Online Appendix

to

Measuring Transformational Change

September 2022

Contents

A Transcript Segmentation Quality

In the following, we provide texture on the quality of our parsing and segmentation routines. That is, we validate the accuracy of our program to correctly parse earnings call transcript and to split individual remarks into individual sentences. Even though, SA claims a transcription accuracy of 99.5%, our programm may produce several types of errors, for example, if transcribers use non-standard HTML-layouts, omit individual remarks, or even entire sections (e.g., Allee & DeAngelis, 2015). On the transcript-level, such errors could lead to three main types of parsing errors:

- *Missing remarks:* The segmentation procedure primarily relies on remarks being delimited by bold strings (usually containing the speaker name and affiliation). If this particular layout information is absent, remarks are erroneously omitted.
- *Missing roles:* The role of a participant (e.g., executive or analyst) will not be assigned correctly, if the participant is not listed in the participant list at the beginning of the transcript, if the participant's name contains grammatical errors, or if names in the participant list are displayed in bold.
- *Erroneous delimitation between sections:* The differentiation between the scripted remarks and the Q&A section will fail, if the delimiting "Question-and-Answer Session" string is wrongly set by the transcribers.

A manual validation of 320 randomly sampled transcripts yields an accuracy of 91.88% (294 of 320 transcripts are segmented without error). An error analysis reveals that the scripted remarks and Q&A section are erroneously parsed in 1.3% of the cases (3 of 320). Similarly, we observe that corrupt participant lists at the beginning of the transcript lead to all analysts being classified as "Unknown" in 1.3% of the cases (3 of 320). Lastly, we find that 0.03% (6 of 20,527) of the remarks are omitted due to corrupt HTML-tags and that "Unknown" roles couldn't be resolved due to speaker names not being stated in bold or containing substantial grammatical errors for 0.097% (20 of 20,527) of the remarks.

Moreover, we derive sentences using the Python *spacy* library. In particular, we employ a pretrained dependency parser that performs sentence-boundary-detection (https://spacy.io/usage/linguistic-features#sbd). We manually evaluate the model's quality on a sample of 2,000 sentences and find 0.647% of the sentences to be erroneously segmented, e.g., mostly due to punctuation errors.

B Labeling Transformation Statements

We employ human coders to identify transformation statements in the Q&A section of earnings conference calls. We present them with the task of classifying 7,869 sentences by firm executives as either transformation statements ($T = 1$) or concerned with topics other than transformational change ($T = 0$), to derive a training dataset for our deep learning classifier. We perform this task at the sentence-level because single remarks often comprise of multiple sentences and, thus, commingle a diverse set of different topics. Further, labeling individual sentences reduces the cognitive burden imposed on the coders and ensures a more accurate identification of our construct. Moreover, we adhere to best practices from the fields of content analysis (Krippendorff, 2019), computational lingusitics (Artstein & Poesio, 2008), and data-centric artificial intelligence (Bernstein, 2022). Importantly, we acknowledge that the coding task is hard and inherently ambiguous, given the innate complexity of the theoretical notion of transformational change (see section 2 of the paper). Therefore, our main focus is on a compiling a high-quality set of training labels that trades-off diversity (i.e., allowing for varying interpretations) with consensus (i.e., reweighting training labels based on inter-coder consistency). We enforce both through a carefully designed coding guideline as well as an oversampling scheme during model training.

B.1 Design of Coding Guideline

We formalize our coding task in terms of a comprehensive coding guideline (see appendix C). The guideline was designed and refined in an iterative process that involved various test runs (i.e., hand-labeling of hundreds of individual sentences) conducted by the authors as well as student coders not involved in the final coding task. Importantly, we freeze the guideline's content before starting the main coding task. The guideline prescribes a two-staged process in which coders have to assess (1) whether a sentence refers to transformational change (*necessary condition*) and (2) which sphere of the firm is affected by the change (*sufficient condition*). We enforce the second stage to ensure that our coders internalize the theoretical facts of the construct and contemplate deeply about the contents of each sentence. Finally, we cross-check our guideline with consulting expert(s) to validate our conceptualization of the underlying construct.

B.2 Recruitment and Remuneration

We recruit three graduate students for our coding tasks. This redudnancy later enables us to quantify inter-coder agreement and mitiagte the adverse effect of idiosyncratic biases (Artstein & Poesio, 2008; Krippendorff, 2019).[1] We require each coder to be proficient in the English language and to have succesfully passed graduate-level courses on strategic management and accounting. This way, we ensure that coders exhibit the skills to parse and comprehend the language endemic to earnings calls and apply our coding guideline as intended. All coders are fairly renumerated according to state-wide renumeration laws.[2]

B.3 Training Phase

We train our coders by exposing them to an initial set of 500 sentences by executives.[3] Thereby, we expect the coders to familiarize themselves with the coding guideline and task. Upon completion, we provide the true label for each sentence, alongside a corresponding rationale, to give feedback and gauge the coders' understanding of the task. This feedback is accompanied by a personal meeting to resolve ambiguities and adress questions before proceeding to the main phase. Therafter, we do not further discuss the task to prevent label drift due to retroactively adjusted perceptions of the

[1] For the benefits of repeated labeling in noisy coding tasks also see Sheng et al. (2008).

[2] We did not resort to crowdsourcing to preempt the induction of additional noise into our coding task (e.g., Bentley, 2021). Instead, we rely on a small set of attentive, motivated, and reliabe coders to ensure consistent labels. Moreover, we acknowledge that our first-order goal is not necessarily in easy replicability but in obtaining a high-quality corpus of transformation-realted texts. This practice is well-established in the data science field with domain experts being particularly suitable when the labeling task is more challenging and labeling quality is of vital importance (Muller et al., 2021).

[3] This initial set of training sentences was dilligently curated by the authors and labels were assigned manually and thorugh several rounds of in-depth discussions. Note that this sample is never used during model training and, thus, also serves as an out-of-sample benchmark for our trained classification model.

underlying construct. However, the coders can access the training materials at any time during the main coding taks.

B.4 Main Phase

The main phase of our coding task presents the three coders with 7,869 (as-of 2022-08-05) randomly sampled sentences spoken by firm executive during the Q&A.[4] Sentences are shuffled, so that the coders retrieve sentence in random order, and displayed over a graphical user interface. In addition, we mix in 250 duplicate sentences to assess the intra-coder consitency. The coders are expected to perform the task independently and refrain from communicating with other coders. In the event that coders accidently skip a sentence or employ the “Don’t Know” backdoor-label, we later resubmit the sentences to elicit a definitive label from the coders for each sentence.

B.5 Agreement Statistics and Label Distribution

We report the reliability of the human-labeled sentences and, thus, validity of our coding scheme by computing intra-coder consistency (i.e., stability) as well as inter-coder agreement (i.e., replicability) metrics in the spirit of Krippendorff, ch. 12.2.1 (2019). Before doing so, we note that rather than to ensure perfect alignment among the human coders, our overarching goal is to construct a set of training labels from which our classifier can derive the coders’ shared mental model. We are fully aware of the exploratory nature of this work and the innate complexit and ambiguity of our underlying construct which is why we predict lower agreement statistics a priori. Put differently, high agreement rates would falsely suggest that there is a common and clear-cut theoretical notion of transformational change which is there is arguably not as outlined in section 2 of our paper.[5]

---

[4] Submitted sentences have a minimum length of 75 characters (incl. whitespaces). Thereby, we ensure that our coding sample is not diluted by uniformative sentences that have a low a priori probability of containing meaningful content (e.g., greetings, farewells, transitions between speakers, pleasantries).

[5] For the issue of high agreement statistics in machine learning see Artstein & Poesio, ch. 4.1.4 (2008).

We observe intra-coder consistency rates of 11.6%, 1.2%, and 2%, respectively, based on the 250 duplicate senteces we mixed into the overall sample. Further, we document that 893 (11.35%), 377 (4.79%), and 88 (1.12%) sentences are labeled as transformation statements ($T = 1$) by one, two, and three coders, repectively. Conversely, 6,485 (82.41%) sentences are assigned the complement label ($T = 0$). Next, we compute bootstrapped Krippendorff's α coefficient to quantify the agreement between coders beyond chance. We obtain a coefficient of $\alpha = 0.2767$ and confidence bounds of [0.2422; 0.3075] which are deemed low to moderate by the standards of classical content analysis works (Artstein & Poesio, 2008). However, against the backdrop of our aforementioned goal, we rate these values satisfactory and refer to section 5 of the paper for the validation of the shared understanding of what constitutes a transformation statement instilled in our classification model.

C Coding Guideline

Classifying Business Transformation Statements
in Earnings Conference Calls

This guideline serves as an instruction for coding transformation- vs. non-transformation-related sentences (i.e. statements) by executives of U.S. firms. The goal is to classify sentences into one of several classes based on the definitions and instructions contained in this guideline.

1. Definition of "Business Transformation" ***(necessary condition)***

In the following, you will find three established definitions of *business transformation*. These definitions help you gain a general understanding of the construct and serve as the basis for deciding whether or not a given sentence contains transformation-related content. Please read these definitions carefully and revisit them as often as necessary during the coding task.

Definition 1: "[Transformation] is the deliberate modification of a firm's strategy, structure, resources, or operations to improve alignment with an altered external or internal environment."

Definition 2: "Business transformation is a change approach where both the level of radicalness of changes and the expected value of results are high. It may then impact various dimensions of the organization: strategy, people, processes, information, and technology."

Definition 3: "Transformations almost by definition involve replacing important parts of a company and its strategy, and affect the long-term prospects of the firm."

In addition to these general definitions, the following list details some important characteristics of business transformations which provide more tangible guidance:

» Transformations are characterized by transformative actions which are actively and deliberately initiated by the firm (as opposed to stagnation or standstill).
» Transformations reflect substantial and not mere incremental changes. Hence, transformative actions are material to the firm and are to be distinguished from "business as usual".
» Transformations regularly mark deviations from previous practices (i.e., variation in a firm's business actions *over time*) and/or deviations from established industry norms (i.e., variation in conformity with regard to *peer firms*).
» Transformations are not only limited to firms facing a crisis situation, but are also regularly initiated by firms from a position of strength to anticipate future challenges.
» Transformations can involve a potentially large set or sequence of actions, each of which may address one facet of the transformational change endeavor within the firm.

Taken together, the three definitions, as well as the characteristics outlined in this section, constitute a **necessary condition** for classifying a sentence as transformation-related.

---

2. Definition of Business Transformation Spheres *(sufficient condition)*

Conceptually, business transformations can be broadly categorized into three main spheres: *portfolio, organizational* and *financial*. You will find a definition as well as a non-exhaustive list of exemplary transformative business actions for each of the three spheres below.

a. Portfolio Transformation

Portfolio transformations involve significant changes in the mix of firm assets, of product and service offerings and/or the addressable markets. As such, they regularly affect the assets side of the balance sheet or the revenues in the profit and loss statement.

Exemplary business actions which can reflect portfolio transformations, are product launches (both, within existing and new markets), product eliminations (both, dropping entire product categories or retracting from specific local markets), substantial changes in investment policies, market expansion programs, product mix shifts, expansions into new areas through acquisitions or joint ventures and retractions from existing areas through divestures.

b. Organizational Transformation

Organizational transformations involve significant changes in a firm's organizational design and/or its mode of cooperation with outside parties. As such, they regularly affect the ownership structure, top management team, employee base, operations, internal processes or corporate culture.

Exemplary business actions which can reflect organizational transformations, are CXO changes (both, departures and tenure starts), adjustments to the employee base (both, layoffs and new hire), process changes, adjustments to manufacturing/administrative/sales capacities, advancements of digital capabilities, post-merger or post-acquisition integrations of new business lines, separation of divested business areas, initiatives to change the firm culture and new strategic alliances or partnerships.

c. Financial Transformation

Financial transformations involve significant changes in the capital and/or cost structure of a firm. As such, they regularly affect the liabilities side of the balance sheet or the cost lines in the profit and loss statement.

Exemplary business actions which can reflect financial transformations, are equity-debt restructuring, bond or equity issuance, liquidity-unlocking divestures, cost-cutting, cost efficiency or production efficiency initiatives.

These spheres represent a **sufficient condition** for classifying a sentence as transformation-related. That is, they provide tangible indicators for business transformations but should not be viewed as definitive for a transformational business action. Instead, they should be assessed in conjunction with the necessary condition provided in Section 1.

---

3. Coding Rules

After having internalized the general definition of *business transformation* and familiarized yourself with the three spheres of transformation *(portfolio, organizational, financial)*, you are supposed to classify a sample of English sentences stemming from top-level firm executives. For each sentence, you will be tasked to answer two questions and classify the sentence accordingly.

Question 1:
Does the sentence contain transformation-related content?
(see section 1 of this guide)

no → **No Transformation**

yes → **Transformation**

Question 2:
Which of the three spheres of business transformation are present in the given sentence?
(see section 2 of this guide)

a: Portfolio Transformation
b: Organizational Transformation
c: Financial Transformation

On the first stage, you are tasked to judge whether a sentence contains transformation-related content or not. If you find that a given sentence clearly refers to a business transformation as defined in section 1, you shall classify it as "Transformation" and proceed to the second question. If the executive discusses any other (non-transformation-related) topic, you shall assign the code "No Transformation" and proceed to the next sentence. In rare cases, when you cannot decide upon the classification even after re-consulting the definitions and characteristics above, you may classify the sentence as "Don't Know". See Figure 1 in the Appendix for a screenshot of the web interface for the first stage.

