# Appendix to "Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements"*

David O. Lucca
Federal Reserve Board †

Francesco Trebbi
University of Chicago and NBER‡

## A Introduction

This Appendix companion to Lucca and Trebbi [2009] describes with more detail the methods used to calculate the semantic orientation scores in the paper, and connects the parts of the algorithm to the technology and academic works from which they are derived.

The concept of semantic orientation (SO) seeks to evaluate a given word or phrase's location on a semantic axis over which both direction and intensity of meaning can be defined. Operationally, a semantic axis is defined by two terms of opposite meaning, or antonyms—say, strong/weak, robust/fragile—which define direction and, by some given unit of measurement, intensity. In using their semantic expertise, human beings can subjectively categorize a sentence out of a statement. To a vast majority of readers the statement "*Ernest Hemingway could kill a bear with his bare hands*" will indicate strength rather than weakness, robustness rather than fragility. However, using the fuzzy logic of semantics leaves much potential for disagreements in terms of intensity and sometimes direction. The purpose of the automated SO procedures described in this Appendix is to provide an automated method of assigning such semantic values, which is objective, transparent, and easily replicable. The objectivity and replicability of the scores are relative to a reference corpus of text—in our implementation, the Internet and information from news outlets—on which the semantic orientation scores are based.

For most of the examples in this Appendix, we will think of the word-pair "hawkish" and "dovish" as being the relevant antonymy over which a measure of "hawkishness" of central bank communication can be defined. Specifically we will use the word "hawkish" as indicating increases in policy rates, and the word "dovish" as pointing to future decreases. More broadly we also simultaneously investigate multiple pertinent semantic axes in this Appendix.

*__September 8, 2009__ We are particularly grateful to Isaac Laughlin for outstanding research assistance and his help in writing this technical appendix.

†Division of Monetary Affairs, email: david.o.lucca@frb.gov.

‡Booth School of Business, email: ftrebbi@chicagobooth.edu.

The text that we attempt to measure in the paper is composed of the statements released by the Federal Open Market Committee (FOMC) after its policy meetings in the sample May 1999 to December 2008.[1] Two sentences from FOMC statements which we will use extensively in the examples below are:

"*Though longer-term inflation expectations remain well contained, pressures on inflation have picked up in recent months and pricing power is more evident.*" (March 22, 2005)

"*However, inflation pressures seem likely to moderate over time, reflecting reduced impetus from energy prices, contained inflation expectations, and the cumulative effects of monetary policy actions and other factors restraining aggregate demand.*" (December 12, 2006)

In the following sections we provide a step-by-step description of two different procedures to measure the content of the statements. Multiple linguistic procedures are useful to corroborate the robustness of the analysis, especially for such a new application. A first procedure is based on information retrieval (IR) on the Internet. As we discuss below, this procedure has the main advantage of allowing us to focus directly on the text of FOMC statements. But, since we can only access the content of webpages indirectly through web-engine searches, it also has the drawback of limiting the researcher's control over the dominion of documents under analysis and the search mechanism, inducing substantial measurement error in the measures. A second procedure is based on IR from a specifically selected corpus of business news gathered from the Dow Jones Factiva database. The Factiva approach will not allow for a direct match of FOMC text, because of the much smaller number of documents that we can reference to build our measures; however, the analysis of news issued around FOMC announcements can still capture the content of the announcements, while leaving us complete control over the corpus composition and the search. We first illustrate the Internet-based IR approach and then discuss the implementation on Factiva.

## B Google-based Information Retrieval

The starting point of the Google-based analysis is the collection of texts of FOMC policy statements. Each announcement is divided into "raw phrases" by splitting the text at a punctuation mark (periods or commas, as no other punctuation marks are used in the statements in our sample). Punctuation is a natural first step to disambiguate the meaning of the sentences. Sentences in the FOMC statement are often long enough to contain several pieces of information requiring further separation to assign separate scores. Think of the sentence:

"*Though longer-term inflation expectations remain well contained,/ pressures on inflation have picked up in recent months and pricing power is more evident./*"

[1] To form the corpus of texts, we manually removed the policy action preface and the roll call votes included in the statements. See the data section of the paper for additional detail.

Based on the punctuation-mark splitting rule just described, the text is composed of two raw phrases, the first highlighting moderation of long-term inflationary expectations, and the second pointing to current inflationary pressures. By inspection, each raw phrase in this sentence has a different degree of "hawkishness", and, by cutting it at the comma, we can separately assign scores to each of the two. Starting with the set of raw phrases, three steps are used to obtain semantic orientation from the Internet: parsing, chunking and searching the Internet. The first two steps aim at obtaining reasonably compact, but meaningful, search units while the third is used to calculate the semantic orientation score.

