Text Mining and Sentiment Extraction in Central Bank Documents

Giuseppe Bruno*

Bank of Italy, Economics and Statistics Directorate.

Abstract

The deep transformation induced by the World Wide Web (WWW) revolution has thoroughly impacted a relevant part of the social interactions in our present *global society*. The huge amount of unstructured information available on blogs, forum and public institution web sites puts forward different challenges and opportunities. Starting from these considerations, in this paper we pursue a two-fold goal. Firstly we review some of the main methodologies employed in text mining and for the extraction of sentiment and emotions from textual sources. Secondly we provide an empirical application by considering the latest 20 issues of the Bank of Italy *Governor's concluding remarks* from 1996 to 2015. By taking advantage of the open source software package *R*, we show the following:

1. checking the word frequency distribution features of the documents;
2. extracting the evolution of the sentiment and the polarity orientation in the texts;
3. evaluating the evolution of an index for the readability and the formality level of the texts;
4. attempting to measure the popularity gained from the documents in the web.

The results of the empirical analysis show the feasibility in extracting the main topics from the considered corpus. Moreover it is shown how to check for positive and negative terms in order to gauge the polarity of statements and whole documents. Although *R*, the employed software, has proved suitable and comprehensive for the required tasks, the whole picture presents lights and shadows. Improvements in the documentation and the package arrangement and portability among platforms are suggested in the text.

JEL classification: C83, E58, E66.
Keywords: Text Mining, Wordcloud, Polarity, Sentiment Analysis, Zipf's Law.

1. Bank of Italy, Research Area.
* giuseppe.bruno@bancaditalia.it
This version December 2015
The author wishes to thank Roberto Stok for his helpful suggestions and unceasing cooperation.
The views expressed are the author's only and do not imply those of the Bank of Italy.

1 Introduction and motivation

A big computer, a complex algorithm and a long time does not equal science. – Robert Gentleman

The deep transformation caused by the World Wide Web (WWW) revolution has thoroughly impacted a relevant part of social interactions. According to *www.internetlivestats.com* on August 20 of 2015, we had more the 3.1 billion of Internet users visiting almost 1 billion websites. Moreover there were about 1.5 billion of Facebook active users and approximately 900 million Tweets per day. The final upshot of these numbers is that we end up with a huge amount of unstructured textual information available on blogs, forum and public institution web sites which puts forward different challenges and opportunities. According to recent estimates about 4/5 of the whole set of web pages is composed of textual information. Mining valuable information out of a corpus of many documents, classifying sentiment or even providing a polarity measure on a given topic is an already widely applied technique in different fields such as marketing, social and political networks. Although their popularity in these fields, text mining and sentiment analysis have just recently shown up on the tool repertoire of Central bank's economists (for an overview of the problem see D. Bholat and Schonhardt-Bailey [2]). Text mining consists in building a sound synthesis from a set of documents, often referred as a corpus, that allow to quickly carry out the following tasks:

a) allocate the text to a given semantic category;

b) evaluate the word frequency distribution in the text;

c) verify global statistical features of the considered text;

The sentiment analysis is a follow up task aimed at designing automatic procedures for opinion discovery and summarization system from a set of textual information (Bing [1] presents a thorough and formal analysis on the topic). There we find a clear definition of the sentiment analysis as the computational study of opinions and emotions expressed in textual document. Moreover it is introduced a dichotomy between direct and comparative opinions which are the goal of the investigation. Text minining has also been considered by Hendry and Madeley [6] for the Bank of Canada communications. Here the authors put forth the *Latent Semantic Analysis* (LSA) to extract relevant meanings from Bank of Canada communications [1].

Our work provides some useful insights on the actual capability to start dealing with textual information in order to gauge sentiment orientation and possibly some polarity evaluation for broadening the set of our statistical collection tools. In particular, with reference to our corpus composed of a set of the Bank of Italy Governor's *Concluding Remarks*, this work attempts to answer the following questions:

1) what are the features of the word frequency distributions of the texts?

2) what is the sentiment orientation and the polarity value?

3) what is the readability and formality level?

4) what is the interest/memorability gained from the documents in our corpus?

[1] LSA is a technology widely used in internet search engines. It is based on the application of a Singular Value Decomposition on the term-document matrix obtained from the considered corpus.

The recent literature on these themes (see for example D. Bholat and Schonhardt-Bailey [2] and R. Nyman and Tuckett [12]) witnesses the growing interest shown by other Central Banks on these issues and that one closely linked with the emergence of the *Big Data* theme.

The appearance of web tools for monitoring social media by semantic and sentiment orientation, provides analytical support fostering a wider adoption of text mining techniques for implementing well informed decisions.

This paper is arranged in the following way. After this introduction in section 2 we present the basic theoretical concepts behind the text-mining techniques. Section 4 presents some graphical tools widely employed to show the main concepts contained in the examined document. Section 3 introduces the two relevant computational linguistic statistics allowing to compare and check the stability of these synthetic measure over time. Section 5 presents some results of this statistical analysis carried out on a corpus containing the last 20 editions, from 1996 to 2015, of the *Governor's Concluding Remarks*. Finally section 6 provides some concluding remarks.

2 Analysing a corpus of Text documents

Textual data provides a huge and diversified range of information. Nonetheless this information is encoded in a way that doesn't lend itself to an automatic deciphering step. Since about three decades, computational linguistics has tried to take advantage of large text collections in order to design text analysis algorithms. Text mining is a broad umbrella term illustrating a wide range of technologies for analyzing textual data which are intrinsically unstructured.

Following the suggestions in Hearst [5] we define text mining as the task of harnessing *a large online text collections to discover new facts and trends about the world itself*.

The task of analysing a corpus of text can be grasped by looking at the diagram in figure 1: starting from character encoding and language detection we have a tokenizaton of each text. This is followed by the stop-word remover and the word stemmer / lemmatiser. The stop-word remover has the goal of removing from a text the words appearing with high frequency but not carrying as much meaning, among such words we count, for example, determiners, coordinators and prepositions. The stemmer/lemmatiser are software tools aiming at the reduction of the inflectional forms of a word to a common root by means of a conflation process.

In order to apply the steps envisaged in the scheme shown in the picture 1, it is important to be aware of the need of software tools whose scope depends heavily from the particular languages taken in consideration[2].

For our empirical application we have built a corpus composed of the last 20 issues from 1996 to 2015 of the *Governor's Concluding Remarks* both in English and in Italian. Though this is a small corpus, it allows us to check the validity of the adopted analytical environment. Among the significant measures for the analysis of corpus of texts we have considered the correlation among words in different documents. This measure consists in finding all the Pearson correlation coefficients above a given threshold among a set of words and all the others present in different documents. In our corpus we have chosen to search for all the association of the following list of frequently occurring words:

banca , mercato , capitale , credito , lavoro , rischio [3]

[2]Here we have considered documents written in English and in Italian.
[3]The English translation of this words is respectively: bank, market, capital, credit, work, risk.