On the second stage, you are tasked to judge for each transformation-related sentence (i.e., classified as "Transformation") which of the three transformation spheres described in section 2 is the most prominent. For example, if you conclude that the executive only discusses financial transformation actions, you shall assign it the "Financial Transformation" class. However, spheres are not generally mutually exclusive. If you conclude that two or even three spheres of business transformation are clearly present in a given sentence, you may assign multiple classes. For example, if you are given a sentence which refers to both an organizational and a portfolio transformation, you shall classify it as "Portfolio Transformation" and "Organizational Transformation". Analogous to the first stage, you may classify the sentence as "Don't Know" if you conclude that you cannot allocate the transformation sentence to any of the three spheres at all. Figure 2 in the appendix depicts a screenshot of the web interface for the second stage. Also refer to the Appendix for additional information of how to navigate the web interface.

In order to fully internalize this two-staged coding scheme, you will be tasked to classify a training sample of 500 sentences. You will then receive the true labels for the 500 sentences to validate and potentially revise your conceptual understanding, before you will start with the final sample of earnings call sentences.

4. Frequently Asked Questions

In the following you will find a brief FAQ section. The information below should help you perform the coding task reliably and consistently. Please keep them strictly in mind at all times!

How much should I rely on my previous knowledge/understanding of transformation?
Keep a neutral stance and classify sentences using only the conceptual definitions above. Try to rely as little as possible on your pre-existing knowledge of transformation. In particular, do not consult with others during the classification task and do not refer to external resources (except for the purpose of translating individual words or expressions).

How do I proceed in the event of an ambiguous, hard-to-classify sentence?
Transformation is a contextual construct which is rarely explicitly mentioned in the sentences that you will receive. Oftentimes it is referred to more subtly and may take multiple different forms as outlined in section 1 and 2 of this guideline. For example, you may only observe the transformation outcome, such as a significant change in the mix of firm assets (Portfolio Transformation), whereby the transformative action itself is only implicitly referred or alluded to. This may render it challenging at times to distinguish between actual transformative actions and mere incremental change. Therefore, you must stay attentive and carefully scrutinize each sentence in order to assess how the coding scheme should best be applied.
In general, you must always choose the most likely present class/classes. If you are unsure how to classify a given sentence, please refer to the conceptual descriptions given in this guideline. Revisiting section 1 (2) may help you answering question 1 (2) in the coding scheme above. In case you perceive a sentence as transformation-related but too generic, you may classify it as “Transformation” on the first and as “Don’t know” on the second stage. However, oftentimes this pattern is an indicator to reconsider whether the sentence in fact reflects a transformation in the first place.

How do I account for the source, timing and varying degree of commitment of transformation?
In general, you should assume that a firm discusses its own transformation efforts in its earning calls and, hence, classify the sentence accordingly. Only if the sentence clearly refers solely to a competitor’s or industry’s transformation, you shall not classify it as transformation. Moreover, you should classify a sentence as transformation-related irrespective of the sentence’s tense (i.e. reference to past, present or future transformation) as well as its degree of commitment (i.e. rejected, executed or planned/hypothetical transformative actions).

I haven’t recognized a transformation sentence for quite some time. What am I doing wrong?
You will receive a random sample of sentences from all earnings call transcripts in our data. Thus, it might be the case that transformation sentences are rare. Hence, you should stay attentive throughout the course of the coding task in order to detect transformation statements whenever they occur. As soon as you feel fatigued, please take a break. It is crucial that you stay focused and concentrated throughout the course of the coding task. High-quality classifications are more important than speed of execution and, hence, crucial for the success of this research project!

With regards to the content, some sentences just don’t make sense. How should I classify them?
The sentences are transcripts of spoken words. Some words may have been transcribed incorrectly or words may be accidently omitted. Further, the sentences are produced by splitting paragraphs into smaller units using a segmentation algorithm which may produce imperfect segmentations. Irrespective of these grammatical and syntactical imperfections, you should apply the coding scheme above as if the sentence is a complete statement by an executive.

Appendix: Coding Web Interface

Attached you find an exemplary sentence[6], illustrating how the two-staged classification is supposed to be performed within the accompanying web interface. Classifications should be saved frequently, e.g., every ten sentences, by clicking the "disk"-button in the top left corner (or using the shortcut "CTRL + S"). Classifications must be submitted by clicking the green check mark (or using the shortcut "A" on your keyboard). Further, note that you have access to your most recent classification history in the left "History" panel (last 10 sentences). By clicking on one of the entries in the list you may reclassify a sentence in case that you mis-clicked in the first place. Note that the history will be archived as soon as you save your progress via the "disk"-button.

Figure 1 Coding Web Interface (first stage)

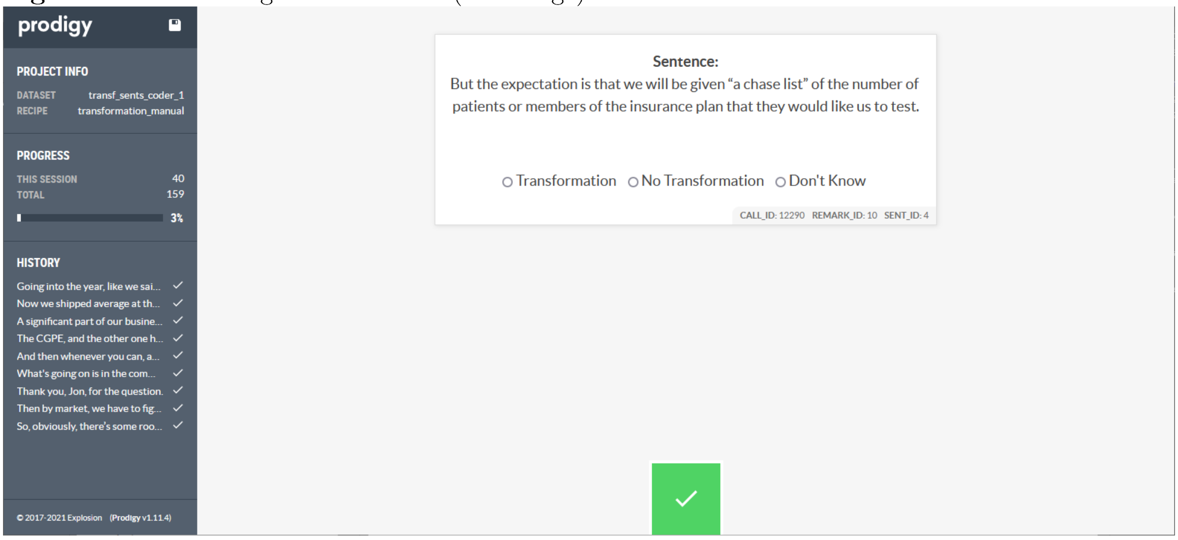
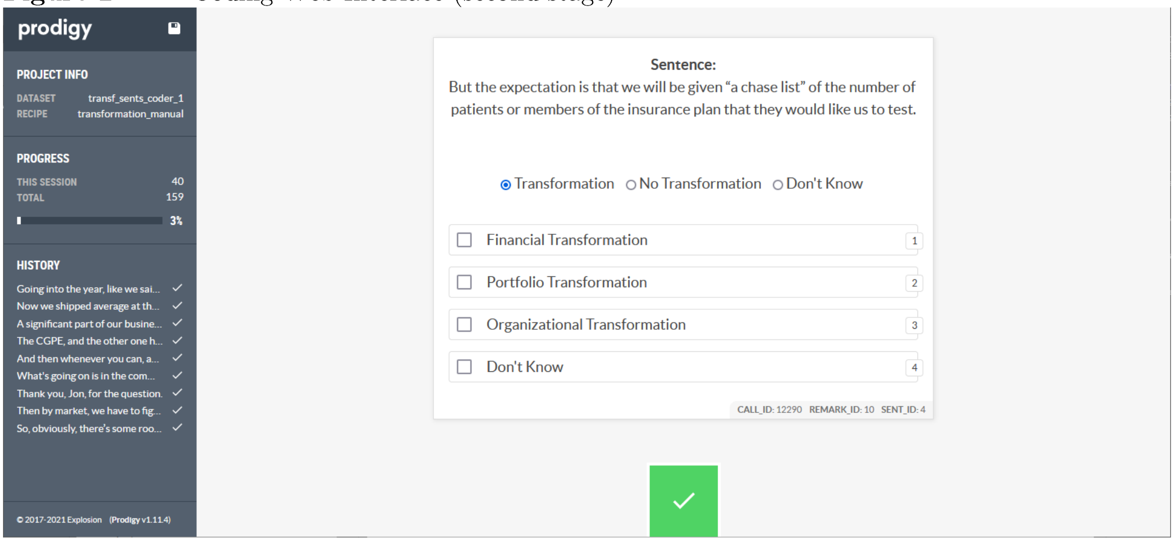


Figure 2 Coding Web Interface (second stage)



[6] Please note that the stated sentence was simply chosen to demonstrate the features of the coding interface.

D Language Models

Our contextual classification model belongs to the family of deep Transformer networks (Vaswani et al., 2017) which have become the de-facto standard in natural language processing since the advent of Google's BERT model (Devlin et al., 2018).[7] In particular, we employ Distill-RoBERTa (Liu et al., 2019; Sanh et al., 2019) which is one of the most prominent sucessors of BERT. In the following, we focus on describing BERT and point to the central improvements introduced by Distill-RoBERTa.

D.1 Model Input

BERT takes as input a raw text sequence $(\omega_1, \dots, \omega_T)$ of length T, whereby each word ω_t stems from the vocabulary $\mathcal{V}$. The sequence is embraced by two special tokens, [CLS] and [SEP], to indicate the start respectively end of the sequence. Next, a subword tokenizer splits the text sequence into single words, except if it encounters infrequent words in which case it resorts to subword tokens.[8] This approach not only caters to the humongous vocabulary $\mathcal{V}$ of a large training corpus, it can also flexibly handle unknown words that may appear out-of-sample and, thus, cope with dynamically evolving language patterns over time. This is in stark contrast to the more rigid practices prevalent in BoW-based approaches where infrequent words are usually omitted via a subjectively set threshold or filter. Eventually, BERT is trained to learn, among other things, static embeddings $\boldsymbol{v}_{\boldsymbol{\omega}_t}$ for each subword $\omega_t \in \mathcal{V}^{30,000}$ during model training.

The model allows for a maximum input sequence of length $T = 512$ (excluding the [CLS] and [SEP] token) and, hence, enables the processing of larger contexts and modeling of long-range dependencies between words. In contrast to BoW-based approaches, tokens are not lowercased but handled as-is to retain the information

7 For a visual as well as more technical introductions to Transformer-based text classification models see (Alammar, 2018a, 2018b) and Rush et al. (2018).

8 BERT employs the so-called WordPiece tokenizer which retains a (sub-)word vocabulary of size 30,000, i.e. $\omega_t \in \mathcal{V}^{30,000}$. In contrast, RoBERTa employs a more flexible *Byte-Pair Encoding* tokenization algorithm which yields a vocabulary incorporating 50,265 unique subword token, including the special beginning and end of sequence token <s> and </s> (as analog to [CLS] and [SEP]).

conveyed through word capitalization. Aside from the subword tokenization procedure the model operates on the raw text inputs with minimal preprocessing. Since training of neural networks is performed by feeding in batches of text sequences to leverage parallelization capabilities in the underlying hardware (e.g., GPU), text sequences are either grouped to be of equal length or otherwise truncated or padded. We present a raw as well as tokenized sentence in OA Tab. 3 to illustrate the various steps.

D.2 Model Architecture

BERT is designed to learn a contextualized representation $\boldsymbol{h}_{\boldsymbol{\omega}_t}$ for each input token ω_t via $\boldsymbol{h}_{\boldsymbol{\omega}_t} = f\left(\boldsymbol{v}_{\boldsymbol{\omega}_1}, \dots, \boldsymbol{v}_{\boldsymbol{\omega}_T}; \boldsymbol{\theta}\right)$, i.e., by mapping $(\omega_1, \dots, \omega_T)$ into $\left(\boldsymbol{h}_{\boldsymbol{\omega}_1}, \dots, \boldsymbol{h}_{\boldsymbol{\omega}_T}\right)$, whereby $f(\cdot)$ depicts the Transformer model, $\boldsymbol{\theta}$ its weights and $\boldsymbol{v}_{\boldsymbol{\omega}_T}$, $\boldsymbol{h}_{\boldsymbol{\omega}_t} \in \mathbb{R}^{d_{hidden}}$.[9] This mapping from $(\omega_1, \dots, \omega_T) \mapsto \left(\boldsymbol{h}_{\boldsymbol{\omega}_1}, \dots, \boldsymbol{h}_{\boldsymbol{\omega}_T}\right)$ lends BERT its name as a so-called text 'encoder'. Further, the above formulation implies that it learns $\boldsymbol{\theta}$, $\boldsymbol{v}_{\boldsymbol{\omega}_t}$ and hence $\boldsymbol{h}_{\boldsymbol{\omega}_t}$ simultaneously.

