whole.

Therefore, the formula for measuring the sentiment of each word in the dictionary $S(w_i)$ ceases to be the dictionary's tag and gives way to the difference between the positive PMI and the negative PMI. The calculation of sentiment for each news article (ISN^{PMI}) is also specific to this methodology, and corresponds to the linear average of the sentiment of the PMI-measured words, as seen below:

$$(2) \quad S(w_i) = \text{PMI}(w_i, POS) - \text{PMI}(w_i, NEG)$$

$$(3) \quad ISN^{PMI} = \overline{S(w)}$$

Finally, the formulas corresponding to the sentiment index for one day or the chosen time window are common to both methodologies and are calculated as the linear average of each of the processed news stories (formulas 2 and 3).

7. IS-NEWS RESULTS AND RELATIONSHIP WITH ECONOMIC INDICATORS

To evaluate the usefulness of the IS-News high frequency indicator, we explore its ability to anticipate confidence indicators such as the Economic Perception Index (IPEC²⁴), the Monthly Business

²² Re-weighting with PMI is an approximation to methods known as vector space models.

²³ For this purpose, we use the calculation algorithm detailed in the previous section to assign, according to the ISN distribution, the positive, neutral, or negative tag.

²⁴ Economic Perception Index (IPEC in Spanish) produced by GfK Adimark Chile. It is calculated by the monthly application of a structured questionnaire to a sample of 1100 persons 18 years old or more, residing in the main cities of Chile.

Confidence Index (IMCE²⁵) and their respective sub-indexes, the Business Confidence Index (ICE²⁶), as well as various activity data, in order to identify those with which it is best related. Also, we explore its capacity to anticipate shocks in the economy, whose occurrence over time is also examined.

In addition, we identify which IS-News calculation formula delivers a higher predictive value, i.e., whether the one that uses exclusively the dictionary (IS-News^{Dict}) or the one that uses the dictionary-based corpus lexicon (IS-News^{PMI}). The results for the period 2015-2020 show the following:

- The Pearson correlation coefficient²⁷ presents, without exception, significantly higher values when IS-News is calculated using the PMI methodology. This is consistent with the findings of [Shapiro, Sudhof, & Wilson \(2020\)](#) in which the predictive accuracy of the indicator using PMI metrics is better than that generated with any other of the available on-the-shelf dictionaries.
- Correlations rise almost widely when they are lag behind confidence and activity indicators in about four weeks, and they tend to decline when the lag grows to two months, or they are measured contemporaneously.
- The highest correlations of the IS-News^{PMI} were found, with a one-month lag, with the Economic Perception Index and its two subindexes, “situation to purchase home items” and “perception of the firms’ current economic situation”, which yield Pearson coefficients of 83%, 92%, and 85%, respectively (figure 7). In other words, the IS-News^{PMI} has the best fit with indicators that reflect the population’s perception of the present and future economic situation as well as of the propensity to consume.
- Regarding the Economic Perception Index, the IS-News succeeds in capturing with intensity the two shocks occurring in the period analyzed, i.e., the social outbreak of October 2019 and the arrival of the pandemic in Chile in March 2020. Figure 8 focuses specifically on the period in which these shocks occurred (the complete series is presented in figure 7), showing that both events begin to manifest in the IS-News approximately in the chronological month prior to the one that appears in the survey results. The Pearson coefficient for this specific period rises to 92% for the Economic Perception Index and to 95% for the sub-index "situation to buy household items." Something similar occurs during the recovery phase, where the IS-News recovers its level before the confidence indicators do so.

²⁵ Monthly Indicator of Business Confidence (IMCE in Spanish) produced by ICARE and Adolfo Ibañez University. It is a synthetic index that is applied to 607 companies in four sectors (Industry, Mining, Trade and Construction) reflecting the weighted sum of those indicators for each sector.

²⁶ Business Confidence Index (ICE in Spanish) produced by Del Desarrollo University. It is measures economic perception from businessmen point of view. It is elaborated from surveys to approximately 300 general managers, business owners or executives, through telephone surveys or emails.

²⁷ A measure of the strength of the linear dependence between two quantitative random variables, irrespective of the scale on which they are displayed. It is a good metric when the samples are large, and its distribution follows a normal curve.

Figure 7

IS-NEWS^{PMI} VS INDICATORS OF CONFIDENCE IN THE ECONOMY (IPEC): 2015-2020

(graphic of IPEC with lag +1; Pearson's R calculated using IPEC+1)

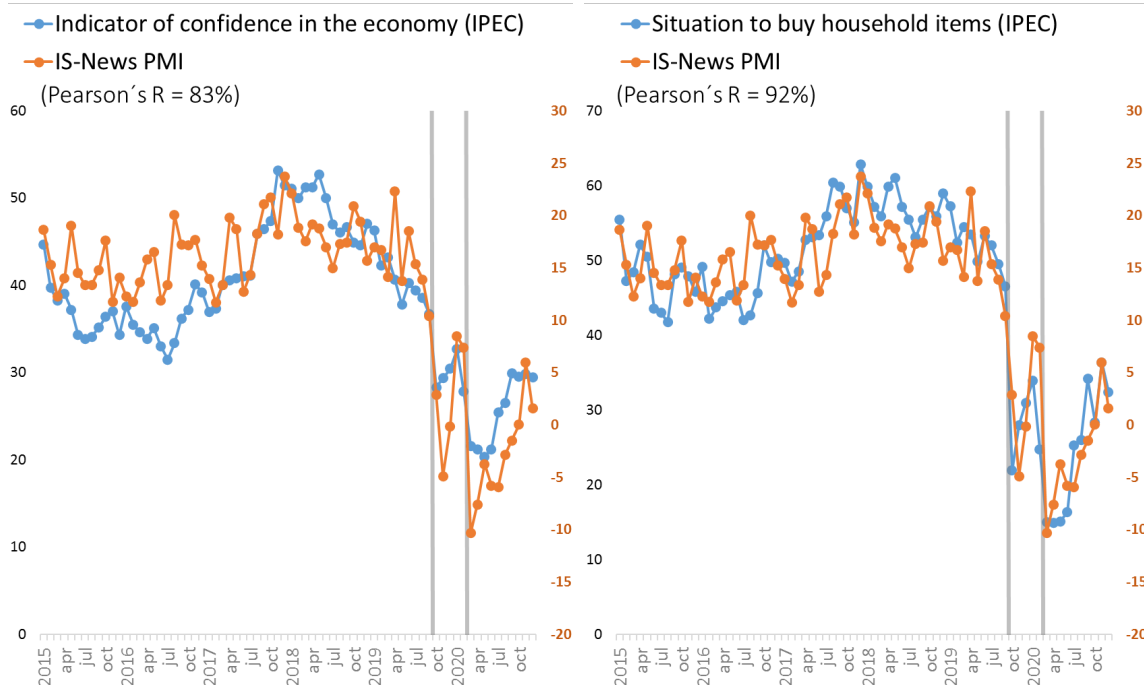
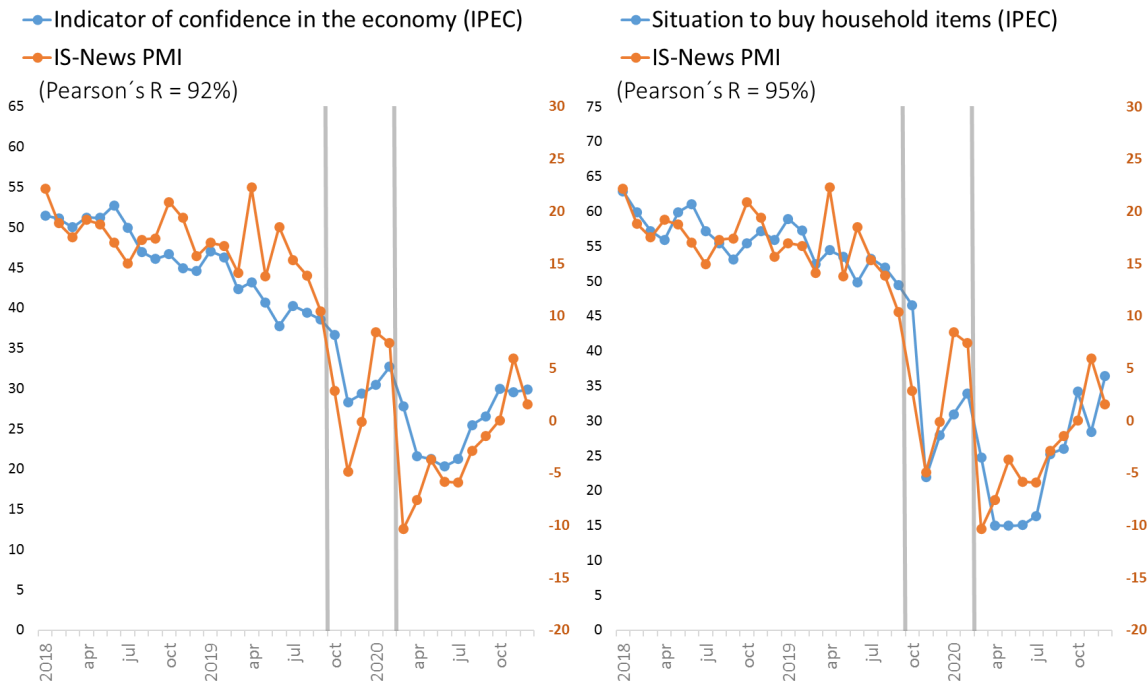


Figure 8

IS-NEWS^{PMI} VS INDICATORS OF CONFIDENCE IN THE ECONOMY (IPEC): 2018-2020

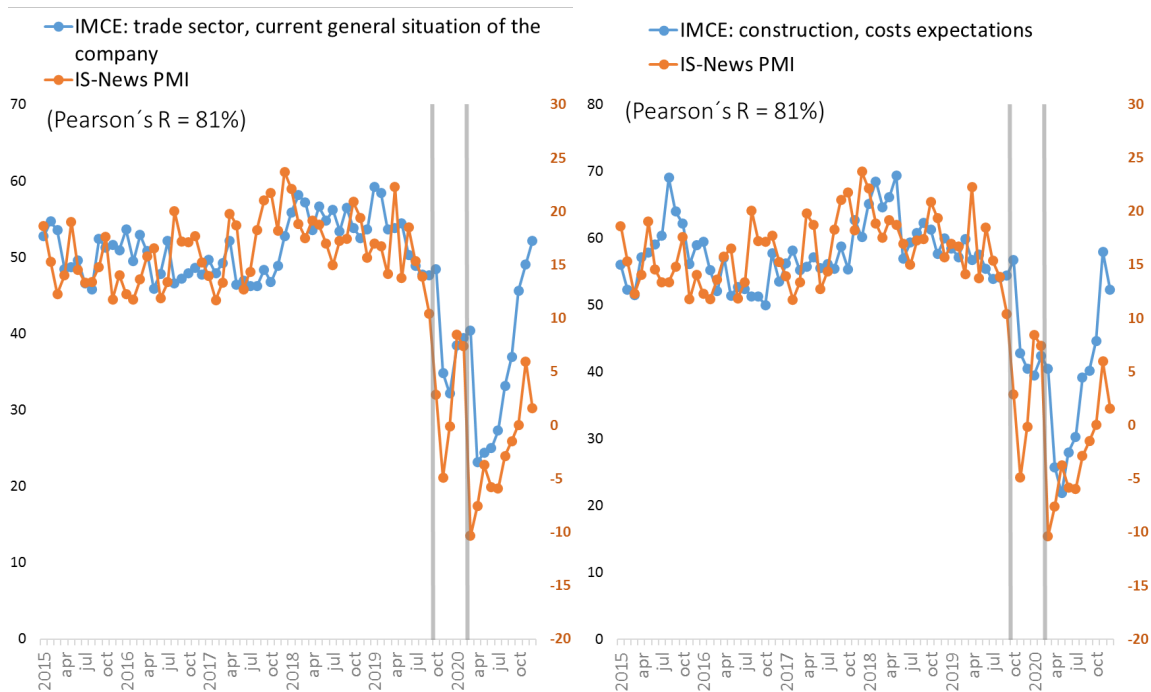
(graphs show contemporaneous series; Pearson's R calculated using IPEC+1)



- When the IS-News^{PMI} is analyzed relative to indicators linked to business confidence surveys, the correlations are weaker than with the Economic Perception Index (IPEC). In fact, the correlation coefficient measured against the Monthly Business Confidence Index (IMCE) and its sectoral indexes, ranges between 50% and 60% and the Business Confidence Index (ICE) stands at 73%. Only a couple of exceptions can be noted in some very specific IMCE sub-indexes in which the Pearson coefficient is higher, one being “general current situation of the firm” in the trade sector, and another being “expected costs” in the construction sector in which, for both, the correlation coefficient climbs to 81% (figure 9).