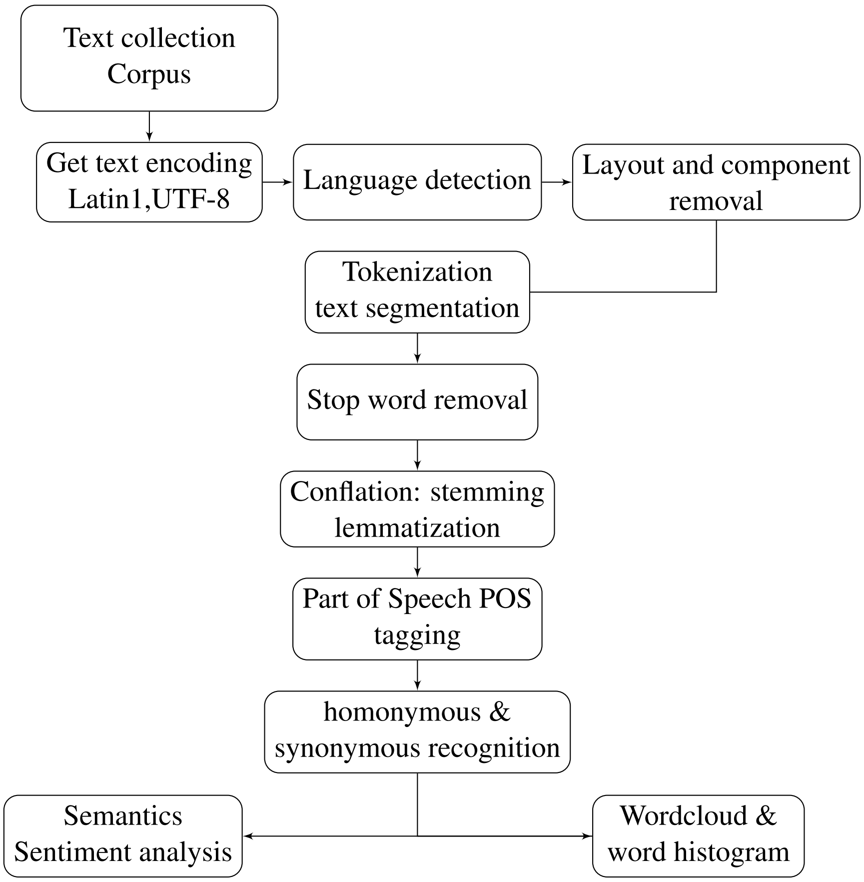


Figure 1: Block diagram describing the steps for text and opinion mining

Table 1: Pearson correlation among words in the documents

banca	**mercato**	**capitale**	**credito**	**lavoro**	**rischio**
consiglieri (0.98)	capacitá (0.98)	ricchezza (0.97)	aumento (0.99)	espansione (0.99)	quote (0.98)
proposta (0.98)	equilibrio (0.98)	strada (0.97)	monetaria (0.97)	shock (0.97)	controllo (0.97)
territoriale (0.98)	europee (0.98)	dotazione (0.96)	produzione (0.97)	effetti (0.96)	ispezioni (0.97)
completamento (0.97)	portata (0.98)	requisiti (0.96)	prospettive (0.99)	evoluzione (0.99)	sistemico (.98)

In table 1 we show the keywords featuring the highest correlation with the six column keywords. Between parentheses the Pearson coefficient is reported. These high values of correlation imply the coexistence of the keywords in the same documents. A deeper analysis could be achieved by splitting each issue of the *Considerazioni Finali* by paragraph or even sentences. Here we have performed a simple test to check the performances and the capabilities of the available analytical instruments.

3 Documents Statistical laws

Quantitative analysis of human languages allows to discover common features of spoken or written text. The most basic statistical property of human languages is the frequency with which different words are used. In this section we present two scaling relations which are the most relevant normative features of computational linguistics based on word occurrence frequency: Zipf's and Heaps' laws.
A relevant principle of information theory states that a text should follow the Zipf's law in order to maximize its capacity to convey information with a limited set of words (see Zipf [15]). Earlier noted by other authors, but popularized by G. K. Zipf, this law states that the frequency of words in a given vocabulary follows a power law. Specifically, if we rank all the occurrences of words in a given text from the most to the less common one, we find that the probability $p(w_i)$ of occurrence of the word w_i can be expressed as:

$$p(w_i) = \frac{1}{K} \cdot i^{-\gamma} \tag{1}$$

where i is the rank of the word w_i, $\gamma \approx 1$ and K the constant for the normalization to a probability distribution. This regularity can be observed in any modern human language when analyzing any corpus of an adequate size. We say that textual data follow the Zipfian distribution when the 2nd ranked word has half the occurrences of the 1st ranked one, the 3rd word has a third of the occurrences of the 1st, the 4th has a quarter, etc.

Word frequency distributions are extremely interesting. They are one of the most basic properties of humans' communicative system and play a critical role in language acquisition and processing. It is remarkable that they can be well-characterized by a simple mathematical law. Despite its simplicity this law constitutes just a normative feature of computational linguistic. As far as we know there are no positive theories of language or discourse production explaining the linguistic roots of the law according to models for evolution of communication. The Zipfian nature of the examined documents is checked with the Kolmogorov-Smirnov non parametric test.

The method of plotting word frequency distributions sometimes has obscured an important fact: word frequencies are not actually so simple. They sometimes show statistically-reliable structure beyond Zipf's law that likely will not be captured with any simple model. It can be observed that Zipf's law works well for the first few ranks, but it usually doesnt fit for bigger ranks. At the same time, the large-scale structure is robustly Zipfian.

Here we do not try to account for any psychological processes of word production, especially the intentionality of choosing words in order to convey a desired meaning. We simply check the adherence of our corpus to the proposed law.

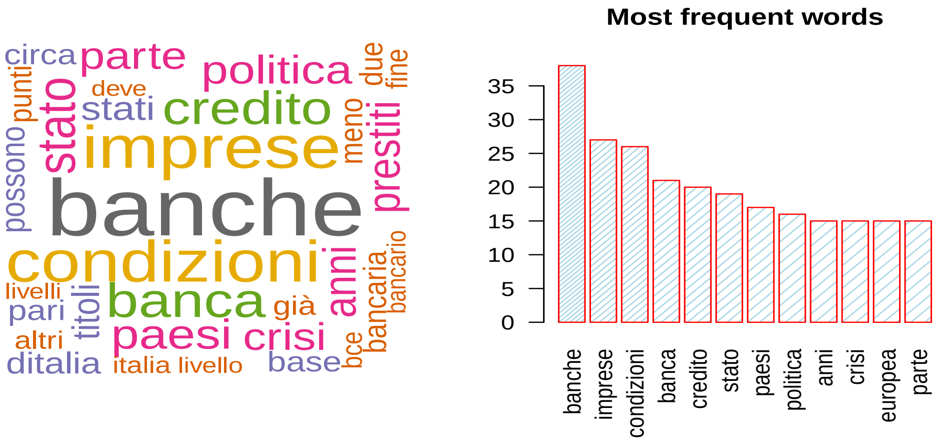
The second regularity law, closely related to a generalization of the Zipf's law, observed in computational linguistic is the Heaps' law (see Leijenhorst and van der Weide [9] and Egghe [4]) which states a relationship between the number of different words and the total number of used words in a given corpus of documents. In particular this relationship says that lexicons sizes are concave increasing power of text's sizes. Heaps' law predicts the vocabulary size in a text consisting of a given number of words. Formally the law can be expressed with the following relationship:

