The Language of Earnings Announcements

Vitaly Meursault*

08 November, 2019

Abstract

This study quantifies and characterizes the information content of earnings announcement language via a statistical model of language that extracts the latent factors most associated with absolute returns around the time earnings announcements are released. The language of earnings announcements explains 11% of the variation in absolute announcement returns out-of-sample. That is comparable to the explanatory power of standard numerical variables. Using the latent factors to recover the features that are important, we show that the information content depends on what is mentioned, how it is mentioned, and where in a document it is mentioned. Findings show that earnings components are more important than bottom line net income. Sentiment and forward-lookingness amplify the information content of all themes, and information content is more concentrated at the beginnings of texts.

*Tepper School of Business, Carnegie Mellon University. Email: vitaly.meursault@gmail.com. I thank my advisor, Bryan Routledge for the mentorship, Chester Spatt, Dokyun Lee, and Stephen Karolyi for their advice; and Madeline Scanlon and Eric Lu for their support. All errors are my own. This work used the Bridges system, which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC) (Nystrom et al. 2015). I gratefully acknowledge the support by PSC staff members Tom Maiden and Rick Costa.

1 Intro

Earnings announcements contain important information for investors. Absolute price changes on announcement days are more than twice the size of changes on other days (Beaver, McNichols, and Wang 2018). Various aspects of earnings press releases are shown to be informative for the markets, including the presence of operational details in a release (Francis, Schipper, and Vincent 2002a) and forward-lookingness (Bozanic, Roulstone, and Van Buskirk 2018). Interestingly, the content of earnings announcements is varied. For example, companies can mention topics such as their segment sales ("Meanwhile, sales for our disk drive analyzer products reached their highest level in more than two years"; LeCroy, 2005, Q3), or relay their excitement about the potential of their products ("We believe that these new and enhanced products represent great potential"; Mobius Management System, 2006, Q1). This paper shows that we can systematically understand what language is important.

To quantify the information content of announcement language, we use a flexible statistical model. Specifically, the deep neural network approach employed here allows extracting information from language (words in order) without prejudging the value of content. The model estimates a series of latent factors based on all document content. These factors reflect word usage, local context, and word order. Here, the latent factors are used to predict absolute returns around the time of an earnings announcement.

Earnings announcement language has high information content. By our estimate, the statistical model of language explains 11% of the variation in absolute announcement returns. Regression analysis of the same returns data with standard financial variables has comparable explanatory power. Combining language with the numerical variables consistently improves the explanatory power of the model. About 13% of the information contained in the language is incremental to my full set of numbers. Overall, language both substitutes for and complements the numbers.

Since the model explains the absolute price changes well, we use it to understand what features of the language of earnings announcements are valuable to investors. To do that, we project the latent factors onto the document space by quantifying, for every sentence, how much the prediction of the model would change if the sentence were absent. This process enables a variety of thought experiments about how language conveys information. For example, we estimate what happens with removal of all sentences mentioning earnings components, forward-looking sentences, sentences with tone markers, quotes, and so on.

The magnitude of absolute price changes depends on the content of earnings announcements. When a document mentions earnings components, the price changes tend to be larger than when it mentions bottom-line net income. Mentions of losses are associated with especially large swings in prices.

It is interesting to see that the way earnings announcements say something can be as important as what they say. Our model encapsulates interactions between the content groups and language features. Sentiment and forward-lookingness amplify the information content of all content groups. Quotes from managers are informative when they talk about earnings components and operations. Markets also respond to discussions of non-GAAP measures.

Starting with Beaver (1968), many studies examine the importance of various items contained in earnings announcements. Examples include earnings value (Ball and Brown 1968), earnings components (Lipe 1986), recurrent and nonrecurrent components of earnings (Fairfield, Sweeney, and Yohn 1996), non-GAAP earnings (Bradshaw and Sloan 2002), and balance sheet items (Collins et al. 1997).

A series of papers discusses specific textual features of earnings announcements. Examples include the provision of operational details (Francis, Schipper, and Vincent 2002b), an emphasis on non-GAAP measures (Bowen, Davis, and Matsumoto 2005), sentiment (Davis, Piger, and Sedor 2012), and prevalence of forward-looking statements (Bozanic, Roulstone, and Van Buskirk 2018). An extensive literature studies financial texts other than earnings announcements. Popular collections of texts include forms 10Q/10K, earnings calls, and financial media. Many papers focus on extracting specific features of text like sentiment (Tetlock 2007; Li 2010), readability (Loughran and McDonald 2014), or prevalence of forward-looking statements (Muslu et al. 2015). It is common to extract textual features either by hand or using rule-based algorithms.

Another literature applies statistical models to financial disclosure texts. In Finance, Kogan et al. (2009) pioneered the use of text regression, in which word and phrase counts are used in predictive tasks. Routledge, Sacchetto, and Smith (2017) and Chebonenko, Gu, and Muravyev (2018) take a similar approach. The current study uses a deep neural network that can model language in a more flexible way.[1]

Statistical models of language can help link disclosure theory to textual data. In models, scholars tend to look at signals with low dimensional value. For example, in Kim and

[1] Topic modeling is a statistical approach alternative to text regression. It involves extracting latent factors (topics) based on word concurrence across documents. The topics can then become inputs to regressions. Hoberg and Lewis (2017) is an example of this approach.

Verrecchia (1991), Kim and Verrecchia (1994), and in Tetlock (2010), the signal is univariate normal. Statistical models of text provide a tool for processing raw text data into a signal. The mapping of text to signal can be a source of disagreement and lead to market frictions.

Statistical models of language can be of practical use to investors. Earnings announcements are informative, but investor attention is constrained. Our model can ease the constraints by extracting the most informative text.

Earnings announcements also attract attention from regulators, often because of the possibly misleading non-GAAP reporting. Our model can estimate the magnitude of price changes associated with non-GAAP measures. This feature can be useful in the regulatory process.

2 Data

Data are the texts of earnings announcements from SEC's EDGAR database combined with information from Compustat and CRSP. The dataset contains 128,317 observations at the firm-quarter level and covers the period from 2005 to 2017.

2.1 Earnings Announcements Corpus

The corpus includes forms 8-K that contain a "Results of Operations and Financial Condition" item according to their metadata and that are filed within five days of an earnings release date as recorded in Compustat. This paper refers to the whole text of the 8-K as the earnings announcement, even if it contains other items besides the earnings press release. The median number of words, punctuation marks and special symbols (collectively referred to as tokens) in an earnings announcement in this corpus is 3,278. Across the sample years, the median size of a document grows from 2,282 to 3,816 tokens.

The Appendix A presents details of corpus construction.

2.2 Compustat and CRSP Data

The usual numerical variables related to the firms' performance come from Compustat and CRSP. Earnings announcements are matched to Compustat observations corresponding to the quarter discussed in the earnings announcement. For example, an earnings announcement released during the second quarter that discusses the first quarter is matched with the Compustat observation for the first quarter. This way earnings announcements are matched to accounting numbers that are public at the time of the release or at least soon after.

Calculation of the variables derived from CRSP data, such as absolute announcement returns and stock volatility, is based on the earnings announcement date from Compustat.

2.3 Data split

We split the data into training, validation, and test sets to measure information content. Since the models used in this paper involve large numbers of parameters and have the potential for overfitting the data, the performance of the models was evaluated on the test set, a subset of data not seen by the models during estimation.

The total number of firms represented in the dataset is 6,202. We randomly split the firms into training (4,462 firms and 91,914 observations), validation (882 firms and 18,431 observations), and test sets (858 firms and 17,972 observations). The training and validation sets are used to estimate the model, and the test set is used for the model evaluation. All the results presented in this paper are obtained using only the test set—that is, the observations generated by the firms whose data were not used for model estimation.

3 Statistical Model of Language

The statistical model of language used here is a CNN-GRU (Convolutional Neural Net with Gated Recurrent Units) that operates by sequentially creating a set of latent factors. These factors reflect word usage, local context, and word order. Minimization of the prediction error drives factor extraction.

3.1 Model Overview

CNN-GRU is a composition of several functions:

$$f(X_{f,t}; w) = \mathrm{L} \circ \mathrm{GRU}_2 \circ \mathrm{GRU}_1 \circ \mathrm{Max}_3 \circ \mathrm{Conv}_3 \circ \mathrm{Max}_2 \circ \mathrm{Conv}_2 \circ \mathrm{Max}_1 \circ \mathrm{Conv}_1 \circ \mathrm{Emb}(X_{f,t}),$$

where $X_{f,t}$ is the document, w is the set of model parameters, Emb is the embedding, Conv is the convolution, Max is max pooling, GRU is the Gated Recurrent Unit, and L is linear layer.

The document is represented as a vector $X = [x_1 \cdots x_n]$, where x_j is the index of the j'th word in the vocabulary, and n is the maximum document length. If a document is shorter than the maximum document length, the extra space is filled with special padding words.

Embedding (Emb) assigns embedding vectors to individual words. The starting values for word embeddings are obtained using the word2vec algorithm (Mikolov et al. 2013) applied to the training set. The embedding vectors are then updated during estimation. An important property of embeddings is that words that likely to be used interchangeably tend to cluster in the embedding space. Figure 1 shows several such clusters, including names of the months, corporate titles, and the words "earnings," "income," and "loss." Thus, embeddings reflect word usage.

The model proceeds by creating a series of latent factors. Convolutions create latent factors that reflect local context. For example, a latent factor might learn to distinguish "net income" from "operating income." Max pooling only prunes the factors, keeping the ones most associated with absolute announcement returns. At the next stage, gated recurrent units create latent factors that take into account word order. For example, the model can learn to distinguish between the phrases "the board of directors dismissed the CEO" and "the CEO dismissed the board of directors." Finally, the linear layer takes the final set of latent factors and runs them through a linear regression to produce the prediction.[2]

All layers of CNN-GRU are estimated simultaneously. The objective function is the sum of squared errors:

$$\min_{w} \sum_{f,t} \left[y - f(X; w)\right]^2 ,$$

[2]Convolutional neural nets were originally used for computer vision tasks, but quickly made headway into natural language processing (Collobert et al. 2011; Kim 2014). Along with other word-order-aware models, convolutional neural networks are becoming one of the standard choices in sentence and document classification tasks (including sentiment analysis) and make some headway into the social sciences, where they are used to, for example, classify political discourse (Bilbao-Jayo and Almeida 2018).

where $f(X, w)$ is the prediction of the model. The weights are estimated using an iterative procedure.[3]

3.2 Model Interpretation

We want to use the model to know what is important in earnings announcements. Intuitively, if a passage is important, removing it should change the model's prediction. CNN-GRU is a nonlinear model containing more than 600,000 parameters. To interpret it, one computes impact scores, which represent the change in model prediction resulting from moving from an empty space to the observed word.

Impact scores were computed using integrated gradient method (Sundararajan, Taly, and Yan 2017).[4] Integrated gradients correspond to Aumann-Shapley values from Game Theory (Aumann and Shapley 1974). They are computed at the word level by propagating the derivatives associated with individual words through CNN-GRU. The same word can have different impact scores depending on its context. Within every document the scores sum to the model prediction. The Appendix B provides details.

Figure 2 provides a sample paragraph with words highlighted according to their impact scores. Red represents wirds with an impact larger than zero. Removing them would drive the model's prediction downwards. Blue words have an impact smaller than zero. Removing them would result in higher $|\hat{R}^a|$. Color density represents the magnitude. In this example, the word "earnings" and the phrase "strategic decision" are associated with low absolute

[3] We use adaptive moment estimation (Adam) algorithm, which is a form of stochastic gradient descent.

[4] Other methods of computing attribution scores include Shapley additive values (SHAP) (Lundberg, Allen, and Lee 2017), DeepLift (Shrikumar, Greenside, and Kundaje 2017), and Relevance Propagation (Arras, Horn, et al. 2017; Arras, Montavon, et al. 2017)

announcement returns, while the phrase “Mike Brooks, chief executive officer, commented” is associated with more repricing activity.

The Appendix presents details of the model’s architecture and training. Table 2 presents an overview of the model, including the number of parameters.

4 Information content of language

If the language of earnings announcements is informative, we should be able to predict the absolute announcement returns well just by looking at the language. First, we show that language has high information content, with information content defined as the ability of language to predict absolute announcement returns. The information content is defined as the ability of the language to predict absolute announcement returns. We get an idea of relevant magnitudes by comparing language to several sets of numerical variables. Second, we show that language both substitutes for and complements the numbers. While the media differ, language and numbers both reflect value-relevant information.