Figure 9

IS-NEWS^{PMI} VS BUSINESS CONFIDENCE INDICATORS (IMCE): 2015-2020
(contemporaneous series; Pearson's R calculated with IMCE+1)



- When measuring the correlations of IS-News^{PMI} with economic activity indexes, it appears that they are also somewhat lower than those observed with the confidence-in-the-economy index. The Pearson coefficient marks 78% with the headline Imacec, 79% with the non-mining Imacec, and 83% with the services Imacec, all of them displaced one month ahead with respect to the IS-News. In the particular case of the Trade Imacec, the correlation coefficient becomes very low (46%), which has to do with the impulse on consumption during the economy's recovery phase of the pandemic (second half 2020), which is factored into neither the IS-News nor the Economic Perception Index (IPEC) and sub-indexes, which show a rather minor rebound (figure 10).

Figure 10
IS-NEWS^{PMI} VS TRADE IMACEC AND CONFIDENCE-IN-THE-ECONOMY INDICATOR, 2018-2020
 (contemporaneous series; Pearson's R calculated using contemporaneous series)

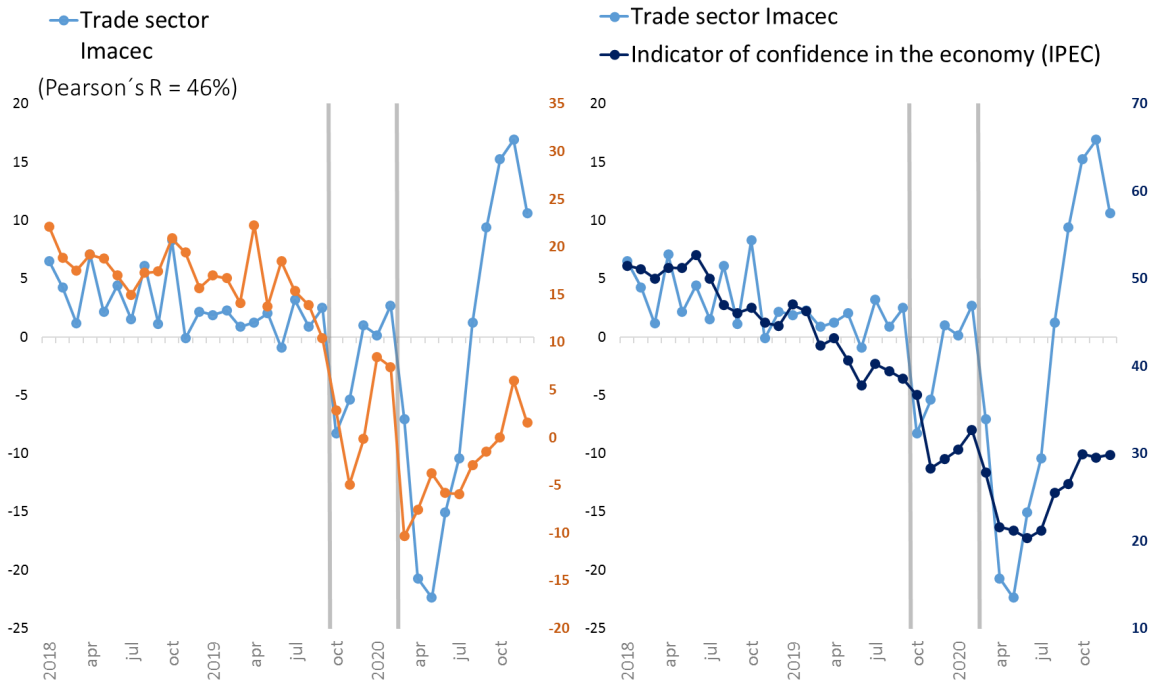
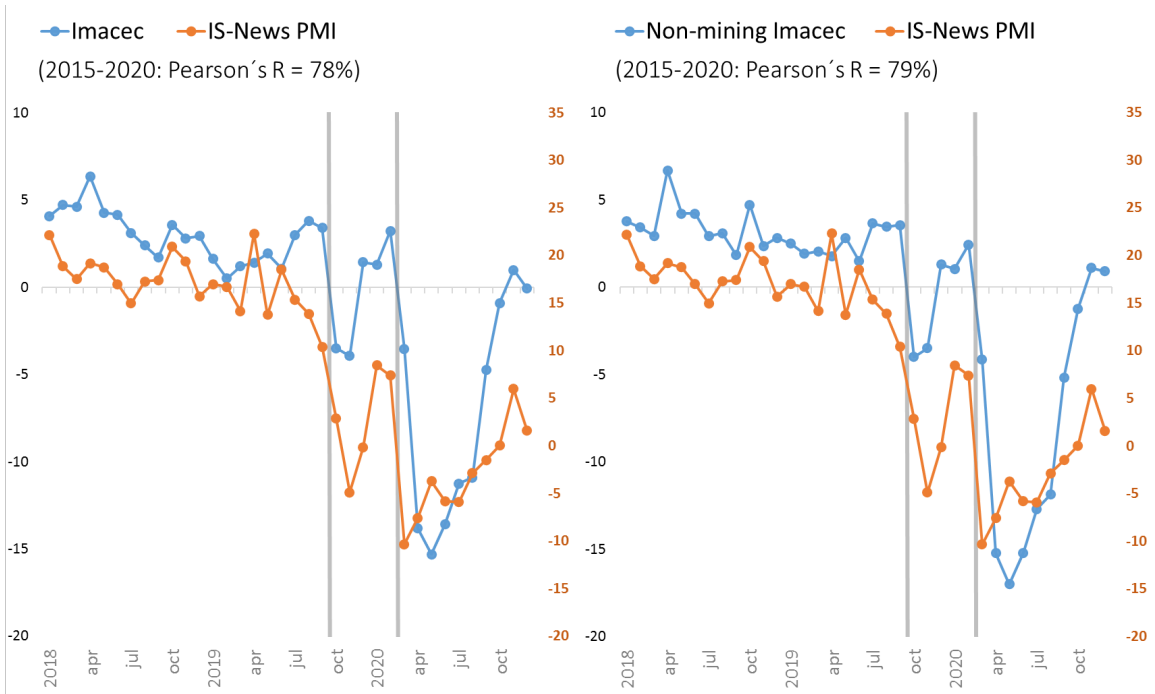


Figure 11
IS-NEWS^{PMI} VS IMACEC AND NON-MINING IMACEC: 2018-2020
 (confidence indices in bases; activity indices in 12-month percent variations; Pearson's R)



- As with the confidence indicators, the IS-News^{PMI} reflects the impact of the social crisis approximately one month in advance of the Imacec, as is also the case with the sanitary crisis. This short-term predictability condition of the IS-News couples with its real-time availability, which ultimately makes it possible to anticipate by roughly two months what the activity indicators will show (figure 11).

In short, the news index developed for the period 2015-2020 shows a high correlation with indicators of citizens' confidence regarding their personal and the country's situation. Its best fit is achieved with expectations indicators referring to consumption intentions, which depend on greater or lower confidence in the conditions of the economy and their own income, and with their perceived economic stability level that allow sustaining said intended consumption over time (figure 12).

This result is consistent with international research findings, which show that the news contributes to the formation of individuals' expectations, which in turn generate information flows that feed the perceptions surveys. Thus IS-News is capable of recording clearly and in advance the confidence shocks that will be later shown in confidence surveys.

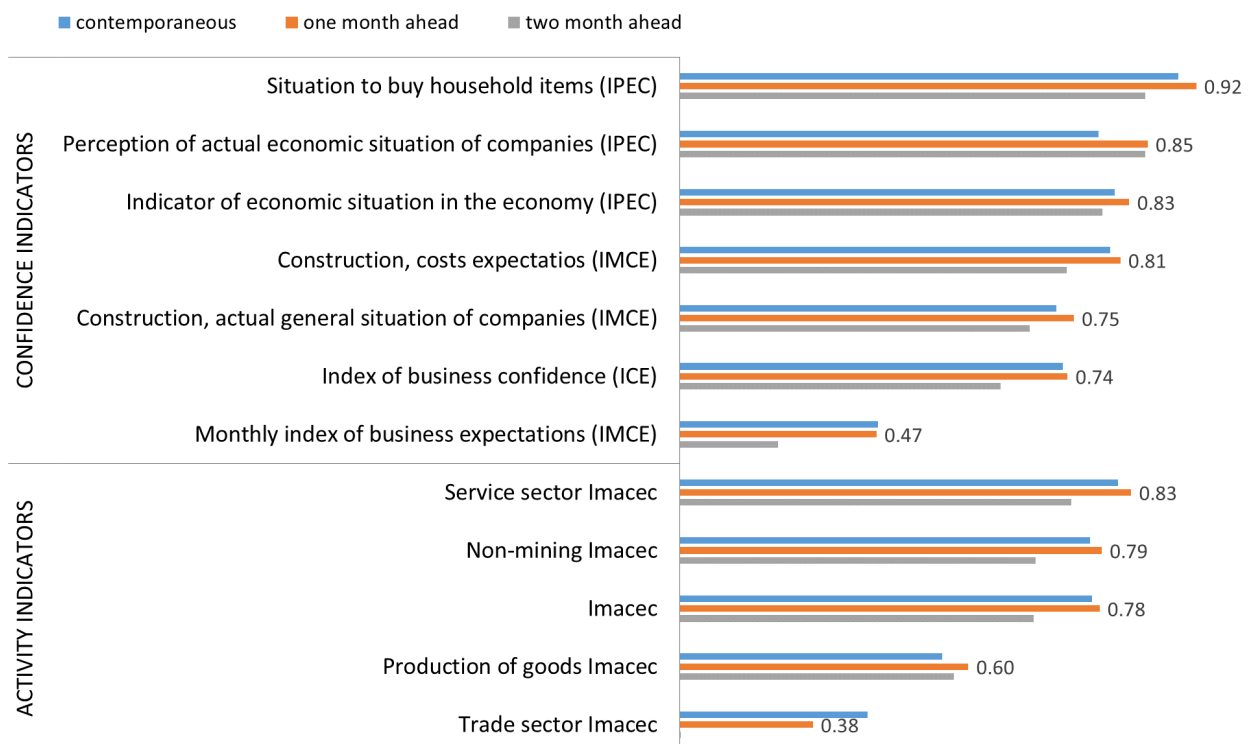
Even though its greatest strength is found with confidence indicators, it also shows satisfactory results with activity indicators, and its greatest fit is found with the services Imacec.