$$V_{size} = K \cdot T_{size}^{\theta} \tag{2}$$

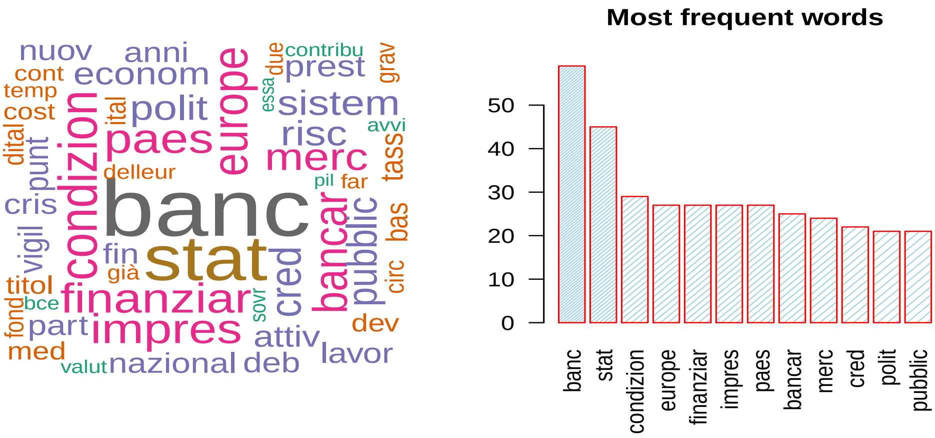
where: V_{size} is the lexicon size and T_{size} is the size of the whole text expressed as the total number of words, with K and $0 < \theta < 1$ constant. With English text corpora, K is usually between 10 and 100, and θ is between 0.4 and 0.6. In our empirical applications we found that the lexicon size grows with the square root of the total number of employed words. This square root evidence empirically confirms the previously given range.
The relation between Zipf's and Heaps' law is a relevant research issue. In Leijenhorst and van der Weide [9] it is shown a formal derivation of the Heaps' law from the Mandelbrot distribution which is a generalization of the Zipf's law.

4 Word frequency distributions and their evolution

The most popular tools to investigate and extract a synthesis from a text or a corpus of texts are the word cloud [4] and some histogram of the frequency of the words employed. A word cloud is an handy tool to show word frequency in a document by resizing the fonts of individual words included in the document proportionally to how often they are employed, to quickly display their relative popularity.



(a) Wordcloud for the "Considerazioni Finali 2005" (without stemming)

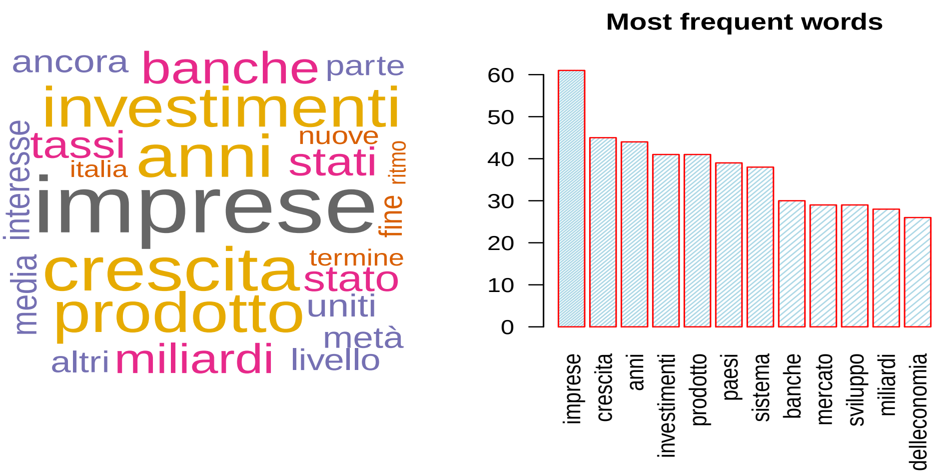


(b) Wordcloud for the "Considerazioni Finali 2005" (with stemming)

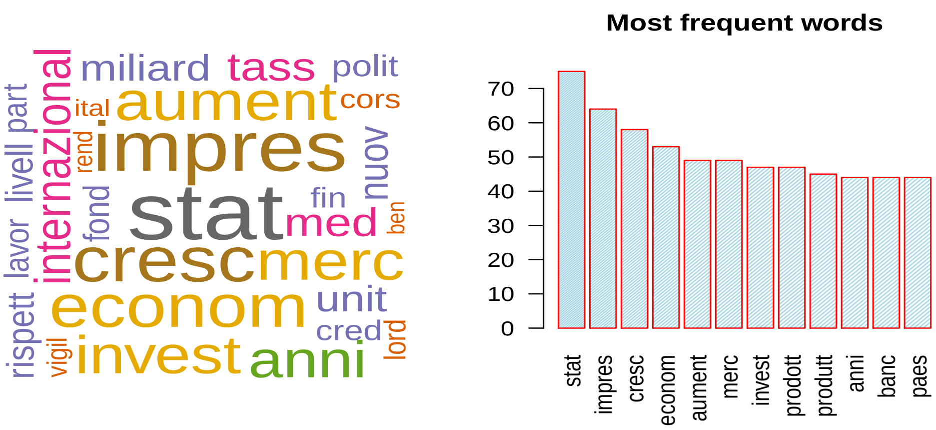
Figure 2

As examples, the pictures 2, 3 and 4 show the more relevant themes considered in the *Governor's Concluding Remarks* for the years 2005, 2010 and 2015. For each year, we show a couple of figures which allow to compare the effect of the stemming on the actual word frequency distribution. The process of stemming allows to extract the root of the words. The ensuing wordcloud provides a more accurate picture of the

[4] Sometimes they are called tag cloud.

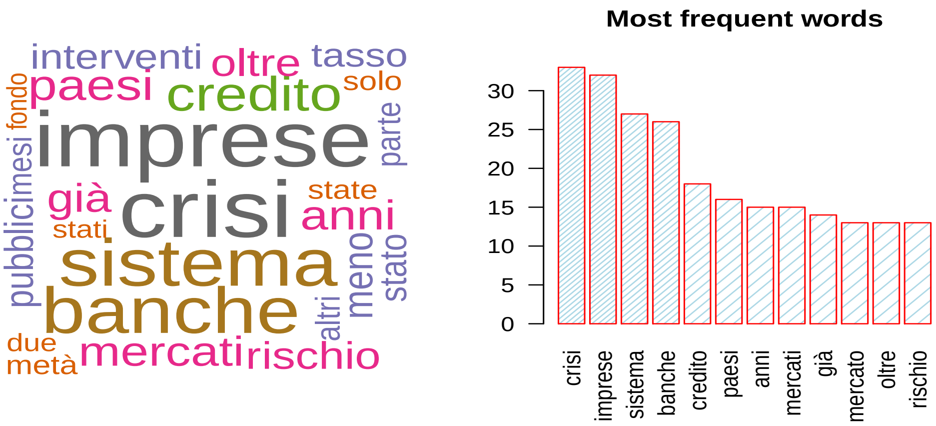


(a) Wordcloud for the "Considerazioni Finali 2010" (without stemming)