We define the information content of a set of variables as the out-of-sample R^2 of a regression model that has the absolute announcement return value on its left.

To understand relative magnitudes, we compare the information content of several sets of variables, estimating the following regressions:

$$\text{Language: } |R^a| = f(X) + \epsilon,$$

$$\text{Numbers: } |R^a| = \alpha + \beta_s N_s \epsilon,$$

$$\text{Combined: } |R^a| = \alpha + \beta_l f(X) + \beta_s N_s + \epsilon,$$

where $|R^a|$ is absolute return for three days centered around the earnings announcement minus the S&P500 return for the same period, $f(x)$ is CNN-GRU, and $s \in$ {Accounting, Market, Lags, Fixed Effects, All} is the subset of numerical variables. Theå *language* regression refers to the CNN-GRU model. The *numbers* and *combined* models are elastic net regressions (Zou and Hastie 2005). Elastic net adds regularization terms to the OLS objective function to improve out-of-sample performance. The *combined* model uses the prediction made by the *language* model as one of the inputs.

Accounting variables are obtained from Compustat (variable definitions in footnotes):[5] earnings per share,[6] log size,[7] earnings to assets,[8] volatility of earnings to assets for the last four quarters, market to book,[9] accruals to assets,[10] special items to assets,[11] loss indicator.[12]

Market variables are obtained from CRSP: stock return and volatility across the fifty trading

[5]All continuous variables in this category are included in the regression in five forms: level in the latest available quarter, change between the last available and the previous quarter, the absolute value of that change, change between the last availaible quarter and the same quarter in the previous year, the absolute value of that change.

[6](ibq - spiq) / cshoq

[7]log(prccq * cshoq)

[8]niq / atq

[9](prccq * cshoq + ltq) / atq

[10](niq - oancfy) / atq

[11]spiq / atq

[12]I[ibq < 0]

days preceding an announcement date (approximately one quarter).

Lags refer to the absolute return around earnings announcements in the previous year (four lags).

Fixed effects include year-quarter and industry indicators.

Figure 3 depicts the performance of different sets of variables. Language has an out-of-sample R^2 of 11%. This is comparable to the results obtained using individual sets of numbers. The best performing single set of numeric variables, *lags* also has an out-of-sample R^2 of 11%. Combining all sets of numerical variables leads to a higher R^2 of 18%.

The information content of language is incremental to that of numbers. Figure 4 shows that combining language with numbers consistently produces a better model. The best model has an out-of-sample R^2 of 18%. It includes the language and all sets of numerical variables. Bootstrap analysis confirms that all differences are significant at the 1% level. Table 1 presents the bootsrapped R2 values with standard errors.

We define the degree of complementarity to compare the information content of the language and numerical variables. The degree of complementarity reveals what percentage of information contained in one subset of variables is complementary to another subset:

$$Comp(A, B) = \frac{R^2(A \cup B) - R^2(B)}{R^2(A)}.$$

A and B are subsets of variables. For example, A could be the language, and B could be accounting variables. $R^2(A)$ denotes the out-of-sample R^2 of a regression that operates on the subset of variables A. $Comp(A, B)$ of 0% means that subset A contains no new information relative to subset B. Conversely, $Comp(A, B)$ of 100% means that all information contained in subset A is new relative to subset B.

Figure 5 depicts the complementarity of language and different subsets of numbers (and vice versa). About one to two thirds of the information content of language is complementary to individual subsets of numerical language. Ten percent of the information contained in the language is new relative to the full set of numerical variables.

A statistical model of language can explain a significant portion of absolute announcement returns. This suggests that language is efficient at conveying the state of a firm. Language and different sets of numbers both substitute and complement each other. This suggests two points: First, language reflects a wide range of company characteristics, and not, for example, merely a firm's industry or size. Second, language and numbers are two different media that reflect the same underlying state of a company.

5 What makes earnings announcements informative?

Some earnings announcements cause substantial price revisions. Others go barely noticed. What makes some earnings announcements more informative than others? In this section, we show that what the press releases say, how they say it, and where they say it all affect informativeness.

Analysis is at the sentence level. For every sentence, we calculate an impact score by summing the impact scores of all its words. The measurement units of impact are absolute returns. The impact reflects the change in model prediction relative to the benchmark condition of the sentence's being absent. We also keep track of a sentence's position in its document to analyze position's effect on sentence impact.

Sentences are classified into content groups using keyword matching. The four groups of interest are earnings, loss, earnings components, and operations. The earnings group includes all sentences containing the words "earnings" or "net income." The loss group includes sentences containing the word "loss." The earnings components group includes sentences mentioning at least one of the common components of earnings.[13] The operations group includes sentences containing at least one of the keywords from Muslu et al. (2015) related to firm operations.

Within the content groups, I identify sentences having specific properties, looking at whether a sentence is positive or negative in tone (sentiment) and at whether it contains forward-looking statements, references to non-GAAP reporting, and is quoted from a firm executive. For sentiment, we use Loughran and McDonald's sentiment word list. For forward-lookingness, the word list from Muslu et al. (2015) is used, and the Stanford NLP Group's Quote Annotator identifies quotes (Manning et al. 2014). We identify sentences related to non-GAAP reporting by looking at words such "adjusted," "non-GAAP" or "pro-forma," as well as the names of standard non-GAAP exclusions.

5.1 Earnings, Earnings Components, and Loss

Earnings announcements can provide different levels of detail to a reader. A firm can focus on the bottom line by discussing earnings: "Recent key business highlights include: fourth

[13] Revenues, expenses, sales, profits, EBT, EBIT, ebitda, depreciation, amortization, SG&A, tax, COGS, working capital, accounts receivable, accounts payable, inventory, allowance, accruals, cash, gains.

quarter revenues of $ _number_, GAAP earnings of $ _number_ per diluted share (...)" (CalAmp Corp., 2007 Q1). It can also focus on the formation of the final number by highlighting earnings components: "The decline in offline sales was primarily due to a loss of sales to a large customer in Q3 _number_" (U.S. Auto Parts Network, 2009 Q1)". In this section, we show that discussions of earnings components are more informative to markets than those of earnings. However, discussions of losses by far exceed them both in informativeness.

Discussions of earnings, components, and losses are all ubiquitous. Ninety-four percent of the earnings announcements studied (8% of all sentences) mention earnings; 100% of announcements (26% of the sentences) mention at least one earnings component; and 82% (4% of the sentences) mention loss. Table 3 shows the words most characteristic of each group of sentences. The most characteristic earnings sentences are the standard headlines and announcements of earnings calls. A wide range of accounting terms characterize earnings components sentences. Loss sentences can talk either about the bottom line or particular aspects of firm performance. In the first case, the loss is mentioned together with net income. In the second case, loss refers to a specific item, such as loss from discontinued operations or impairment.

Figure 6 presents a comparison of the impact of the groups across sentence positions. At the beginning of a document, the components sentences have a larger impact than earnings sentences by around 40% of a standard deviation. The difference becomes smaller as we move toward the end of the document. Around sentence position 100, the earnings overtake the earnings components, but only by a little. At any position, discussions of losses dominate. Their impact can be as much as four times as big as the impact of components sentences.

We also compare the impact of different content groups using regression analysis, estimating the following OLS model:

$$
\begin{aligned}
I_{f,t,s} =& \beta_1 I\{\text{Earnings}\}_{f,t,s} + \beta_2 I\{\text{Loss}\}_{f,t,s} + \beta_3 I\{\text{Components}\}_{f,t,s} + \beta_4 I\{\text{Operations}\}_{f,t,s} + \\
& FE_s + FE_{f,t} + \epsilon,
\end{aligned}
$$

where s is the index of sentence position (from the 1st to the 300th sentence in a document), I is sentence impact, $I\{x\}$ are indicator variables equal to one if a sentence belongs to the content group x, and FE are sentence position and the document (firm-year-quarter) fixed effects. The model is estimated separately for all sentence positions, positions 1 to 100, positions 101 to 200, and positions 201 to 300. Table 4 presents the results of the regression analysis, which confirm the results in Figure 6.

Findings suggest that discussions of earnings components and losses stimulate investors to revise their prior valuations. In contrast, headline discussions of earnings are less important. Perhaps the information contained in them has already been priced. Alternatively, earnings headlines can be boilerplate to the extent of not being useful. If anywhere, investors may find helpful information about the bottom line in the tables at the ends of press releases.

5.2 Sentiment

When presenting financial results, writers can adjust the tone of a document to convey additional information. They can choose to report the numbers neutrally: "Gross margin was $# and operating expenses were $# for the September # quarter (...)" (Lam Research, 2005, Q3). They can sprinkle in some positivity: "Meanwhile, sales for our disk drive analyzer products reached their highest level in more than # years" (LeCroy, 2005, Q3). Or, they can invoke negative sentiment: "Grouping these retirement benefits together, and discussing changes in this volatile net expense is helpful in analyzing the operational performance of the company" (Carpenter Technology Corp, 2005, Q3).

Words associated with a positive or negative tone are widespread in earnings announcements. Using Loughran-McDonald Sentiment Word Lists,[14] which identify sets of positive and negative sentiment, or tone, markers, we categorize 11% of sentences as positive and 13% of sentences as negative. The remainder are classified as neutral. All the documents contain at least one sentence with a sentiment marker.

Table 5 presents the most characteristic words for positive, negative and neutral sentences. As expected, the list largely follows the sentiment dictionary. However, it is notable that the word most characteristic of positive sentences is not itself a sentiment word, but the pronoun "our." Another pronoun, "we," is also close to the top. It is also worth noting that many measures of firm performance tend not to be accompanied by sentiment words. Examples include "cash," "revenues," "sales," and "earnings."

Figure 7 shows the impact of sentiment within different content groups. For all content groups, sentences with tone markers predict higher absolute returns. A positive tone has a smaller impact than a negative tone. The results are the weakest among the earnings sentences, and the strongest among the operations sentences. Table 6 presents the results of regression analysis that confirm the results in Figure 7.

[14] https://sraf.nd.edu/textual-analysis/resources/#LM%20Sentiment%20Word%20Lists

Tone is impactful. It amplifies the potential of disclosure to move the markets. However, the tone of an announcement doesn't move the markets on its own. An especially impactful exposure combines the right content with the right sentiment in the right place. Negative presentation of the details behind a company's bottom line at the beginning of an earnings announcement predicts substantial price revisions.

5.3 Forward-looking statements

Managers use earnings announcements to transmit expectations about the future: "We believe strong momentum built in the third quarter will lead to similar solid revenue growth for the fourth quarter and are therefore revising our full year revenue guidance upward." (Liveperson, 2006, Q3). We show that forward-looking statements are associated with extensive price revisions.

Our algorithm for identifying forward-looking sentences is based on Muslu et al. (2015). A sentence is identified as forward-looking if it contains one or more expressions containing words such as "intend," "expect" and "anticipate," contains the word "will," or directly references a moment in the future, such as "next quarter."

Forward-looking sentences are highly prevalent. Almost all (97%) of the earnings announcements contain forward-looking statements, and 5% of sentences are classified as such.Table 7 shows the words most characteristic of forward-looking sentences. Often forward-looking sentences talk about risk factors and potential impactful events. Some aspects of firms' operations tend not to be discussed in a forward-looking context; "income," "loss," and "taxes," among other words, are characteristic of non-forward-looking sentences.

Forward-looking sentences are associated with large absolute announcement returns. Figure 8 shows that across all content groups, the difference between the impact of forward-looking sentences and the average sentence is massive. The effect of forward-lookingness is especially powerful for discussions of earnings components. Table 7 presents the results of regression analysis that confirm the results in Figure 8.

The results in this section suggest that investors pay attention to forward-looking sentences and revise their prior valuations based on them. Discussions of earnings components at the beginning of a document are especially attention-grabbing.

5.4 Quotes

Earnings announcements tend to give prominence to the quotes from the companies' executives. These often include performance highlights: "According to Maier, CEO, 'Sequentially sales were similar to the June quarter. However, demand increased, resulting in a positive book to bill ratio' "(Linear Technologies, 2005, Q3). We show that the quotes are only informative when they discuss earnings components or firm operations, not bottom-line earnings or losses.

To identify quotes, we use the Stanford CoreNLP quote extraction tool (Manning et al. 2014). We only consider phrases containing no less than ten words to exclude definitions.