Tagging is the process of taking raw text—the set of raw phrases from the statements—and identifying the grammatical structure of that text. Structural elements that are identified in this step are the parts of speech (POS) in the English language, for example, nouns, verbs, adjectives and adverbs. The step is implemented using an off-the-shelf algorithm in computational linguistics.

After identifying these grammatical elements, a code is used to join words into longer elements called "chunks" based on their POS tags. Chunks do not necessarily coincide with entire raw phrases, but rather, with their relevant semantic building blocks. These chunks are in turn used to form "search units" by joining consecutive chunks. This allows us to generate search units of measurable semantic meaning. While the phrase "*inflation expectations*" does not carry an identifiable semantic orientation (at least in terms of our metric), the phrase "*inflation expectations remain well contained*" does. The number of chunks joined to form a search unit will have an effect on the amount of separation. Within the raw phrase "*pressures on inflation have picked up in recent months and pricing power is more evident"*, for instance, one may wish to separately consider "*pressures on inflation have picked up in recent months*" and "*pricing power is more evident*".

Based on the search units, the searching process collects data on the Internet to calculate their semantic orientation through search-hit counts on the Google search engine (the information retrieval protocol for the Google-based scores). By averaging all search-units' scores we obtain a value for the semantic orientation for each phrase. Finally, averaging across phrases we obtain a measure of semantic orientation of a statement. We now describe each step separately.

## B.1 Tagging

The problem of the algorithmic tagging of text belongs to the linguistic subfield of natural language processing (NLP). Researchers in NLP have proposed several methods to identify POS and other higher-level grammatical constructs. The family of probabilistic taggers and parsers are a frequently employed approach to this problem. Such approaches rely on already-tagged corpora of text for training the parsers through machine-learning algorithms. A host of off-the-shelf tools are available to make the implementation of complex taggers and parsers easier, and in our implementation, we use a parser that has already been trained on large corpora of text. In particular, we utilize the tagger included in the parser developed by the Stanford Natural Language Processing Group, and, specifically, the tagger based on an unlexicalized probabilistic context-free grammar (Klein and Manning [2003a,b]).[2] This grammar is trained on text from the standard LDC Penn Treebank Wall Street Journal sections 2-21 augmented with some additional data. We apply the tagger to each sentence of the FOMC statements and store the tagged sentences in the customary NLP format. An example of a phrase tagged for POS is shown in Table 1. The POS and corresponding tags in the customary NLP format are shown in Table 2.

**Table 1:** Example of a sentence tagged for parts of speech

| | | | | | | |
|---|---|---|---|---|---|---|
| *Though* | *longer-term* | *inflation* | *expectations* | *remain* | *well* | *contained...* |
| $IN$ | $JJ$ | $NN$ | $NNS$ | $VBP$ | $RB$ | $VBN$ |

**Table 2:** Part of speech tags in the Penn Treebank Project

| Num | Tag | Description | Num | Tag | Description |
|---|---|---|---|---|---|
| 1 | CC | Coordinating conjunction | 19 | PRP$ | Possessive pronoun |
| 2 | CD | Cardinal number | 20 | RB | Adverb |
| 3 | DT | Determiner | 21 | RBR | Adverb, comparative |
| 4 | EX | Existential there | 22 | RBS | Adverb, superlative |
| 5 | FW | Foreign word | 23 | RP | Particle |
| 6 | IN | Preposition or subordinating conj. | 24 | SYM | Symbol |
| 7 | JJ | Adjective | 25 | TO | *to* |
| 8 | JJR | Adjective, comparative | 26 | UH | Interjection |
| 9 | JJS | Adjective, superlative | 27 | VB | Verb, base form |
| 10 | LS | List item marker | 28 | VBD | Verb, past tense |
| 11 | MD | Modal | 29 | VBG | Verb, gerund or present participle |
| 12 | NN | Noun, singular or mass | 30 | VBN | Verb, past participle |
| 13 | NNS | Noun, plural | 31 | VBP | Verb, non-3rd person singular present |
| 14 | NNP | Proper noun, singular | 32 | VBZ | Verb, 3rd person singular present |
| 15 | NNPS | Proper noun, plural | 33 | WDT | Wh-determiner |
| 16 | PDT | Predeterminer | 34 | WP | Wh-pronoun |
| 17 | POS | Possessive ending | 35 | WP$ | Possessive wh-pronoun |
| 18 | PRP | Personal pronoun | 36 | WRB | Wh-adverb |

Source: University of Pennsylvania, Department of Linguistics.