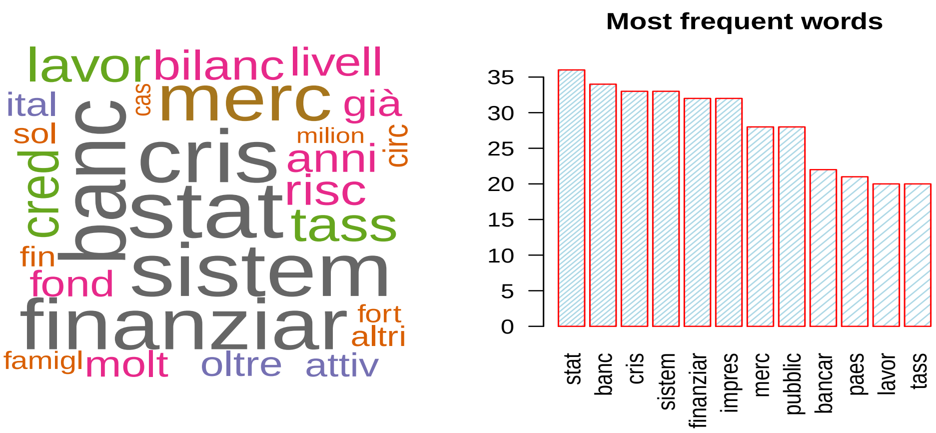


(b) Wordcloud for the "Considerazioni Finali 2010" (with stemming)

Figure 3



(a) Wordcloud for the "Considerazioni Finali 2015" (without stemming)



(b) Wordcloud for the "Considerazioni Finali 2015" (with stemming)

Figure 4

relative frequency of the word stems. From these set of figures we can see how the focus of the words has shifted from the Banks in 2005 to the firms and the State in 2010 to arrive, in 2015, to financial system, crises and banks. Although they don't provide a metric space for numerical comparisons, from a qualitative standpoint reading and interpretation of word-clouds is much faster with respect to the histograms. The word-cloud employs the technique of tagging for classifying the importance of words. The social media have already extensively used the tagging technique as a massively distributed and participative method of expressing emotions and classifying information. Here we will restrict ourselves considering the tagging as a means of classifying the relative importance of words.

5 An empirical application with a Bank of Italy corpus

In order check the actual behavior of the available text mining procedures we have taken into account a corpus composed of documents relevant for central banking issues. In order to allow comparisons we have chosen a set of publicly available documents. With this goal in mind we have taken the stream of the issues from 1996 to 2015 of the Bank of Italy *Governor's Concluding Remarks* [5]. The *Concluding Remarks* is a synthesis of the key developments of the economic picture for Italy, the euro area and the rest of the world for the previous year and the first quarter of the current year. This is a corpus of about 181000 words and $1.2MB$ in size. Though of very limited size, this corpus is very helpful in checking and evaluating the analytical tools available in the *R* package. In particular this empirical section has a twofold purpose:

- to present the result of some analyses carried out on corpus of documents relevant for a Central Bank;
- to show the suitability of the analytical instruments available on the statistical package *R*;

A very relevant consideration to take into account here is the language of the corpus. As a matter of fact many computational tools are already quite developed for the English, German and Chinese languages while they are still at a rather primitive level for the Italian.

Our approach is somewhat similar to Moniz and De Jong [10] who chose to analyze the interest rate minutes of the Bank of England in order to design an analytical engine that is able to evaluate and predict the impact of central bank communication on formation of investors' interest rate expectations. On the other hand, here we just consider the linguistic features and the polarity of the sentiment expressed in the texts.

The first analysis consider the word frequency distributions of the documents contained in our corpus. The following picture shows the frequency distribution of the ranked word frequency and the Vocabulary size vs number of terms for the corpus of the Bank of Italy *Governor's Concluding Remarks*

Figure 6: Zipf plot of word frequency vs rank

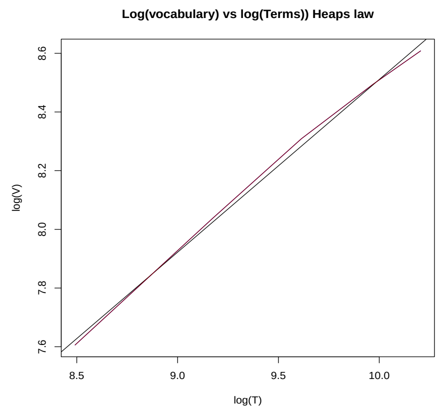


Figure 7: Heaps plot of Vo size vs # terms

In the following table we show the results of a two-sample Kolmogorov-Smirnov test on our corpus:

1) hypothesis test that the rank-frequency distribution of the corpus is based on the Zipf's nature:

$D = 0.2922, p - value < 2.2 \cdot e^{-16}$

[5] In Italian the document is titled: *Le considerazioni finali*.

2) hypothesis test that the Vocabulary size follows the Heaps' distribution:

$D = 0.2, p - value = 1$

The first Kolmogorov-Smirnov test rejects the null hypothesis that the our corpus follow the Zipf's distribution. The second test cannot reject the null hypothesis that the Vocabulary size of our corpus adheres to the Heaps' distribution. The rejection of the Zipf's hypothesis might be consequence of the limited number of documents taken into account.

Aside from these statistical analysis we have also employed the following computational linguistic tools:

- Sentiment analysis;
- Readability index;
- Formality index;
- Memorability index.

The previous tasks have been carried out by taking into account the English version of our the corpus. All the algorithms for carrying out these analyses have been taken from the *R* package **qdap**.

5.1 Polarity evaluation

All the previous analyses have been carried out by splitting the whole *Concluding Remarks* text on its component statements.

The algorithm employed to assign a polarity value for each sentence utilizes the sentiment dictionary proposed in Hu and Liu [8]. This algorithm builds a context cluster for each one of the polarized word [6]. A context cluster is simply a bunch of consecutive words around a polarized word [7]. In a context cluster the words are tagged as neutral (x_i^0), negator (x_i^N), amplifier (x_i^a), or de-amplifier (x_i^d). Neutral words only affect the word count (n). These tagged words can act as valence shifters increasing or decreasing the effect of the polarized word.

In order to get a polarity measure for a given sentence, finally the polarity for each context cluster (x_i^T) is achieved by summing the value of polarized words and normalizing this sum by the square root of the word count.

$$\delta_i = \frac{x_i^T}{\sqrt{n}} \tag{3}$$

where:

$$x_i^T = \sum (1 + c \cdot (x_i^A - x_i^D)) \cdot w \cdot (-1)^{\sum x_i^N} \tag{4}$$

c is a weight employed for smoothing the effects of amplifiers or de-amplifiers words, this user-defined value defaults to 0.8, $w = \pm 1$ is the sign of the polarized word.

$$x_i^A = \sum (w_{neg} \cdot x_i^a)$$
$$x_i^D = \max\left(x_i^{D'}, -1\right)$$

$$x_i^{D'} = \sum (-1) \cdot \left(w_{neg} \cdot x_i^a + x_i^d\right)$$

and finally

$$w_{neg} = \left(\sum x_i^N\right) \mod 2$$

[6] As default these words belong to a list of about 6800 words including mainly nouns, adjectives and adverbs.