Quotes are prevalent, and sentences containing them differ lexically from other sentences. Approximately 75% of the studied documents include at least one quote, and 5% of all sentences are quotes. Table 9 shows the words most characteristic of the sentences containing quotes compared to a subsample of sentences not containing quotes. The quotes are more likely to identify the speaker with the company, as manifested by the use of pronouns "we" and "our." The words "growth," and "strong" are also characteristic of quotes. In contrast, the word "loss" rarely appears in them.

Figure 9 shows the normalized impact of sentences containing quotes by content group. Quotes have a large impact when they mention earnings components and operations. However, there is no additional impact for quotes discussing earnings. Finally, the quotes talking about loss have a smaller impact than average loss sentences in most document positions. Table 10 presents the results of regression analysis that confirm the results in Figure 9.

The direct speech of executives allows the companies to broadcast strategy to the public. The results in this section suggest that quotes are only sometimes informative for the markets. Executives can mention important details behind a quarter's bottom line. But the markets tend to disregard what they have to say about general performance and losses.

5.5 Non-GAAP measures

Earnings announcements are one of the primary vehicles for publishing non-GAAP results. Non-GAAP measures are obtained by excluding certain items from the nearest GAAP measure. For example, the earnings announcement by Maxim Integrated Products in 2005, Q3, notes: "We are showing pro forma (non-GAAP) consolidated statements of income, which are adjusted to reflect the GAAP results to exclude all stock-based compensation expense."

We identify sentences related to non-GAAP reporting by looking for keywords such as "adjusted," "non-GAAP," and "pro forma." as well as the list of common charges and gains. Ninety-one percent of documents contain at least one such non-GAAP sentence, and 12%

of sentences are classified as non-GAAP. Table 11 shows the words most characteristic of sentences mentioning non-GAAP measures. The list largely reflects the nature of the non-GAAP measures as modifying the most similar GAAP measure.

Figure 10 shows the impact of non-GAAP sentences within different content groups. I most cases, non-GAAP sentences are associated with more repricing than an average sentence within the same group. Loss sentences follow a slightly different pattern. Non-GAAP loss sentences at the beginning of a document are associated with more repricing than an average sentence, but around sentence position 50 non-GAAP sentences start having a smaller impact than average. Table 12 presents the results of regression analysis that confirm the results in Figure 10.

The high impact of non-GAAP sentences suggests that they convey information to investors. Since non-GAAP measures provide investors with additional details about firm performance, this result is consistent with our argument in Section 5.1 that argues that how a firm arrives at earnings numbers is more important than the discussion of earnings themselves.

5.6 Comparing different themes

This section shows that all the content groups contribute to the model's prediction by reporting a set of regressions. The left-hand-side variable is the model's prediction. The right-hand side is the total impact of sentences belonging to a given group. Formally, for the four content groups $g \in \{\text{Earnings, Loss, Components, Operations}\}$, we estimate

$$|\hat{R}_{f,t}| = \beta_g I_{f,t,g} + \epsilon$$

where $I = \sum_{s \in g} \mathrm{IG}_{f,t,s}, \mathrm{IG}_s$ is the sentence impact, and g is the set of sentences belonging to a given group.

The R^2 values of these regressions are used to quantify the influence of a given group of sentences on the model prediction. Figure 11 presents the R^2 for different groups compared to the percentage of sentences belonging to a given group. Earnings and loss groups each account for 20% of prediction, while representing only 6% and 4% of sentences, respectively. Thirty-one percent of sentences discuss earnings components and account for 50% of the model's predictions. Finally, 72% of the prediction can be attributed to the 47% of sentences discussing operations.

These results suggest that the importance of earnings announcements cannot be reduced to a single content group. Earnings announcements are important because of the full range of detail they contain. And while discussions of firm operations can explain the model's predictions well, the source of about a quarter of the model's predictive power lies beyond them.

6 Discussion

Quantifying the information content of earnings announcements' language furthers understanding of information transmission in financial markets. Scholars have long known that earnings announcements are associated with large market movements (Beaver 1968). The amount of information in earnings announcements has increased over time (Beaver, McNichols, and Wang 2018). In contrast to that, the informativeness of earnings value has gone down (Shao, Stoumbos, and Zhang 2018). The findings in this paper help reconcile these facts. We show that the language of earnings announcements goes far beyond the summary discussion of earnings and explains a significant portion of the total information content released during the announcement days.

Operational details and discussions earnings components account for a large fraction of the information content of language. This result expands on the work of Francis, Schipper, and Vincent (2002b), who argues that the expansion of nonearnings content drives the historically upward trend in the informational content of earnings announcements.

Sentiment and forward-lookingness amplify information content. Sentences with markers for these characteristics are associated with larger price revisions. The magnitude of amplification depends on the content and the position of a sentence. Our results on sentiment contribute to the large literature starting with (Tetlock 2007). The results on forward-lookingness are in line with the literature that finds significant associations between market reactions and the prevalence of forward-looking sentences in 10-K's (Li 2010; Muslu et al. 2015) and earnings announcements (Bozanic, Roulstone, and Van Buskirk 2018).

We also contribute to the literature about non-GAAP reporting by documenting a link between mentions of non-GAAP measures in the language of an announcement and absolute returns around the time of the announcement. This is important because, while the link of non-GAAP measures with returns is strong (Bhattacharya et al. 2003), managers can emphasize them over the conventional GAAP measures (Bowen, Davis, and Matsumoto 2005) and potentially bias less sophisticated investors (Bhattacharya et al. 2007). Additionally, the short-term reaction to non-GAAP measures might not be reflective of long-term value, as results in Doyle and Lundholm (2003) suggest.

7 Conclusion

A statistical model that links the language of earnings announcements to absolute announcement returns is useful. We show that the information content of language is comparable to that of numbers. Language is inherently nonlinear. Various themes and language properties come together to make it informative. We show that the magnitude of price revision correlates with what is said, where it is said, and whether it was a quote. A statistical model of language provides a unified framework within which to study familiar features such as sentiment and forward-lookingness in the context of everything else that a document says.

The results of this paper are relevant for the literature on earnings announcement information content, as well as aspects of voluntary reporting, such as sentiment, the use of forward-looking statements, and non-GAAP measures. Our findings support the view that earnings announcements are informative to investors because they provide details about firm operations beyond the bottom line. Market reactions are especially strong when details are combined with amplifiers such as sentiment and forward-lookingness.

8 Appendix A: Earnings Announcement Dataset Construction

I download all 8-K's using a script by Bill McDonald.[15] Next, I select 8-K's that are likely to contain an earnings announcement press-release. Following the procedure described in Bozanic, Roulstone, and Van Buskirk (2018), the 8-K's that include a "Results of Operations and Financial Condition" item and are filed within five days of the earnings announcement date from Compustat are selected.

Images and tags are removed from the text. Non-word tokens (such as punctuation marks) are preserved. All numbers are replaced with "number token." All text is set to lower case, but a special token is added before capitalized words. Repeated tokens are replaced with a single one, but a special token is added to indicate repetition. The earnings announcements are tokenized and split into sentences using SpaCy.[16] All tokens containing numbers are changed to "number_token." We include the 30,000 most frequent tokens into the model dictionary, changing all other tokens to "unknown_token."

[15] https://sraf.nd.edu

[16] https://spacy.io

Appendix B

8.1 Convolutional Neural Net with Gated Recurrent Units (CNN-GRU)

CNN-GRU operates by sequentially creating a set of latent factors. These factors reflect word usage, local context, and word order. Minimization of prediction error drives factor extraction. The implementation of CNN used in this paper is written in Keras and is based on the code of Patty Ryan from Microsoft.[17] Table 2 presents an overview of the model, including the number of parameters.

CNN-GRU is a composition of several functions (layers):

$$f(X_{f,t}; w) = \mathrm{L} \circ \mathrm{GRU}_2 \circ \mathrm{GRU}_1 \circ \mathrm{Max}_3 \circ \mathrm{Conv}_3 \circ \mathrm{Max}_2 \circ \mathrm{Conv}_2 \circ \mathrm{Max}_1 \circ \mathrm{Conv}_1 \circ \mathrm{Emb}(X_{f,t}),$$

where $X_{f,t}$ is the document, Emb is the embedding, Conv is the convolution, Max is max pooling, GRU is the gated recurrent unit, and L is linear layer.

8.2 Document

A document is represented as a vector $X = [x_1 \cdots x_n]$, where x_j is the index of the j'th word in the vocabulary, and n is the maximum document length. If a document is shorter than the maximum document length, the extra space is filled with special padding words.

[17] https://github.com/SingingData/StockPerformanceClassification

8.3 Embedding

Embedding (Emb) assigns embedding vectors to individual words. The starting values for word embeddings are obtained using *word2vec* algorithm (Mikolov et al. 2013) applied to the training set. The embedding vectors are then updated during estimation. An important property of embeddings is that words often used interchangeably tend to cluster in the embedding space. Figure 1 shows several such clusters, including names of the months, corporate titles, and words "earnings," "income," and "loss." We denote the embedding of the word x_i as $e(x_i) = e_i \in \mathbb{R}^E$, where E is the embedding size. The document is thus represented as

$$E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} e_{1,1} & e_{1,2} & \dots & e_{1,d(E)} \\ e_{2,1} & e_{2,2} & \dots & e_{2,d(E)} \\ \vdots & & & \\ e_{n,1} & e_{n,2} & \dots & e_{n,d(E)} \end{bmatrix} = \text{Emb}(X_{f,t}),$$

where $d(E)$ is the embedding size (model hyperparameter).

8.4 Convolution

Convolution (Conv(E)) applies a set of parameters, called a filter ($w^c \in \mathbb{R}^{h^c \cdot d(E)}$) and an intercept parameter b^c, to the window of h^c words to produce latent factors:

$$c_i = w^c \cdot e_{i:i+h^c-1} + b^c,$$

We apply $d(c)$ filters to create a matrix of latent factors:

$$C^1 = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{s(C^1)} \end{bmatrix} = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,d(c)} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,d(c)} \\ \vdots & & & \\ c_{s(C^1),1} & c_{s(C^1),2} & \cdots & c_{s(C^1),d(c)} \end{bmatrix} = \mathrm{Conv}^1(E),$$

where $s_(C^1) = n - h^c + 1$.

8.5 Max Pooling

Max pooling ($\mathrm{Max}(C)$) with window size h^m is applied to each column of matrix C. We implement exponential unit non-linearity after max pooling.

$$m_{i,j} = \begin{cases} \max\{c_{k:m,j}\}, & \text{for } \max\{c_{k:m,j}\} \geq 0 \\ e^{\max\{c_{k:m,j}\}} & \text{for } \max\{c_{k:m,j}\} < 0 \end{cases}.$$

where $k = (i-1) \cdot h^m$ and $m = (i-1) \cdot h^m + h^m$. In the end we obtain

$$M^1 = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_{s(M^1)} \end{bmatrix} = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,d(c)} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,d(c)} \\ \vdots & & & \\ m_{s(M^1),1} & m_{s(M^1),2} & \cdots & m_{s(M^1),d(c)} \end{bmatrix} = \mathrm{Max}^1(C^1),$$

where $s(M^1) = (n - h^c + 1)/h^m$.

8.6 Repeating Convolutions and Max Pooling

Convolution and max pooling operations are repeated two more times:

$$C^2 = \text{Conv}^2(M^1),$$

$$M^2 = \text{Max}^2(C^2),$$

$$C^3 = \text{Conv}^3(M^2),$$

$$M^3 = \text{Max}^3(C^3),$$

with the last max pooling layer having a larger filter size than the previous ones.

8.7 Gated Recurrent Unit

A gated recurrent unit (GRU) is a kind of deep neural network commonly used to model long-range dependencies in language. It creates another set of latent factors. These factors take into account the order in which information is presented in a document.