Finally, the predictive capacity of IS-News becomes particularly acute in periods suffering activity and confidence shocks, where correlations of the generality of the indicators increase and very high fits are achieved, especially with expectations ones.

Figure 12

IS-NEWS^{PMI} VS ACTIVITY AND CONFIDENCE INDICATORS: 2015-2020

(confidence indices in bases; activity indices in 12-month percent variations; Pearson's R)



7.1 MEASURING THE DICTIONARY'S ROBUSTNESS

Generating a reliable IS-News-type index requires the use of a dictionary with a sufficiently large number of terms to tolerate repeated iterations without altering its evolution. In this way, testing the robustness of the dictionary is equivalent to demonstrating that none of its terms modifies or conditions the trajectory of the indicator, which remains statistically unaltered in the face of testing modifications in its content.

The first robustness test is based on an exercise similar to the one developed by [Correa, Garud, Londono, & Mislant \(2017\)](#), in which a sensitivity analysis is performed consisting of randomly deleting words from the dictionary, with numerous computational iterations with each extraction exercise. This test allows demonstrating that the dictionary does not have to be exhaustive to be complete and reliable.

Following this line of testing, we proceed to randomly extract 10% and 20% of unique words in the dictionary, using both the simple method and the stratified method (it does not modify the original proportion of positive and negative terms) and calculate the indicator with the remaining words. This stress test is successfully applied, but with some limitations, due to the significant processing cost in each iteration.²⁸ For this reason, the test is applied only on a part (40%) of the total database and is iterated a suboptimal number of times (20), instead of the 1,000 performed by [Correa, Garud, Londono, & Mislant \(2017\)](#) (figure 13).

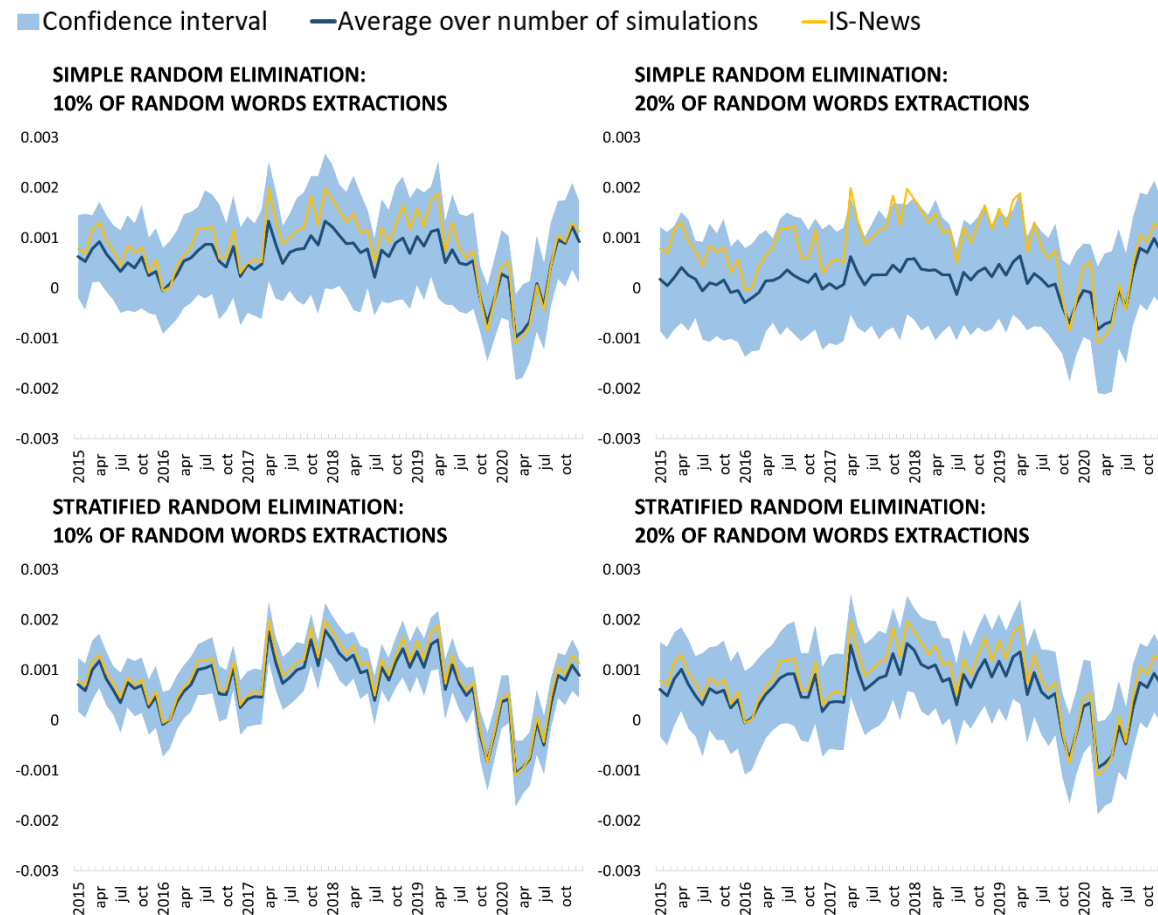
Under these conditions, the results show that the removal of terms does not statistically alter the IS-News, when it is performed in a stratified random fashion (without modifying the original proportion of positive and negative terms in the dictionary), at a 95% confidence level.

Meanwhile, when the terms are deleted in a simple, unstratified way, there is a change in the original structure of the dictionary that generates a reading bias, and the result for IS-News yields in general statistically different versions from the original. Only when the term removal is 10%, i.e., the lowest of the tests, is the robustness procedure satisfactory, suggesting that, when term removal tends to be lower, the structural change in the dictionary tends to be marginal.

These exercises would suggest, in first instance, that the IS-News dictionary is robust, in the sense that there is no evidence that certain words or groups of words are critical for the production of the index, given that by removing up to 20% in a stratified manner, it is possible to construct indicators without statistical differences with the original version. However, a robustness test requires making iterations many more times, which is inapplicable with the available technologies. Therefore, a complementary robustness test is presented below in section 7.2, which increases the iterations up to a thousand times.

²⁸ Processing the base of roughly 220,000 economic and opinion news articles (which yields the best results for IS-News) takes approximately 2.7 days for 100% of the stories, and 1.0 day for 40%, using a server with two Intel® Xeon® Gold 5118 processors and 192 GB of Random-Access Memory (RAM).

Figure 13

ROBUSTNESS TESTS ON THE DICTIONARY**7.2 COMPLEMENTING WITH A TIMELESS ROBUSTNESS MEASURE**

The complementary test, known as timeless robustness, is less demanding in technological terms, but maintains the requirement in statistical terms, being able to increase the number of iterations in the following way:

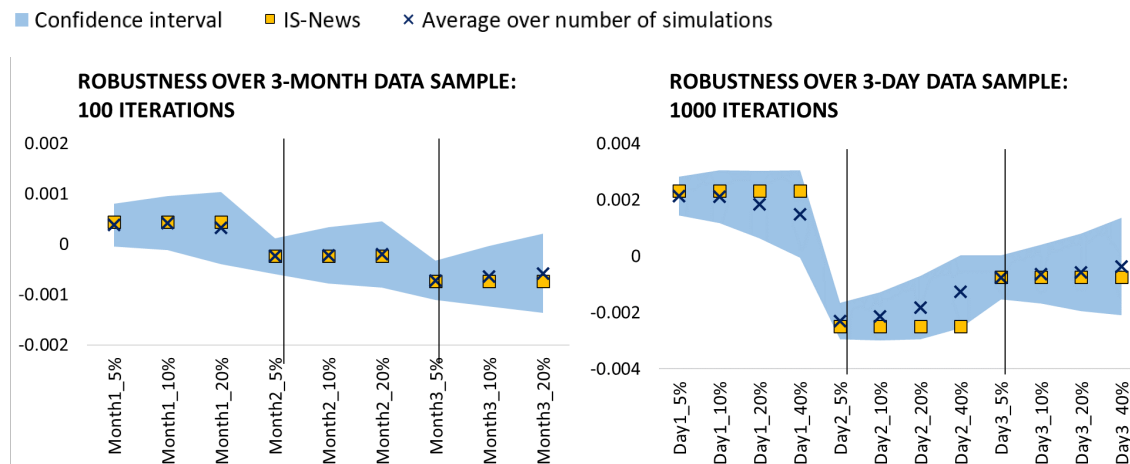
- To go up to 100 iterations, three months of news data are randomly selected, 5%, 10% and 20% of unique words are randomly removed from the dictionary²⁹ and the indicator is recalculated for each of the three randomly selected months with the remaining words.
- To go up to 1,000 iterations, the news base is further reduced by randomly selecting three days' worth of news (60 pieces of news per day on average), randomly removing 5%, 10%, 20% and up to 40% of unique words from the dictionary²⁸ and recalculating the indicator for each of the three randomly selected days with the remaining words.
- Figure 14 presents the results of these robustness tests. It shows that, for both the three-month data sample and the three-day sample, the tested term removals from the dictionary do not statistically alter the IS-News. In the first case, the indicator in its original version (generated with 100% of the words) is within the 95% confidence interval, calculated with the 100 iterations according to

²⁹ Via stratified method.

the removal of 5%, 10% and 20% of terms from the dictionary. In the second case, where up to 40% of dictionary terms are removed in the 1000 iteration calculation procedure, the original indicator also remains within the respective 95% confidence intervals.

Figure 14

TIMELESS ROBUSTNESS TESTS COMPLEMENTARY TO THE DICTIONARY



8. CONCLUSIONS

This research uses the latest methodologies to perform sentiment analysis on voluminous news databases. By computationally reading the country's leading economic and opinion newspapers, we have generated a time series that shows a high correlation with confidence-in-the-economy indicators, and somewhat lower correlations with indicators of economic activity.

The goodness of the results obtained is grounded on the quality of the database constructed, and on the use of a Spanish dictionary, generated from the same news database, extensive in the number of tagged words, complex in its variety of grammatical forms, tested on a sample of manually tagged stories and robust to statistical testing. This Spanish dictionary is an important contribution to text mining research in Spanish, not only because of its size, but also because it allows structuring in various ways the information coming from the written press.

The application of the PMI methodology succeeds in bringing significant improvements to the time series, showing the relevance of incorporating contexts into the process of formulating a manually tagged dictionary. In this sense, it corrects the weaknesses of sentiment analysis methods based solely on dictionaries, allowing the use of a dynamic lexicon, which feeds back as the news base keeps expanding.

In any case, it is important to emphasize that this research made use of the complete database to calculate the PMI, meaning that the valuation of the tagged dictionary words was adjusted according to the usage detected in the entire corpus of news articles. During the period considered, there were two major negative shocks, namely the social outbreak and the pandemic, which may have generated some bias in the application of the PMIs. Therefore, it is proposed here as part of some complementary research, to calculate the optimal and moving PMIs for a time series that uses tagged dictionaries.

The high correlations yielded by the results confirm the international evidence, in the sense that it is possible to computationally capture in real time the level of optimism or pessimism present in the news. There is also evidence that this way of capturing information makes it possible to anticipate confidence shocks, such as those that occurred in 2019 and 2020 in the Chilean economy.

The advantages of having a real-time press indicator are diverse. It provides information that is independent of other sources, has low implementation costs once the methodology has been installed, and can serve as a warning signal of an event of internal or external shock. Moreover, it enables additional applications such as topic analysis, to interpret the unfolding phenomena that accompany the business cycle, as well as the use of bags of words, to obtain intensity measurements.