[7] For the English syntax, a context cluster defaults to a sequence of 4 words before and two words after the polarized word.

This amounts to say that the polarity for a given context cluster is computed by the algebraic sum of amplifiers and de-amplifiers suitably considering the number of negator elements.

There exist two main approaches to the problem of extracting sentiment automatically.

The first one is a lexicon-based approach which involves calculating orientation for a document from the semantic orientation of words or phrases in the document (see Turney [14]). The second one is a text classification approach involves building classifiers from labeled instances of texts or sentences (Pang et al. [11]), essentially a supervised classification task. The latter approach could also be described as a statistical or machine-learning approach. In this empirical application, we have followed the first method in which we employ dictionaries of words annotated with the words semantic orientation, or polarity.

In our empirical application we have evaluated the evolution of the polarity over the flow of the *Concluding Remarks* of the last twenty years. As example, in the picture 16 we show the polarity of eight of the past issues of the *Concluding Remarks*: the whole set of eight pictures show a substantial prevalence for a neutrality with a small optimism towards the end of the speech for some years.

5.2 Readability assessment

Readability assessment provides a measure of the effort required to a reader to understand a text. Most of the classical readability metrics are linear model of few superficial features of words and sentences. In our empirical application we have chosen to employ the default function available in the **qdap** R package for computing the automated readability index ARI (Senter and Smith [13]). It produces an approximate representation of the US grade level needed to comprehend the text.

The formula for calculating the automated readability index (ARI) is given in equation 5:

$$ARI = 4.71 \cdot \left(\frac{N_{char}}{N_{words}}\right) + .5 \cdot \left(\frac{N_{words}}{N_{sentences}}\right) - 21.43 \tag{5}$$

where: N_{char} is the total number of characters in the text,

N_{words} is the total number of words and

$N_{sentences}$ is the total number of sentences in the text.

This index tends to reward the employment of short words and short sentences. ARI is a simple empirical derivation for the English language which provides a useful comparison tool over time and among different documents belonging to the same class. In the picture 25 we show the readability evolution for eight of the past issues of the *Concluding Remarks*,

Values of the ARI index in the range $12 - 15$ are usually classified as readable for people with a College degree.

5.3 Formality assessment

Another relevant feature allowing to numerically gauge the degree of the context-dependence of a document is the formality of a document. There are some different definitions of the for the formality. Here we have chosen the the definition suggested in Heylighen and Dewaele [7] where the Formality score is calculated according the equation:

$$F = 50 \cdot (\frac{n_f - n_c}{N} + 1) \tag{6}$$

where: $f = \{nouns, adjective, preposition, article\} \qquad n_f = |f|,$

$c = \{pronoun, adverb, verb, interjection\} \qquad n_c = |c|$

$N = \sum(f + c + conjunctions)$

This Formality score gets higher when statements make more use of nouns and adjectives rather than pronouns and adverbs. In the picture 34 we show the formality evolution for eight of the past issues of the *Concluding Remarks*,

Reference values for the Formality measure taken from Heylighen and Dewaele [7] indicates values of 70 as highly formal. The last issues of the *Concluding Remarks* feature a systematically higher values for the Formality index which is in the range $[70 \div 75]$. This result confirms our intuition of highly formal documents leaving few room to individual interpretation.

5.4 Scraping Google search hits

Because of its aggregation of millions of searches, Google search data features a great potential in tracking and forecasting massive social phenomena. Google search results can be employed as a useful indicator of public opinion. For example, Davidowitz [3], a fellow analyst at Google has used search data to research several topics, such as quantifying the effect of racism on the 2008 presidential election. The finding was that Obama did worse in states with higher racist query volume. An arbitrary, though comparable way to measure the interest gained by the issuance of the *Concluding Remarks* has been devised. The measure is based on the computation of the Google Hits for each one of the sentences contained in the considered document. The memorability of eight of the past issues of the *Concluding Remarks* is shown in the figure 43:

Memorability is expressed as natural logarithms. Values in the range $12 - 13$ mean that each sentence has a *Google-like* popularity of

$$e^{12} \leq Google_{popularity} \leq e^{13} \quad \rightarrow \quad 162000 \leq Google_{popularity} \leq 443000 \tag{7}$$

Timing considerations are quite relevant for these results. Here we have tallied these *google hits* for statements the are partially repeated and span a very wide time interval. To provide an approximate form of a yardstick, we have applied the same measurement to the following two analogous documents:

1) The Federal Reserve Board *Semiannual Monetary Policy Report to the Congress* held on July 15, 2015;

2) The European Central Bank *Account of the monetary policy meeting of the Governing Council of the E.C.B.* held on October 22, 2015; [8]

The figure 46 provides a quick description of the analysis on these documents. Even if the behavior of the two curves is quite different and the *Google* measured popularity seems higher for the Federal Reserve Board testimony, the shown numbers go from 10 to 13 and therefore are very close to those achieved by the considered *Concluding Remarks*. These results should be used to further dig up specialized press releases in order to evaluate some form of impact of the news on other financial agents.

6 Concluding Remarks

In this paper we have presented the main computational linguistic methodologies for mining the relevant information from a corpus of documents and evaluating some summary statistics. These methodologies employ a tool-set taken from the information retrieval technology.

We have shown that these technologies can be quite helpful in quickly analyzing huge amount of documents and automatically extracting sentiment or opinion orientation and gauging polarity of these sentiments. By taking advantage of some packages available on the CRAN[9] repository, we have written some *R* procedures implementing different algorithms for text mining and sentiment analysis. These *R* procedures have been

[8]These documents can be found respectively at http://www.federalreserve.gov/newsevents/testimony/yellen20150715a.htm and https://www.ecb.europa.eu/press/accounts/2015/html/mg151119.en.html

[9]Comprehensive R Archive Network, https://cran-r-project.org

applied for the analysis of an homogenous corpus of documents based of the Bank of Italy *Governor's concluding Remarks* of the last 20 years.

The main conclusions are:

1) the consideration of the last 20 editions of the *Concluding remarks* has shown us that these documents tend to stay pretty much neutral over all the extension of the speech. The results of these analyses allow to trace back the sentiment shift over time in relation to the different short term economic conditions.

2) The ARI readability index shows a readability in the range $12 - 15$ which approximately implies a college degree.

3) The examined corpus of documents, as it could be expected, has shown a quite high average level of the Formality score.

4) The English versions of the *Concluding Remarks* have shown a number of hits in the the range 10^5. This measure is comparable with that achieved with similar reports from the ECB and the FRB.

The present strand of research looks quite promising especially for the possibility to quickly provide institutional answers more closely connected to the social emotions and preferences of the different economic agents.

References

[1] Bing, L. (2010). Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing, Second Edition. Taylor and Francis Group, Boca.*

[2] D. Bholat, S. Hansen, P. S. and C. Schonhardt-Bailey (2015). Text Mining for Central Banks. *Centre for Central Banking Studies 33*, 1–19.