First, the model retrieves a static embedding $\boldsymbol{v}_{\boldsymbol{\omega}_t}$ for each element in the tokenized input sequence $(\omega_1, \dots, \omega_T)$. Each embedding is then augmented by positional information stemming from a set of trainable positional embeddings as described in Vaswani et al. (2017):

$$\boldsymbol{x}_{\boldsymbol{\omega}_t} = \boldsymbol{v}_{\boldsymbol{\omega}_t} + \boldsymbol{e}_{\boldsymbol{\omega}_t} \tag{1}$$

where $\boldsymbol{v}_{\boldsymbol{\omega}_t}$, $\boldsymbol{e}_{\boldsymbol{\omega}_t}$, $\boldsymbol{x}_{\boldsymbol{\omega}_t} \in \mathbb{R}^{d_{hidden}}$. The injected information $\boldsymbol{e}_{\boldsymbol{\omega}_t}$ inform the model about the position of each token in the text sequence, essentially precluding the model from viewing the input sequence as a BoW. Next, the augmented embeddings $\boldsymbol{x}_{\boldsymbol{\omega}_t}$ are packed into a matrix $\boldsymbol{X} \in \mathbb{R}^{T \times d_{hidden}}$ and fed into a Transformer block (Vaswani et al., 2017) which consists of a *multi-headed self-attention sub-layer*, paired with a *dense feed-forward*

[9] Note that due to the beginning and end of sequence token [CLS] and [SEP] the effective sequence length extends to $T + 2$, i.e., $\left(\omega_{[CLS]}, \omega_1, \dots, \omega_T, \omega_{[SEP]}\right)$. However, in the light of simplicity I keep on referring to the original sequence throughout this section, highlight the special token only when required.

neural network plus some additional features. Finally, the output of the Transformer block is routed into a final classification layer, also referred to as the *model head*.

Multi-Headed Self-Attention Sub-Layer. The multi-headed self-attention sub-layer forms the heart of the Transformer model. Its central idea is to aid the model in learning dependencies between different tokens in the input sequence $(\omega_1, \dots, \omega_T)$ in a highly efficient and parallelizable manner. More precisely, when processing token ω_t, the attention mechanism enables the model to learn a contextualized representation $\boldsymbol{h}_{\boldsymbol{\omega_t}}$ by attending to other token surrounding ω_t as well as itself (hence *self-attention*). Depending on the length of the input sequence, attention can expand beyond sentence boundaries and thus facilitate long-range dependency learning. In their original paper, Vaswani et al. (2017) propose scaled-dot product attention as the concrete mechanism through which the model achieves context-awareness. Thereby, they first decompose $\boldsymbol{X}$ into a set of three, low-dimensional matrices, $\boldsymbol{Q}$, $\boldsymbol{K}$ and $\mathbf{V}$ referred to as the queries, keys and values, which are learned by the network:

$$\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{W}^{\boldsymbol{Q}},\ \boldsymbol{K} = \boldsymbol{X}\boldsymbol{W}^{\boldsymbol{K}},\ \boldsymbol{V} = \boldsymbol{X}\boldsymbol{W}^{\boldsymbol{V}} \tag{2}$$

where $\boldsymbol{W}^{\boldsymbol{Q}}$, $\boldsymbol{W}^{\boldsymbol{K}}$ and $\boldsymbol{W}^{\boldsymbol{V}}$ are weight matrices of dimensionality $\mathbb{R}^{d_{hidden} \times d_k}$ for the first two and $\mathbb{R}^{d_{hidden} \times d_v}$ for the latter matrix. Second, an attention matrix is computed via:

$$Att(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \sigma\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\boldsymbol{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{3}$$

where $\sigma(\cdot)$ denotes the softmax function which projects the value onto the [0; 1]-interval, $\sqrt{d_k}$ is a normalization term which is intended to stabilize the learning process of the model and $Att(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) \in \mathbb{R}^{T \times d_v}$. Intuitively, eq. (3) first evaluates the affinity respectively alignment between ω_t (*query*) and every other token in the sequence (*keys*), captured by the dot-product $\boldsymbol{Q}\boldsymbol{K}^{\boldsymbol{T}}$, to produce a vector of attention scores. This bi-directionality allows the model to evaluate the left as well as the right context of ω_t when encoding $(\omega_1, \dots, \omega_T)$, i.e. mapping it to $(\boldsymbol{h}_{\boldsymbol{\omega_1}}, \dots, \boldsymbol{h}_{\boldsymbol{\omega_T}})$, with T determining the length of the context to which the model can ultimately pay attention to. Next, the distribution of attention scores is scaled by $\sqrt{d_k}$ to smoothen the distribution and ensure stable gradients during training. $\sigma(\cdot)$ then squishes the scaled scores into the [0; 1]-interval, yielding a probability distribution over $(\omega_1, \ldots, \omega_T)$ that sums to one and describes how strong ω_t draws its contextual meaning from neighboring token in the sequence, as well as itself. Due to the s-shaped anatomy of the softmax, strong (weak) similarities between token in $(\omega_1, \ldots, \omega_T)$ are amplified (attenuated). Lastly, the embeddings in $\boldsymbol{V}$ (*values*) are multiplied by the probabilities to obtain a context embedding which can be interpreted as an attention-weighted representation of ω_t over $(\omega_1, \ldots, \omega_T)$. Importantly, for any given ω_t this representation may vary depending on the context $(\omega_1, \ldots, \omega_{t-1}, \omega_{t+1}, \ldots, \omega_T)$ which greatly improves the fidelity of the generated embeddings vis-à-vis static word embedding models. Further, this approach obviates the need for subjectively chosen stop words lists, n-gram widths, stemming schemes or other manually set text filters because the attention mechanism is designed to automatically learn informative dependencies between token in the input sequence during training time.

Vaswani et al. (2017) apply this mechanism h times, leading to so-called *multi-head attention* with h prescribing the number of heads. The authors not only find that this approach allows the model to learn multiple attention patterns (i.e., different $\boldsymbol{W}_i^{\boldsymbol{Q}}$, $\boldsymbol{W}_i^{\boldsymbol{K}}$ and $\boldsymbol{W}_i^{\boldsymbol{V}}$ matrices with $i = 1, \ldots, h$), similar to the idea of a human reader paying attention to different parts of a text sequence when interpreting a specific sub-sequence, but they also emphasize that multi-headed self-attention is highly parallelizable, allowing the model to scale to large model sizes and training corpora.

In a last step, the h attention matrices are concatenated to produce the output of the attention sub-layer:

$$MultiHead(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = (Att_1 \oplus \ldots \oplus Att_h)\, \boldsymbol{W}^{\boldsymbol{O}} \tag{4}$$

where $Att_i \in \mathbb{R}^{T \times d_v}$, $MultiHead(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) \in \mathbb{R}^{T \times h*d_v}$ and $\boldsymbol{W^O} \in \mathbb{R}^{h*d_v \times d_{hidden}}$ is a weight matrix that learns the optimal way for the model to reconcile the various attention heads.[10] BERT *base* postulate that $d_{hidden} = 768$, $h = 12$ and $d_k = d_v = \frac{d_{hidden}}{h} = 64$.

Feed-Forward Sub-Layer. Subsequent to the multi-headed attention layer, the intermediate representations of $(\omega_1, \ldots, \omega_T)$, here denoted by $\widetilde{\boldsymbol{X}}$, are post-processed by a wide feed-forward neural network which is applied to every embedding in $\widetilde{\boldsymbol{X}}$ in parallel:

$$FFN(\widetilde{\boldsymbol{X}}) = g(\widetilde{\boldsymbol{X}}\boldsymbol{W_1} + \boldsymbol{b_1})\boldsymbol{W_2} + \boldsymbol{b_2}, \tag{5}$$

where $g(\cdot)$ is a non-linear activation function and $\boldsymbol{W_1} \in \mathbb{R}^{d_{hidden} \times d_{ff}}$, $\boldsymbol{W_2} \in \mathbb{R}^{d_{ff} \times d_{hidden}}$, $\boldsymbol{b_1} \in \mathbb{R}^{d_{ff}}$ and $\boldsymbol{b_2} \in \mathbb{R}^{d_{hidden}}$ are weight matrices and bias vectors respectively. BERT employs a Gaussian Error Linear Unit (GELU) activation function for $g(\cdot)$ with $d_{ff} = 4 * d_{hidden} = 3{,}072$ in the model's base variant. This design permits the network to model non-linearities and interactions present in $(\omega_1, \ldots, \omega_T)$.

Additional Features. Following Vaswani et al. (2017), Devlin et al. (2018) implement a residual connection as well as a layer normalization step following each of the two aforementioned sub-layers (as well as the initial embedding and the final output layer). The former adds up the input and output of the same layer element-wise to facilitate the effective training of a deep, multi-layer neural network. It further enables each layer to pass on positional information, to learn a residual part of the mapping from ω_t to $\boldsymbol{h}_{\boldsymbol{\omega_t}}$ and, hence, to internalize distinct linguistic information encoded in the input sequence, such as part-of-speech, morphology, syntax or semantics. The latter normalizes each 768-dimensional embedding by subtracting its mean and dividing by its standard deviation to improve the efficiency of the model and increase its robustness against strong fluctuations in the values that are propagated through the network. Third, the model leverages dropout to prevent the model from overfitting by deactivating hidden units in each layer with a probability of 10%. For additional design choices, refer to the appendix in Liu et al. (2019).

---

[10] Also see Alammar (2018a) for a visual presentation of the bespoken attention mechanism.

Eventually, BERT consists not only of one but of a whole stack Transformer blocks, each in turn consisting of a stack of the aforementioned components. BERT comes in two flavors, a *base* and a *large* variant, with 12 and 24 blocks, respectively. Although, the seminal paper illustrates that the larger variants learn more expressive functions $f(\cdot)$ to map $(\omega_1, \dots, \omega_T)$ into a contextual embedding space, i.e. $f(\cdot): (\omega_1, \dots, \omega_T) \mapsto (\boldsymbol{h}_{\omega_1}, \dots, \boldsymbol{h}_{\omega_T})$, this paper builds upon the Distill-RoBERTa base variant with 82 million parameters to render computations with the available hardware resources tractable.[11] Lastly, the output $(\boldsymbol{h}_{\omega_1}, \dots, \boldsymbol{h}_{\omega_T})$ of the core model is passed into the previously mentioned head for decoding the contextualized text sequence and performing sequence classification.

Model Head. Up to this point, BERT reflects a "Transformer Encoder" (Devlin et al., 2018, fn. 4) that encodes ω_t as a 768-dimensional, general-purpose embedding $\boldsymbol{h}_{\omega_t}$. This unified architecture implies that it needs some additional functionality to leverage $\boldsymbol{h}_{\omega_t}$ and solve a desired target task (e.g., text classification). This capability is induced by training a final feed-forward network with dropout, referred to as the *model head*:

$$FFN(\boldsymbol{h}_{<CLS>}) = g(\boldsymbol{h}_{<CLS>}\boldsymbol{W}_{\boldsymbol{dense}} + \boldsymbol{b}_{\boldsymbol{dense}})\boldsymbol{W}_{\boldsymbol{proj}} + \boldsymbol{b}_{\boldsymbol{proj}}, \tag{6}$$

where $\boldsymbol{h}_{<CLS>} \in \mathbb{R}^{768}$ is the embedding that corresponds to the first element in the extended input sequence $(\omega_{<CLS>}, \omega_1, \dots, \omega_T, \omega_{<SEP>})$ and can be viewed as a contextualized embedding which distills information encoded in the entire sequence (*sequence encoding*). Moreover, $g(\cdot)$ is a non-linear $tanh$ activation function and $\boldsymbol{W}_{\boldsymbol{dense}} \in \mathbb{R}^{768 \times 768}$, $\boldsymbol{W}_{\boldsymbol{proj}} \in \mathbb{R}^{768 \times C}$, $\boldsymbol{b}_{\boldsymbol{dense}} \in \mathbb{R}^{768}$ and $\boldsymbol{b}_{\boldsymbol{proj}} \in \mathbb{R}^{C}$ are weight matrices and bias vectors from a dense and a projection layer, respectively. In particular, the outer projection layer projects the 768-dimensional output into a C-dimensional space where C is the number of distinct labels. If $C = 1$ the head enables regression ($\boldsymbol{h}_{<s>} \mapsto \hat{y} \in \mathbb{R}$), i.e. predicting a continuous label for the text, whereby if $C \geq 2$ the head enables (multi-

[11] As the name suggests, Distill-RoBERTa is a compressed and, thus, computationally efficient version of the original RoBERTa model with competitive performance on various NLP tasks (Sanh et al., 2019).

class) classification ($\boldsymbol{h}_{<s>} \mapsto \hat{y} \in \mathbb{R}^{C}$), i.e. classifying the text into C discrete classes. In the case of regression, the model optimizes for a simple mean-squared error loss whereas in the case of classification the model employs a cross-entropy loss to measure the deviation of the model prediction $\hat{y}$ from the true label y.