This notwithstanding, the high correlations between the IS-News and the confidence indicators do not allow inferring that these variables are mutually affecting themselves, meaning that the consumption decisions are affected by the news and vice versa. This would be part of a complementary investigation to this paper.

Finally, the production of the IS-News can serve as a basis for the construction of a model that estimates the way in which sudden confidence-disrupting events detected through the press alter investment and consumption behavior. The early identification of individuals' optimism or pessimism through the IS-News is an intermediate stage between the release of information by the media and the construction of a predictive model about the behavior of economic variables. Modeling this information also represents a way of extending this research and adding value to the fact of having structured information collected from the press.

APPENDIX

Table 1

LEMNAS OF THE TAGGED VERBAL TERMS THAT ORIGINATE THE DICTIONARY³⁰

									(1 from 6)				
abalar	0	bautizar	0	desafiar	0	entrevistar	0	interceptar	0	prevenir	0	sacar	0
abanderar	0	beber	0	desafinar	0	entrevistar	0	interconectar	0	prever	0	saciar	0
abandonar	-1	bendecir	0	desaforar	0	entristecer	0	interesar	0	primar	0	sacrificar	0
abaratar	0	beneficiar	1	desagradar	0	entrometer	0	interferir	0	priorizar	0	sacudir	0
abarcar	0	besar	0	desagregar	0	entronizar	0	intermediar	0	privar	0	saldar	0
abastecer	0	bifar	0	desahogar	0	enturbiar	0	internacionali	0	privatizar	0	salir	0
abatir	0	blandir	0	desahuciar	0	entusiasmar	1	internacionali	0	privilegiar	0	salpicar	0
abdicar	0	blanquear	0	desalentar	-1	enumerar	0	internar	0	probar	0	saltar	0
ablandar	0	blindar	0	desalinizar	-1	enunciar	0	interpelar	0	proceder	0	saludar	0
abocar	0	bloquear	-1	desalinizar	0	envaletonar	0	interponer	0	procesar	0	salvaguardar	0
abogar	0	blumar	0	desalojar	0	envasar	0	interpretar	0	proclamar	0	salvar	1
abolir	0	boicotear	-1	desanclar	0	envajecer	0	interrogar	0	procrear	0	sanar	1
abonar	0	bombarde	0	desangrar	0	envenenar	0	interrumpir	0	procurar	0	sancionar	-1
abordar	0	bombear	0	desanimar	0	enviar	0	intervenir	0	prodigar	0	sanear	1
aborrecer	0	bonificar	0	desaparecer	-1	envidiar	0	intimidar	-1	producir	0	sangrar	0
abortar	0	bordar	0	desaprobar	0	enviudar	0	intoxicar	0	preferir	0	saquear	0
abrazar	0	bucear	0	desaprovecha	0	envolver	0	introducir	0	profesar	0	satisfacer	0
abreviar	0	borrar	0	desarmar	0	equilibrar	0	intuir	0	profesionaliza	0	saturar	0
abrigar	0	botar	0	desarrollar	0	equipar	0	inundar	0	profundizar	0	secar	0
abrir	0	bregar	0	desarticular	0	equiparar	0	inutilizar	0	programar	0	secuenciar	0
abrochar	0	brillar	0	desatar	0	equivaler	0	invadir	-1	progresar	0	secuestrar	0
abrumar	0	brindar	0	desatender	-1	equivocar	0	invalidar	0	prohibir	-1	secundar	0
absolver	0	bromear	0	desautorizar	0	erguir	0	inventar	0	proliferar	0	sedar	0
absorber	0	brotar	0	desayunar	0	erigir	0	invertir	1	prolongar	0	seducir	0
abstener	0	bucear	0	desbanca	0	erogar	0	investigar	0	promediar	0	segmentar	0
abuchear	0	burlar	0	desbandar	0	erosionar	-1	investir	0	prometer	0	segregar	0
abultar	0	buscar	0	desbaratar	0	erradicar	0	invisible	0	promocionar	0	seguir	0
abundar	0	cabalgar	0	desbloquear	0	errar	-1	invitar	0	promover	0	seleccionar	0
aburrir	0	caber	0	desbocar	0	esbozar	0	invocar	0	promulgar	0	sellar	0
abusar	-1	cachar	0	desbordar	0	escabullir	0	involver	0	pronosticar	0	sembrar	0
acabar	0	caducar	0	descabezar	0	escalar	0	inyectar	0	pronunciar	0	sensibilizar	0
acallar	0	caer	-1	descalificar	-1	escandalizar	0	ir	0	propagar	0	sentar	0
acampar	0	cagar	0	descansar	0	escanear	0	ironizar	0	propender	0	sentenciar	0
acaparar	0	calamar	0	descarbonizar	0	escapar	0	irradiar	0	propiciar	0	sentir	0
acariciar	0	calar	0	descargar	0	escarbar	0	irritar	0	propinar	0	señalar	0
acarrear	0	calcular	0	descarrilar	0	escasear	0	irrogar	0	proponer	0	señalar	0
acatar	0	caldear	0	descartar	0	escatimar	0	irrumper	0	proporcionar	0	separar	0
acceder	0	calefactar	0	descender	0	escenificar	0	izar	0	propugnar	0	sepultar	0
accidental	0	calentar	0	descentralizar	0	escindir	0	jactar	0	propulsar	0	ser	0
accionar	0	calibrar	0	descifrar	0	esclarecer	0	jalar	0	prorrogar	0	serpentear	0
acechar	0	calificar	0	desclasificar	0	escoger	0	jerarquizar	0	proscribir	0	servir	0
acelerar	1	callar	0	descolgar	0	escolar	0	joder	0	perseguir	0	sesgar	0
acentuar	0	calmar	0	descolocar	0	escoltar	0	jubilar	0	prospectar	0	sesionar	0
aceptar	0	calzar	0	descomponer	0	escorder	0	judicializar	0	prosperar	0	significar	0
acercar	0	cambiar	0	descomprimir	1	escribir	0	jugar	0	prostituir	0	silenciar	0
acertar	0	caminar	0	desconcentrar	0	escribir	0	juntar	0	protagonizar	0	simbolizar	0
achacar	0	campear	0	desconcertar	0	escrutar	0	juramentar	0	proteger	1	simpatizar	0
achicar	0	camuflar	0	desconectar	0	escuchar	0	jurar	0	protestar	-1	simplificar	0
acicatear	0	canalizar	0	desconfiar	-1	escudar	0	justificar	0	protocolizar	0	simular	0
aclarar	0	cancelar	-1	descongelar	0	escudriñar	0	juzgar	0	proveer	0	sincerar	0
aclarar	0	canjear	0	descongestión	0	esculpir	0	labocar	0	provenir	0	sincronizar	0
acoger	0	cansar	0	desconocer	-1	escupir	0	laborar	0	provisionar	0	sindicar	0
acometer	0	cantar	0	descontaminar	0	escurrir	0	labrar	0	provocar	0	singularizar	0
acomodar	0	capacitar	0	descontar	0	esforzar	0	ladrar	0	proyectar	0	sintetizar	0
acompañar	0	capear	0	descontrolar	0	esfumar	0	lamentar	-1	publicar	0	sintonizar	0
acondicionc	0	capitalizar	0	describir	0	esgrimir	0	lamer	0	publicitar	0	sistematizar	0
aconsejar	0	captar	0	descubrir	0	esmerar	0	languidecer	0	puer	0	sitiar	0

³⁰ The list contains the lemmas of the tagged verbs with which the dictionary is created. Adjectives and adverbs were also tagged. The full lexicon can be found [here](#).