## B.2 Chunking

Once the POS tags have been identified and recorded, we proceed in constructing "chunks" of text. We use the off-the-shelf method implemented in the MontyLingua package by Liu [2004] to separate POS-tagged text into verb, noun, and adjective chunks (VX, NX, and AX respectively).[3]

For example, consider the phrase "*The committee believes*". The POS tagger would generate the POS-tagged text "*The*/DT *committee*/NN *believes*/VB". Then the module separates the

[2] The Stanford part of speech tagger is available at `http://nlp.stanford.edu/software/tagger.shtml`.

[3] In particular the MontyREChunker, available as a Python module, or as Java library.

tagged text into chunks and outputs it in a form: "(NX *The*/DT *committee*/NN NX) (VX *believes*/VB VX)". All the chunking procedures we implemented are standard in NLP. In particular the MontyLingua chunker relies on a regular-expression-based approach also implemented with greater generality in the NLTK package (Bird and Loper [2006]). The code for creating the chunks is available upon request as a Python script.

The rules for generating the chunks are somewhat complex. They are presented here in ascending order of complexity, with the hope that the dedicated reader will identify some of the patterns in the simpler cases making the more complex ones easier to understand. Those familiar with regular expressions may recognize the structure of these rules.

Referring to Table 2 for the POS tags, an adjective chunk is defined as zero or more consecutive words tagged as RB, RBR, RBS, JJ, JJR or JJS, immediately followed by one or more consecutive words tagged as JJ, JJR or JJS. Note that this rule, and those to follow, can result in chunks of arbitrary length.

A verb chunk is defined as zero or more consecutive words tagged as RB, RBR, RBS or WRB, which may optionally be followed by a single word tagged as MD, which may again be followed by zero or more consecutive words tagged as RB, RBR, RBS or WRB. After this group, a word tagged as VB, VBD, VBG, VBN, VBP or VBZ must appear. Following this, there may be zero or more consecutive words tagged as VB, VBD, VBG, VBN, VBP, VBZ, RB, RBR, RBS or WRB. This group may be optionally be followed by a word tagged as RP. The preceding rules describe the mandatory conditions which must be met to qualify as a verb phrase, but any phrase that meets this condition, may be lengthened if immediately followed by a group of words whose first word is tagged as TO, then as zero or more words tagged as RB followed by a word tagged as VB or VBN which may optionally be followed by a word tagged RP.

There are three distinct cases which make up the larger rule describing a noun case. The first case may begin with an optional word tagged PDT, and is then followed by a word tagged as DT, PRP$ WDT, or WP, which is followed by zero or more consecutive words tagged as VBG, VBD, VBN, JJ, JJR, JJS, CC, NN, NNS, NNP, NNPS, or CD. This case requires that the phrase end with one or more consecutive words tagged with NN, NNS, NNP, NNPS or CD. The second noun case begins with an optional word tagged PDT, which is then followed by zero or more consecutive words tagged as JJ, JJR, JJS, ",", CC, NN, NNS, NNP, NNPS, or CD. This case requires that the phrase end with one or more consecutive words tagged as NN, NNS, NNP, NNPS or CD. Finally the third case, marks any word tagged as EX, PRP, WP or WDT as a noun chunk unto itself. The full rule describing noun chunks is an extension of the three rules above. Any words satisfying any of the three cases will qualify as a noun chunk. Additionally, two groups of words satisfying any of the three cases, will be combined into a single noun chunk as long as there is a word tagged as POS joining the two.

Once chunks are generated from the chunker, we decide how many chunks to use for creating search units. After some experimentation with the number of chunks we consider a maximum of five chunks to be the best format (that is, the maximum of an entire raw phrase or five chunks if smaller). All results were also replicated for a maximum of three chunks to a maximum of seven chunks.