D.3 Model Pretraining

BERT is pretrained on a 16GB general-domain corpus of English Wikipedia articles and publicly available books (Devlin et al., 2018). In contrast, RoBERTa samples its training sequences $(\omega_1, \dots, \omega_T)$ from a substantially larger corpus of 160GB, including the BERT corpus as well as news articles, web content and story-like texts (Liu et al., 2019). As such, both models are domain-agnostic in the sense that they are trained to internalize features and patterns of general-domain English language. Moreover, RoBERTa is trained longer as well as using larger batch sizes and sequence lengths relative to BERT.

Whereas BERT is pretrained on a masked language modeling and next sentence prediction objective, Liu et al. (2019) find that training speed can be increased by discarding the latter without forfeiting predictive accuracy. In addition, they dynamically corrupt token in the input sequence when fed into the model during pretraining, instead of once during text pre-processing. More specifically, they employ the following scheme:

1. Sample token from the input sequence $(\omega_1, \dots, \omega_T)$ with a probability of 15%.
2. Replace the sampled token ω_t by a corrupted token which equates to the special <MASK> token with 80% probability, to another random token from the input sequence with 10% probability, to sampled token ω_t with 10% probability.
3. Try to predict ω_t for the corrupted token by paying attention to the other token in the corrupted input sequence. Inversely, uncorrupted tokens are skipped and hence do not contribute to the computation of training loss and gradients.

The pretrained models are made publicly available and can be downloaded, for example, on the Huggingface website (e.g., https://huggingface.co/bert-base-cased).

D.4 Model Fine-Tuning

The bespoken pretraining scheme enables the model to learn parameters $\boldsymbol{\theta}$ that produce meaningful but generic text representations $\boldsymbol{h}_{\boldsymbol{\omega}_t}$. That is, they capture the contextual meaning of words in a given sentence, but may not be particularly useful for discriminating between transformation statements and discussions of other topics (i.e., our *target task*). Therefore, to attune the model to our specific text classification task, we must substitute the model head that was learned during pretraining for a head that reflects the target task. For example, BERT's masked language modeling head is designed for token classification, i.e., for classifying $\boldsymbol{h}_{\boldsymbol{\omega}_t}$ into one of $N =$ 30,000 classes whereas a sequence classification head that is designed to classify transformation statements must map $\left(\boldsymbol{h}_{\boldsymbol{\omega}_1}, \dots, \boldsymbol{h}_{\boldsymbol{\omega}_T}\right)$ into $N = 2$ classes ($T = 0$ or $T = 1$). Thus, a second training run is performed to learn a new model head and concurrently adjust the other pretrained parameters $\boldsymbol{\theta}$ to better reflect the target task. This process is referred to as *fine-tuning.*

As described in section 3 of the paper, we fine-tune our model using the hand-labelled sentences compiled by the human coders. To find the optimal model configuration we perform a grid search over the learning rate ([2e-4, 1e-4, 5e-5, 3e-5]), batch size ([8, 16, 32, 64]), and number of epochs ([2, 3, 4]) and select the best performing hyperparameter candidates based on the ROC-AUC on the validation set.[12]

[12] The other hyperparameters (e.g., warm-up steps, learning rate schedule, weight decay, etc.) are fixed and inspired by the specifications presented in the original papers (Devlin et al., 2018; Liu et al., 2019).

E Text Preprocessing Routines

In section 5 of the paper, we construct various text-based measures that build on the bag-of-words (BoW) model of text representation. In contrast to our contextual classifier, which handles raw text inputs, these approaches require us to apply a common suite of preprocessing steps to leverage text as data and render numerical computations feasible (e.g., Gentzkow et al., 2019).

Let $\mathcal{D}$ denote a text document, e.g., Q&A section filtered for exectuives' remarks or Item 1 section, which is composed of a sequence of words $\omega_1, \ldots, \omega_N$ that stem from the corpus vocabulary $\mathcal{V}$. First, we split each document into $t_1, \ldots, t_M$ token, i.e., computational units, based on whitespace characters. Second, we reduce each token to its normal form using lemmatization, remove stop words to eliminate uniformative content, and convert token to lowercase.[13] These operations reduce the dimensionality $\mathcal{V}$ and, thus, ultimately the length of each document to $M \ll N$ while increasing its information density. Third, we employ named entity recognition to identify and mask named entities, in particular numbers, percentages, dates, currencies, quantities, persons, organizations, as well as countries, cities, and states. We argue that a firm's transformational change endeavours should be measured generically and indepently of any referrals to named entities. Finally, we identify collocations, i.e., statistically significant coocurrences of adjacent words, to account for local dependencies. Hence, an individual token t_m can reflect both single words (unigrams) as well as longer local contexts in the form of multi-word-phrases, such as bi-, tri-, or fourgrams.[14] We summarize the preprocessing steps in OA Tab. 5.

---

[13] We use the Loughran-McDonald stop word list (https://sraf.nd.edu/textual-analysis/stopwords/) to eliminate words that are commonly deemed uniformative in business narratives.

[14] We leverage the Python *spacy* library for tokenization, lemmatization, stop word removal, and named entitiy recognition and integrate functionalities from *gensim* to model collocations. We implement two passes of collocation modeling to identify statistically significant neighbours (bigrams) and longer local contexts (i.e., tri- or even fourgrams) in the first and second iteration, respectively. We restrict our collocation analysis to words that occur at least ten times in our sample and identify statistically

F Wordlist Approaches

F.1 Transformation Wordlist From Textbook Indices

First, we attempt to identify language patterns of transformational change by pooling the indices of six popular transformation textbooks. We draw these books from different time periods (2009-2021) and a set of inter-related domains (e.g., corporate restructuring, organizational change, strategy) to obtain vocabulary that covers our entire sample period and reflects the multi-faceted nature of our construct. This approach rests on the assumption that the authors are subject matter experts and only disclose index terms that convey strong signals of a firm's transformational change endeavors. Since we can discard the latter after an initial review of the respective index lists, we contrast the extracted vocabulary with the index entries of three popular economics, two general management, and two corporate finance textbooks to eliminate general business terms unrelated to transformational change. See OA Tab. 6 for an overview of the employed textbooks.

Next, we manually remove all remaining index entries that refer to named entities, such as persons, organizations, countries, specific industries, or legislations, and apply the common suite of text preprocessing steps (see Online Appendix E) to both the resulting wordlist entries and our transcript sample. Finally, we remove generic terms not intercepted by the complementary textbooks by dropping n-grams that occur in at least 80% of all transcripts. We expect those terms to contribute little towards explaining firm-level variation in firms' transformational change endeavors. Our final transformation vocabulary comprises of 1,116 entries.

Finally, we calculate a firm-level score as

$$TC_{ijk}^{IDX} = \frac{1}{|\mathcal{D}_{ijk}|} \sum_{t}^{|\mathcal{D}_{ijk}|} \mathbb{1}[t \in \mathbb{W}] \times w(t) \tag{7}$$

significant collocations based on the bigram scoring function outlined in Mikolov et al. (2013b) (using a threshold of 25 and 10 in round one and two, respectively).

where $|\mathcal{D}|$ gives the number of tokens per document $\mathcal{D}$ (i.e., transcript length) and i, j, and k index firms, years, and quarters, respectively. Further, $\mathbb{W}$ denotes our wordlist with entries $\tau_1, \ldots, \tau_{1116}$ and $\mathbb{1}[\cdot]$ is the indicator function that takes on a value of 1 if token t maps into the transformation vocabulary, i.e., $t \in \{\tau_1, \ldots, \tau_{1116}\}$ (and 0 otherwise). By summing over all token per document and scaling by the document's length, we compute the relative frequency for each entry in our transformation vocabulary. In a last step, $w(t)$ assigns a weight to each frequency count that is proportional to the frequency with which $\tau_1, \ldots, \tau_{1116}$ appear at least once per index. Thereby, we account for the fact that individual wordlist entries may be more or less reflective of our construct and simultaneously attenuate the influence of potential false positives.

F.2 Transformation Wordlist From Word Embedding Models

Second, we derive a transformation wordlist by training and querying a word embedding model to retrieve words and multi-word-terms that are semantically related to our construct. That way, we minimize the amount of discretion while explicitly accounting for our domain of interest, that is, earnings conference calls. Li et al. (2021a) and Li et al. (2021b) argue that this approach proves more effective compared to predefined wordlists when proxying for inherently subtle and contextual constructs.

Conceptually, a word embedding model "embeds" the semantics of a token t_m, i.e., word or n-gram, in a dense, d-dimensional vector $\boldsymbol{v}_{\boldsymbol{t_m}}$ of real numbers where $d \ll M$ with M being the number of unique tokens in the corpus. Hence, all words and multi-word-terms are projected into a vector space in which words with similar meaning are located in close proximity to each other. This allows us to perform a semantic search by providing a token as input to the model (*seed*) and, in turn, retrieving tokens that are semantically related to the query, i.e., on average used in similar linguistic contexts (*semantic neighbors*).[15]

[15] We apply our standard set of preprocessing steps (see Online Appendix E) and provide details on the inner workings of the model, training procedure, parameter choices, and search mechanism in Online Appendix G. Moreover, we make the trained model available under

Concretely, we prompt the trained model with a carefully curated set of eleven seed tokens: "organizational_change", "change_management", "transformation", "transform", "transformative", "pivot", "restructuring", "restructure", "reorganization", "reorganize", and "redesign".[16] The model converts each token into its corresponding vector representation, i.e., embedding. Analogue to Li et al. (2021b), we compute the element-wise average over our seed embeddings as

$$\bar{\boldsymbol{v}}_{\boldsymbol{seed}} = \frac{1}{11}\sum_{i=1}^{11} \boldsymbol{v}_{\boldsymbol{t_i}} \tag{8}$$

to model the average local context in which executives discuss transformational change. This design choice alleviates the influence of a single seed on the final vocabulary and reflects the multi-faceted nature of our construct. Finally, we extract the 300 token that are closest in vector space to our average seed embedding, resulting in a final wordlist with 300 entries.[17]

Analogue to the previous approach, we calculate a firm-level score as

$$TC_{ijk}^{EMB} = \frac{1}{|\mathcal{D}_{ijk}|}\sum_{t}^{|\mathcal{D}_{ijk}|} \mathbb{1}[t \in \mathbb{W}] \times w(t) \tag{9}$$

https://huggingface.co/spaces/simonschoe/Call2Vec to allow future research to easily construct own vocabularies gleaned from the language employed in earnings call transcripts.

16 A number of considerations feed into the design of this list. First, we follow Li et al. (2021b), who glean their seed words from defintions of their construct available in the existing literature, and require the seeds to align with the theoretical notion of transformational change as outlined in section 2 of the paper. Second, we aim for a narrow list to reduce the risk of capturing polysemous terms and, hence, the inclusion of potential false positives. For example, we experimented with terms such as "transition", "renewal", or "turnaround" and discovered that these occur frequently in contexts other than transformational change, such as transportation, contract design, and the maintenance of oil refineries, respectively. Third, if applicable, we indclude the different parts-of-speech of each seed, in particular noun, verb, and adjective, to include the various ways in which transformational change may be discussed. Besides, we aim for a unique seed list that covers different regions of the semantic vector space, that is, we desire minimal semantic overlap. For example, we observed that the bigrams "change_management" and "orgnaizational_change" reside in the same semantic neighbourhood. Since we compute a 'representative seed' by averaging over the embeddings of the individual seed words, having two semantically identical seeds overweights that particular region of the the search space.

17 We experimentally determine this threshold through various preliminary analyses and human evaluations to reduce the number of potential false positives. We do not manually filter the resulting vocabulary to restrict the extent to which discretion may influence the final wordlist.

while adjusting the choice of $w(t)$ to the underlying method. That is, each count is now weighted by the respective cosine similarity scores of $\tau_1, \dots, \tau_{300}$, i.e., proximity to the average seed embedding within the d-dimensional embedding space.

F.3 Benchmark of Construct Validity

To begin with, we perform face validity checks to assess the credibility of the two alternative transformations wordlists vis-à-vis our contextual classifier. First, we randomly sample and inspect remarks from each quintile of the non-zero region of the TC^{IDX} and TC^{EMB} distribution as well as from the large set of remarks with a score of zero, respectively. Inspecting OA Tab. 7, panel A and B, we observe several failure modes of the wordlist approaches that are primarily due to the focus on indiviual words in lieu of context. First, the wordlists partly flag false positive discussions that describe standstill rather than transformational change (e.g., "we are in the process of pulling our guidance together", "but I think the mix really won't change"). Second, they consider generic words as stand-alone signals of transformational change (e.g., 'companies', 'cost', 'revenue', or 'contract', and 'leadership' or 'organization', respectively). Third, especially the index-based wordlist tends to identify transformation statements only due to correlated measurement error. For example, we observe that it correctly flags discussions on new product launches (market expansions and leaderhsip changes) [novel partnerships] as transformative because of terms like 'share' and 'product pipeline' ('strategies') ['customer', 'management']. More problematically though, we expect that this measurement error likewise leads to false positives, i.e., references of those terms in rather generic contexts. Fourth, both wordlists cannot cope with contextual references to the magnitude of a change (e.g., "a significant opportunity for several years in this market" is missed while "We don't see any change in that momentum as we go into the second half" is classified as a transformation statement). Fifth, we stress that the sole focus on individual word ocurrences does not allow for the discrimination between different topics, e.g., discussions of financial results versus operational changes and strategic initiatives. And sixth, because our contextual classifier operates on sub-word tokens, we note that it is capable of handling typos or transcription artifacts. In contrast, any wordlist approach requires a 1-to-1 match with the respective wordlist entry.