(2 from 6)													
acontecer	0	capturar	0	descuidar	-1	espantar	0	lanzar	0	pugnar	0	situar	0
acopiar	0	caracteriz	0	desdecir	0	esparcir	0	largar	0	pujar	0	sobornar	0
acoplar	0	carecer	-1	desdeñar	0	especializar	0	láser	0	pulir	0	sobrar	0
acordar	0	cargar	0	desdibujar	0	especificar	0	lastimar	0	pulsar	0	sobrecargar	0
acorralar	0	caricaturiz	0	desdramatizar	0	especular	0	lastrar	0	pulular	0	sobredimensior	0
acortar	0	carretear	0	desear	0	esperar	0	latir	0	pulverizar	0	sobreestimar	0
acosar	0	casar	0	desechar	0	espetar	0	lavar	0	puntualizar	0	sobrellevar	0
acostar	0	castigar	-1	desembarcar	0	espigar	0	leer	0	purgar	0	sobrepasar	0
acostumbrar	0	catalizar	0	desembocar	0	esposar	0	legalizar	0	purificar	0	sobreponderar	0
acotar	0	catalogar	0	desembolsar	0	esquiar	0	legar	0	quebrantar	-1	sobreponer	0
acrecentar	0	catapultar	0	desempeñar	0	esquivar	0	legislar	0	quebrar	0	sobrerreaccior	0
acreditar	0	catar	0	desempolvar	0	estabilizar	0	legitimar	0	quedar	0	sobresalir	0
acribillar	0	catastrar	0	desencadenar	0	establecer	0	lesionar	-1	quejar	-1	sobreseer	0
activo	0	categorizc	0	desenmascara	0	estacionar	0	levantar	1	quemar	0	sobrevvalorar	0
actualizar	0	causar	0	desenredar	0	estafar	0	liberalizar	0	querellar	0	sobrevenir	0
actuar	0	cautelar	0	desentender	0	estallar	-1	liberar	0	querer	0	sobrevivir	0
acuchillar	0	cautivar	0	desenterrar	0	estampar	0	librar	0	quintuplicar	0	sobrevolar	0
acudir	0	cavar	0	desentrañar	0	estancar	-1	licenciar	0	quitar	0	socavar	-1
acumular	0	cazar	0	desenvolver	0	estándar	0	licitar	0	racionalizar	0	sociabilizar	0
acunar	0	cebar	0	desequilibrar	-1	estandarizar	0	licuar	0	racionar	0	socializar	0
acunar	0	ceder	0	desertar	0	estar	0	liderar	0	radicalizar	0	socorrer	0
acusar	0	cegar	0	desesperar	0	estafizar	0	lidiar	0	radicar	0	sofisticar	0
adaptar	0	cejar	0	desestabilizar	-1	esterilizar	0	ligar	0	ralentizar	-1	sofocar	0
adecuar	0	celebrar	0	desestimar	0	estigmatizar	0	limar	0	ramificar	0	soler	0
adelantar	0	cenar	0	desfigurar	0	estilar	0	limitar	0	raptar	0	solicitar	0
adelgazar	0	censo	0	desfilar	0	estimar	0	limpiar	0	rascar	0	solidarizar	0
adentrar	0	censurar	0	desfondar	0	estimular	1	linchar	0	rasgar	0	solidificar	0
adeudar	-1	centralizar	0	desgarrar	0	estipular	0	liquidar	0	rasguñar	0	sollozar	0
adherir	0	centrar	0	desgastar	0	estirar	0	listar	0	raspar	0	soltar	0
adicionar	0	ceñir	0	desglosar	0	estorbar	0	litigar	0	rastrear	0	solucionar	1
adiestrar	0	cercar	0	deshacer	0	estrangular	-1	llagar	0	ratificar	0	solventar	0
adivinar	0	cercenar	0	deshidratar	0	estrechar	0	llamar	0	rayar	0	someter	0
adjudicar	0	cernir	0	deshonrar	0	estrellar	0	llegar	0	razonar	0	sonar	0
adjuntar	0	cerrar	0	deshumanizar	0	estremecer	0	llenar	0	reabrir	0	sondear	0
administrar	0	certificar	0	designar	0	estrenar	0	llevar	0	reaccionar	0	soñar	0
admirar	0	cesar	0	desilusionar	0	estresar	0	llorar	-1	reacomodar	0	sopesar	0
admitir	0	chantajec	0	desincentivar	-1	estribar	0	lloriquear	0	reactivar	1	soplar	0
adoctrinar	0	charlar	0	desinfectar	0	estropear	0	lover	0	readecuar	0	soportar	0
adolescer	0	chatear	0	desinflar	0	estructurar	0	localizar	0	reafirmar	0	sorprender	0
adoptar	0	chequear	0	desinformar	0	estrujar	0	lograr	0	reagendar	0	sortear	0
adorar	0	chillar	0	desintegrar	0	estudiar	0	luchar	-1	reagrupar	0	soslayar	0
adormecer	0	chocar	0	desistir	0	eternizar	0	lucir	0	reajustar	0	sospechar	0
adornar	0	chupar	0	deslegitimar	0	etiquetar	0	lucrar	0	realizar	0	sostener	0
adosar	0	cicatrizar	0	desligar	0	evacuar	0	machacar	0	realzar	0	soferrar	0
adquirir	0	cifrar	0	deslindar	0	evadir	0	madurar	0	reanimar	1	star	0
adscribir	0	cimentar	0	deslizar	0	evaluar	0	magister	0	reanudar	0	suavizar	0
aducir	0	circular	0	deslumbrar	0	evangelizar	0	malentender	0	reaparecer	0	subarrendar	0
adueñar	0	circundar	0	desmantelar	-1	evaporar	0	malgastar	0	rearmar	0	subastar	0
adujar	0	circunscrit	0	desmarcar	0	evidenciar	0	malinterpretar	0	rearticular	0	subcontratar	0
adular	0	cifra	0	desmayar	0	evitar	0	maltratar	-1	reassignar	0	subdividir	0
adulterar	0	clamar	0	desmejorar	0	evocar	0	malversar	0	reasumir	0	subestimar	0
advertir	-1	clarificar	0	desmembrar	0	evolucionar	0	manchar	0	reavivar	0	subir	1
afanar	0	clasificar	0	desmentir	0	exacerbar	-1	mancillar	0	rebajar	0	sublimar	0
afectar	0	claudicar	0	desmenuzar	0	exagerar	0	mandar	0	rebalsar	0	subordinar	0
afeitar	0	clausurar	0	desmerecer	0	exaltar	0	manejar	0	rebasar	0	subponderar	0
aferrar	0	clavar	0	desmitificar	0	examinar	0	maniar	0	rebatir	0	subrayar	0
afianzar	1	clonar	0	desmontar	0	exasperar	0	manifestar	0	rebautizar	0	subrogar	0
aficionar	0	coaccionar	0	desmoronar	-1	excavar	0	maniobrar	0	rebelar	0	subsancar	0
afilar	0	coadyuvanc	0	desmovilizar	0	exceder	0	manipular	0	rebosar	0	subscribir	0
afiliar	0	coartar	0	desnaturalizar	0	exceptuar	0	mantener	0	rebotar	0	subsidiar	0
afinar	0	cobijar	0	desnudar	0	excitar	0	manufacturar	0	rebrotar	0	subsistir	0
afirmar	0	cobrar	-1	desobedecer	0	exclamar	0	mapear	0	rebuscar	0	subvalorar	0
afligir	0	cocer	0	desocupar	0	excluir	0	maquillar	0	recabar	0	subvencionar	0
aflojar	0	cocinar	0	desordenar	0	exculpar	0	maravillar	0	recaer	0	subvertir	0
aflorar	0	codear	0	desorientar	0	excusar	0	marcar	0	recalar	0	subyacer	0
aforar	0	codificar	0	despachar	0	exhalar	0	marchar	0	recalcar	0	suceder	0
afrontar	0	coexistir	0	desparramar	0	exhibir	0	marchitar	0	recalcular	0	sucumbir	0
agachar	0	cofinanciar	0	despedir	-1	exhortar	0	marear	0	recalendariza	0	sudar	0
agarrar	0	coger	0	despegar	0	exhumar	0	marginar	0	recalficar	0	sufragar	0
agasajar	0	cohabitar	0	despejar	1	exigir	0	maridar	0	recapacitar	0	sufrir	-1
agilizar	0	cohesionar	0	despenalizar	0	exiliar	0	masacrar	0	recapitalizar	1	sugerir	0
agitar	0	coincidir	0	desperdiciar	0	eximir	0	masificar	0	recargar	0	suicidar	0
aglomerar	0	cojear	0	desperdigar	0	existir	0	masticar	0	recaudar	0	sujetar	0

(3 from 6)													
aglutinar	0	colaborar	0	despertar	0	exonerar	0	matar	-1	recepcionar	0	sumar	0
agobiar	-1	colapsar	-1	despilfarrar	-1	expandir	1	materializar	0	recetar	0	sumergir	0
agolpar	0	colar	0	desplazar	0	expedir	0	matizar	0	rechazar	0	suministrar	0
agonizar	0	coleccion	0	desplegar	0	expende	0	matricular	0	recibir	0	sumir	-1
agotar	0	colegir	0	desplomar	-1	experimentar	0	maximizar	0	reciclar	0	supeditar	0
agradar	0	colgar	0	despojar	0	expirar	0	mear	0	recitar	0	superar	1
agradecer	1	colindar	0	despreciar	0	explayar	0	mecer	0	reclamar	-1	superponer	0
agrandar	0	colisionar	0	desprender	0	explicar	0	mediar	0	recluir	0	supervigilar	0
agravar	-1	colmar	0	desprestigiar	0	explicitar	0	medir	0	reclutar	0	supervisar	0
agraviar	0	colocar	0	desproporcion	0	explorar	0	meditar	0	recobrar	0	suplantar	0
agredir	-1	colonizar	0	despuntar	0	explosar	0	mejorar	1	recoger	0	suplementar	0
agregar	0	colorear	0	desregular	0	exponer	0	mellar	0	recolectar	0	suplicar	0
agrupar	0	coludir	-1	destacar	0	exportar	0	memorizar	0	recomendar	0	suplir	0
aguantar	0	comanda	0	destapar	0	expresar	0	mencionar	0	recomenzar	0	suponer	0
aguar	0	combatir	-1	desteñir	0	expimir	0	menguar	0	recomer	0	suprimir	0
aguardar	0	combinar	0	desterrar	0	expropiar	0	menoscabar	0	recompensar	0	surcar	0
agudizar	-1	comentar	0	destilar	0	expulsar	-1	menospreciar	0	recomponer	0	surgir	0
ahogar	0	comenzar	0	destinar	0	extender	0	mentir	-1	recomprar	0	surtir	0
ahondar	0	comer	0	destituir	0	exteriorizar	0	merecer	0	reconciliar	0	suscitar	0
ahorcar	0	comercial	0	destrabar	1	exterminar	0	mermar	0	reconducir	0	suscribir	0
ahorrar	0	comerciar	0	destronar	0	externalizar	0	merodear	0	reconectar	0	suspender	-1
ahuyentar	0	cometer	-1	destrazar	0	extinguir	0	meter	0	reconfigurar	0	sustanciar	0
aislar	0	comisiona	0	destruir	-1	extirpar	0	mezclar	0	reconfortar	0	sustentar	0
ajustar	0	compade	0	desvalorizar	0	extorsionar	0	migrar	0	reconocer	0	sustituir	0
ajusticiar	0	comparar	0	desvanecer	0	extractar	0	militar	0	reconquistar	0	sustraer	-1
alabar	1	comparec	0	desvelar	0	extrañar	0	militarizar	0	reconsiderar	0	susurrar	0
alardear	0	compartir	0	desvestir	0	extraer	0	mimar	0	reconstituir	0	tachar	0
alargar	0	compatibi	0	desviar	0	extralimitar	0	minar	-1	reconstruir	1	talar	0
alarmar	-1	compensc	0	desvincular	-1	extrañar	0	minimizar	0	reconvertir	0	tallar	0
albergar	0	competer	0	desvirtuar	-1	extrapolar	0	mirar	0	recopilar	0	tambalear	-1
alcanzar	0	competir	0	detallar	0	extraviar	0	mitigar	0	recordar	0	tantear	0
alegar	0	compilar	0	detectar	0	extremar	0	modelar	0	recorrer	0	tapar	0
alegrar	1	complace	0	detener	-1	exudar	0	moderar	0	recortar	-1	tapizar	0
alejar	0	complejizc	0	detentar	0	fabricar	0	modernizar	0	recrear	0	tararear	0
alentar	0	compleme	0	deteriorar	-1	facilitar	0	modificar	0	recriminar	0	tardar	0
alertar	-1	completar	0	determinar	0	facturar	0	modular	0	recrudecer	-1	tasar	0
aliar	0	complicar	-1	detestar	0	facturar	0	mojar	0	rectificar	0	tatar	0
alienar	0	complotar	0	defonar	0	faenar	0	moldear	0	recubrir	0	teclear	0
aligerar	0	componer	0	devaluar	0	fallar	0	moler	0	recular	0	tecnificar	0
alimentar	0	comporta	0	devascular	-1	fallecer	-1	molestar	0	recuperar	1	tejer	0
alineal	0	comprar	0	develar	0	fallir	0	monetizar	0	recurrir	0	telar	0
alistar	0	comprenc	0	devenegar	0	falsar	0	monitorear	0	recusar	0	telefonar	0
alivianar	0	comprimir	0	devenir	0	falsificar	0	monopolizar	0	redactar	0	televisar	0
aliviar	1	comprobc	0	devolver	0	faltar	-1	montar	0	redefinir	0	temblar	0
allanar	0	compromi	0	devorar	0	familiarizar	0	morder	0	redescubrir	0	temer	-1
allegar	0	computar	0	diagnosticar	0	fantasear	0	morigerar	0	redestinar	0	temperar	0
almacenar	0	comulgar	0	dialogar	0	farrear	0	morir	-1	redimir	0	templar	0
almorzar	0	comunica	0	dibujar	0	fascinar	0	mostrar	0	redireccionar	0	tender	0
alojar	0	concebir	0	dictaminar	0	fastidiar	0	motejar	0	redirigir	0	tener	0
alquilar	0	conceder	0	dictar	0	favorecer	1	motivar	1	rediseñar	0	tensar	0
altamar	0	concentrc	0	diezmar	0	fechar	0	mover	0	redistribuir	0	tensionar	-1
alterar	-1	conceptu	0	difamar	0	felicitar	1	movilizar	0	redituar	0	tentar	0
alternar	0	concernir	0	diferenciar	0	feriar	0	mudar	0	redoblar	0	teñir	0
alucinar	0	concertar	0	diferir	0	fertilizar	0	multar	-1	redondear	0	teorizar	0
aludir	0	concesior	0	dificultar	-1	festejar	1	multiplicar	0	reducir	-1	tergiversar	0
alumbrar	0	concientiz	0	difuminar	0	fiar	0	murmurar	0	redundar	0	terminar	0
alzar	0	conciliar	0	difundir	0	fichar	0	mutar	0	reeditar	0	testear	0
amagar	0	concitar	0	digerir	0	fidelizar	0	mutilar	0	reelegir	0	testificar	0
amainar	0	concluir	0	digitalizar	0	figurar	0	nacer	0	reembolsar	0	testimoniar	0
amalgamar	0	concorda	0	digital	0	fijar	0	nacionalizar	0	reemplazar	0	tildar	0
amamantar	0	concretar	0	dignificar	0	filmar	0	nadar	0	reemprender	0	timbrar	0
amanecer	0	conculcar	0	dilapidar	0	filtrar	0	narrar	0	reencantar	0	tipificar	0
amar	0	concurrir	0	dilatar	0	finalizar	0	naturalizar	0	reencarnar	0	tirar	0
amargar	0	concurrir	0	dilucidar	0	financiar	0	naufragar	0	reencontrar	0	tiritar	0
amarrar	0	condecir	0	diluir	0	finjir	0	navegar	0	reenfocar	0	titubear	0
amasar	0	condecor	0	dimensionar	0	finiquitar	0	necesitar	0	reenviar	0	titular	0
ambicionar	0	condena	-1	diminuir	0	firmar	0	negar	-1	reequilibrar	0	tocar	0
ambientar	0	condensa	0	dimitir	0	fiscalizar	0	negociar	0	reescribir	0	tolerar	0
amedrentar	-1	condicion	0	dinamitar	0	flamear	0	neutralizar	0	reestablecer	0	tomar	0
amenazar	-1	condonar	0	dinamizar	1	flanquear	0	nevar	0	reestructurar	0	tonificar	0
amenizar	0	conducir	0	direccionar	0	flaquear	0	nidificar	0	reevaluar	0	topar	0
ameritar	0	conectar	0	dirigir	0	flexibilizar	0	ningunear	0	reexaminar	0	torcer	0
amilanar	0	confeccion	0	dirimir	0	floreecer	0	nivelar	0	referir	0	tornar	0