### B.3 Google-based measure of Semantic orientation

In constructing the Google-based semantic orientation scores, we follow the information retrieval approach of Turney [2001, 2002] and Turney and Littman [2002]. Their method requires calculating the pointwise mutual information (PMI) between a given search unit and the two ends of a given semantic axis. We defer to Chapter 5 of Manning and Schütze [1999] for a detailed derivation of the information theoretic foundations of PMI, but in synthesis the goal is to compare the odds that two terms $x$ and $b$ appear jointly than to occur independently. This comparison produces a measure of association between concepts $x$ and $b$. Operationally, the chunks defined above form the basis for the search units. For a search unit $x$ and a benchmark word $b$ —"hawkish" or "dovish", for example—the PMI is defined as:

$$PMI(x, b) = log[p(x\&b)/p(x)p(b)],$$

where $p(x\&b)$ represents the probability that concepts $x$ and $b$ occur jointly, and $p(x)p(b)$ represent the probability that they occur independently. Calculating the PMI of a given search unit $x$ against both benchmark words lets us to calculate the semantic orientation of $x$, defined as:

$$SO(x) = PMI(x, hawkish) - PMI(x, dovish). \tag{1}$$

Because we cannot directly access the universe of webpages on the Internet, we use Google hit counts on joint searches of a given search unit $x$ and each word in the antonymy "hawkish-dovish", as empirical estimates of the PMI's in (1). Based on these hit-counts, we obtain an empirical estimate of the Google-based semantic orientation score as:

$$\begin{aligned} \overline{GSO}^h(x) &= \log\left(\frac{hits(hawkish, x)}{hits(dovish, x)}\frac{hits(dovish)}{hits(hawkish)}\right) \qquad (2) \\ &= \log\left(\frac{hits(hawkish, x)}{hits(dovish, x)}\right) + \xi \end{aligned}$$

where $\xi$ is a constant.[4] Note that the antonymy "hawkish-dovish" is often used in other contexts than discussions of monetary policy. Because their frequency of occurrence might differ from those in discussions of monetary policy, the constant $\xi$ is probably measured with error making it hard to identify the level of the score. We directly abstract from this level by defining the Google semantic

[4]To be precise, $\xi = \log\left(\frac{hits(dovish)}{hits(hawkish)}\right)$ is a constant term at the moment a set of Google searches over all $x$'s is launched.

orientation score (GSO) as:

$$GSO^h(x) = \overline{GSO}^h(x) - \xi. \tag{3}$$

in what follows avoiding to interpret the score's level. This lack of identification is, however, rather inconsequential for our empirical analysis as we use the score in first differences (for which the constant drops out). As an example, we report in Table 3 results of the analysis of the statement of December 12, 2006:

"*Economic growth has slowed over the course of the year, partly reflecting a substantial cooling of the housing market. Although recent indicators have been mixed, the economy seems likely to expand at a moderate pace on balance over coming quarters. Readings on core inflation have been elevated, and the high level of resource utilization has the potential to sustain inflation pressures. However, inflation pressures seem likely to moderate over time, reflecting reduced impetus from energy prices, contained inflation expectations, and the cumulative effects of monetary policy actions and other factors restraining aggregate demand. Nonetheless, the Committee judges that some inflation risks remain. The extent and timing of any additional firming that may be needed to address these risks will depend on the evolution of the outlook for both inflation and economic growth, as implied by incoming information.*"

The formulation for (3) can also be extended to multiple semantic dimensions at the small cost of increasing the number of searches in Google. We consider a set of candidate antonymies that were verified on a thesaurus.[5] The extended set of antonymies that includes a total of 12 words is listed in Table 4.

In order to obtain a score for the extended set of antonymies, we computed the sum of Google hits for each benchmark word $b$ separately for each column of Table 4 (positive and negative) so as to obtain a total number of hits of words referring to rate increases, and a total pointing to rate declines. Then, we constructed the Google-based semantic orientation score for the expanded set of antonymies as:

$$GSO^e(x) = \log\left(\frac{\sum_{b \in Positive} hits(b,x)}{\sum_{b \in Negative} hits(b,x)}\right). \tag{4}$$

which is a formulation similar in spirit to (3), but clearly not equivalent to an average of the semantic orientation scores obtained separately for each antonymy. The operator (4) accounts for the fact the words in each set of extended of antonymies are synonyms—at least in terms of our metric—and should therefore be treated as substitutes. It also has the advantage of drastically reducing the problem of not finding hits for some of the words in each set in conjunction, which would make (4) undefined due to the presence of the logarithm.