[3] Davidowitz, S. S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics 118*, 26–40.

[4] Egghe, L. (2007). Untangling Herdan's Law and Heaps' Law: Mathematical and Informetric Arguments. *Journal of the Americal society for Information Science and Technology 58*, 702–709.

[5] Hearst, M. A. (1999). Untangling Text Data Mining. *ACL '99 Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, 3–10.

[6] Hendry, S. and A. Madeley (2010). Text Mining and the information content of Bank of Canada Communications. *Bank of Canada Working Paper*, 2–53.

[7] Heylighen, F. and J. Dewaele (2002). Variation in the Contextuality of Language: an Empirical Measure. *Foundation of Science*, 293–340.

[8] Hu, M. and B. Liu (2004). Mining opinion features in customer reviews. In *Proceedings of the 19th National Conference on Artifical Intelligence*, AAAI 2004, pp. 755–760. AAAI Press.

[9] Leijenhorst, D. C. and T. P. van der Weide (2005). A formal derivation of Heaps' Law. *Information Science 170*, 263–272.

[10] Moniz, A. and F. De Jong (2014). Predicting the impact of central bank communications on financial market investors interest rate expectations. In V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, and A. Tordai (Eds.), *The Semantic Web: ESWC 2014 Satellite Events*, Volume 8798 of *Lecture Notes in Computer Science*, pp. 144–155. Springer International Publishing.

[11] Pang, B., L. Lee, and S. Vaithyanathan (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in NLP.*, pp. 79–86.

[12] R. Nyman, P. O. and D. Tuckett (2015). Measuring financial sentiment to predict financial instability: A new approach based on text analysis. *University College London*.

[13] Senter, R. J. and E. A. Smith (1967). Automated readability index. *Aerospace Medical Research Laboratories*.

[14] Turney, P. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of review. *Proceedings of the 40th Meeting of the Association for Computational Linguistics*, 417–424.

[15] Zipf, G. K. (1949). *Human behaviour and the principle of least effort: an Introduction to Human Ecology*. Cambridge, Mass.: Addison-Wesley Press.