(4 from 6)													
aminorar	0	conferir	0	discernir	0	flotar	0	nombrar	0	refichar	0	torpedear	0
amnistiar	0	confesar	0	disciplinar	0	fluctuar	0	nominar	0	refinanciar	0	torturar	0
amoldar	0	confiar	1	discontinuar	0	fluir	0	normalizar	0	refinar	0	toser	0
amonestar	0	configurar	0	discrepar	0	focalizar	0	normar	0	reflejar	0	totalizar	0
amordazar	0	confinar	0	discriminar	0	fomentar	1	notar	0	reflexionar	0	trabajar	0
amortiguar	0	confirmar	0	disculpar	0	fondear	0	notificar	0	reflotar	0	trabar	-1
amortizar	0	confiscar	0	discurrir	0	forjar	0	nublar	0	reforestar	0	traducir	0
amparar	0	confluir	0	discutir	0	formalizar	0	nutrir	0	reformalizar	0	traer	0
ampliar	0	conformar	0	diseminar	0	formar	0	obedecer	0	reformar	0	traficar	0
amplificar	0	confrontar	0	disentir	0	formular	0	objetar	0	reformular	0	tragar	0
amputar	0	confundir	-1	diseñar	0	fortalecer	1	obligar	0	reforzar	0	traicioner	-1
analizar	0	congelar	0	disfrazar	0	forzar	-1	obrar	0	refrendar	0	tramar	0
anclar	0	congeniar	0	disfrutar	1	fotografiar	0	obsequiar	0	refrescar	0	tramitar	0
andar	0	congestio	0	disgustar	0	fracasar	-1	observar	0	refrigerar	0	tranquilizar	1
anegar	0	congratula	0	disimular	0	fracccionar	0	obsesionar	0	refugiar	0	trancar	0
anestesiarse	0	congrega	0	disipar	0	fracturar	0	obstaculizar	-1	refundar	0	transcender	0
anexar	0	conjugarse	0	disminuir	-1	fragmentar	0	obstruir	-1	refundir	0	transcribir	0
angustiar	0	conjurar	0	disociar	0	fraguar	0	obtener	0	refutar	0	transcurrir	0
anhelar	0	conllevar	0	disolver	0	frecuentar	0	obviar	0	regalar	0	transferir	0
anidar	0	conmemo	0	dispar	0	fregar	0	ocasionar	0	regar	0	transformar	0
animar	0	conminar	0	disparar	-1	frenar	-1	ocultar	-1	regenerar	0	transgredir	-1
aniquilar	0	conmocion	0	dispensar	0	fructificar	0	ocupar	0	regir	0	transigir	0
anocheecer	0	conmover	0	dispersar	0	fruncir	0	ocurrir	0	registrar	0	transitar	0
anotar	0	conmutar	0	disponer	0	frustrar	0	odiar	0	reglamentar	0	transmitir	0
ansiar	0	connotar	0	disponibilizar	0	fugar	-1	ofender	0	regresar	0	transparentar	0
anteceder	0	conocer	0	disputar	0	fumar	0	ofertar	0	regular	0	transparentar	0
anteponer	0	conquistar	0	distanciar	0	fumigar	0	oficializar	0	regularizar	0	transportar	0
anticipar	0	consagrar	0	distar	0	funcionar	0	oficiar	0	rehabilitar	0	trascender	0
antojar	0	conseguir	0	distender	0	fundamentar	0	ofrecer	0	rehacer	0	trascorrir	0
anular	0	consensuc	1	distinguir	0	fundar	0	ofuscar	0	rehuir	0	trasformar	0
anunciar	0	consentir	0	distorsionar	0	fundir	0	oler	0	rehusar	0	trasgredir	0
añadir	0	conservar	0	distraer	0	fungir	0	olfatear	0	reimpulsar	1	trasladar	0
añorar	0	considerar	0	distribuir	0	fusilar	0	olvidar	-1	reinar	0	traslapar	0
apabullar	0	consignar	0	disuadir	0	fusionar	0	omitir	0	reincidir	0	traslucir	0
apaciguar	1	consistir	0	diversificar	0	fustigar	0	ondear	0	reincorporar	0	transmitir	0
apadrinar	0	consolar	0	divertir	0	galardonar	0	opacar	0	reingresar	0	transparentar	0
apagar	0	consolidar	0	dividir	-1	ganar	1	operar	0	reiniciar	0	traspasar	0
apalancar	0	conspirar	-1	divisar	0	garantizar	1	opinar	0	reinscribir	0	trasplantar	0
aparcar	0	constar	0	divorciar	0	gastar	0	oponer	-1	reinsertar	0	trastocar	0
aparecer	0	constatar	0	divulgar	0	gatillar	0	oprimir	0	reinstalar	0	trasuntar	0
aparejar	0	constituir	0	doblar	0	generalizar	0	optar	0	reinstaurar	0	tratar	0
aparentar	0	constreñir	0	doblegar	0	generar	0	optimizar	1	reintegrar	0	trazar	0
apartar	0	construir	0	doctorar	0	germinar	0	orar	0	reinterpretar	0	trepas	0
apasionar	0	consultar	0	documentar	0	gestar	0	orbitar	0	reintroducir	0	trepidar	0
apedrear	0	consumar	0	doler	-1	gesticular	0	ordenar	0	reinventar	0	triangular	0
apelar	-1	consumir	0	domar	0	gestionar	0	ordenar	0	reinvertir	1	tributar	0
apenar	0	contabilizar	0	domesticar	0	girar	0	organizar	0	reiterar	0	triplicar	0
apersonar	0	contactar	0	dominar	0	globalizar	0	orientar	0	reivindicar	0	triunfar	1
apetecer	0	contagiar	-1	donar	0	glorificar	0	originar	0	rejuvenecer	0	trizar	0
apilar	0	contaminar	-1	dormir	0	gobernar	0	orinar	0	relacionar	0	tropezar	0
aplacar	0	contar	0	dosificar	0	golpear	0	orquestrar	0	relajar	0	trotar	0
aplanar	0	contemplar	0	dotar	0	golpear	-1	osar	0	relanzar	0	truncar	0
aplastar	0	contender	0	dramatizar	0	gozar	0	oscilar	0	relatar	0	tuitear	0
aplaudir	1	contener	1	drenar	0	grabar	0	oscurecer	0	relativizar	0	tumbar	0
aplazar	0	contentar	0	drogar	0	graduar	0	ostentar	0	releer	0	turnar	0
aplicar	0	contestar	0	duchar	0	graficar	0	otorgar	0	relegar	0	ubicar	0
apodar	0	contextuar	0	dudar	0	grafificar	0	ovacionar	0	relevar	0	ufanar	0
apoderar	0	continuar	0	duplicar	0	gravar	0	oxigenar	0	relicitar	0	ultimar	0
aportar	1	contraata	0	durar	0	gravitar	0	pacificar	0	rellenar	0	ungir	0
apostar	0	contradec	0	echar	0	gritar	0	pactar	0	relocalizar	0	unificar	0
apoyar	1	contraer	-1	eclipsar	0	guardar	0	padecer	0	relucir	0	uniformar	0
apreciar	0	contrapes	0	economizar	0	guiar	0	pagar	0	remar	0	unir	0
aprehender	0	contrapor	0	edificar	0	gustar	0	palear	0	remarcar	0	universalizar	0
apremiar	0	contrariar	0	editar	0	haber	0	paliar	0	rematar	0	untar	0
aprender	0	contrarres	0	educar	0	habilitar	0	palidecer	0	remecer	0	urbanizar	0
apresar	0	contrastar	0	efectuar	0	habitar	0	palpar	0	remediar	0	urdir	0
aprestar	0	contratar	0	egresar	0	habituarse	0	parafrasear	0	rememorar	0	urgir	0
apresurar	0	contraver	0	ejecutar	0	hablar	0	paralizar	-1	remitir	0	usar	0
apretar	0	contribuir	1	ejecutoriar	0	hacer	0	parapetar	0	remodelar	0	usufructuar	0