Once scores are computed for $\{k = h, e\}$ and for all search units $x$, we average the scores $GSO^k(x)$

[5] We believe this set of six antonymies captures accurately most of the debate about central bank stance. We also experimented with a small set of alternative, more exotic, and much less relevant antonymies (both semantically and in terms of context in which the words are generally used) without much effect on results other than increasing noise to signal ratios. Examples are antonymies: "jack" versus "lax" or "jack" versus "slash".

across all $x$'s in a statement released $t$, thus obtaining an average assessment of the stance of the statement at that date. The score for the statement released at $t$ is then:

$$GSO^k(t) = \sum_{x \text{ at } t} \frac{GSO^k(x)}{\#(x \text{ at } t)}.$$

As a proxy to the unexpected shock to the statement, we subtract from the score at meeting $t$ the semantic orientation score at meeting $t-1$:

$$\Delta GSO^k(t) = GSO^k(t) - GSO^k(t-1).$$

See the main text of the paper for further discussion on this approximation.

### B.4 Remarks on the Google searches

Some important issues arose in the actual implementation of the Google-based IR described above. Because generating hit counts for searches is a computationally intensive task, the Google search engine only delivers approximate counts, rather than actual ones, using a proprietary algorithm (unavailable to the public). In order to obtain comparable and relatively error-free measures we employ a few strategies, which we now discuss, as they might be unfamiliar to normal users.

The main issue in terms of accuracy was to avoid that the web-traffic optimizer embedded in Google randomly switched our searches across Google data centers mixing together information (textual corpora) of different centers making the hit counts hard to compare across searches. In fact, Google servers store multiple caches of web information, which are located in the different data centers and are not necessarily identical in the type of information they carry, and, on occasion, can produce drastically different hit counts. Our solution to this problem was to force Google to operate on a single IP address (74.125.95.104), which we picked randomly among the Google data centers' IP addresses, making sure all our searches were implemented on the same corpus of text.[6] This protocol excluded the use of the much faster Google University Research program, which does not allow such "IP forcing" option, and it required focusing on periods of low web traffic in the U.S. (nights, weekends or holidays mostly).[7] We employed courtesy times of 5 seconds between searches, a commonly employed method for reducing the demands on the unwitting datacenter. The hit count information was scraped directly from the resulting pages using a code in Python.

In our experience Google also appeared to produce extremely noisy counts for very short search

[6] Results were also tested across different data centers for consistency. No detectable difference in precision were evident once we made sure of searching the same data center.

[7] While the hit values we employ are obtained by standard Googling, access to Google's research program allows for an automated execution of this process via their API. The program is only open to university researchers, and applicants must apply and be approved before being granted access to the program. Details about obtaining a login id for the program, as well as sample code are available at `http://research.google.com`. Searches are returned via xml, which includes a field for the estimated number of hits. We are grateful to Hal Varian, chief economist at Google, for help with the Google Research Program.

**Table 3:** Search units and Google semantic orientation score in the Dec 12, 2006 FOMC statement

| Phrase id | Search unit "x" | Chunks in phrase | Chunks in "x" | Hits ("x",hawkish) | Hits ("x",dovish) | $GSO^h$("x") |
|---|---|---|---|---|---|---|
| 337 | Economic growth has slowed over the course of the year | 4 | 4 | 490 | 327 | 0.4 |
| 337 | partly reflecting a substantial cooling of the housing market | 3 | 3 | 225 | 225 | 0 |
| 338 | recent indicators have been mixed | 3 | 3 | 102 | 633 | -1.82 |
| 338 | the economy seems likely to expand at a moderate pace | 8 | 5 | 112 | 317 | -1.04 |
| 338 | seems likely to expand at a moderate pace on balance | 8 | 5 | 213 | 166 | 0.24 |
| 338 | likely to expand at a moderate pace on balance over coming | 8 | 5 | 225 | 161 | 0.33 |
| 338 | expand at a moderate pace on balance over coming quarters | 8 | 5 | 225 | 9 | 3.21 |
| 339 | Readings on core inflation have been elevated | 4 | 4 | 129 | 485 | -1.32 |
| 339 | the high level of resource utilization has the potential to sustain | 6 | 5 | 177 | 584 | -1.19 |
| 339 | resource utilization has the potential to sustain inflation pressures | 6 | 5 | 271 | 96 | 1.03 |
| 340 | inflation pressures seem likely to moderate over time | 5 | 5 | 229 | 135 | 0.52 |
| 340 | reflecting reduced impetus from energy prices | 3 | 3 | 40 | 98 | -0.89 |
| 340 | contained inflation expectations | 2 | 2 | 142 | 101 | . |
| 340 | the cumulative effects of monetary policy actions and other factors restraining aggregate demand | 4 | 4 | 216 | 167 | 0.25 |
| 341 | the Committee judges that some inflation risks remain | 3 | 3 | 194 | 107 | 0.59 |
| 342 | The extent and timing of any additional firming that may be needed to address these risks | 9 | 5 | 150 | 122 | 0.2 |
| 342 | any additional firming that may be needed to address these risks will depend | 9 | 5 | 154 | 127 | 0.19 |
| 342 | that may be needed to address these risks will depend on the evolution | 9 | 5 | 153 | 131 | 0.15 |
| 342 | may be needed to address these risks will depend on the evolution of the outlook | 9 | 5 | 157 | 133 | 0.16 |
| 342 | these risks will depend on the evolution of the outlook for both inflation and economic growth | 9 | 5 | 159 | 130 | 0.2 |
| 342 | as implied by incoming information | 2 | 2 | 340 | 202 | 0.52 |