In addition, we list the top 150 entries per wordlist in

OA Tab. 8, panel A and B, to provide texture on the signals that each wordlist captures. We note that, by construction, the deductive wordlist contains a sizable number of arguably generic terms, despite the carefully designed filters discussed in section F.1, while the terms gleaned from the embedding model tend to be more specific and attuned to our focal construct. Consistent with the varying sizes of both lists (1,116 versus 300), we conjecture that the former is more susceptible to false positives while the latter leads to a potentially higher number of false negatives (e.g., “So we're looking in a case which is built very broadly across the entire company” or “So it's certainly in our view a huge target for us”).

Next, we extend Tab. 5 of the paper and report the predictive performance of the two alternative approaches vis-à-vis our contextual text classifier in OA Tab. 9. Notably, we find that large language models consistently outperform the two wordlist approaches, in parts by a sizable margin. In particular, we observe that the index-based wordlist produces an excessive amount ob false positives, as indicated by the relatively low precision value, while the embedding-based wordlist misses a considerable number of actual transformation statements, as indicated by the relatively low recall value. These findings align with the previously discussed anecdotal failure cases and stress that both wordlists are inferior to our text classifier, albeit for slightly different reasons.

In a final set of tests, we assess the merits of the two alternative approaches for inferential analysis. We rerun our quantitative validation tests (see section 5.4 of the paper) while inserting TC^{IDX} and TC^{EMB} as the main independent variable, respectively. We standardize each score a priori to ensure comparability. Consistent with OA Tab. 10, we observe that our main TC measure exhibits a stronger association with proxies conceptually related to transformational change (i.e., strategic variation and economic dissimilarity) (panel A), shows stronger and consistently positive associations with variables that we would expect to correlate with a firm’s transformational change endeavors (panel B), and is also consistently more predictive of future changes in a firm’s resource allocation (panel C), all relative to the two wordlist-based scores.

G Word Embedding Models

The seminal word embedding model, Word2Vec, was proposed by Google researchers in 2013 (Mikolov et al., 2013a; Mikolov et al., 2013b). It consists of two shallow neural networks which learn continuous word representations, so-called word embeddings, from very large text corpora by solving a fixed-window word prediction task. Although the authors devise two different training schemes, continuous bag-of-words and skip-gram, we detail the latter due to its wider dissemination in the literature.

Let $t_m \in \mathcal{V}$ be a unique token (e.g., word) in the vocabulary $\mathcal{V}$ (i.e., the list of all unique token). For each t_m, the skip-gram model learns a fixed d-dimensional embedding $\boldsymbol{v}_{\boldsymbol{t_m}} \in \mathbb{R}^d$ by sliding a local context window of length c across the entire text corpus.[18] Within each sliding window, the model attempts to predict the context token $t_{-c}, t_{-c+1}, \dots, t_{c-1}, t_c$ based on the center token t_m which translates to maximizing the following expression:[19]

$$\max_{v_{t_m}} \frac{1}{M} \sum_{m=1}^{M} \sum_{-c \leq j \leq c, j \neq 0} log\, p\left(t_{m+j} \middle| t_m\right) \tag{10}$$

where M denotes the size of the training corpus and $\boldsymbol{v}_{\boldsymbol{t_m}}$ is learned through the process of backpropagation. Eq. (10) reflects the idea that the meaning of individual words can be inferred by evaluating the contexts in which they occur.

After training, the embedding vector $\boldsymbol{v}_{\boldsymbol{t_m}}$ of a given token t_m can be extracted from the neural network's weight matrix which spans a vector space in which words with similar contexts are located in close proximity to each other. For example, Mikolov et al. (2013c) demonstrate that the trained model yields unique embeddings such that $\boldsymbol{v}$("quick") – $\boldsymbol{v}$("quickly") + $\boldsymbol{v}$("slowly") corresponds to $\boldsymbol{v}$("slow") and $\boldsymbol{v}$("Berlin") –

[18] Note that the dimensions cannot be intuitively interpreted. Instead embeddings predicate on the idea (which is backed by the empirical evidence) that the entirety of linguistic word information, such as tense, frequencies, dependencies and topics, are encoded in the values of the resulting embeddings $\boldsymbol{v}_{\boldsymbol{t_m}}$.

[19] This training scheme was later refined to improve the embeddings' quality as well as the model's efficiency, e.g., by downsampling frequent words due to their low information content or negative sampling of counterexamples from a 'noise distribution' (Mikolov et al., 2013b). For a thorough and intuitive explanation of these computational tricks see Rong (2014).

$\boldsymbol{v}$("Germany") + $\boldsymbol{v}$("France") corresponds to $\boldsymbol{v}$("Paris"). To compare the similarity between two individual token or retrieve token that are semantically related to a given query, we employ the cosine similarity measure. It compares two vectors, e.g., the two word embeddings $\boldsymbol{v}_{\boldsymbol{t}_i}$ and $\boldsymbol{v}_{\boldsymbol{t}_j}$, by their scaled dot product:

$$cos(t_i, t_j) = \frac{\boldsymbol{v}_{\boldsymbol{t}_i} * \boldsymbol{v}_{\boldsymbol{t}_j}}{\left\|\boldsymbol{v}_{\boldsymbol{t}_i}\right\| * \left\|\boldsymbol{v}_{\boldsymbol{t}_j}\right\|} = \frac{\sum_d v_{t_i}^d * v_{t_j}^d}{\sqrt{\sum_d \left(v_{t_i}^d\right)^2} * \sqrt{\sum_d \left(v_{t_j}^d\right)^2}} \tag{11}$$

where $cos(t_i, t_j)$ lies on the $[0;1]$-interval and d is the vector dimensionality.[20]

We employ Facebook's fastText model which introduces various refinements to the seminal Word2Vec framework (Joulin et al., 2016a; Joulin et al., 2016b; Bojanowski et al., 2017).[21] We implement it using the Python *gensim* library, train it on our entire sample of Q&A sections filtered for executives' remarks, and rely on prior works when setting the training parameters (e.g., Mikolov et al., 2013b; Li et al., 2021b). In particular, we set $d = 300$, $c = 5$, drop tokens that appear less than twenty times in our sample to omit textual or firm-specific artiefacts, and choose an initial learning rate of 0.025. We train our model for 50 epochs, i.e., passes over the transcript sample, to ensure that the embeddings vectors converge.[22] The final model has a vocabulary size of $\mathcal{V} = 56{,}070$, which reflects the number of unique word embeddings.

---

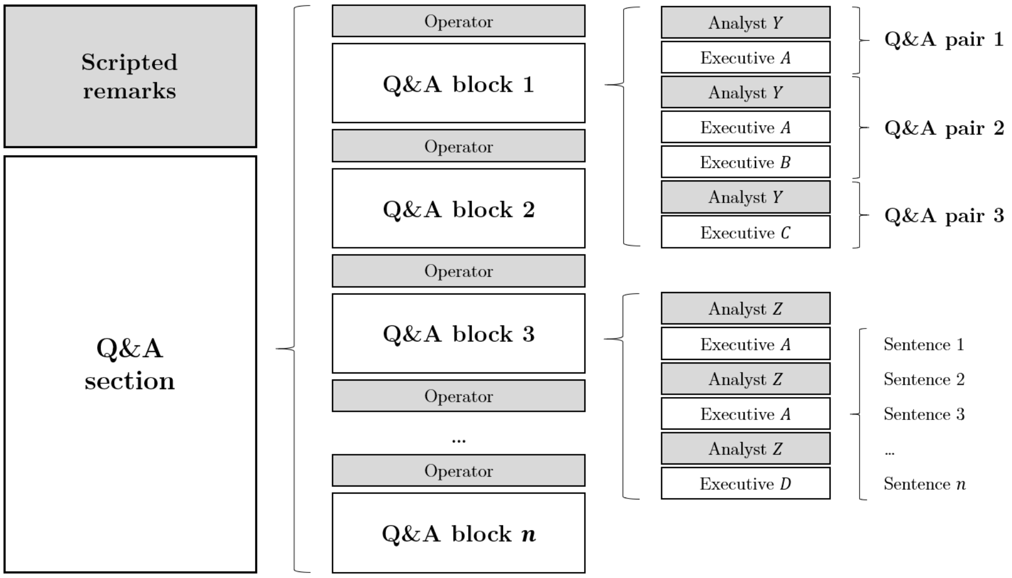
[20] For applications of the cosine measure to gauge the similarity of business narratives in accounting, finance, and economics research see, for example, Brown & Tucker, 2011; Peterson et al., 2015; Hoberg & Phillips, 2016; Lee, 2016; Kelly et al., 2018; Engle et al., 2020.

[21] For example, the authors achieve efficiency gains by leveraging several compression techniques, such as vocabulary pruning, vector quantization or text hashing, and produce higher quality word embeddings by leveraging character n-grams (*sub-words*) which enable information sharing among words to better represent rare as well as out-of-vocabulary words.

[22] We experiment with alternative training regimes. First, we train the model for 10, 20, 30, 40, and 50 epochs, respectively, and manually validate the model quality. That is, after each 10 consecutive epochs, we extract the ten most similar token for the words "earning", "equity", "innovation", "risk", "competition", "factory", "decrease", "move", "exaggerate", "avoid", "consolidate", "guidance", "sustainable", "remarkable", and "negative". We document that the model output is fairly similar when trained for 30, 40, and 50 epochs. We settle with the 50-epoch model to ensure that is has succesfully converged and the resulting embedding space is stable. Second, we experiment with a context window of $c = 8$ (e.g., Song, 2021). We document an 87%-overlap with our base configuration and conclude that our final transformation vocabulary is robust to this design choice.

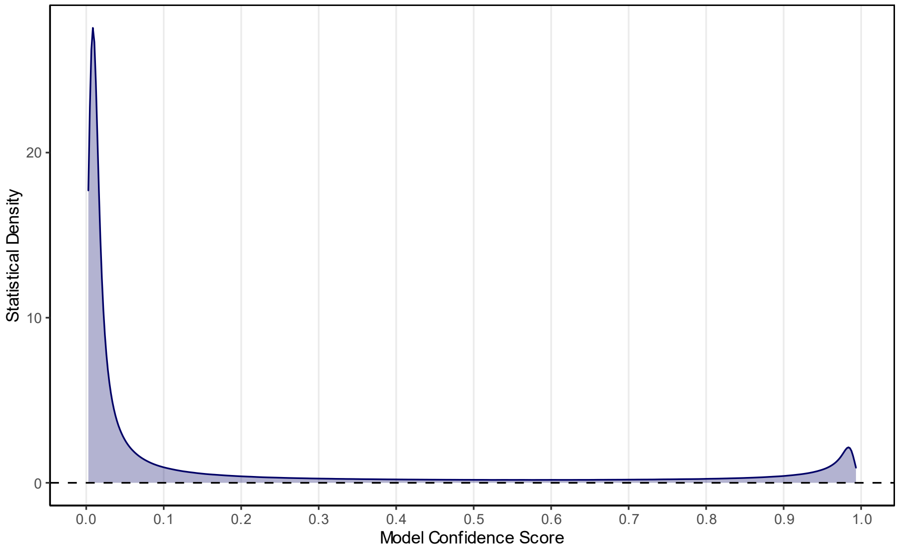
H Additional Tables and Figures

OA Fig. 1 Earnings Call Anatomy



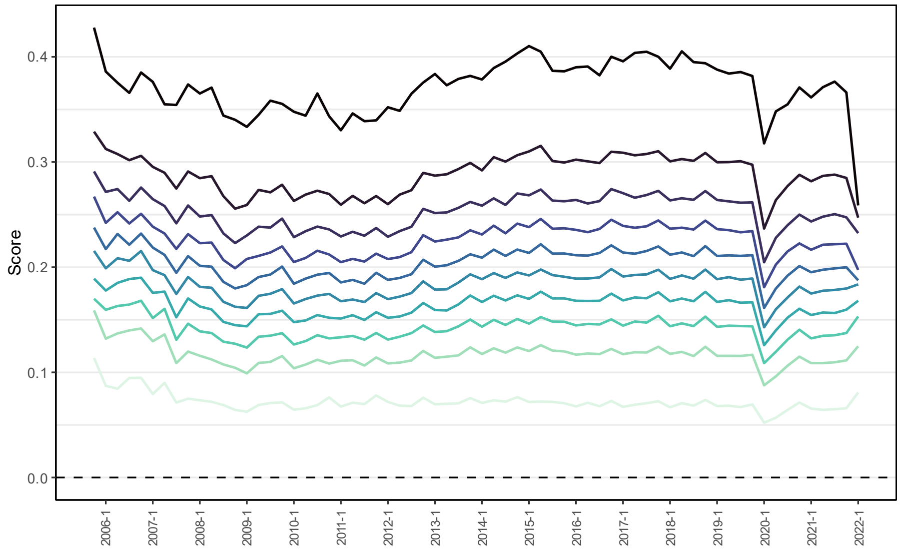
This figure presents the anatomy of a prototypical earnings conference call. It can be subdivided in a scripted remarks and Q&A section. The Q&A section comprises of various information exchanges between a single analyst and one or multiple executives, which we refer to as Q&A blocks. Blocks are segmented by intersecting operator remarks, calling upon the respective analysts to come forward with their questions. Each Q&A block can consist of one or several Q&A pairs, i.e., a single question by a specific analyst and one or multiple answers by the firm's executives. Q&A pairs are segmented by intersecting analyst remarks. Finally, each remark can encompass one or multiple sentences. Portions of the call that are later removed from our analyses (but exploited for parsing and identification) are displayed in grey.