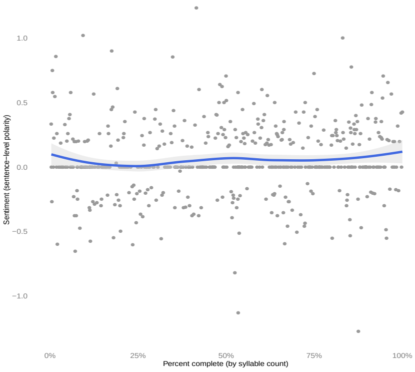


Figure 8: Polarity of Concluding Remarks 1999

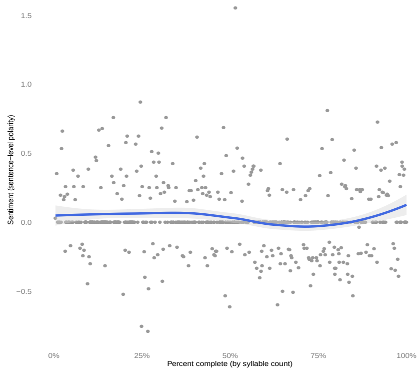


Figure 9: Polarity of Concluding Remarks 2000

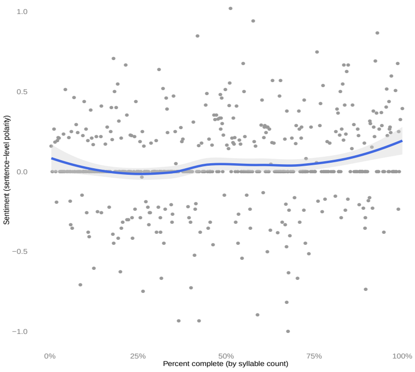


Figure 10: Polarity of Concluding Remarks 2004

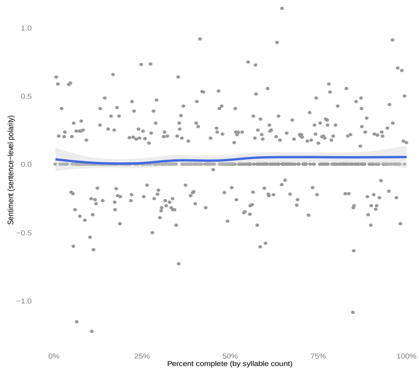


Figure 11: Polarity of Concluding Remarks 2005

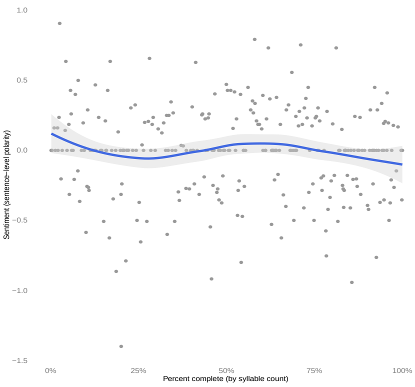


Figure 12: Polarity of Concluding Remarks 2009

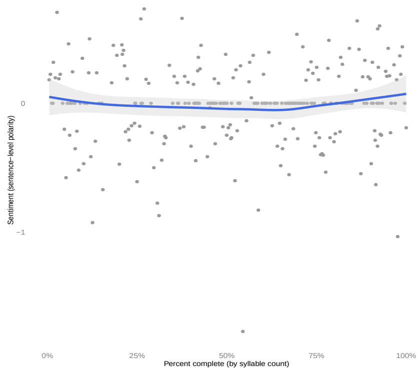


Figure 13: Polarity of Concluding Remarks 2010

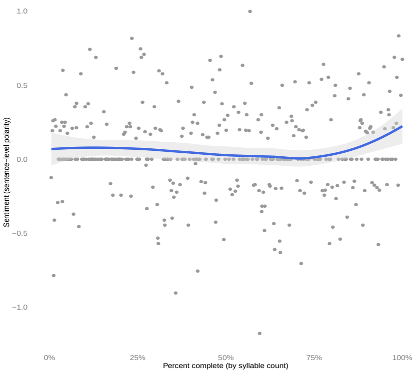


Figure 14: Polarity of Concluding Remarks 2014

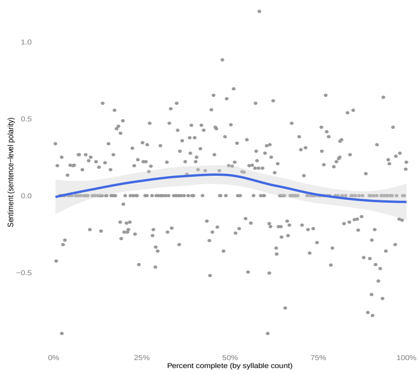


Figure 15: Polarity of Concluding Remarks 2015

Figure 16

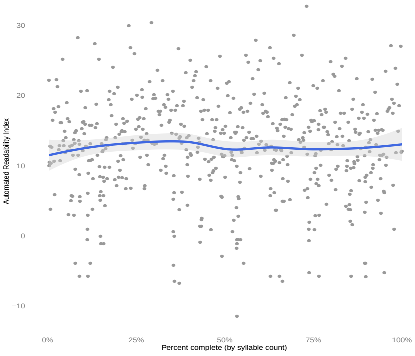


Figure 17: Readability of Concluding Remarks 1999

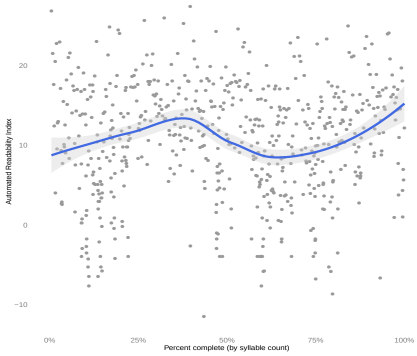


Figure 18: Readability of Concluding Remarks 2000

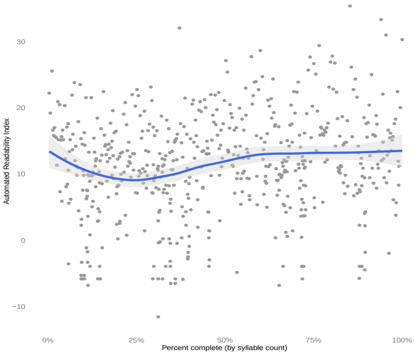


Figure 19: Readability of Concluding Remarks 2004

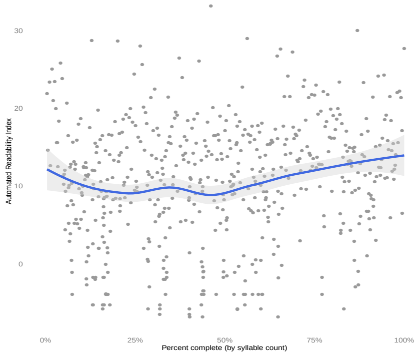


Figure 20: Readability of Concluding Remarks 2005

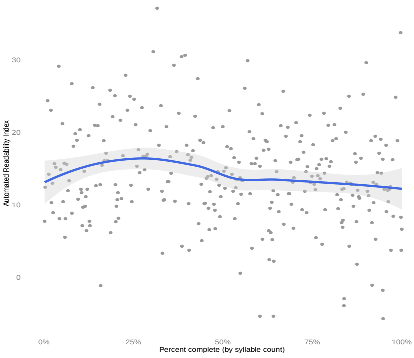


Figure 21: Readability of Concluding Remarks 2009

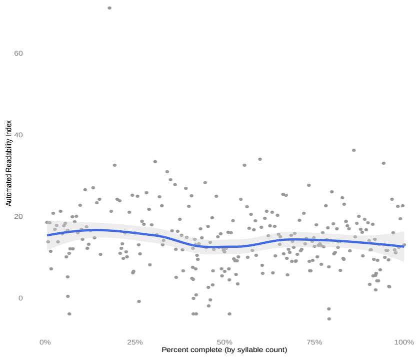


Figure 22: Readability of Concluding Remarks 2010

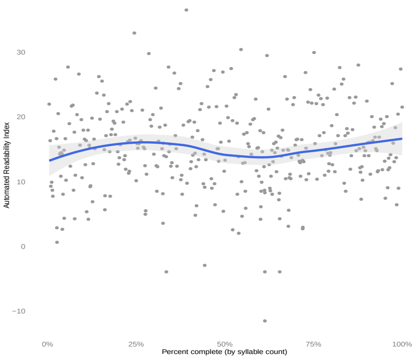


Figure 23: Readability of Concluding Remarks 2014

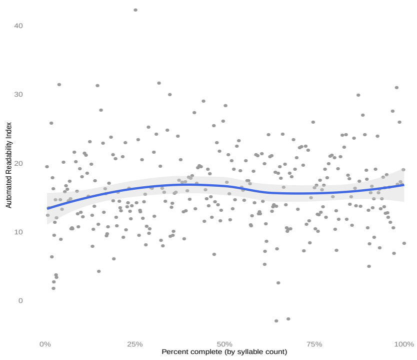


Figure 24: Readability of Concluding Remarks 2015

Figure 25

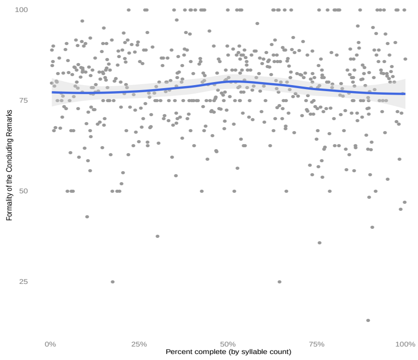


Figure 26: Formality of Concluding Remarks 1999

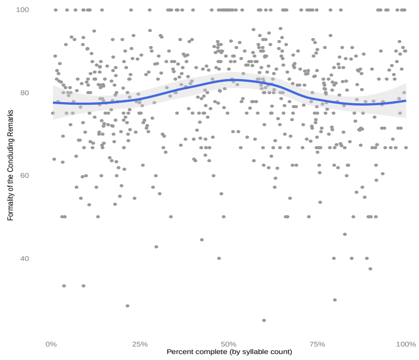