The latent factors are called hidden states in this case. Their computation proceeds in several steps:

$$\begin{aligned}
\text{Update gate: } u_s &= \sigma(W^u h_{s-1} + U^u c_s + b^u), \\
\text{Reset gate: } r_s &= \sigma(W^r h_{s-1} + U^r c_s + b^r), \\
\text{Hidden state update: } \tilde{h}_s &= \tanh(W^h(r_s \otimes h_{s-1}) + U^h c_s + b^h), \\
\text{New hidden state: } h_s &= (1 - u_s) \otimes h_{s-1} + u_s \otimes \tilde{h}_s,
\end{aligned}$$

where $\sigma()$ is sigmoid function. The update gate controls which parts of the previous hidden state are updated or preserved. The sigmoid function ensures that the output is between zero and one. The reset gate that controls which parts of the previous hidden state are used to compute new content. The reset gate selects useful parts of the previous hidden state. We use that and current input to compute the hidden state update. Finally, we compute the new hidden state. The update gate controls both what is kept from the previous hidden state, and what is taken from the hidden state update. Applying $d(h)$ different sets of weights yields

$$H^1 = \begin{bmatrix} h_0 \\ h_1 \\ h_2 \\ \vdots \\ h_{s(M^3)} \end{bmatrix} = \begin{bmatrix} h_{0,1} & h_{0,2} & \dots & h_{0,d(h)} \\ h_{1,1} & h_{1,2} & \dots & h_{1,d(h)} \\ h_{2,1} & h_{2,2} & \dots & h_{2,d(h)} \\ \vdots & & & \\ h_{s(M^3),1} & h_{L,2} & \dots & h_{s(M^3),d(h)} \end{bmatrix} = \mathrm{GRU}^1(M^3).$$

8.8 Repeat GRU

We run the *GRU* again with H^1 as input to obtain the final set of latent factors.

$$H^2 = \mathrm{GRU}^2(H^1).$$

8.9 Linear layer

Finally, we apply another set of weights to obtain the prediction:

$$\hat{y} = \bar{H}^2 w^l + b^l,$$

where $\bar{H}^2$ is vector that consists of stacked columns of the matrix H^2.

8.10 Integrated gradients

The integrated gradients are defined in the following way:

$$\mathrm{IG}_i(x) = (x_i - x_i') \times \int_{\alpha=0}^{1} \frac{\partial f(x' + \alpha \times (x - x'))}{\partial x_i} d\alpha,$$

where x is a vector of word embeddings in order, x' is the vector of baseline inputs, α is the placeholder for a point on the straight line path (in R^n) from the baseline x' to the input x, and f is the deep neural network. The computation of the integrated gradients involves Riemman approximation of the integral (Sundararajan, Taly, and Yan 2017).

The integrated gradients capture the change in the prediction of a deep neural network that occurs when the input changes from the baseline to the actual input. We pick the embedding vector of the padding token as a baseline since it represents the absence of a word.

References

Arras, Leila, Franziska Horn, Grégoire Montavon, Klaus-Robert Müller, and Wojciech Samek. 2017. “"What is relevant in a text document?": An interpretable machine learning approach.” *PLOS ONE* 12 (8): e0181142.

Arras, Leila, Grégoire Montavon, Klaus-Robert Uller, and Wojciech Samek. 2017. “Explaining Recurrent Neural Network Predictions in Sentiment Analysis.” *Proceedings of the*

8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 159–68.

Aumann, Robert J., and Lloyd S. Shapley. 1974. *Values of non-atomic games.*

Ball, Ray, and Philip Brown. 1968. "An empirical evaluation of accounting income numbers." *Journal of Accounting Research* 6 (2): 159–78.

Beaver, William H. 1968. "The information content of annual earnings announcements." *Empirical Research in Accounting: Selected Studies* 6: 67–92.

Beaver, William H., Maureen F. McNichols, and Zach Z. Wang. 2018. "The information content of earnings announcements: new insights from intertemporal and cross-sectional behavior." *Review of Accounting Studies* 23 (1): 95–135.

Bhattacharya, Nilabhra, Ervin L. Black, Theodore E. Christensen, and Chad R. Larson. 2003. "Assessing the relative informativeness and permanence of pro forma earnings and GAAP operating earnings." *Journal of Accounting and Economics* 36: 285–319.

Bhattacharya, Nilabhra, Ervin L. Black, Theodore E. Christensen, and Richard D. Mergenthaler. 2007. "Who trades on pro forma earnings information?" *The Accounting Review* 82 (3): 581–619.

Bilbao-Jayo, Aritz, and Aitor Almeida. 2018. "Political discourse classification in social networks using context sensitive convolutional neural networks." *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*, 76–85.

Bowen, Robert M., Angela K. Davis, and Dawn A. Matsumoto. 2005. "Emphasis on pro forma versus GAAP earnings in quarterly press releases: determinants, SEC intervention, and market reactions." *The Accounting Review* 80 (4): 1011–38.

Bozanic, Zahn, Darren T. Roulstone, and Andrew Van Buskirk. 2018. "Management earnings forecasts and other forward-looking statements." *Journal of Accounting and Economics* 65 (1): 1–20.

Bradshaw, Mark T., and Richard G. Sloan. 2002. "GAAP versus the street: an empirical assessment of two alternative definitions of earnings." *Journal of Accounting Research* 40 (1): 41–66.

Chebonenko, Tatiana, Lifeng Gu, and Dmitriy Muravyev. 2018. "Text sentiment's ability to capture information: evidence from earnings calls." *Working Paper.*

Collins, Daniel W, Edward L Maydew, Ira S Weiss, and Ira S Weiss. 1997. "Changes in the value-relevance of earnings and book values over the past forty years." *Journal of Accounting and Economics* 24: 67.

Collobert, Ronan, Jason Weston, Leon Bottou, Koray Kavukcuoglu, and Pavel Kuksa. 2011. "Natural language processing (almost) from scratch." *Journal of Machine Learning Research* 12: 2493–2537.

Davis, Angela K., Jeremy M. Piger, and Lisa M. Sedor. 2012. "Beyond the numbers: measuring the information content of earnings press release language." *Contemporary Accounting Research* 29 (3): 845–68.

Doyle, Jeffrey T, and Russell J Lundholm. 2003. "The predictive value of expenses excluded from pro forma earnings." *Review of Accounting Studies* 8: 145–74.

Fairfield, Patricia M, Richard J Sweeney, and Teri Lombardi Yohn. 1996. "Accounting classification and the predictive content of earnings." *The Accounting Review* 71 (3): 337–55.

Francis, Jennifer, Katherine Schipper, and Linda Vincent. 2002a. "Earnings announcements and competing information." *Journal of Accounting and Economics* 33 (3): 313–42.

———. 2002b. "Expanded disclosures and the increased usefulness of earnings announcements." *The Accounting Review* 77 (3): 515–46.

Hoberg, Gerard, and Craig Lewis. 2017. "Do fraudulent firms produce abnormal disclosure?" *Journal of Corporate Finance* 43: 58–85.

Kim, Oliver, and Robert E Verrecchia. 1991. "Trading volume and price reactions to public announcements." *Journal of Accounting Research* 29 (2): 302–21.

———. 1994. "Market liquidity and volume around earnings announcements." *Journal of Accounting and Economics* 17: 41–67.

Kim, Yoon. 2014. "Convolutional neural networks for sentence classification." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–51.

Kogan, Shimon, Dimitry Levin, Bryan R Routledge, Jacob S Sagi, and Noah Smith. 2009. "Predicting risk from financial reports with regression." *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL*, 272–80.

Li, Feng. 2010. "The information content of forward-looking statements in corporate filings—a naïve bayesian machine learning approach." *Journal of Accounting Research* 48 (5): 1049–1102.

Lipe, Robert C. 1986. "The information contained in the components of earnings." *Journal of Accounting Research* 24: 37–64.

Loughran, Tim, and Bill McDonald. 2014. "Measuring readability in financial disclosures." *The Journal of Finance* 69 (4): 1643–71.

Lundberg, Scott M, Paul G Allen, and Su-In Lee. 2017. "A unified approach to interpreting model predictions." *Advances in Neural Information Processing Systems 30*, 4765—–4774.

Manning, Christopher D, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J Bethard, and David Mcclosky. 2014. "The Stanford CoreNLP Natural Language Processing Toolkit." *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations.*

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Distributed representations of words and phrases and their compositionality." *NIPS'13 Proceedings of the 26th International Conference on Neural Information Processing Systems* 2: 3111–9.

Muslu, Volkan, Suresh Radhakrishnan, K R Subramanyam, and Dongkuk Lim. 2015. "Forward-looking MD&A disclosures and the information environment." *Management Science* 61 (5).

Nystrom, Nicholas A., Michael J. Levine, Ralph Z. Roskies, and J. Ray Scott. 2015. "Bridges." *Proceedings of the 2015 XSEDE Conference on Scientific Advancements Enabled by Enhanced Cyberinfrastructure - XSEDE '15*, 1–8.

Routledge, Bryan R, Stefano Sacchetto, and Noah A Smith. 2017. "Predicting Merger

Targets and Acquirers from Text." *Working Paper*.

Shao, Shuai, Robert Stoumbos, and X Frank Zhang. 2018. "The power of accounting information in explaining stock returns." *Working Paper*.

Shrikumar, Avanti, Peyton Greenside, and Anshul Kundaje. 2017. "Learning Important Features Through Propagating Activation Differences." *Proceedings of Machine Learning Research* 70: 3145–53. {\%}5C.

Sundararajan, Mukund, Ankur Taly, and Qiqi Yan. 2017. "Axiomatic Attribution for Deep Networks." *ICML'17 Proceedings of the 34th International Conference on Machine Learning* 70: 3319–28.

Tetlock, Paul C. 2007. "Giving content to investor sentiment: the role of media in the stock market." *The Journal of Finance* 62 (3): 1139–68.

———. 2010. "Does public financial news resolve asymmetric information?" *Review of Financial Studies* 23 (9): 3520–57.

Zou, Hui, and Trevor Hastie. 2005. "Regularization and variable selection via the elastic net." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67 (2): 301–20.

Figures

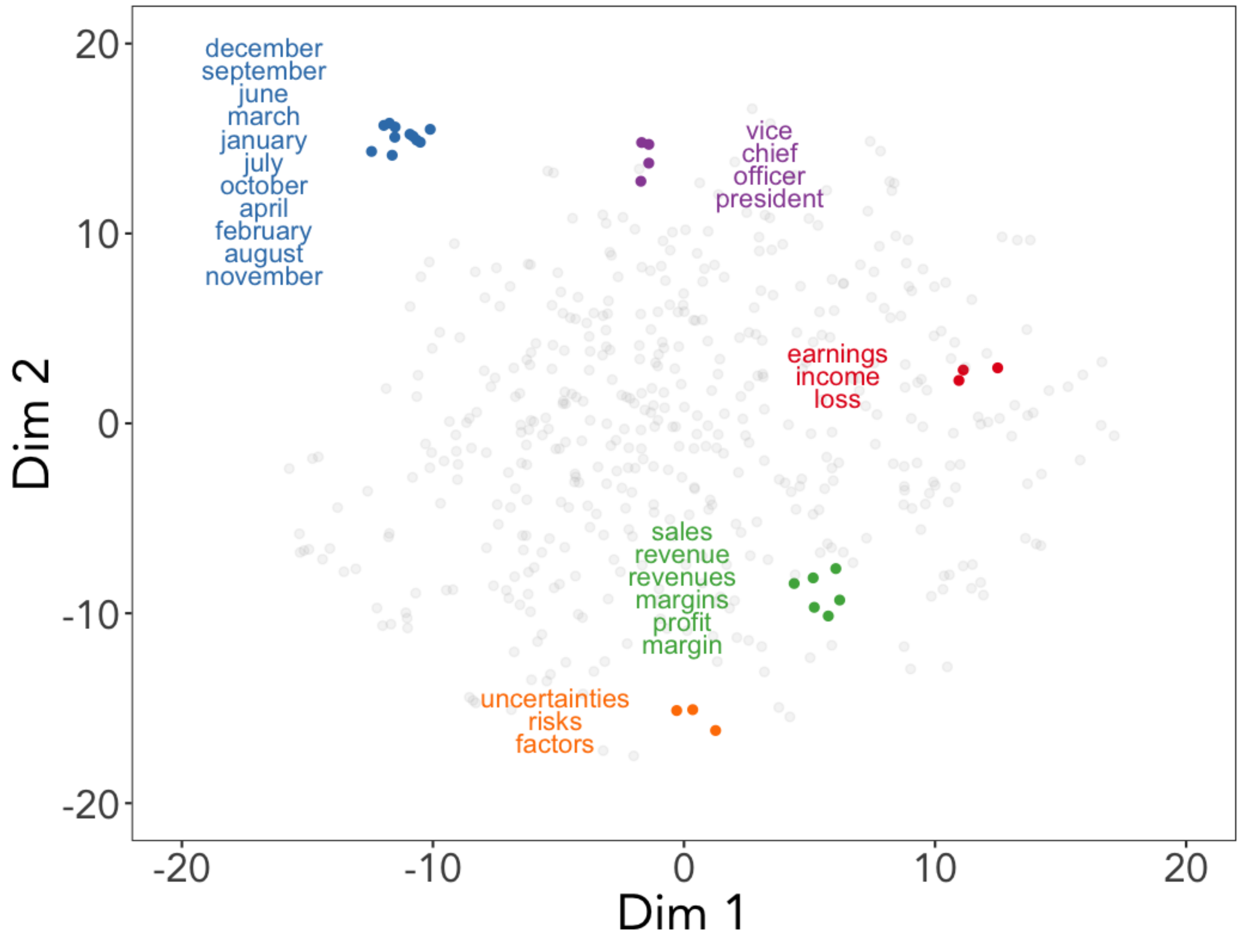

Figure 1: Embeddings example. Words that tend to be used interchangeably cluster together in the embedding space. We perform Principal Component Analysis on 200-dimensional word embeddings and visualize the first two principal components. The clusters are picked manually. The embeddings are intially obtained using the word2vec algorithm (Mikolov, 2013) on earnings announcements. They are further updated during the estimation of CNN-GRU.

mike brooks , chairman and chief executive officer , commented , “ while our third quarter sales were in - line with our projections our earnings were lower than we anticipated due to a combination of factors . xbos xfld 1 during the quarter we experienced a num_tok basis point decline in gross margin as a result of significant pricing pressure and an increase in product costs . xbos xfld 1 in addition , we made the strategic decision to increase our retail operating expenses in order to capitalize on the near - term prospects created by the bankruptcy of a key competitor . xbos xfld 1 we are committed to driving further top - line gains in both our wholesale and retail divisions while at the same time evaluating all our opportunities in an effort to return to more normalized margins beginning in num_tok . (...)."