OA Fig. 2 Distribution of $P(T|s)$



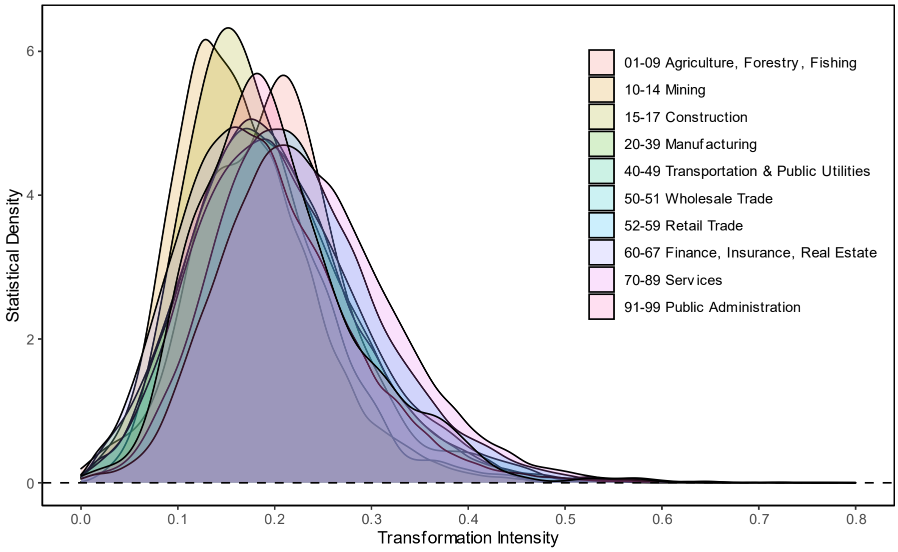
This figure presents the statistical distribution over $P(T|s)$, that is, over the predicted values (confidence scores) generated by our contextual text classification model for all 'meaningful' sentences (i.e., sentences with at least 75 characters). Note that the distribution is almost bimodal. That is, 61% of the probability mass resides within the (0; 0.05]-interval while 7% resides within the [0.95; 1]-interval. The remainder disperses almost uniformly over the (0.05; 0.95)-interval.

OA Fig. 3 Time-Series of TCS by Deciles



This figure presents the time-series of quarterly means of our TCS measure, grouped by deciles. The color intensity increases proportionally with the decile number.

OA Fig. 4 TCS by SIC Code



This figure presents the statistical distribution of TCS, grouped by SIC1 industry code. It complements Tab. 3 of the main paper.

OA Tab. 1 Randomly Sampled Predicted Sentences

Panel A: Predicted Transformation Statements

Sentence (s)	$P(T\|s)$
Now, given the digital transformation for almost every enterprise customers, we do see more and more customers that are very interested.	0.9580
But of course, it's clear that the green initiatives will lead to higher standards and the higher standards then indirectly will give us more business because that will mean that they will upgrade or build new projects according to more stringent standard, environmental standards, a lead standard or whatever, which drives technology up, which is good news for us.	0.9657
As the markets emerge -- for example, if you take our play into -- and as we said in Investor Day and other events, that we are going to put about $1 billion to work in the vehicle repair and maintenance side and technical information side and we're well underway and SRS will be the beginning of that platform, as well as some of these other quietly sampled emerging platforms around the world, that segment has 4.5x more transactions on an annual basis than you do in the collision side of the business.	0.9853
But if we can accomplish what we want to do with a big operation, with a big acquisition, and then we get the benefit of a rising pricing deck, well that will be a seller again.	0.9789
We continue to have lots of discussions in the U.S. with plan sponsors who are thinking about close outs.	0.9601
As I mentioned we've got a couple of online gambling companies that are using this now outside of the US to support their customers globally, so we're getting traction on some of the smaller, more global companies because these large enterprises just have so much restructuring to do to take advantage of the significant cost savings of these technologies.	0.9895
If we look at the plans for 2018, it is to introduce 650 new products, which is an absolute all- time high.	0.9652
We have a lot of the products for the printers and we want to expand the customer base from our traditional base of American customers to a more global presence in Asia and in Japan.	0.9801
On the R&D side, I think, I mean, we are where we have kept, as I said before, very stable R&D, and we are gradually steering more and more of our R&D efforts into the, what you call new technology, electromobility, automation and also connectivity.	0.9641
Now Accord networks is a company in video, and he led the sales team, and the marketing group at Accord, and he took it from start up, sound familiar, it's from start up to $60 million company in two years.	0.9712

Panel B: Predicted Equivocal Statements

Sentence (s)	$P(T\|s)$
As is, I would expect it to consolidate more and more around leading players, leading brands taking more and more share.	0.5302
And we just have such an opportunity right now with the CPT code, with the direct to patient advertising and even some things that we're doing across the other businesses particular also international which by the way grew high or mid-single digit for the fourth quarter, and going to improve from there for the full year.	0.4810
With our current Tier 1 suppliers that has all been done, they already have all the tooling, they have the manufacturing process, they can actually build the thing themselves.	0.4656
They're finding some new opportunities to reduce the capital intensity in Metals & Minerals business for sure.	0.4946
We have been doing kind of an integrated campaign, so it's TV, online, we do the Google Ad Words - all those different elements together.	0.5026
We are continuing to make some investments within our digital print business and that's for 2017.	0.5016
And given the improvements that we've made we think that we are poised to completely put that issue to rest.	0.5333
We initiated some settlement discussions, and based on that and in consultation with our independent auditors, we thought now is the time to book this number.	0.5019
Obviously, CTS has majority of the people, so majority of the savings would go to CTS.	0.4833
It's hard for me to exactly quantify how much is the non-seasonal impact, but if we compare last year, working reduction was about $110 million, and this quarter, $160 million.	0.5374

Panel C: Predicted Non-Transformation Statements

Sentence (s)	$P(T\|s)$
So that turned out to be beneficial for us, and I think, we'll just see how the market and interest rates move over the course of the year,	0.0115
As you know, Marlboro remains the strongest brand by far in the cigarette category; it's the iconic brand in the category.	0.0144
Let me jump in and see if Gary wants to offer additional color after I take a crack at this.	0.0060
So for the most part I would say that's how we're looking to future production.	0.0216

So it's not about our capability to continue to work and develop the next model year for the Sierra and Silverado, but (indiscernible) which is a van program.	0.0887
We saw that trend on customers buying more promotional products throughout all of last year.	0.0072
I don't want to be naive and say we don't really watch our share price because that would be untrue.	0.0102
We had a liner aligner class that go during the frac, so that well has been repaired, it has been completed.	0.0432
So in that sense, once we establish Oraxol in any indication, let's say breast cancer, it's logical to think, it will be better than oral paclitaxel in any other indication that IV paclitaxel is being used.	0.0457
You talk about the -- you mentioned the effects on the 28-nanometer contributions.	0.0055

This table present 30 randomly sampled sentences, alongside the model's predicted confidence scores $Pr(T|s)$. Sentences are sampled from three distinct regions of the probability distribution. Panel A lists sentences from [0.95; 1]-interval, i.e., sentences that are predicted as transformation statements ($T = 1$) with high probability. Panel B lists sentences from [0.45; 0.55]-interval, i.e., sentences where the model has difficulties discriminating between the two labels. Panel C lists sentences from [0; 0.05]-interval, i.e., sentences that are predicted as non-transformation-related ($T = 0$) with high probability.

OA Tab. 2 Randomly Sampled Predicted Sentences (Word Importances)

Panel A: Predicted Transformation Statements

Sentence (s)	$P(T\|s)$
#s Now , given the digital transformation for almost every enterprise customers , we do see more and more customers that are very interested . #/s	0.9580
#s But of course , it 's clear that the green initiatives will lead to higher standards and the higher standards then indirectly will give us more business because that will mean that they will upgrade or build new projects according to more stringent standard , environmental standards , a lead standard or whatever , which drives technology up , which is good news for us . #/s	0.9657
#s As the markets emerge -- for example , if you take our play into -- and as we said in Investor Day and other events , that we are going to put about $ 1 billion to work in the vehicle repair and maintenance side and technical information side and we 're well underway and S RS will be the beginning of that platform , as well as some of these other quietly sampled emerging platforms around the world , that segment has 4 . 5 x more transactions on an annual basis than you do in the collision side of the business . #/s	0.9853
#s But if we can accomplish what we want to do with a big operation , with a big acquisition , and then we get the benefit of a rising pricing deck , well that will be a seller again . #/s	0.9789
#s We continue to have lots of discussions in the U . S . with plan sponsors who are thinking about close outs . #/s	0.9601
#s As I mentioned we 've got a couple of online gambling companies that are using this now outside of the US to support their customers globally , so we 're getting traction on some of the smaller , more global companies because these large enterprises just have so much restructuring to do to take advantage of the significant cost savings of these technologies . #/s	0.9895
#s If we look at the plans for 2018 , it is to introduce 650 new products , which is an absolute all - time high . #/s	0.9652
#s We have a lot of the products for the printers and we want to expand the customer base from our traditional base of American customers to a more global presence in Asia and in Japan . #/s	0.9801
#s On the R & D side , I think , I mean , we are where we have kept , as I said before , very stable R & D , and we are gradually steering more and more of our R & D efforts into the , what you call new technology , electrom ob ility , automation and also connectivity . #/s	0.9641
#s Now Accord networks is a company in video , and he led the sales team , and the marketing group at Accord , and he took it from start up , sound familiar , it 's from start up to $ 60 million company in two years . #/s	0.9712

Panel B: Predicted Equivocal Statements

Sentence (s)	$P(T\|s)$
#s As is , I would expect it to consolidate more and more around leading players , leading brands taking more and more share . #/s	0.5302
#s And we just have such an opportunity right now with the C PT code , with the direct to patient advertising and even some things that we 're doing across the other businesses particular also international which by the way grew high or mid - single digit for the fourth quarter , and going to improve from there for the full year . #/s	0.4810
#s With our current Tier 1 suppliers that has all been done , they already have all the tool ing , they have the manufacturing process , they can actually build the thing themselves . #/s	0.4656
#s They 're finding some new opportunities to reduce the capital intensity in Met als & Miner als business for sure . #/s	0.4946
#s We have been doing kind of an integrated campaign , so it 's TV , online , we do the Google Ad Words - all those different elements together . #/s	0.5026
#s We are continuing to make some investments within our digital print business and that 's for 2017 . #/s	0.5016
#s And given the improvements that we 've made we think that we are poised to completely put that issue to rest . #/s	0.5333
#s We initiated some settlement discussions , and based on that and in consultation with our independent aud itors , we thought now is the time to book this number . #/s	0.5019
#s Obviously , C TS has majority of the people , so majority of the savings would go to C TS . #/s	0.4833
#s It 's hard for me to exactly quantify how much is the non - season al impact , but if we compare last year , working reduction was about $ 110 million , and this quarter , $ 160 million . #/s	0.5374

Panel C: Predicted Non-Transformation Statements

Sentence (s)	$P(T\|s)$
#s So that turned out to be beneficial for us , and I think , we 'll just see how the market and interest rates move over the course of the year , #/s	0.0115

#s As you know , Mar l boro remains the strongest brand by far in the cigarette category ; it 's the iconic brand in the category . #/s	0.0144
#s Let me jump in and see if Gary wants to offer additional color after I take a crack at this . #/s	0.0060
#s So for the most part I would say that 's how we 're looking to future production . #/s	0.0216
#s So it 's not about our capability to continue to work and develop the next model year for the Sierra and Silver ado , but (indisc ern ible) which is a van program . #/s	0.0887
#s We saw that trend on customers buying more promotional products throughout all of last year . #/s	0.0072
#s I don 't want to be naive and say we don 't really watch our share price because that would be untrue . #/s	0.0102
#s We had a liner align er class that go during the fr ac , so that well has been repaired , it has been completed . #/s	0.0432
#s So in that sense , once we establish O rax ol in any indication , let 's say breast cancer , it 's logical to think , it will be better than oral pac lit ax el in any other indication that IV pac lit ax el is being used . #/s	0.0457
#s You talk about the -- you mentioned the effects on the 28 - nan ometer contributions . #/s	0.0055

The table illustrates the context-awareness property of our deep learning text classification model. In particular, it lists the 30 sentences presented in OA Tab. 1, supplemented by word importance scores that are indicated by colour coding. The sentences are tokenized with tokens being delimited by whitespaces. Green and red colours present signals that prompt the model to classify sentences as transformation statements ($T = 1$) and non-transformation-related ($T = 0$), respectively. The intensity of the colour indicates the strength of the signal. Word importance scores (so-called *attributions*) are computed via the integrated gradients method.