Figure 27: Formality of Concluding Remarks 2000

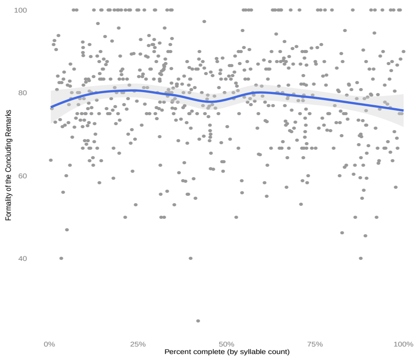


Figure 28: Formality of Concluding Remarks 2004

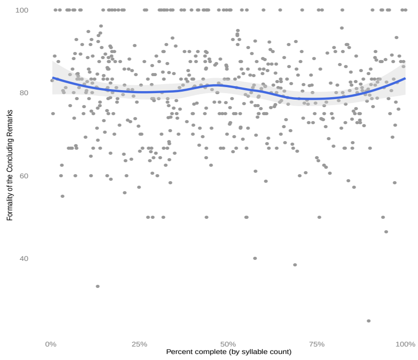


Figure 29: Formality of Concluding Remarks 2005

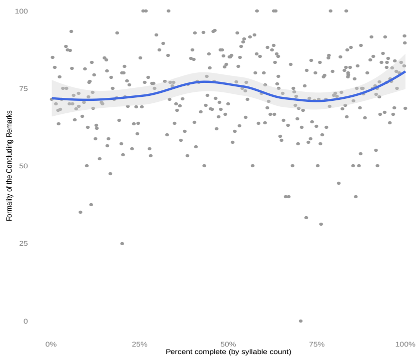


Figure 30: Formality of Concluding Remarks 2009

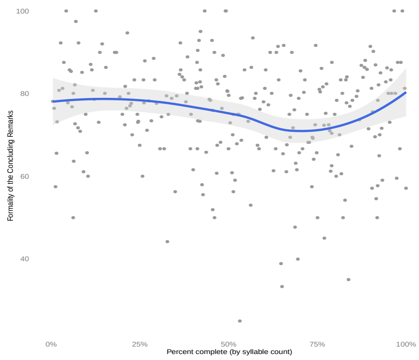


Figure 31: Formality of Concluding Remarks 2010

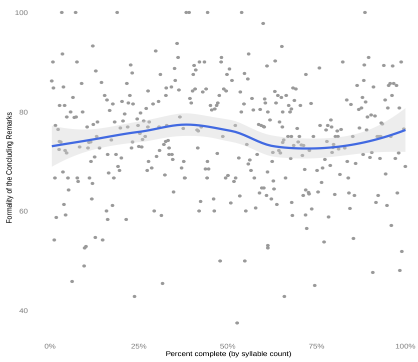


Figure 32: Formality of Concluding Remarks 2014

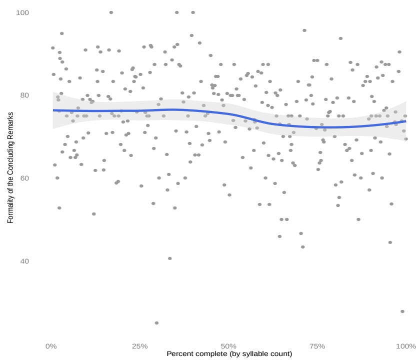


Figure 33: Formality of Concluding Remarks 2015

Figure 34

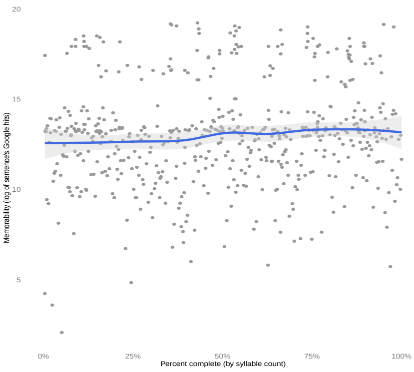


Figure 35: Memorability of Concluding Remarks 1999

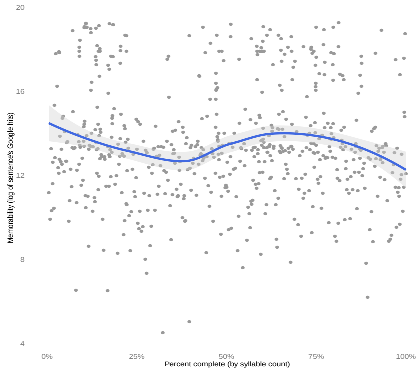


Figure 36: Memorability of Concluding Remarks 2000

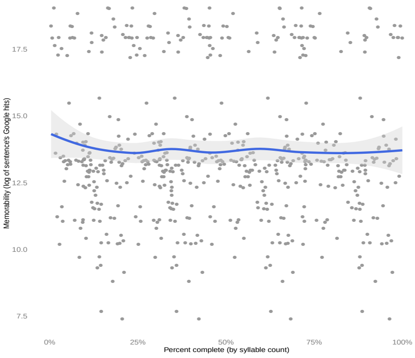


Figure 37: Memorability of Concluding Remarks 2004

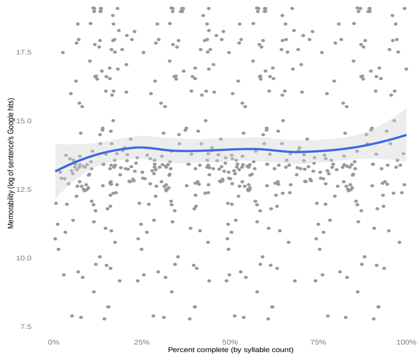


Figure 38: Memorability of Concluding Remarks 2005

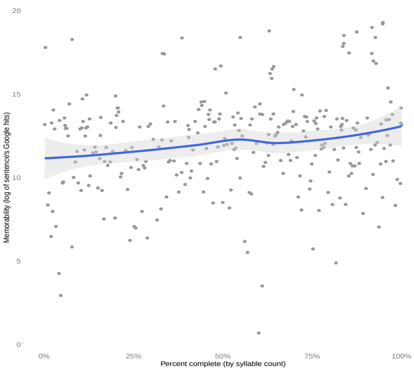


Figure 39: Memorability of Concluding Remarks 2009

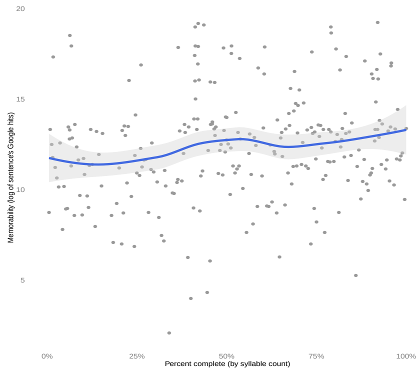


Figure 40: Memorability of Concluding Remarks 2010

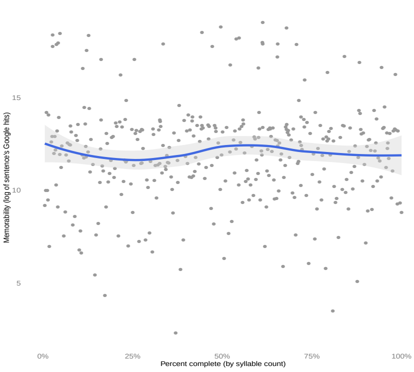


Figure 41: Memorability of Concluding Remarks 2014

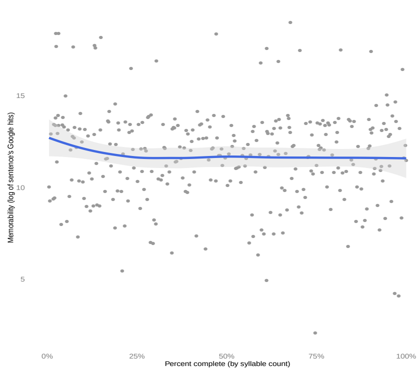


Figure 42: Memorability of Concluding Remarks 2015

Figure 43

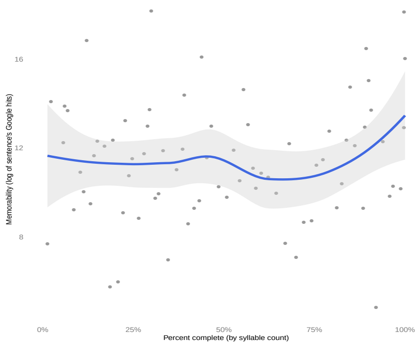


Figure 44: Memorability of July 15, 2015 FRB testimony

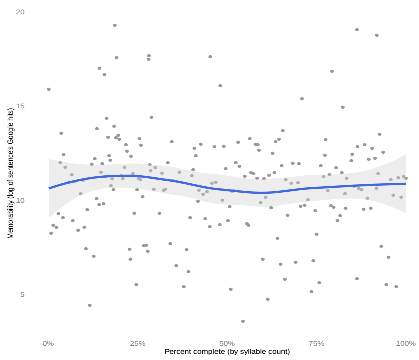


Figure 45: Memorability of November 19, 2015 ECB press release

Figure 46