Figure 2: Integrated gradients example. The model associates the words in red with high absolute announcements returns, and the words in blue – with low absolute announcement returns. The color density corresponds to the magnitude of the association. The weights are computed using the integrated gradients method (Sundararajan et al., 2017), which is one way to attribute model predictions to input words.

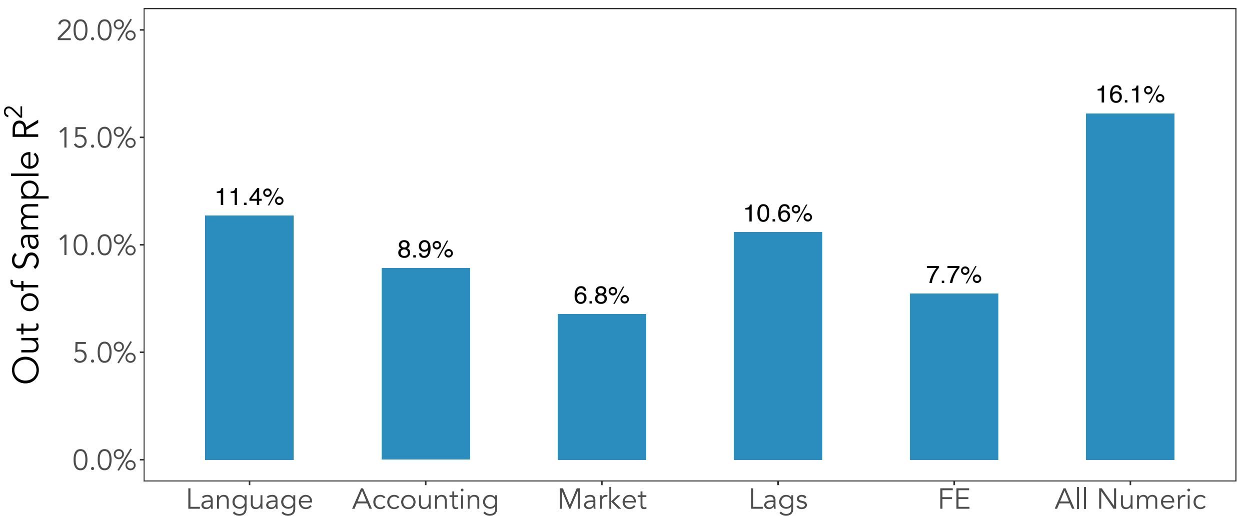

Figure 3: Out-of-sample R^2 of regressions using different subsets of variables. The "language" bar corresponds to the performance of the deep neural network. Other bars correspond to the performance of the regularized linear regression (elastic net) model trained using only numerical variables. The dependent variable is absolute announcement returns. Accounting variables include various accounting measures from Compustat. Market-based variables include stock return and volatility for the last quarter. Lags refer to four lags of the dependent variable. The fixed effects are industry and year-quarter. For every accounting variable the model includes level, the difference between a current quarter and the previous one, the difference between a current quarter and the same quarter of the previous year, and the absolute value of both differences.

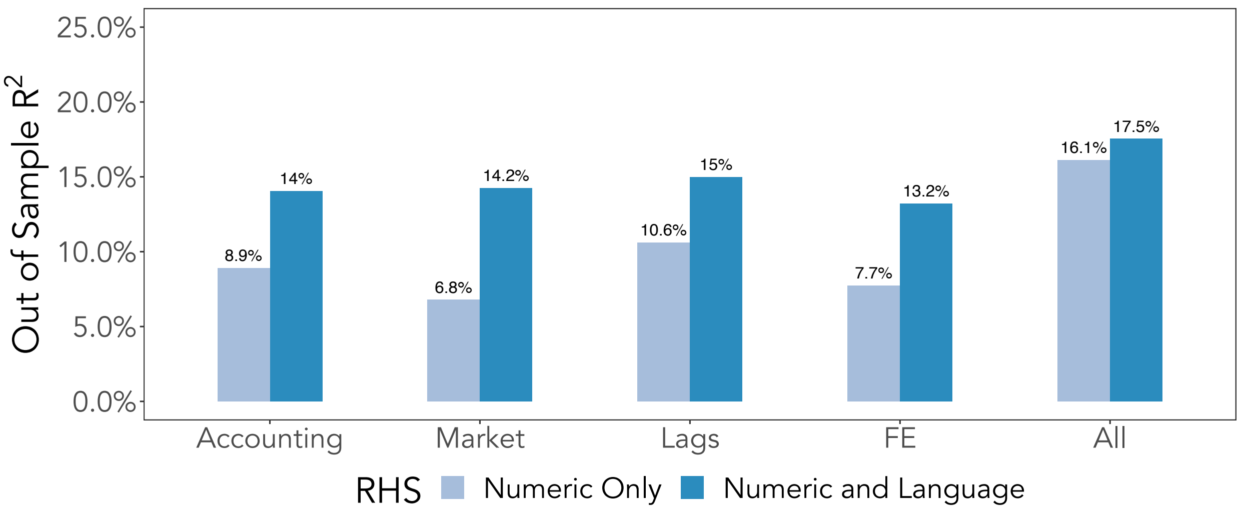

Figure 4: Incremental out-of-sample R^2. The values in the "Numeric only" bars correspond to the performance of the regularized linear regression (elastic net) model trained using only numerical variables. The values in "Numeric and language" bars correspond to the performance of the elastic net model trained using both numerical variables and the predictions of the deep neural network. The dependent variable is absolute announcement returns. Accounting variables include various accounting measures from Compustat. Market-based variables include stock return and volatility for the last quarter. Lags refer to four lags of the dependent variable. The fixed effects are industry and year-quarter. For every accounting variable the model includes level, the difference between a current quarter and the previous one, the difference between a current quarter and the same quarter of the previous year, and the absolute value of both differences.

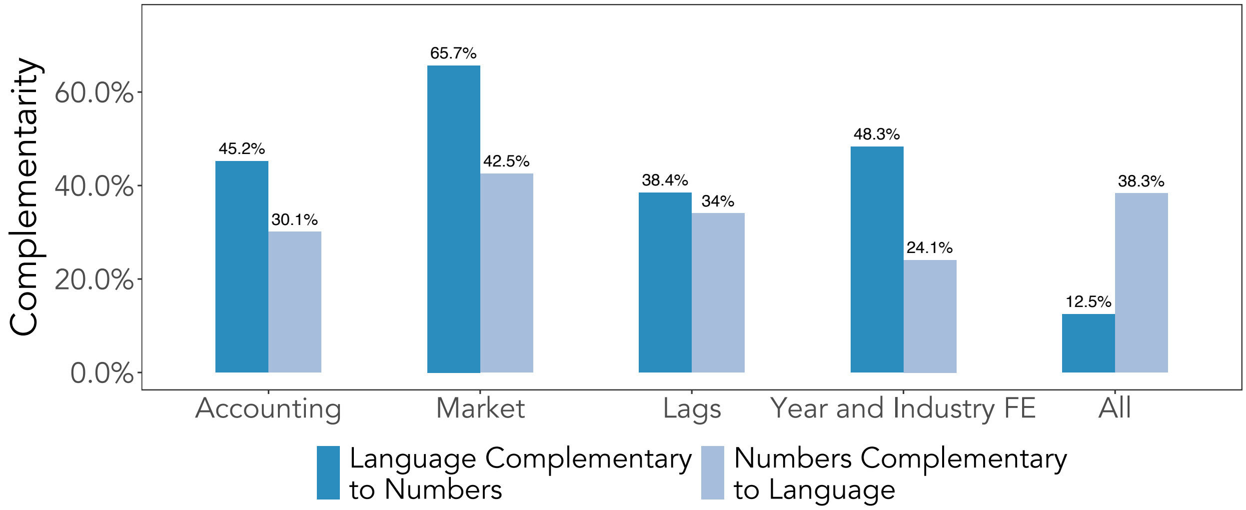

Figure 5: Complementarity. Complementarity of a set of variables A relative to the set of variables B measures the percentage of information contained in A that is new relative to B. It is defined as $Comp(A, B) = (R^2(A \cup B) - R^2(B))/R^2(A)$. "Language complementary to numbers" bars show the complementarity of language relative to numeric variables. "Numbers complementary to language" show the complementarity of numbers relative to language. The dependent variable is absolute announcement returns. Accounting variables include various accounting measures from Compustat. Market-based variables include stock return and volatility for the last quarter. Lags refer to four lags of the dependent variable. The fixed effects are industry and year-quarter. For every accounting variable the model includes level, the difference between a current quarter and the previous one, the difference between a current quarter and the same quarter of the previous year, and the absolute value of both differences.

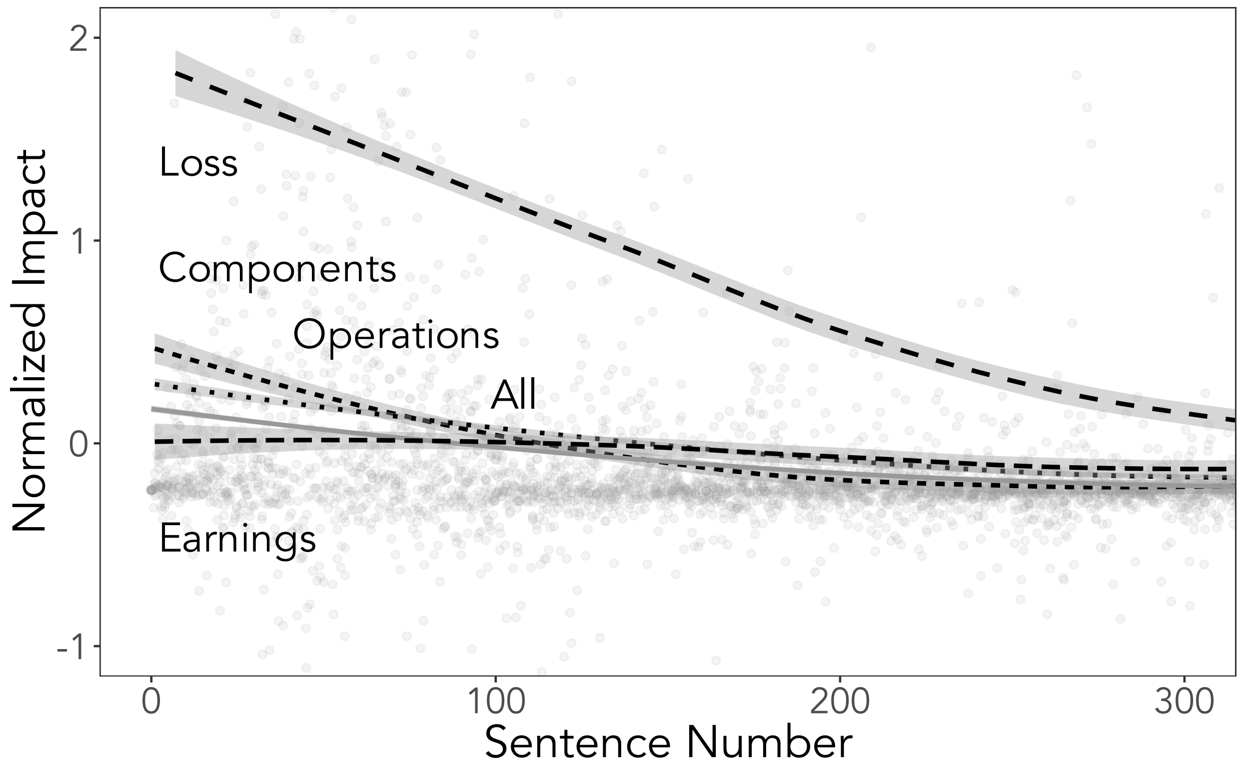

Figure 6: Impact by content group and sentence position. Sentence impact is calculated as a sum of integrated gradient weights for all words within a given sentence. The dots represent individual sentences. The lines represent average impact sentences by group across sentence positions. Smoothing across sentence positions is done using LOESS regression. Content groups are identified using keyword matching. The *earnings* group includes sentences mentioning earnings or net income. The *loss* group includes sentences mentioning loss. The *components* group includes sentences mentioning earnings components such as sales, revenues, and expenses. The *operations* group includes the three previous groups plus sentences mentioning various aspects of firm operations, following Muslu et al. (2015). *All* refers to the average impact of all sentences in a given position.