**Table 4:** Semantic antonymies

| **Positive:** Implying higher rates | **Negative:** Implying lower rates |
|---|---|
| *Hawkish* | *Dovish* |
| *Hike* | *Cut* |
| *Tight* | *Ease* |
| *Raise* | *Lower* |
| *Increase* | *Decrease* |
| *Boost* | *Loose* |

units. Think for instance at the recurring phrases "*the committee believes*" or "*the committee judges*" (frequent in the statements of the second half of 2003), which are obviously a excessively short to have a clear semantic orientation along our metric.[8] All searches below four words (even if identified as a search unit by the chunking rules) were dropped due to the impossibility of assigning a clear semantic orientation to such text.

Robustness across various search engines is a potentially important issue, and we have also attempted to implement the IR on Yahoo!.[9] In comparing results from Google and Yahoo!, we found similar, although not identical, hit-counts. In the paper we only report results for Google, as this search engine is generally thought of having better coverage of the universe of webpages.

# C Factiva-based Information Retrieval

In addition to the Internet-based IR protocol, we also consider one based on a large sample of business news and sources. As discussed above the Internet-based semantic scores are only indirectly measured on webpages via Google searches. The IR used in this implementation, instead, is more in line with a traditional corpus-based computational linguistic approach, in that we implement the IR directly on the content of the corpus. In our analysis we employ Factiva, a provider of business news and a content delivery tool. Factiva includes a collection of more than $25,000$ news sources including, among many others, The Wall Street Journal, the Financial Times, as well as newswires from Dow Jones, Reuters, and the Associated Press.

## C.1 Background and sources

We construct the Factiva-based semantic scores starting from the collection of all documents available in the Dow Jones Factiva database on days in which FOMC statements are released, on the

[8]Other short sentence such as "*the increase in energy prices*" (October 3, 2000), although apparently suggestive of inflationary pressures, on a second reading lack the necessary semantic connotation to clarify any policy stance.

[9]Altavista, at one time another major search engine, is now based on Yahoo!'s cache (see Turney [2001]). Since Yahoo does not offer researchers an API to their searching interface, the results had to be obtained by scraping off from the html.

days preceding the releases, and on the following days. This three-day window around the announcement date lets us focus on information as pertinent as possible to the measurement of the announcement (The information ahead of the release of the statement is used below to form a measure of expectations.)

From Factiva we collected all newspaper, magazine, Internet wire and newsletter entries available worldwide in English.[10] All documents were subdivided into sentences using a parser and assigned an unique identifier. The original text documents range from very short pieces of newswire information to long newspaper articles and commentaries. This constitutes the original corpus of text on which we operate our Factiva searches. The total number of sentences is $1,302,977$. The original documents were recorded from the Factiva database as XML documents, simplifying the task of parsing the text in sentences and of identifying the times in which the news were released.

**Table 5:** Top 10 news sources in Factiva data

| Factiva Source | Freq. | Percent |
|---|---|---|
| *Reuters News* | $184,225$ | 31.37 |
| *Market News International* | $94,416$ | 16.08 |
| *AFX Asia* | $60,076$ | 10.23 |
| *Dow Jones International* | $49,947$ | 8.50 |
| *AFX International Focus* | $47,133$ | 8.02 |
| *Associated Press News* | $43,845$ | 7.47 |
| *Dow Jones Capital Markets* | $36,882$ | 6.28 |
| *Dow Jones Commodities* | $26,917$ | 4.58 |
| *Agence France Presse* | $22,178$ | 3.78 |
| *The Wall Street Journal* | $21,721$ | 3.70 |
| Total Sentences from Top 10 | $587,340$ | 100.00 |