OA Tab. 3 Subword Tokenization

Panel A: Raw Input Sequence
Therefore, we've been able to attract 160,000 apps on our app store so quickly and we can have a app store filled with app that's are interesting and users willing to pay.
Panel B: Tokenized Input Sequence
[CLS], Therefore, ,, we, ', ve, been, able, to, attract, 160, ,, 000, apps, on, our, app, store, so, quickly, and, we, can, have, a, app, store, filled, with, app, that, ', s, are, interesting, and, users, willing, to, pay, ., [SEP], <pad>, <pad>, <pad>, <pad>, <pad>, <pad>, <pad>, <pad>

This table presents the raw (panel A) as well as the tokenized input sequence (panel B) after employing the BERT tokenizer. Note that the obligatory beginning- ([CLS]) and end-of-sequence ([SEP]) token are added and that the sequence is padded to a maximum sequence length of 50 (using the special <pad> token).

OA Tab. 4 Capital Market Reaction to Transformation Change

	$CAR_{0,1}$	$CAR_{0,1}$	$CAR_{0,5}$	$CAR_{0,5}$
Transformation Change	0.0300***	0.0300***	0.0299***	0.0286***
	(8.01)	(6.94)	(6.46)	(5.42)
Ln(Assets)		0.00246*		0.0124***
		(1.95)		(7.78)
Ln(Market Value)		-0.0190***		-0.0327***
		(-18.93)		(-24.76)
Revenue Growth		0.0127***		0.0134***
		(12.02)		(10.73)
Return on Assets		0.110***		0.118***
		(7.80)		(7.50)
Return on Assets Growth		0.0656***		0.0830***
		(7.84)		(7.74)
Total Shareholder Return		0.00000836		0.00000202
		(0.50)		(0.12)
Leverage		-0.0167***		-0.0354***
		(-4.53)		(-7.39)
Market-to-Book		-0.0000334		-0.0000641
		(-0.90)		(-1.37)
Return Volatility		-0.0382		-0.0326
		(-0.91)		(-0.67)
Spread		0.101*		0.00404
		(1.76)		(0.05)
Volume		-1.58e-10**		-1.57e-10**
		(-2.56)		(-2.12)
Firm FE	yes	yes	yes	yes
Year-Quarter FE	yes	yes	yes	yes
Adj. R-Sq.	0.0282	0.0475	0.0270	0.0527
N	142,535	107,810	142,516	107,797

This table presents results for the regression of TC on firm's quarterly cumulative abnormal returns following the earnings conference call. TC measures the probability-weighted share of executive statements devoted to the construct of transformational change. $CAR_{0,1}$ and $CAR_{0,5}$ are calculated as the excess of the predicted returns following the event based on the market model for the time windows [0; 1] and [0; 5], respectively. Ln(Assets) is calculated as the logarithm of total assets, Ln(Market Value) is calculated as the logarithm of firms' market value on the event day, Revenue Growth is calculated as the difference in revenue quarter-over-quarter scaled by prior quarter's revenue, Return on Assets is calculated as net income scaled by total assets, Return on Assets Growth is calculated as the difference in Return on Assets quarter-over-quarter scaled by prior quarter's Return on Assets, Total Shareholder Return is calculated as the difference in Market Value quarter-over-quarter for the day of the event scaled by prior quarter's Market Value, Leverage is calculated as the sum of short- and long-term debt scaled by total assets, Market-to-Book is calculated as Market Value scaled by book value of equity, Return Volatility is calculated as the standard deviation for the stock's return for 252 trading days prior to the event day, Spread is calculated as percentage bid-ask spread on the event day, and Volume is calculated as the stock's trading volume on the event day. All final control variables are winsorized on the 1% level and standard errors are clustered by firm. t statistics are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

OA Tab. 5 Text Preprocessing Steps

Panel A: Raw Text

"Sure, Alan. This quarter, because of the success we had with the bottom line, that ties directly into the way that we're compensating our management team, which is bottom line profitability. So the variable compensation expense is what drove the majority of the corp G&A increase. And then we also -- this quarter, there were some shares that were issued as well so that there was some share-based compensation expense that hit as well."

Panel B: Tokenization

Sure, Alan, This, quarter, because, of, the, success, we, had, with, the, bottom, line, that, ties, directly, into, the, way, that, we, 're, compensating, our, management, team, which, is, bottom, line, profitability, So, the, variable, compensation, expense, is, what, drove, the, majority, of, the, corp, G&A, increase, And, then, we, also, this, quarter, there, were, some, shares, that, were, issued, as, well, so, that, there, was, some, share, based, compensation, expense, that, hit, as, well

Panel C: Lemmatization

Sure, Alan, This, quarter, because, of, the, success, we, *have*, with, the, bottom, line, that, *tie*, directly, into, the, way, that, we, *be*, *compensate*, our, management, team, which, is, bottom, line, profitability, So, the, variable, compensation, expense, is, what, *drive*, the, majority, of, the, corp, G&A, increase, And, then, we, also, this, quarter, there, *be*, some, *share*, that, *be*, *issue*, as, well, so, that, there, *be*, some, share, *base*, compensation, expense, that, hit, as, well

Panel D: Stop Word Removal

Sure, Alan, This, quarter, success, bottom, line, tie, directly, way, compensate, management, team, bottom, line, profitability, variable, compensation, expense, drive, majority, corp, G&A, increase, also, this, quarter, share, issue, well, share, base, compensation, expense, hit, well

Panel E: Named Entity Masking and Lowercasing

sure, [person], [date], [date], success, bottom, line, tie, directly, way, compensate, management, team, bottom, line, profitability, variable, compensation, expense, drive, majority, corp, [org], increase, also, [date], [date], share, issue, well, share, base, compensation, expense, hit, well

Panel F: Collocation Modeling

sure, [person], [date], [date], success, bottom_line, tie, directly, way, compensate, management_team, bottom_line, profitability, variable, compensation, expense, drive, majority, corp, [org], increase, also, [date], [date], share, issue, well, share, base, compensation, expense, hit, well

This table presents the various preprocessing steps we apply to the raw transcript data. It illustrates each step based on a randomly sampled remark. It lists the raw text (panel A), tokenized text (panel B), tokenized and lemmatized text (panel C, normalized tokens are indicated by italics), preprocessed text after the removal of stop words (panel D), lowercased text with masked named entities (panel E, entities are demarked by special entity token), and finally the text after considering statistically significant word co-occurrences (panel F).

OA Tab. 6 Choice of Textbook Indices

Panel A: Transformation Textbooks
Vance (2009): Corporate Restructuring: From Cause Analysis to Execution. Springer.
Smith, King, Sidhu, & Skelsey (Eds.) (2014): The Effective Change Manager's Handbook: Essential Guidance to the Change Management Body of Knowledge. Kogan Page Ltd.
Balogun, Gustaffsson, & Hailey (2016): Exploring Strategic Change (4th ed.). Pitman.
Gaughan (2018): Mergers, Acquisitions, and Corporate Restructurings (7th ed.). Wiley & Sons.
Johnson (2018): Reinvent Your Business Model: How to Seize the White Space for Transformative Growth. Harvard Business Review Press.
Tjemkes & Mahalache (Eds.) (2021): Transformative Strategies: Strategic Thinking in the Age of Globalization. Routledge.

Panel B: Control Group
Hill & McShane (2008): Principles of Management. McGraw-Hill/Irwin.
Hitt, Black, & Porter (2012): Management (3rd ed.). Prentice Hall.
Keat, Young, & Erfle (2013): Managerial Economics: Economic Tools for Today's Decision Makers (7th ed.). Pearson.
Damodaran (2015): Applied Corporate Finance: A User's Manual (4th ed.). John Wiley & Sons.
Hirschey & Bentzen (2016): Managerial Economics (14th ed.). Cengage Learning.
Froeb, McCann, Shor, & Ward (2018): Managerial Economics: A Problem-Solving Approach (5th ed.). Cengage Learning.
Brealey, Myers, & Allen (2020): Principles of Corporate Finance (13th ed.). McGraw-Hill Education.

This table lists the textbooks that served as the starting point for our deductively constructed wordlist. Panel A enumerates the books from which gleaned our transformation vocabulary while panel B presents our control book groups whose indexes served as a filter to remove generic terms.

OA Tab. 7 Randomly Sampled Remarks

Panel A: Transformation Wordlist Derived from Textbook Indices

Q	Remark	TCS^{IDX}	TCS
0	Hey. Thanks, Anthony. Good talking to you.	0.0000	0.0000
0	No. Good question. Pre-pandemic levels what was originally due pre-pandemic not adjusted for new deferrals, et cetera.	0.0000	0.0021
0	Sorry. Sorry. But Gustavo got tired, so let's have (inaudible) helping us in this answer. Thank you.	0.0000	0.0014
1	Eli, this is Jim, I'll touch on the gas turbine. I think what we are seeing right now does suggest that we'll see some incremental improvement, at least as we move into the early part of fiscal 2015, but we are in the process of pulling our guidance together right now for next year, and we'll be able to give you definitive comments on that at our next quarter.	0.0303	0.0094
1	That's roughly the same I mean, the origination business if you look at like a March run rate, for example, I think origination business made $6 million, $7 million for the month. So we're definitely seeing an upper trajectory there in the origination business and it's performing quite nicely, but I think the mix really won't change. I think it will still be in that 70% to 30%.	0.0222	0.0050
1	I'd like to extend what Brad saidearlier about with cooking, we are firing in all cylinders. The mall is goingon all cylinders in the month of September, that's in the third quarter. So wehave been some stores that would have been in over trends, had they been and ifyou extrapolate what they were doing in the first couple of few weeks of themonth. On the conventional side, it seems to be the attendance, the trade showorganizers were extremely happy and I am surprised that one of them even got alot of traffic, it wasn't really a trade show, it was an investment promotionfair. But the other one I think got about 30000 attendees and that isfantastic. The rest of them were doing gamebusters when the people were there. The rooms we see a higher room rate, ADRpotential then what we, originally when we originally came up with the visionof the strip anticipated and I dare say that there was a possibility that wecan equal the rates we are getting in Las Vegas at The Venetian Macao andperhaps even exceed them. The good thing about this is, since this is theoriginal vision was mine and I can tell you that it's clear that we are cookingon all cylinder, the convention, the Congress Center, the restaurants, theshopping, the anticipation of the shows, the sports arena, everything being soldout. We are truly cooking on all cylinders and I would say that the vision ofthe Cotai strip with the opening of the Venetian Macao is in emphatically	0.0171	0.1415
2	Well Q1 was clearly and hopefully the high watermark expense wise. We have got a history of lower operating expenses year-on-year. And I would just caution, that that's likely not to be the case, as we move forward. I mentioned, that we had 2% year-on-year growth, when you take out separation expenses over the second quarter. And that to me, given the environment, is kind of probably going to be a typical quarter for us, would be my guess.	0.0698	0.0082
2	Data storage is really going to depend on a little bit of data storage market recovery and going back to some basic capacity expansion. The announcements we saw said that should happen and we will see. The metrology, the trend there is less seasonality although in metrology Q3 has historically been a weak bookings quarter for Veeco and yet we delivered a substantial sequential increase. In metrology what is really paying off for us is a couple of years of hard work on the new product pipeline. The products are coming out. They are exceptional products. Our win rates are increasing and I think we are taking share. That I expect we will continue to do that. We have historically had seasonality for solid orders in Q4 and then a little bit of fall off in Q1. In solar there is one important factor and that is we have been building systems but we haven't shipped them and we are not going to take revenue until we get full acceptance in the factory. Their big systems, the new CIGS system weighs 8 tons. It has to be shipped on a ship so we add about six weeks of transit time versus our traditional air shipping. The point I am going to get to is we have been seeing revenue in solar from our sources business but not substantial revenue on the systems side. We will start to see that ramp up a little bit in Q1 and more in Q2. We are going to count on next year being a bigger contributor to solar systems revenue that we really don't have this year. We do have $30 million of backlog going into the year. I think that is going to give us a boost in the solar area and what we have seen in the solar trade shows and things is CIGS is really gaining momentum. A lot of companies that went after other technologies like a more silicon are flipping because the efficiencies are just too low and CIGS is gaining momentum. The Dow announcement of the solar shingle was pretty exciting to see. We have known about that for a couple of years but to actually see it come out as a real announcement from them was exciting because it just represents a huge opportunity and you can't do it with crystalline silicone solar cells or other technologies. It is a CIGS application.	0.0505	0.1736
2	Sure. Thank. You bet. So in the last call I announced very specific strategies for Europe, North America, Asia/Pacific Rim and prior to that I announced executive leaders for each one of those continents and regions. Let me give a little bit of color on those markets and what we were doing there. In Europe, Europe was 0.6% on an increase for additional business. Our data point has moved to 1.3% to 1.6% depending on, who you are talking to or what piece of information you are reading. That sales leader is required to generate new business with mobile, food and electronics with our fixed products and that is in fact has been happening quite nicely.	0.0745	0.2410