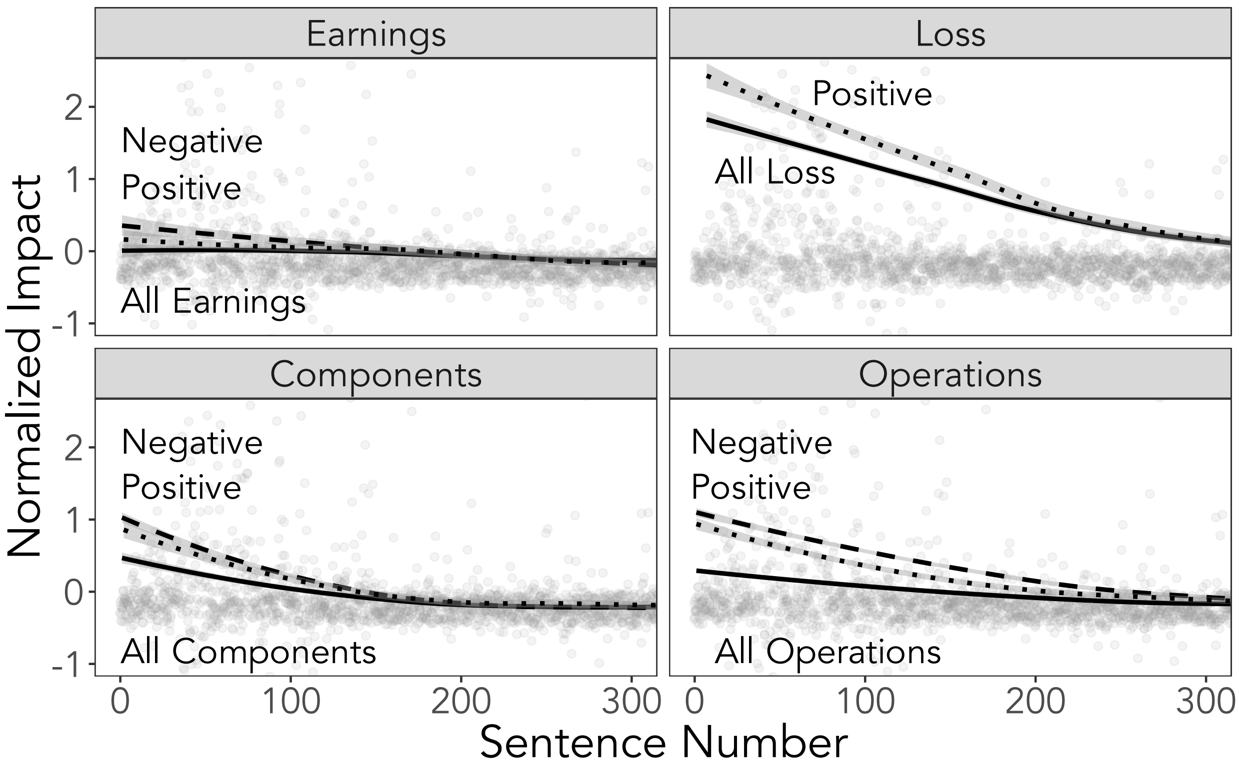

Figure 7: Impact of sentiment by content group and sentence position. Sentence impact is calculated as a sum of integrated gradient weights for all words within a given sentence. The dots represent individual sentences. The lines represent the average impacts of positive, negative and all sentences by group across sentence positions. Smoothing across sentence positions is done using LOESS regression. Positive and negative sentiment sentences are identified using keyword matching with Loughran and McDonald's word list. Content groups are identified using keyword matching. The *earnings* group includes sentences mentioning earnings or net income. The *loss* group includes sentences mentioning loss. The *components* group includes sentences mentioning earnings components such as sales, revenues, and expenses. The *operations* group includes the three previous groups plus sentences mentioning various aspects of firm operations, following Muslu et al. (2015). *All* refers to the average impact of all sentences in a given position.

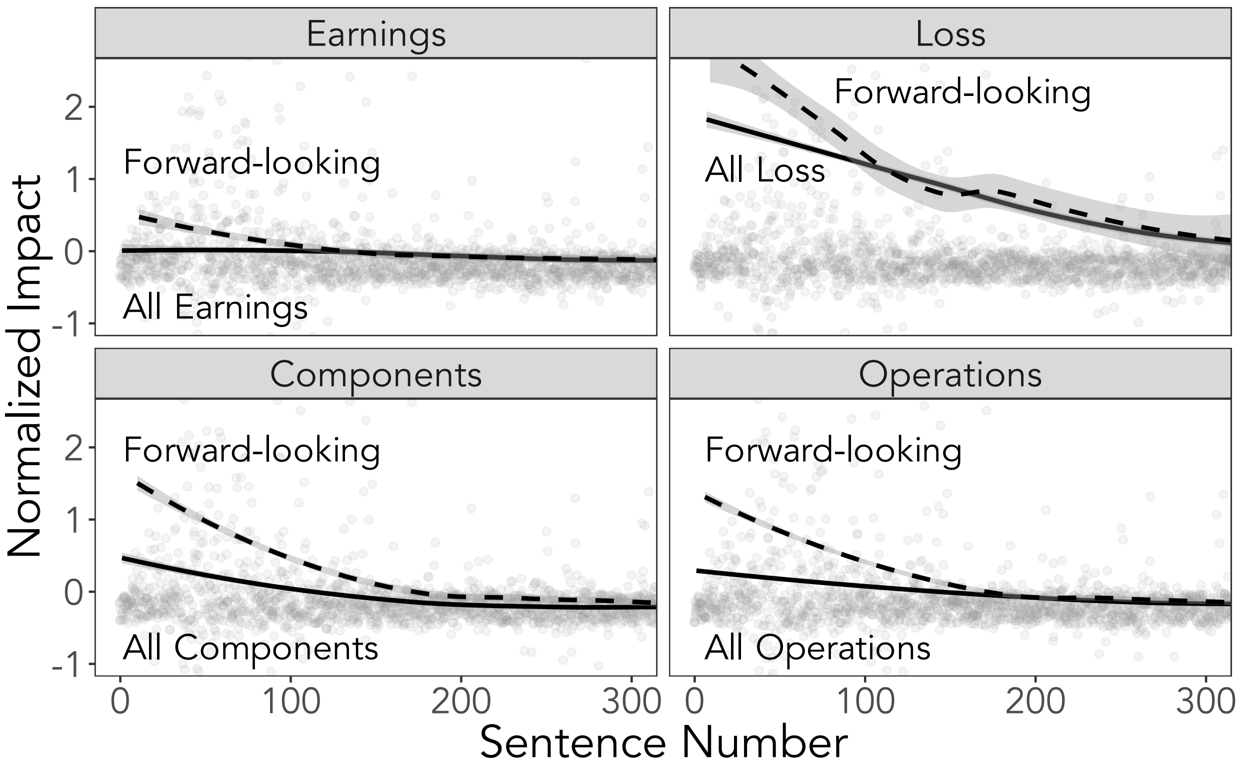

Figure 8: Impact of forward-lookingness by content group and sentence position. Sentence impact is calculated as a sum of integrated gradient weights for all words within a given sentence. The dots represent individual sentences. The lines represent the average impact of forward-looking and all sentences by group across sentence positions. Smoothing across sentence positions is done using LOESS regression. Forward-looking sentences are identified using keyword matching with the word list from Muslu et al. (2015). Content groups are identified using keyword matching. The *earnings* group includes sentences mentioning earnings or net income. The *loss* group includes sentences mentioning loss. The *components* group includes sentences mentioning earnings components such as sales, revenues, and expenses. The *operations* group includes the three previous groups plus sentences mentioning various aspects of firm operations, following Muslu et al. (2015). *All* refers to the average impact of all sentences in a given position.

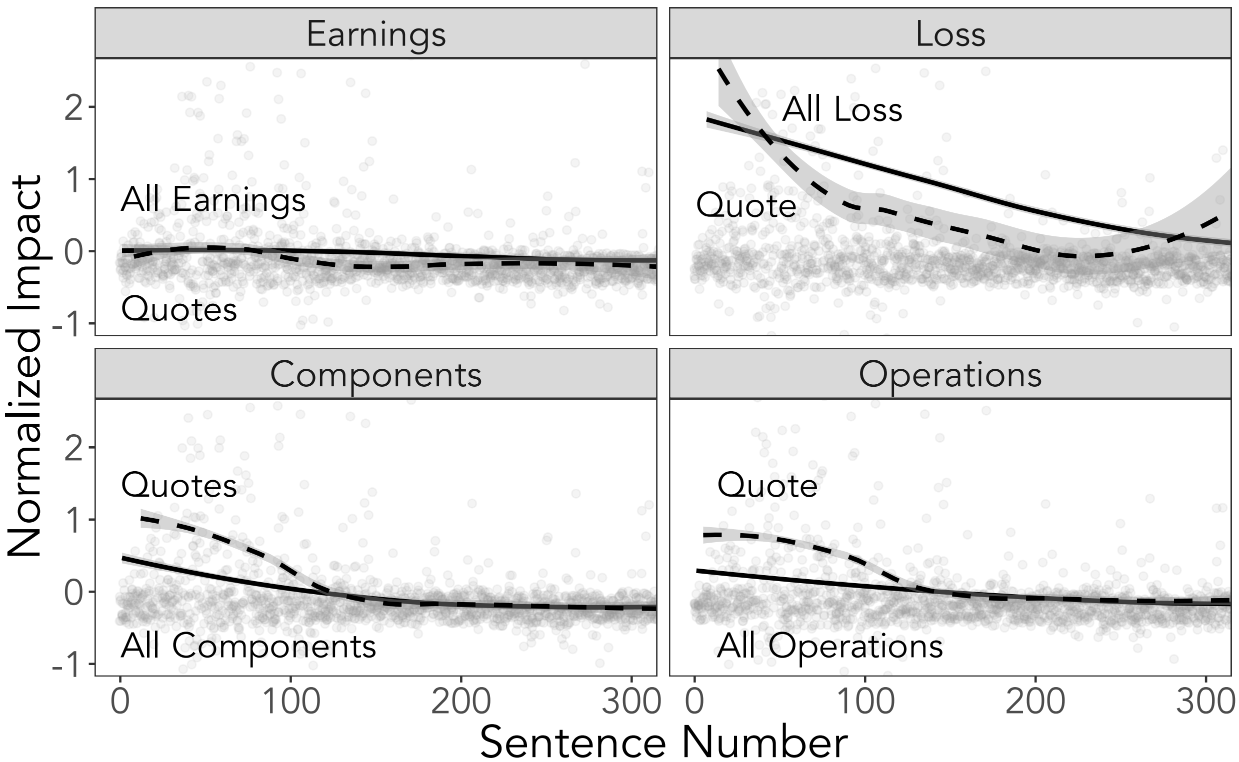

Figure 9: Impact of quotes by content group and sentence position. Sentence impact is calculated as a sum of integrated gradient weights for all words within a given sentence. The dots represent individual sentences. The lines represent the average impact of sentences containing quotes and all sentences by group across sentence positions. Smoothing across sentence positions is done using LOESS regression. Sentences containing quotes are identified using Stanford NLP Toolbox (Manning et al., 2015). Content groups are identified using keyword matching. The *earnings* group includes sentences mentioning earnings or net income. The *loss* group includes sentences mentioning loss. The *components* group includes sentences mentioning earnings components such as sales, revenues, and expenses. The *operations* group includes the three previous groups plus sentences mentioning various aspects of firm operations, following Muslu et al. (2015). *All* refers to the average impact of all sentences in a given position.