## C.2 Factiva Implementation

Since the Factiva corpus is much smaller than the Google one, which includes billions of web pages, exact matching of portions of the statement proves impossible to implement with reasonable coverage. However, an alternative and conceptually affine methodology is available. Since our objective is to generate a semantic orientation index for the stance of the announcements, we first subdivide the complete list of sentences into two groups: Sentences relevant for monetary policy (39.9 percent of the sentences in the Factiva sample) and sentences not apparently relevant (60.1

[10]Some cleaning up of the Factiva sources was implemented. For instance many Australian newswires apparently add the word Fed in the title to all Federal government-related news. We therefore drop, as sources, the "Australian Associated Press" and " The Western Australian". For similar reasons we also exclude "Federal Government Broadcast", the "Journal of Animal Sciences" and "Marine Biology" as sources.

percent of the sample). Relevance was assigned based on the presence of a direct match within a sentence of the words presented in Table 6 along with strictly equivalent words.[11,12]

**Table 6:** Factiva Word Matches for Relevance of Sentence

| |
|---|
| *Rates,* |
| *Policy,* |
| *Policies,* |
| *Statement,* |
| *Announcement,* |
| *Fed,* |
| *FOMC,* |
| *Federal Reserve.* |

The presence of any of these matches and their coincident timing with the statement release arguably identifies sentences related those announcements. In order to precisely assign the sentences to the period before or after the statement we collect the exact time and date of the news release and dropped entries for which the exact time of the release was not available.

### C.3 Factiva Semantic Orientation

We now define a measure of policy stance before and after the release of the statement based on "relevant" sentences in the Factiva corpus. In particular, we construct a semantic measure, $FSO^e$, using joint frequencies for sentences matching the relevance criterion, that is including at least one word in the list of Table 6, and words in the expanded set of antonymies of Table 4. We compute these frequencies using indicator variables for matches on the "positive" set of antonymies (any match for words listed on the left column of Table 4) and indicator variables for matches on the "negative" side of the semantic antonymies set (right column of Table 4) for each "relevant" sentence $s$. Accordingly we also define the score $FSO^h$ on a list of antonymies collapsed to the word-pair "hawkish-dovish". Table 7 reports an illustrative sample of Factiva sentences for the statement of March 22, 2005, with matches for dummies corresponding to the words in the list of "positive" antonymies.[13]

Given the high degree of control over the corpus in the Factiva IR, we apply several refinements in order to improve the measurement of the semantic scores. First, we consider both literal matches to the list in Table 4 as well as strictly equivalent words.[14] Further, because some sentences could

[11] For example, the verb "announced" is considered equivalent to the word "announcement". We achieve this by matching the roots, rather than the complete spelling, of the words in the list. We use the same procedure in matching roots of the words in the expanded set of antonymies below.

[12] We drop sentences that present an absurdly high number of matches, as they very frequently identify bad entries in Factiva (such as mis-parsed tables and data series). The threshold we employ is above 5 matches in the same sentence, a clear break in the empirical distribution and corresponding to the 99th percentile of the distribution of matches per sentence.

[13] These sentences are the actual title lines of four different news articles. Sample is selected for illustrative purposes out of the entire set of sentences for March 22, 2005, which includes all the sentences in the article bodies as well.

[14] See footnote 11.

**Table 7:** Example of relevant sentences in the Factiva corpus and corresponding values of the "positive"-matching dummies

| Sentence | Source and Date | Hawkish | Tight | Raise | Hike | Increase | Boost |
|---|---|---|---|---|---|---|---|
| DJ Australian dollar slips with Fed Hike in view; Bonds down | Dow Jones Chn. 21/3/05 0:54 | 0 | 0 | 0 | 1 | 0 | 0 |
| MARKET BEAT: CSFB Predicts "measured" Fed stance | Dow Jones Intl. 22/3/05 8:31 | 0 | 0 | 0 | 0 | 0 | 0 |
| MARKET BEAT: Hawkish Positioning From Fed | Dow Jones Intl. 23/3/05 15:05 | 1 | 0 | 0 | 0 | 0 | 0 |