(5 from 6)													
aprobar	0	controlar	0	ejemplificar	0	hackear	0	parar	0	remontar	0	usurpar	0
aprontar	0	controver	0	ejercer	0	halagar	0	parchar	0	remover	0	utilizar	0
apropiar	0	conválida	0	ejercitar	0	hallar	0	parear	0	reemplazar	0	vacar	0
aprovechar	0	convence	0	elaborar	0	hartar	0	parecer	0	remunerar	0	vaciar	0
aproximar	0	convenir	0	electricificar	0	heredar	0	parir	0	renacer	0	vacilar	0
apuntalar	0	converge	0	elegir	0	herir	-1	participar	0	rendir	0	vacunar	0
apuntar	0	convergir	0	elegar	0	hervir	0	partir	0	renegar	0	vadear	0
apuñalar	0	conversar	0	eliminar	-1	hilar	0	pasar	0	renegociar	0	vagar	0
apurar	0	converfir	0	elogiar	1	hilvanar	0	pasear	0	renovar	1	valer	0
aquejar	0	convidar	0	eludir	-1	hinchar	0	pastar	0	rentabilizar	0	validar	0
aquejar	0	convivir	0	emanar	0	hipotecar	0	pastear	0	rentar	0	valorar	0
aquilatar	0	convocar	0	embarazar	0	hojear	0	patentar	0	renunciar	-1	valorizar	0
arañar	0	convulsior	0	embarcar	0	homenajear	0	patinar	0	reñir	0	vanagloriar	0
arbitrar	0	cooperar	0	embargar	0	homogeneizc	0	patrocinar	0	reordenar	0	vapulear	0
archivar	0	cooprar	0	embellecer	0	homologar	0	patrullar	0	reorganizar	0	variar	0
arder	0	coordinar	0	embestir	0	honrar	0	pausar	0	reorientar	0	vaticinar	0
arengar	0	copar	0	embocar	0	horadar	-1	paufear	0	repactar	0	velar	0
argüir	0	copiar	0	embolsar	0	hornear	0	pavimentar	0	reparar	0	vencer	1
argumentar	0	coquetea	0	emborrachar	0	horrorizar	0	pecar	0	repartir	0	vender	0
armar	0	corear	0	emboscar	0	hospedar	0	pedalear	0	repasar	0	venerar	0
armonizar	0	coronar	0	emerger	0	hospitalizar	0	pedir	0	repatriar	0	vengar	0
arraigar	0	corregir	-1	emigrar	0	hostigar	0	pegar	0	repeler	0	venir	0
arrancar	0	correlacic	0	emitir	0	huir	0	peinar	0	repensar	0	ventilar	0
arrasar	0	correr	0	emocionar	0	humanizar	0	pelar	0	repercutir	0	ver	0
arrastrar	-1	correspon	0	empacar	0	humedecer	0	pelear	-1	repetir	0	veranear	0
arrebatar	0	corretear	0	empadronar	0	humillar	-1	pelegrar	-1	replantar	0	verbalizar	0
arreciar	0	corroborar	0	empalmar	0	hundir	-1	penalizar	-1	replantear	0	verificar	0
arreglar	0	corroer	-1	empantanar	0	hurgar	0	penar	0	replegar	0	versar	0
arremeter	0	corrompe	0	empañar	-1	hurtar	0	pender	0	replicar	0	verter	0
arrender	0	costar	0	empapar	0	hustear	0	penetrar	0	repoblar	0	vestir	0
arrepentir	0	cortejar	0	empaquetar	0	idealizar	0	pensar	0	reponer	0	vetar	0
arrestar	0	cosechar	0	emparejar	0	idear	0	pensionar	0	reportar	0	viabilizar	0
arribar	0	coser	0	emparentar	0	identificar	0	percatar	0	reportear	0	viajar	0
arriesgar	0	costar	0	empatar	0	ideologizar	0	percibir	0	reposar	0	vibrar	0
arrimar	0	costear	0	empatizar	0	idolotrar	0	perder	-1	reposicionar	0	viciar	0
arrinconar	0	costillar	0	empeccinar	0	ignorar	0	perdonar	0	repostular	0	vigilar	0
arrodillar	0	cotejar	0	empeñar	0	igualar	0	perdurar	0	reprender	0	vigorizar	1
arrojar	0	cotizar	0	empeorar	-1	iluminar	0	perecer	0	representar	0	vincular	0
arrojar	0	crear	1	empezar	0	ilusionar	0	peregrinar	0	reprimir	0	violar	-1
arrollar	0	crecer	1	empinar	0	ilustrar	0	perfeccionar	0	reprobar	0	violentar	0
arrugar	0	creer	0	emplazar	0	imaginar	0	perfilar	0	reprochar	-1	virar	0
arruinar	-1	criar	0	emplear	0	imitar	0	perforar	0	reproducir	0	visar	0
articular	0	criminaliza	0	empobrecer	-1	impactar	0	perjudicar	-1	reprogramar	0	visibilizar	0
asalarciar	0	crispar	0	empoderar	0	impar	0	permanecer	0	repudiar	0	visitar	0
asaltar	0	cristalizar	0	emprender	0	impartir	0	permear	0	repuntar	1	vislumbrar	0
asar	0	criticar	0	empujar	0	impedir	-1	permitir	0	requerir	0	visualizar	0
ascender	0	crucificar	0	emular	0	imperar	0	pernoctar	0	requisar	0	vitorear	0
asear	0	crujir	0	enajenar	0	impermeabiliz	0	perpetrar	-1	resaltar	0	vitrinear	0
asediar	0	cruzar	0	enaltecer	0	implantar	0	perpetuar	0	resarcir	0	vivir	0
asegurar	0	cuadrar	0	enamorar	0	implementar	0	perseguir	0	resbalar	0	vocear	0
asemejar	0	cuadriplic	0	enarbolar	0	implicar	0	perseverar	0	rescatar	0	vociferar	0
asentar	0	cuadriplic	0	encabezar	0	implorar	0	persistir	0	rescindir	0	volar	0
asentir	0	cuajar	0	encadenar	0	imponer	0	personalizar	0	resentir	-1	volcar	0
asesinar	-1	cuantifica	0	encajar	0	imponer	0	personificar	0	reseñar	0	voltear	0
asesorar	0	cubrir	0	encaminar	0	imposibilitar	0	persuadir	0	reservar	0	volver	0
asestar	0	cuestionar	-1	encandilar	0	impregnar	0	perfenecer	0	resfriar	0	vomitir	0
aseverar	0	cuidar	1	encantar	1	impresionar	0	perturbar	0	resguardar	1	votar	0
asfixiar	-1	culminar	0	encañonar	0	imprimir	0	pesar	0	residir	0	vulnerar	-1
asignar	0	culpar	-1	encapsular	0	improvisar	0	pescar	0	resignar	0	yocer	0
asilar	0	cultivar	0	encaramar	0	impugnar	0	pesquisar	0	resignificar	0	zafar	0
asimilar	0	cumplir	1	encarar	0	impulsar	1	picar	0	resistir	0	zambullir	0
asir	0	cundir	0	encarcelar	0	imputar	0	picotear	0	resolver	0	zanjar	0
asistir	0	curar	0	encarecer	0	inaugurar	0	pifiar	0	resonar	0	zarpar	0
asociar	0	cursar	0	encargar	0	incautar	0	pillar	0	respaldar	1	zigzaggear	0
asolar	0	custodiar	0	encarnar	0	incendiar	0	pilotar	0	respear	0		
asomar	0	dañar	-1	encasillar	0	incentivar	1	pilotear	0	respetar	1		
asombrar	0	dar	0	encausar	0	incidir	0	pinchar	0	respirar	0		
aspirar	0	datar	0	encauzar	0	incinerar	0	pintar	0	responder	0		
asumir	0	deambula	0	encender	0	incitar	0	piratear	0	responsabiliza	0		
asustar	-1	debatir	0	encerrar	0	inclinat	0	pisar	0	resquebrajar	0		

(6 from 6)											
atacar	-1	deber	0	enchufar	0	incluir	0	pisotear	0	restablecer	1
atajar	0	debilitar	-1	encoger	0	incomodar	0	placer	0	restar	-1
atañer	0	debutar	0	encomendar	0	incorporar	0	plagar	0	restaurar	1
atar	0	decaer	-1	encontrar	0	incrementar	1	planear	0	restituir	0
atascar	-1	decantar	0	encriptar	0	increpar	0	planificar	0	restringir	-1
atemorizar	-1	decapitar	0	encuadrar	0	incriminar	0	plantar	0	resucitar	0
atender	0	decepcio	-1	encubrir	-1	incrustar	0	plantear	0	resultar	0
atener	0	decidir	0	encuestar	0	incubar	0	plasmear	0	resumir	0
atentar	-1	decir	0	encumbrar	0	inculcar	0	platear	0	resurgir	0
atenuar	0	declamar	0	enderezar	1	inculpar	0	platicar	0	retar	0
aterrar	0	declarar	0	endeudar	0	incumbir	0	plebiscitar	0	retardar	0
aterrizar	0	declinar	0	endosar	0	incumplir	-1	plegar	0	retener	0
aterrorizar	0	decodificar	0	endulzar	0	incurrir	0	poblar	0	retirar	0
atesorar	0	decomisar	0	endurecer	-1	incursionar	0	podar	0	retocar	0
atestiguar	0	deconstru	0	enemistar	0	indagar	0	poder	0	retomar	0
atiborrar	0	decorar	0	energizar	0	indemnizar	0	polarizar	-1	retorcer	0
atinar	0	decrecer	0	enfatar	0	independizar	0	polemizar	0	retornar	0
atisbar	0	decretar	0	enfermar	0	indexar	0	polinizar	0	retractar	0
atizar	-1	dedicar	0	enfilear	0	indicar	0	politizar	0	retraer	0
atomizar	0	deducir	0	enfocar	0	indignar	0	pololear	0	retransmitir	0
atormentar	0	defender	0	enfrascar	0	individualizar	0	ponderar	0	retrasar	0
atracar	0	defenestrar	0	enfrentar	0	inducir	0	poner	0	retratar	0
atraer	0	definir	0	enfriar	0	indultar	0	pontificar	0	retribuir	0
atrapar	0	deformar	0	enfurecer	0	industrializar	0	popularizar	0	retroalimentar	0
atrascar	0	defraudar	-1	enganchar	0	infectar	0	porfiar	0	retroceder	-1
atravesar	0	degenerar	0	engañar	-1	inferir	0	portar	0	retrotraer	0
atrever	0	degradar	0	engendrar	0	infiltrar	0	posar	0	retrucar	0
atribuir	0	degustar	0	englobar	0	inflamar	0	poseer	0	retuitear	0
atrincherar	0	dejar	0	engordar	0	inflar	0	posibilitar	0	retumbar	0
atrofiar	0	delatar	0	engrandecer	0	infligir	-1	posicionar	0	reubicar	0
atropellar	0	delegar	0	engrosar	0	influnciar	0	posponer	0	reunir	0
auditar	0	deleitar	0	engullir	0	influir	0	postear	0	reutilizar	0
augurar	0	deliberar	0	enjuiciar	0	informar	0	postergar	-1	rev alidar	0
aular	0	delimitar	0	enlazar	0	infringir	-1	postular	0	revalorizar	1
aumentar	1	delinear	0	enlodar	0	infundir	0	potenciar	1	revaluar	0
aunar	0	delinquir	0	enloquecer	0	ingeniar	0	practicar	0	revelar	0
auscultar	0	demandar	0	enmarcar	0	ingerir	0	precarizar	-1	revender	0
ausentar	0	demarcar	0	enmascarar	0	ingresar	0	precaver	0	reventar	0
auspicar	0	demasiar	0	enmendar	0	inhabilitar	0	preceder	0	reverdecer	0
autodefinir	0	democrat	0	enmudecer	0	inhalar	0	preciar	0	reversar	0
autogestionar	0	demoler	0	enojar	0	inhibir	0	precipitar	0	revertir	0
automatizar	0	demonizar	0	enorgullecer	1	iniciar	0	precisar	0	revestir	0
autorizar	0	demorar	-1	enquistar	0	injuriar	0	predecir	0	revisar	0
autorregular	0	demostrar	0	enraizar	0	inmiscuir	0	predeterminar	0	revisitar	0
auxiliar	0	denegar	0	enrarecer	0	inmolar	0	predicar	0	revisitar	0
avalar	0	denigrar	0	enredar	0	inmortalizar	0	predisponer	0	revitalizar	1
avaluar	0	denominar	0	enrielar	0	inmovilizar	0	predominar	0	revivir	0
avanzar	0	denostar	-1	enriquecer	0	inmunizar	0	prefabricar	0	revocar	0
avasallar	0	denotar	0	enrolar	0	inmutar	0	preferir	0	revolotear	0
avecinar	0	densificar	0	enrostrar	0	innovar	0	pregonar	0	revolucionar	0
avenir	0	denunciar	-1	ensalzar	0	inquieta	0	preguntar	0	revolver	0
aventajar	0	deparar	0	ensamblar	0	inquirir	0	prejuzgar	0	rezar	0
aventurar	0	depender	0	ensanchar	0	inscribir	0	premiar	1	ridiculizar	0
avergonzar	0	deplorar	0	ensañar	0	insertar	0	prender	0	rigidizar	0
averiguar	0	deponer	0	ensayar	0	insinuar	0	preocupar	-1	rivalizar	0
avisar	0	deportar	0	enseñar	0	insistir	0	prepagar	0	robar	-1
avistar	0	depositar	0	enseñorear	0	inspeccionar	0	preparar	0	robustecer	1
avivar	0	depreciar	0	ensombrecer	-1	inspirar	0	presagiar	0	rociar	0
avizorar	0	deprimir	-1	ensuciar	-1	instalar	0	prescindir	0	rodar	0
ayudar	1	depurar	0	entablar	0	instar	0	prescribir	0	rodear	0
azotar	0	derivar	0	entender	0	instaurar	0	presenciar	0	rogar	0
azucar	0	derogar	0	enterar	0	instigar	0	presentar	0	romper	0
bailar	0	derramar	0	enterrar	0	institucionalizar	0	presentir	0	roncar	0
bajar	-1	derretir	0	entonar	0	instituir	0	preservar	0	rondar	0
balancear	0	derribar	0	entorpecer	-1	instruir	0	presidencializar	0	rotar	0
balear	0	derrocar	0	entrampar	0	instrumentalizar	0	presidir	0	rotular	0
bañar	0	derrochar	0	entrañar	0	insultar	0	presionar	-1	rozar	0
barajar	0	derrotar	0	entrar	0	integrar	0	prestar	0	rubricar	0
barrer	0	derrumbar	-1	entrecruzar	0	intensificar	0	prestigiar	0	rugir	0
bartender	0	desacelerar	-1	entregar	0	intentar	0	presumir	0	rumiar	0
basar	0	desaconsejar	0	entrelazar	0	interactuar	0	presuponer	0	rumorear	0
bastar	0	desacordar	0	entremezclar	0	intercalar	0	presupuestar	0	saber	0
batallar	0	desacreditar	-1	entrenar	0	intercambiar	0	pretender	0	saborear	0
batir	0	desactivar	0	entretener	0	interceder	0	prevalecer	0	sabotear	0