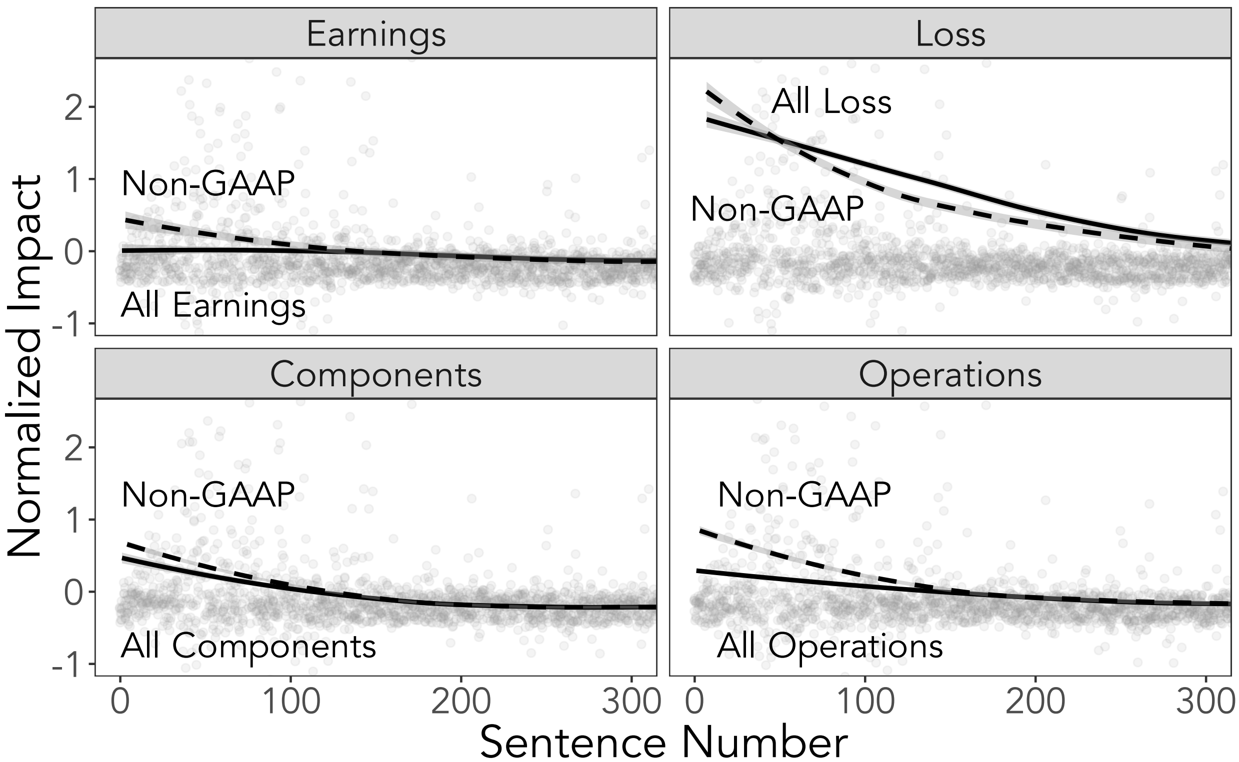

Figure 10: Impact of non-GAAP sentences by content group and sentence position. Sentence impact is calculated as a sum of integrated gradient weights for all words within a given sentence. The dots represent individual sentences. The lines represent the average impact of non-GAAP and all sentences by group across sentence positions. Smoothing across sentence positions is done using LOESS regression. Non-GAAP sentences are identified using keyword matching. Content groups are identified using keyword matching. The *earnings* group includes sentences mentioning earnings or net income. The *loss* group includes sentences mentioning loss. The *components* group includes sentences mentioning earnings components such as sales, revenues, and expenses. The *operations* group includes the three previous groups plus sentences mentioning various aspects of firm operations, following Muslu et al. (2015). *All* refers to the average impact of all sentences in a given position.

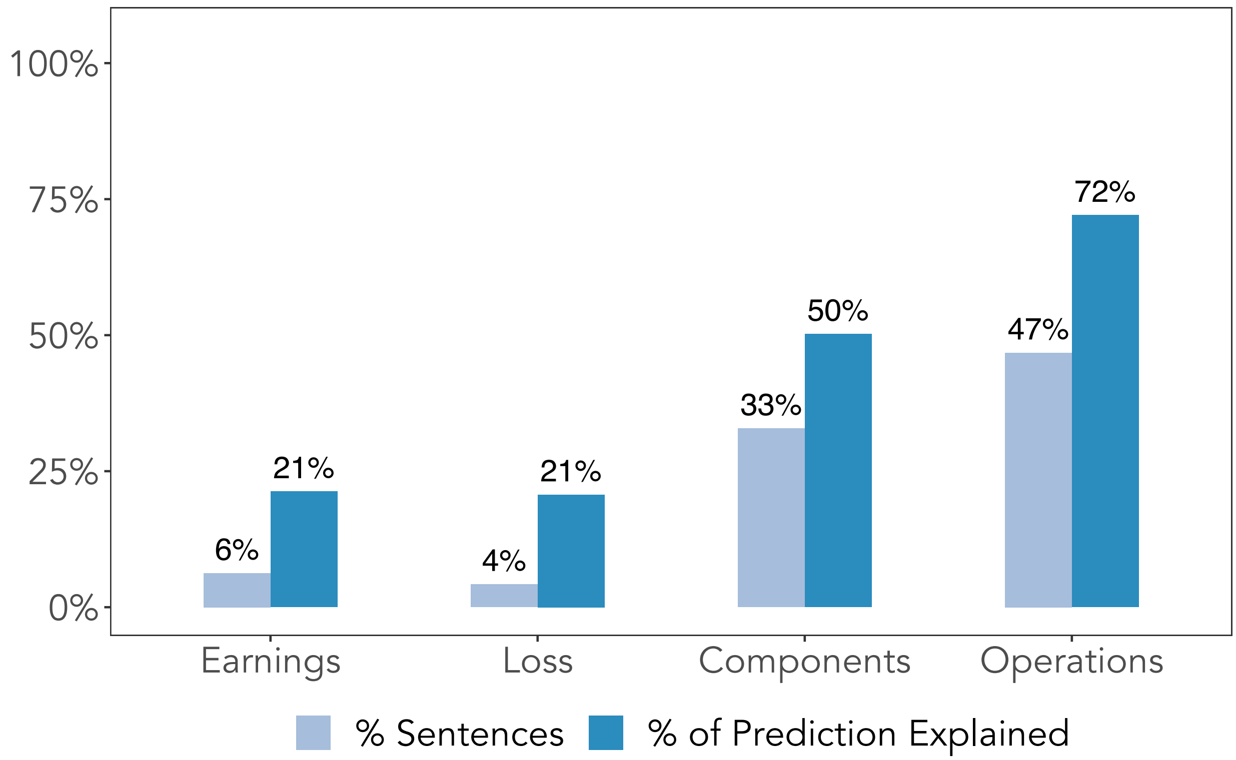

Figure 11: Theme prevalence and the percentage of the prediction explained. The light blue bars correspond to the percentage of sentences identified as belonging to a given group. The dark blue bars represent the R^2 of regressions with model prediction on the left-hand side and total impact of a given content group on the right-hand side. The sentence group impact is the sum of integrated gradient weights of all words in the sentences belonging to the group.

Tables

Table 1: Bootstrapped out-of-sample R^2 for different models. The values in the "No Language" column correspond to the performance of the regularized linear regression (elastic net) model trained using only numerical variables. The values in the "With Language" column correspond to the performance of the elastic net model trained using both numerical variables and the predictions of the deep neural network. The dependent variable is absolute announcement returns. Accounting variables include various accounting measures from Compustat. Market-based variables include stock return and volatility for the last quarter. Lags refer to four lags of the dependent variable. The fixed effects are industry and year-quarter. For every accounting variable the model includes level, the difference between a current quarter and the previous one, the difference between a current quarter and the same quarter of the previous year, and the absolute value of both differences. Standard errors are computed using bootstrap (10,000 iterations) and are in parentheses. The asterisks represent the significance levels as follows: 1% (**), 5% (**) and 10% (*).

Variables	No Language	With Language	Difference
Language		11.51%	
		(0.44%)	
Acc	8.97%	14.13%	5.16%***
	(0.45%)	(0.49%)	(0.34%)
Market	6.79%	14.31%	7.52%***
	(0.41%)	(0.49%)	(0.39%)
Lags	10.68%	15.06%	4.39%***
	(0.49%)	(0.51%)	(0.32%)
Fixed Effects	7.69%	13.25%	5.56%***
	(0.41%)	(0.47%)	(0.35%)
All	16.18%	17.60%	1.42%***
	(0.56%)	(0.54%)	(0.22%)

Table 2: Dimensions of CNN-GRU. This tables presents the operations (layers) of CNN-GRU. For each layer, the associated parameters and hyperparameters are listed. Parameters are estimated, while hyperparameters are picked using the validation set.

Layer	Notation	Output dim	Num of par.	Par. Notation	Hyperparameters
Input	X	12,000	0		
Embedding	Emb	12,000 × 200	6,000,400	$e_{i,j}$	d(E) = 200
Convolution	Conv^1	11996 × 128	128,128	w^c, b^c	$h^c = 5, d(C) = 128$
Max Pooling	Max^1	2,399 × 128	0		$h^m = 5$
Convolution	Conv^2	2,395 × 128	82,048	w^c, b^c	$h^c = 5, d(C) = 128$
Max Pooling	Max^2	479 × 128	0		$h^m = 5$
Convolution	Conv^3	475 × 128	82,048	w^c, b^c	$h^c = 5, d(C) = 128$
Max Pooling	Max^2	13 × 128	0		$h^m = 35$
GRU (Bidir)	GRU^1	13 × 256	198,144	$\{W, U, B\}^{\{u,r,h\}}$	128
GRU (Bidir)	GRU^2	13 × 128	123,648	$\{W, U, B\}^{\{u,r,h\}}$	64
Linear	L	1	1665	w^l, b^l	

Total parameters: 6,616,081
Parameters excluding embedding: 615,681

Table 3: Most characteristic words (tokens) for earnings, earnings componenents, and loss sentences. The top 25 words (tokens) are displayed. The words are ranked according to their log-odds ratio with a Dirichlet prior (Monroe, Colaresi, and Quinn, 2008). The earnings, earnings components, and loss sentences are identified using keyword matching.

Earnings

Word	L/O
earnings	227
share	151
per	145
diluted	144
income	94
eps	92
release	78
net	76
gaap	73
call	60
adjusted	57
conference	52
common	52
basic	47
press	41
exhibit	39
shares	37
guidance	36
weighted	35
announcing	34
today	33
dated	31
or	31
a.m	31
reported	30

Components

Word	L/O
sales	99
cash	97
revenue	82
the	59
revenues	59
current	56
equivalents	56
assets	53
expenses	50
and	50
cost	48
increased	44
loans	43
percent	43
total	42
increase	41
term	41
due	40
primarily	39
loan	39
by	39
payable	37
accounts	34
higher	33
flow	33

Loss

Word	L/O
loss	369
num_tok	108
income	90
net	77
operations	71
discontinued	60
attributable	56
basic	53
comprehensive	53
share	51
continuing	49
per	45
common	45
gain	43
shares	36
impairment	35
accumulated	33
or	32
before	30
diluted	29
noncontrolling	28
benefit	25
from	25
stockholders	23
taxes	23

Table 4: Normalized impact of sentences belonging to earnings and components content groups by sentence position. The analysis is performed at the sentence level. The left-hand variable is sentence impact (absolute announcement returns attributed to the sentence by the deep neural net). The impact is normalized by subtracting the mean and dividing by the standard deviation. The right-hand variables are indicators that are equal to one if a sentence is categorized as an earnings sentence, components, or loss sentence. The model includes document and sentence position fixed effects. Standard errors are clustered at the sentence position level. Standard errors are in parentheses. The asterisks represent the significance levels as follows: 1% (**), 5% (**) and 10% (*).