possibly relate to the action taken at the meeting, rather than to future actions and the content of the announcement, we remove all instances in the past tense for verbs, thus avoiding discussions of the most recent or past actions (We also control for policy actions in the empirical analysis). Because of their opposite meaning, we also count direct negations of the words in each the set of antonymies as belonging to the opposite set. It is important to notice that the possibility to impose these refinements and to pinpoint the semantic orientation of given sentences, rather entire webpages, is only available in the Factiva IR, for which we can access the underlying documents, but not in the Google IR implementation. After applying these refinements, our data contain a list of relevant sentences with their corresponding exact time and date of release and dummy variables corresponding to matches with each of the words in the list of antonymies as in Table 7. Let $I[\sigma, \mathbf{R}, \mathbf{P}]$ be an indicator function that takes value 1 if sentence $\sigma$ contains at least a word from list or relevant words $\mathbf{R}$ and at least a word from list of "positive words" $\mathbf{P}$, and 0 otherwise. Analogously define $I[\sigma, \mathbf{R}, \mathbf{N}]$, for the list of "negative" words $\mathbf{N}$. To obtain measures of unexpected changes in the stance of policy around the announcement, we then calculate the two indicator functions for all sentences $\sigma$ in the set of documents released ahead $\mathbf{T}_{t-}$ and after $\mathbf{T}_{t+}$ the announcement released at $t$. We compute the Factiva semantic orientation score after the announcement $t$ as:

$$FSO^{e}_{t+} = \log\left(\frac{\sum\limits_{\sigma\in\mathbf{T}_{t+}} I\left[\sigma, \mathbf{R}, \mathbf{P}\right]}{\sum\limits_{\sigma\in\mathbf{T}_{t+}} I[\sigma, \mathbf{R}, \mathbf{N}]}\right) = \tag{5}$$

$$= \log\left(\frac{\text{Prob}_{t+}\left(\text{Relevant } \cap \text{Positive}\right)}{\text{Prob}_{t+}\left(\text{Relevant } \cap \text{Negative}\right)}\right). \tag{6}$$

where $\mathbf{P}$ and $\mathbf{N}$ are the antonymies defined in the two columns of Table 7. Restricting the sets to the "hawkish-dovish" word pair we define the score after announcement $t$ based on this antonymy:

$$FSO^h_{t+} = \log\left(\frac{\sum\limits_{\sigma\in\mathbf{T}_{t+}} I\left[\sigma, \mathbf{R}, \text{hawkish}\right]}{\sum\limits_{\sigma\in\mathbf{T}_{t+}} I[\sigma, \mathbf{R}, \text{dovish}]}\right) = \tag{7}$$

$$= \log\left(\frac{\text{Prob}_{t+}\left(\text{Relevant } \cap \text{hawkish}\right)}{\text{Prob}_{t+}\left(\text{Relevant } \cap \text{dovish}\right)}\right). \tag{8}$$

We then measure the unexpected change in policy stance for the date $t$ policy announcement as the difference between the scores computed on sentences released ahead and after the announcement. That, is for each $k = \{h, e\}$ these differences are:

$$\Delta FSO^k(t) = FSO^k(t^+) - FSO^k(t^-)$$

For the occasional unscheduled meeting we lack sufficient data for assessing expectations about the statement ahead of the announcement. Hence, we employ the $t-1$ post-statement measure as opposed to the $t$ pre-statement values to generate the $FSO$ movements. The possibility to construct semantic measures based on the exact time of the documents in the Factiva IR greatly improves the measurement of announcement shocks. Similar calculations are impossible in the Google IR, for which a rigorous division of the corpus based on date of release or posting of the webpage is not possible using the current technology.

## References

Bird, S., and E. Loper (2006): "Natural Language Toolkit," Available at: `nltk.sourceforge.net`.

Klein, D., and C. Manning (2003a): "Accurate unlexicalized parsing," in *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pp. 423–430. Association for Computational Linguistics Morristown, NJ, USA.

——— (2003b): "Fast exact inference with a factored model for natural language parsing," *Advances in Neural Information Processing Systems*, pp. 3–10.

Liu, H. (2004): "MontyLingua: An end-to-end natural language processor with common sense.," .

Lucca, D., and F. Trebbi (2009): "Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements," Mimeo, University of Chicago.

Manning, C., and H. Schütze (1999): *Foundations of statistical natural language processing.* MIT Press.

Turney, P. D. (2001): "Mining the Web for synonyms: PMI-IR versus LSA on TOEFL," *Proceedings of the Twelfth European Conference on Machine Learning*, pp. 491–502.

——— (2002): "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 417–424.

Turney, P. D., and M. L. Littman (2002): "Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word Corpus," National Research Council, Institute for Information Technology, Technical Report ERB-1094.