Table 2

SUMMARY OF LEMMAS TAGGED IN IS-NEWS DICIONARY

UNIQUE LEMMAS								
Part of speach	Negative (-1)	Positive (+1)	Decreasers (*0,5)	Increasers (*1,5)	Negations (* -1)	Subtotal	Neutral	Total
Verbs	179	78	-	-	-	257	2575	2832
Adjectives	26	10	7	24	-	67	4090	4157
Adverbs	2	1	11	31	5	50	578	628
TOTAL	207	89	18	55	5	374	7243	7617

SUMMARY OF WORD TAGGED IN IS-NEWS DICIONARY

Lemmas and their derivations or inflected forms*

WORDS								
Part of speach	Negative (-1)	Positive (+1)	Decreasers (*0,5)	Increasers (*1,5)	Negations (* -1)	Subtotal	Neutral	Total
Verbs	3442	1662	-	-	-	5104	2575	7679
Adjectives	90	51	37	84	-	262	4090	4352
Adverbs	2	1	14	31	5	53	578	631
TOTAL	3534	1714	51	115	5	5419	7243	12662

*The derivations constitute conjugations in plural words, verbe tenses, among others.

REFERENCES

- Aguilar, P., Ghirelli, C., Pacce, M., & Urtasun, A. (2021). Can News Help Measure Economic Sentiment? An Application in COVID-19 Times. *Economics Letters*, Volume 199.
- Altig, D., Baker, S. R., Barrero, J. M., Bloom, N., Bunn, P., Chen, S., . . . Thwaites, G. (2020). Economic uncertainty before and during the COVID-19 pandemic. *Journal of Public Economic*, 191, 104274.
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *Quarterly Journal of Economics*, 131(4): 1593–1636.
- Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., van der Goot, E., Halkia, M., . . . Belyaeva, J. (2013). Sentiment Analysis in the News. *Proceedings of the 7th International Conference on Language Resources and Evaluation*, (pp. pp. 2216-2220). Malta.
- Banbura, M., Giannone, D., & Reichlin, L. (2011). Nowcasting. In M. P. Clements, & D. F. Hendry, *The Oxford Handbook of Economic Forecasting*. New York: Oxford University Press.
- Becerra, J. S., & Cruces, A. (2021). Análisis de Texto y Sentimiento del Informe De Estabilidad Financiera. *Banco Central de Chile*, Documento de Trabajo N°930.
- Becerra, J. S., & Sagner, A. (2020). Twitter-Based Economic Policy Uncertainty Index for Chile. *Central Bank of Chile*, Working paper N°833.
- Bernabeu, N. (2002). *Breve Historia de la Prensa*. Proyecto QuadraQuinta.
- Buckman, S., Shapiro, A., Sudhof, M., & Wilson, D. (2020). News Sentiment in the Time of COVID-19. *Federal Reserve Bank of San Francisco*, Economic Letter 2020-08.
- Cerda, R., Silva, Á., & Valente, J. (2016). Economic Policy Uncertainty Indices for Chile. *Latin American Center for Economic and Social Policies (CLAPES-UC)*, Working Paper.
- Correa, R., Garud, K., Londono, J. M., & Misláng, N. (2017). Constructing a Dictionary for Financial Stability. *IFDP Notes*. Washington: Board of Governors of the Federal Reserve System.
- Cruz Cornejo, C. (2017). Análisis de Sentimientos Enfocado a la Calidad de Servicio de la Banca Española. *Universidad Complutense de Madrid*. Trabajo fin de máster.
- Cruz, M. d., Peralta, H., & Ávila, B. (2020). Análisis de Sentimiento Basado en el Informe de Percepciones de Negocios del Banco Central de Chile. *Banco Central de Chile*, Documento de Trabajo N°862.
- Diaz Rangel, I., Sidorov, G., & Suárez Guerra, S. (2014). Creación y Evaluación de un Diccionario Marcado con Emociones y Ponderado para el Español. *Onomázein*, 31-46.
- Gorodnichenko, Y., Pham, T., & Talavera, O. (2021). The Voice of Monetary Policy. *National Bureau of Economic Research*(Working Paper 28592), Working Paper 28592.
- Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and Deliberation Within FOMC: A Computational Linguistics Approach. *The Quarterly Journal of Economics*, Volume 133, pp. 801-870.
- International Monetary Fund. (2018). *Regional Economic Outlook, Western Hemisphere*.
- Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2020). Making Text Count: Economic Forecasting Using Newspaper Text. *Bank of England*, Staff Working Paper No. 865.

- Larsen, V., & Thorsrud, L. (2015). The Value of News. *BI Norwegian Business School, CAMP Working Paper Series No 6*.
- Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, Volume 66, pp. 35-65.
- Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The Evolution of Sentiment Analysis - A Review of Research Topics, Venues, and Top Cited Papers. *Computer Science Review*, Volume 27, pp. 16-32.
- Mechulam Burstin, N., & Salvia Varela, D. (2018). Construcción de Léxicos Dinámicos para el Análisis de Sentimientos. *Universidad de la República, Uruguay. Tesis de grado*.
- Medhat, W., Hassan, A., & Korasky, H. (2014). Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal*, Volume 5, pp 1093-1113.
- Moreno Bernal, Á. I., & González Pedraz, C. (2020). Análisis de Sentimiento del Informe de Estabilidad Financiera. *Banco de España, Documentos de Trabajo N.º 2011*.
- Pérez Albertos, A. (2018). Leyendo Más Allá de las Palabras: Uso del Análisis de Textos para Medir el Sentimiento de Mercado en España. *Universidad Pontificia Comillas. Tesis de grado*.
- Rambaccussing, D., & Kwiatkowski, A. (2020). Forecasting With News Sentiment: Evidence With UK Newspapers. *International Journal of Forecasting*, Volume 36, pp1501-1516.
- Reis, J., Benevenuto, F., Vaz de Melo, P. O., Prates, R., Kwak, H., & An, J. (2015). Breaking the News: First Impressions Matter on Online News. *Proceedings of the Ninth International AAAI Conference on Web and Social Media*, (pp. 357-366).
- Saberi, B., & Saad, S. (2017). Sentiment Analysis or Opinion Mining: A Review. *International Journal on Advanced Science, Engineering Information Technology*, pp 1660-1666.
- Sánchez Sánchez, I. (2012). Breve Historia de la Prensa en el Siglo XIX. En A. González-Calero, *Cultura en Castilla-La Mancha en el Siglo XIX* (págs. 283-300).
- Shapiro, A. H., Sudhof, M., & Wilson, D. (2020). Measuring News Sentiment. *Federal Reserve Bank of San Francisco*, Working Paper 2017-01.
- Song, M., & Kyung-Shik, S. (2019). Forecasting Economic Indicators Using a Consumer Sentiment Index: Survey-Based Versus Text-Based Data. *Journal of Forecasting*, Volume 38, pp. 504-518.
- Thorsrud, L. A. (2016). Nowcasting Using News Topics. Big Data Versus Big Bank. *Advances in Short-Term Forecasting. ECB Workshop*. Frankfurt am Main.

<p align="center">Documentos de Trabajo Banco Central de Chile</p>	<p align="center">Working Papers Central Bank of Chile</p>
<p align="center">NÚMEROS ANTERIORES</p>	<p align="center">PAST ISSUES</p>
<p>La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: www.bcentral.cl/esp/estpub/estudios/dtbc.</p>	<p>Working Papers in PDF format can be downloaded free of charge from: www.bcentral.cl/eng/stdpub/studies/workingpaper.</p>
<p>Existe la posibilidad de solicitar una copia impresa con un costo de Ch\$500 si es dentro de Chile y US\$12 si es fuera de Chile. Las solicitudes se pueden hacer por fax: +56 2 26702231 o a través del correo electrónico: bcch@bcentral.cl.</p>	<p>Printed versions can be ordered individually for US\$12 per copy (for order inside Chile the charge is Ch\$500.) Orders can be placed by fax: +56 2 26702231 or by email: bcch@bcentral.cl.</p>

DTBC – 938

Utilización de noticias de prensa como indicador de confianza económica en tiempo real

María del Pilar Cruz, Hugo Peralta, Juan Pablo Cova

DTBC – 937

Monetary Policy Spillover to Small Open Economies: Is the Transmission Different under Low Interest Rates?

Cao et al.

DTBC – 936

Domestic Linkages and the Transmission of Commodity Price Shocks

Damian Romero

DTBC – 935

Firm Export Dynamics and the Exchange Rate: A Quantitative Exploration

Bernabe Lopez-Martin

DTBC – 934

The Labor Earnings Gap, Heterogeneous Wage Phillips Curves, and Monetary Policy

Mario Giarda

DTBC – 933

The impact of climate change on economic output in Chile: past and future

Karla Hernández, Carlos Madeira

DTBC – 932

Risk modeling with option-implied correlations and score-driven dynamics

Marco Piña, Rodrigo Herrera

DTBC – 931

Nowcasting Chilean household consumption with electronic payment data

Marcus P. A. Cobb

DTBC – 930

Sentimiento en el Informe de Estabilidad Financiera del Banco Central de Chile J.

Sebastián Becerra, Alejandra Cruces

DTBC – 929

Inequality, Nominal Rigidities, and Aggregate Demand

Sebastian Diz, Mario Giarda, and Damián Romero

DTBC – 928

Rational Sentiments and Financial Frictions

Paymon Khorrami, Fernando Mendo

DTBC – 927

Saving Constraints, Inequality, and the Credit Market Response to Fiscal Stimulus

Jorge Miranda-Pinto, Daniel Murphy, Kieran James Walsh, Eric R. Young

DTBC – 926

Liquidez del sector corporativo chileno: estimación de disponibilidad de caja bajo escenarios de estrés

Jorge Fernández, Fernando Pino, María Ignacia Valencia

DTBC – 925

Procyclical fiscal policy and asset market incompleteness

Andrés Fernández, Daniel Guzman, Ruy E. Lama, Carlos A. Vegh

DTBC – 924

Reserve Accumulation and Capital Flows: Theory and Evidence from Non-Advanced Economies

Juan Pablo Ugarte

DTBC – 923

Estimates of the US Shadow-Rate

Rodrigo Alfaro, Marco Piña