	Normalized Impact			
	All Sentences	Sent 1 to 100	Sent 101 to 200	Sent 201 to 300
Earnings	0.012	-0.088^{***}	0.160^{***}	0.107^{***}
	(0.013)	(0.017)	(0.008)	(0.005)
Loss	1.027^{***}	1.249^{***}	1.054^{***}	0.510^{***}
	(0.016)	(0.016)	(0.018)	(0.019)
Components	0.166^{***}	0.264^{***}	0.082^{***}	-0.010^{***}
	(0.012)	(0.021)	(0.005)	(0.004)
Operations	0.066^{***}	0.119^{***}	0.007^{***}	0.028^{***}
	(0.010)	(0.017)	(0.002)	(0.002)
FE	Yes	Yes	Yes	Yes
Observations	2,729,526	1,539,447	877,454	246,080
R^2	0.102	0.096	0.145	0.113
Adjusted R^2	0.096	0.087	0.132	0.096

Table 5: Most characteristic words (tokens) for negative, positive, and neutral sentiment sentences. The top 25 words (tokens) are displayed. The words are ranked according to their log-odds ratio with a Dirichlet prior (Monroe, Colaresi, and Quinn, 2008). The sentiment sentences are identified using Loughran and McDonald's word list.

Positive

Word	L/O
our	102
gain	100
benefit	93
gains	83
effective	79
strong	78
on	63
improved	62
we	59
tax	56
ability	56
improvement	55
and	53
the	49
rate	47
profitability	46
growth	44
positive	43
sale	42
favorable	41
improvements	40
efficiency	39
opportunities	38
new	38
pleased	38

Negative

Word	L/O
loss	237
losses	102
restructuring	85
discontinued	83
impairment	75
operations	74
loan	72
allowance	64
income	60
charges	56
decline	55
loans	54
declined	49
net	47
continuing	45
litigation	44
provision	43
charge	43
or	41
nonperforming	40
comprehensive	38
risks	37
attributable	37
accumulated	37
related	36

Neutral

Word	L/O
num_tok	199
cash	119
quarter	82
revenue	79
compared	73
revenues	73
equivalents	72
sales	66
earnings	60
ended	57
months	56
ebitda	55
period	54
current	53
same	52
increased	52
total	51
year	50
were	50
flow	49
first	47
payable	45
adjusted	45
release	45
fiscal	41

Table 6: Normalized impact of sentences with sentiment markers by content group and sentence position. The analysis is performed at the sentence level. The panels correspond to the subsets of sentences belonging to the corresponding content groups. The left-hand variable is sentence impact (absolute announcement returns attributed to the sentence by the deep neural net). The impact is normalized by subtracting the mean and dividing by the standard deviation. The right-hand variables are indicators that are equal to one if a sentence contains positive or negative tone markers from the Loughran and McDonald's word list. The models include document and sentence position fixed effects. Standard errors are clustered at the sentence position level. Standard errors are in parentheses. The asterisks represent the significance levels as follows: 1% (**), 5% (**) and 10% (*).

	Normalized Impact			
	All Sentences	Sent 1 to 100	Sent 101 to 200	Sent 201 to 300
Panel A: Earnings				
Positive	−0.033**	−0.077***	0.025	0.014
	(0.016)	(0.024)	(0.015)	(0.020)
Negative	0.147***	0.224***	0.072***	−0.042**
	(0.013)	(0.021)	(0.015)	(0.018)
Observations	172,042	97,947	56,876	14,564
R^2	0.185	0.256	0.314	0.330
Panel B: Loss				
Positive	0.372***	0.595***	0.295***	0.067
	(0.025)	(0.049)	(0.030)	(0.043)
Observations	115,126	44,201	55,712	13,041
R^2	0.198	0.307	0.221	0.280
Panel C: Components				
Positive	0.366***	0.417***	0.343***	0.163***
	(0.009)	(0.013)	(0.014)	(0.017)
Negative	0.530***	0.620***	0.516***	0.187***
	(0.010)	(0.013)	(0.011)	(0.013)
Observations	391,269	210,671	141,578	32,929
R^2	0.180	0.205	0.208	0.206
Panel D: Operations				
Positive	0.281***	0.394***	0.143***	0.069***
	(0.010)	(0.010)	(0.008)	(0.007)
Negative	0.562***	0.693***	0.536***	0.204***
	(0.010)	(0.010)	(0.010)	(0.009)
Observations	1,283,806	695,960	450,050	113,835
R^2	0.114	0.131	0.111	0.083

Table 7: Most characteristic words (tokens) for forward-looking and non-forward-looking sentences. The top 25 words (tokens) are displayed. The words are ranked according to their log-odds ratio with a Dirichlet prior (Monroe, Colaresi, and Quinn, 2008). The forward looking sentences are identified using keywords from Muslu et al. (2015).

Forward-Looking

Word	L/O
future	135
will	121
our	96
be	88
expected	84
to	71
ability	69
expect	65
that	65
the	64
expects	59
risks	59
we	58
looking	53
forward	52
including	50
statements	50
and	49
range	49
or	48
are	46
may	44
could	44
company	43
any	43

Non-Forward-Looking

Word	L/O
num_tok	199
net	53
income	52
loss	31
total	31
ended	28
was	27
per	26
months	26
cash	25
compared	24
taxes	24
share	24
quarter	23
expense	23
assets	22
liabilities	22
diluted	21
interest	21
operating	19
common	19
amortization	19
period	19
were	19
equity	19

Table 8: Normalized impact of forward-looking sentences by content group and sentence position. The analysis is performed at the sentence level. The panels correspond to the subsets of sentences belonging to the corresponding content groups. The left-hand variable is sentence impact (absolute announcement returns attributed to the sentence by the deep neural net). The impact is normalized by subtracting the mean and dividing by the standard deviation. The right-hand variable is an indicator that is equal to one if the sentence contains forward-lookingness markers from the Muslu et al. (2015) word list. The models include document and sentence position fixed effects. Standard errors are clustered at the sentence position level. Standard errors are in parentheses. The asterisks represent the significance levels as follows: 1% (**), 5% (**) and 10% (*).

	Normalized Impact			
	All Sentences	Sent 1 to 100	Sent 101 to 200	Sent 201 to 300
Panel A: Earnings				
Forward-looking	0.102***	0.107***	0.080***	0.065**
	(0.016)	(0.026)	(0.021)	(0.030)
Observations	172,042	97,947	56,876	14,564
R^2	0.184	0.255	0.314	0.329
Panel B: Loss				
Forward-looking	0.369***	0.560***	0.032	−0.274**
	(0.053)	(0.087)	(0.059)	(0.133)
Observations	115,126	44,201	55,712	13,041
R^2	0.195	0.302	0.218	0.280
Panel C: Components				
Forward-looking	0.632***	0.801***	0.350***	0.158***
	(0.027)	(0.038)	(0.020)	(0.032)
Observations	391,269	210,671	141,578	32,929
R^2	0.149	0.182	0.152	0.182
Panel D: Operations				
Forward-looking	0.418***	0.594***	0.153***	0.044***
	(0.022)	(0.028)	(0.010)	(0.010)
Observations	1,283,806	695,960	450,050	113,835
R^2	0.080	0.099	0.058	0.061

Table 9: Most characteristic words (tokens) for sentences containing and not containing quotes. The top 25 words (tokens) are displayed. The words are ranked according to their log-odds ratio with a Dirichlet prior (Monroe, Colaresi, and Quinn, 2008). The sentences containing quotes are identified using Stanford NLP Toolbox (Manning et al., 2014).

Quote

Word	L/O
we	73
our	64
said	44
officer	34
chief	34
pleased	33
strong	32
continue	31
president	30
growth	29
executive	29
continued	27
expect	26
mr	26
commented	26
chairman	25
ceo	23
while	23
solid	22
have	22
remain	19
will	19
very	19
stated	19
are	19

No Quote

Word	L/O
num_tok	75
net	20
income	19
loss	13
other	13
assets	11
total	11
gaap	10
liabilities	10
tax	10
or	10
ended	10
taxes	10
expense	9
per	9
amortization	9
share	9
diluted	9
non	9
stock	9
common	9
interest	8
cash	8
activities	8
months	8

Table 10: Normalized impact of quotes by content group and sentence position. The analysis is performed at the sentence level. The panels correspond to the subsets of sentences belonging to the corresponding content groups. The left-hand variable is sentence impact (absolute announcement returns attributed to the sentence by the deep neural net). The impact is normalized by subtracting the mean and dividing by the standard deviation. The right-hand variable is an indicator that is equal to one if the sentence is classified as containing a quote by Stanford NLP Toolbox (Manning et al., 2014). The models include document and sentence position fixed effects. Standard errors are clustered at the sentence position level. Standard errors are in parentheses. The asterisks represent the significance levels as follows: 1% (**), 5% (**) and 10% (*).

	Normalized Impact			
	All Sentences	Sent 1 to 100	Sent 101 to 200	Sent 201 to 300
Panel A: Earnings				
Quote	−0.011	−0.011	−0.077*	−0.028
	(0.028)	(0.037)	(0.046)	(0.082)
Observations	172,042	97,947	56,876	14,564
R^2	0.184	0.255	0.314	0.329
Panel B: Loss				
Quote	−0.081	0.014	−0.377***	−0.129
	(0.078)	(0.111)	(0.125)	(0.393)
Observations	115,126	44,201	55,712	13,041
R^2	0.194	0.300	0.218	0.280
Panel C: Components				
Quote	0.587***	0.652***	0.093***	−0.129*
	(0.026)	(0.026)	(0.033)	(0.074)
Observations	391,269	210,671	141,578	32,929
R^2	0.145	0.177	0.148	0.180
Panel D: Operations				
Quote	0.689***	0.794***	0.157***	0.043*
	(0.021)	(0.017)	(0.020)	(0.026)
Observations	1,283,806	695,960	450,050	113,835
R^2	0.086	0.106	0.058	0.061

Table 11: Most characteristic words (tokens) for non-GAAP and other sentences. The top 25 words (tokens) are displayed. The words are ranked according to their log-odds ratio with a Dirichlet prior (Monroe, Colaresi, and Quinn, 2008). The non-GAAP sentences are identified using keyword matching.

Non-GAAP

Word	L/O
gaap	274
adjusted	246
non	184
ebitda	141
charges	134
gain	134
restructuring	112
measures	107
impairment	106
charge	87
goodwill	81
measure	79
reconciliation	78
financial	76
items	71
related	70
accordance	68
recurring	63
integration	61
compensation	60
special	58
extinguishment	57
eps	57
income	55
forma	54

Other

Word	L/O
num_tok	94
sales	89
quarter	69
at	50
revenues	49
increased	49
compared	48
year	47
revenue	47
increase	44
were	44
the	42
first	40
same	38
higher	36
decreased	33
decrease	33
equivalents	32
was	32
lower	32
percent	32
payable	31
growth	31
due	31
period	30

Table 12: Normalized impact of non-GAAP sentences by content group and sentence position. The analysis is performed at the sentence level. The panels correspond to the subsets of sentences belonging to the corresponding content groups. The left-hand variable is sentence impact (absolute announcement returns attributed to the sentence by the deep neural net). The impact is normalized by subtracting the mean and dividing by the standard deviation. The right-hand variable is an indicator that is equal to one if the sentence contains one of the non-GAAP keywords. The models include document and sentence position fixed effects. Standard errors are clustered at the sentence position level. Standard errors are in parentheses. The asterisks represent the significance level of 1% (***), 5% (**) and 10% (*).

	Normalized Impact			
	All Sentences	Sent 1 to 100	Sent 101 to 200	Sent 201 to 300
Panel A: Earnings				
Non-GAAP	0.109***	0.219***	0.014	0.001
	(0.012)	(0.018)	(0.011)	(0.013)
Observations	172,042	97,947	56,876	14,564
R^2	0.185	0.257	0.313	0.329
Panel B: Loss				
Non-GAAP	−0.325***	−0.179***	−0.485***	−0.168***
	(0.024)	(0.045)	(0.025)	(0.035)
Observations	115,126	44,201	55,712	13,041
R^2	0.198	0.301	0.226	0.282
Panel C: Components				
Non-GAAP	0.212***	0.246***	0.210***	0.046***
	(0.007)	(0.014)	(0.009)	(0.010)
Observations	391,269	210,671	141,578	32,929
R^2	0.142	0.171	0.154	0.181
Panel D: Operations				
Non-GAAP	0.075***	0.109***	0.077***	0.020***
	(0.004)	(0.009)	(0.004)	(0.005)
Observations	1,283,806	695,960	450,050	113,835
R^2	0.075	0.090	0.058	0.061