	North America also comes with a new executive leader for the company that we announced in the last call. That team is also assigned new business in Canada with one of the relationships I just mentioned throughout the United States where we have got some new integrators there as well and Mexico where that leader has just returned from, as the mater of fact we have got some good news coming with that. The third piece is Asia-Pacific, RIM, obviously China is the highlight there. There has always been the 20% growth number there, that number has actually changed to about 5% to 10% increased and frankly, we are fine with that, because we continue to go after the high volume complex work there. And as I just announced one continue follow on order there. We have got two new relationships that we are launching in Asia and working on two more in that particular area. So I'm very happy with those three plans. In addition, we have learned a lot more about Brazil and we've got some forward activity that I will be talking about in follow-on calls there. So I think that we did some things and I think we got lucky at the same time and those three elements are progressing nicely.		
3	So let me take the first question in the round on guidance. So let me start with we feel terrific about the first half because the balance looks right. The growth part of the business is doing what we said it would do and the Primary Products business performed really well in a tougher market. When I look at the second half, I think there are 3 key things for us to think through. The first is we expect to see continued progress on the solutions side of the business. We don't see any change in that momentum as we go into the second half. When I look at the Primary Products business, I think about 2 things. The contract round is coming up and we're still very early in the contract round and we need to see how that plays out. And secondly, we want to see how the industrial starch market plays out as it starts to correct for what happens in the first half. So we're being a little bit cautious about Primary Products in the second half. And then the third thing I'd add is, if you remember, we are lapping a one-off upside on Sucralose that has some impact on the year-on-year comps. That's really the sort of in the round. But I don't know, Imran, whether you'd want to add anything to that or - and to maybe take the cogen question so...	0.0877	0.0135
3	Correct, correct. And then, the thing, the thing - the reason I mentioned all those words about our big data sets is, when you go out to the commercial world, if you don't have those big data sets, you don't know what you're looking at. So we've got a an incredible asset when we decide it's time to take it commercial and you'll always have a few early adapters that I believe that it's getting closer to time for going out in the commercial space. That's why I wanted to do more profitable and then we can afford to go back out and look for those other opportunities.	0.0952	0.2367
3	Happy to do it. Asyou see the increase there, remember that we were on the road we were talkingabout two things affecting the overall value. One would be a 25% increase in August by the end of this year and anadditional 25% next year. The otherthing we talked about is seeing the recovery value drop from our model, 80 to 70 by the end of the year although we arecurrently running at 85. If you look at the increase in September, the increase isthe largest, as you pointed out, in the HELOC in the 80 to 90 and 90 to 100, sothe areas that while we were on the road and the last time that we did the callthat we circled up. I think we are still comfortable by and large with the rangeof about 25% growth rate given that we are continuing to see recovery ratespretty consistent at about 85%. So itmay be a little higher on one and a little lower on the other. As you go into next year, and we're still ofthe belief there will be another drop from 70 to 60 and another increase of25%. I think the most important takeaway, however, is that if younoticed in the guidance that Rob just gave, he talked about the bank businessand what he thought it would earn, and then he said in an effort to be prudent,considering another $0.10 of possiblecredit securitization which could come from either securities impairments orsome of these trend lines that you have seen; notwithstanding the fact thatwhat we're currently seeing, I think we're reasonably comfortable with what wehave set. But again, it's a market influx, and that was the identification of the additional $-.10 or $65 million. All of that by the way is pre-tax.	0.0892	0.0123
4	Yes, good question. So we have a channels group with some very capable business development executives who are always looking for different mechanisms to take our products and our IP to market. And it's a wide range, everything from processors to major retailers, to -- I mean, there are different ways of taking our IP to market. Probably a great example is our eBay partnership. So we entered into this partnership with eBay, and as you know, they have tremendous, tremendous reach with a tremendous number of retailers who use them for eBay Enterprise is an infrastructure backbone end support for eBay retailers. And that's completely separate from their Marketplace division. This is eBay Enterprise. And one of the things that they wanted to do as part of providing capabilities to their retailer customers is improve the analytics that they provide. And so we entered into this partnership with them, where we took our campaign management platform and some of our analytics and did some additional work to tailor them to soothe eBay's needs. And eBay is now taking out our IP, our analytics IP and some of our campaign management IP out to their customers. And it's just -- it's a lovely partnership because they get best-in-class analytics and campaign management, and we get distribution to all kinds of retailers that we never ever would have gotten near.	0.1121	0.4443
4	I would characterize it as stabilization at a very high level, quite frankly, versus the rates that we have contracted [inaudible] fortunately. But as we see rates coming up for renewal, we don't see any moderating operation in price just yet. Say in oil field [inaudible] goods, cement, drilling fluids, probably year over year about a 10% rise in costs there. One thing to bear in mind - this is just a general observation of the industry - one thing to bear in mind, our projects, most of our costs have been contracted in and locked in so we're not	0.1475	0.0619

	experiencing that type of inflation just yet. Obviously, for new projects going forward is where we're going to [inaudible] that [picture] cost increase.		
4	Okay, now we actually haven't seen any changes in the competitive landscapes in the core markets that we've been involved in geographically.	0.1111	0.0246
5	Sure, Neal. Thank you for the question. At this point in time, we have not provided which exact assets, although I will tell you we are working multiple options across the portfolio. We don't see any one solution necessarily. We see several that are possible for us. I also think it's important to drive your attention to a few things, the strength of the portfolio and current commodity prices and our current ability to reinvest in the portfolio points us towards evaluating where is the lower EBITDA assets at present? What is the forecasted funding level that we see with this prolonged period of depressed prices? And how can we - is the asset idle or can we accelerate some activity there which would accelerate value? I personally am a fan of the JV structure if it's properly handled. And we would do exactly that. It builds underlying cash flow and accelerates value into the current term. But we also know that some of the assets, it may be better to just - to completely exit the asset just because we're not going to be investing there for some time. When we have stated that we would do these things in the past, that's exactly how we performed and I expect that we will be sharing more in the coming months on how we plan to advance these initiatives and we'll be providing more color as the opportunities mature, Neal.	0.2019	0.1502
5	Yes. We set a 10-year target of EUR 1 billion of revenue and 20-plus EBIT margin. That still stands. This is a target range for the next 3 years. Obviously, as you get localized fiber cement manufacturing and a few other things, we can continue to drive that margin to that 20% long-term target.	0.2973	0.0186
5	Jesse, it's Nate. Maybe just one thing to add to Kelly's comments is that when you think about a market that's in correction like commodities, kind of the risk reward people will tend to go short. And they'll become that much more dependent upon distribution in terms of end market services. So instead of buying trucks or railcars, it's maybe units and trucks. And so that still supports our distribution business incredibly well in terms of kind of moving through the process. So, we expect velocity to be good in part based on customers managing their risk and again, probably wanting to stay short at least for the time being.	0.1695	0.0148

Panel B: Transformation Wordlist Derived from Word Embedding Model

Q	Remark	TCS^{EMB}	TCS
0	Yes, on taxes, I mean, the effective tax rate for the quarter was about 39%, and you're right. It's more favorable this quarter than last quarter. A big variable that impacts the effective tax rate for us is the foreign exchange movement, because those really are balance sheet translation effects. They are not taxed, and so depending upon the size of the loss or the size of the gain that you have in any particular quarter, it obviously moves the effective tax rate. So I would just say I'm not -- that really is the primary driver between the quarters. The other thing that has occurred here is, as we earn more out of our Downstream segment, typically, it is coming from lower tax jurisdictions and so you get a mix effect.	0.0000	0.0077
0	We're feeling good about the business though and we're very focused on organic growth and one of the things I would just say is that our case is built on a very broad base. There are I mean we basically do not accept that any part of our business is mature. And so even within the longest standing solution sets we've got, we are asking for innovation and new forms of value for our customer. So we're looking in a case which is built very broadly across the entire company.	0.0000	0.1200
0	So for the growth investments, we were about $2.5 million in the first quarter. And I'm predicting in that range of $7 million for the full year.	0.0000	0.0358
1	I think you can see from the public announcements that's made in relation to some of the financing that they've achieved on the base station side with key vendors that they - and the statements they've made about their ambitions to expand their network that Sprint represents a significant opportunity for several years in this market. We expect historically our activity with Sprint and with ClearWire who's now part of - the network is now part of Sprint has seen that there is a fairly high percentage of the backhaul is microwave and we certainly believe that the need to go to least fiber, which was a factor in Sprint in the past is probably less strong in the future as those contracts expire or probably less strong than now the SoftBank have ownership of Sprint. So we see that the opportunity in Sprint is significant from their public statements and multi-year in nature. We have a very strong position in their historical network and we believe that we have, as I've said many times on this call, leadership products in the area of high capacity spectrally efficient radios for microcell applications. So it's certainly in our view a huge target for us and whilst I cannot go into any specific details about the relationship, I can tell you that we've had a relationship with Sprint historically and we expect to continue that relationship in the future.	0.0046	0.5650
1	We hedge forward for a very limited amount of time as Ian mentioned. We hedge forward for a few months, a fair percentage of our needs, and basically that combined with our FIFO counting and the flow-through of lead through our inventory aligns our pricing changes with the cost of lead. Now, it took some time to get that balance, and what we said in this was this was the first quarter where we had relatively close balance between pricing and cost flowing through the P&L, because as lead was spiking up so quickly last year, even	0.0045	0.0238

	variable pricing took time to adjust and flow through, so Q4 was particularly painful before all the pricing hit. Q1 was better as lead stabilized, and then pricing caught up with it so, now we have had lead movement downward, so obviously there are some transient effects there that may help, but what we wanted to be clear with our shareholders and folks on the call was we view lead fluctuations as transient. There may be some transient benefits or penalties, but with lead coming down, there are some benefits. Any hedges you haven't placed when lead is coming down hurts you obviously because you've locked in your cost basically, but again our objective is to align those costs with our pricing.		
1	Yes. I think this year, we are contemplating 50% plus revenue growth for Venmo on top of what it did this year. Obviously, they are continuing to add incremental services. Eventually, you are going to see Venmo have a lot of the capabilities that the PayPal super app has because that consumer base loves Venmo, or wants to live more and more of their financial life on the app. We have over 83 million people using Venmo right now in the United States. Think about that as a percentage of the population. Yes, it's one like out of every 3.5 people in the U.S. is using Venmo right now. We obviously have plans to take that overseas. We are going to be rolling out Venmo with Amazon, with Starbucks, with DoorDash, I mean there are quite a number of very high-profile merchants that are going to be implementing pay with Venmo on top of all of the things that they did last year. And so we have got a lot of promise yet to go, as I mentioned. I feel like we are in the beginning stages of the monetization progress, team is executing extremely well against both their objectives, but really importantly, against the roadmap as well. And so I think you just continue to see Venmo grow. And by the way, when it grows, obviously, its ARPU grows and as I mentioned in my remarks, it starts to add to our ability to grow take rates certainly that adds to an upward pressure on take rate as opposed to a downward one. I think that was our final question. I do want to thank everybody for all of your great questions. We really appreciate it. Realize there is a lot of information that we put out today. We want to thank you for your time as well. And we look forward to speaking with all of you soon, and meeting with you in person. And again, thank you for your time. Take care, and bye-bye.	0.0028	0.0942
2	Yes. Crude sourcing is key for us. Being a waterborne, the crude mix is a monthly question for us and ongoing optimization. We really worked and improved on the crude sourcing in the past few months. With more people out there handling directly supply with the suppliers, and developing those relationships, and trying more and more grades in our refinery. The second quarter has been a simple case of low price ANS that we were happy to take. And we build the rest from optimization availability. Going into the third quarter, this is really when we see our crude sourcing improvement getting accretive, and being able to overcome the shortage, the seasonal shortage in ANS availability in third quarter. So, very happy with crude sourcing. Will continue to be a key focus for our business, and drive our numbers.	0.0059	0.0246
2	The markets in which we operate in, especially like the ninemajor markets, we feel like we've got very strong opportunities to continue togo deeper. The majority of our focus ties to that and looking other things thatwould prompt to us enter a new market beyond good diversity, good economics,opportunities over time, stronger barriers to entry, things like that isfrankly, just a big enough beachhead for us to establish our own operatingteam. And more often than not that's a tough thing to achieve. Andwhat we're more focused on doing is just going deeper into either existingmarkets or sub-markets, and we feel like we've got great opportunities tocontinue to do that.	0.0073	0.4542
2	Well it depends on how many customers are out there but lets say you have a company with 20 employees and that's probably the size that can justify 1,000 to 2,000, you're probably going to have somewhere in the neighborhood of around 600 a month for their phone service, maybe even slightly higher. They're also going to need probably at least a T1 line which is anywhere from another four to $700 a month. Now you're up to 1,000 right there. Then the hosting of their web services can be anywhere from another $30 a month to $100 a month depending on the amount of traffic they've got going over. Now you're up to 11 or 1,200 a month. Then you add in SEO which is a constant thing. It is like the yellow pages. SEO is required all the time. As algorithms change you have to change the search engine optimization. That can anywhere from 500,000 a month. We like holding got that and then you add all these other services like operator services, live operator services I might add, bridge, conference bridge, video conferencing and so on, its very to see how a customer with 15 or 20 employees can be up to 1,000 or two a month. And that's a lot less than one operator would cost. A single operator answering the phone with all the benefits and insurance and so on can run 30 to $40,000 a year.	0.0073	0.0105
3	Well, the change in backlog is -- just on a revenue base, in backlog is -- obviously, we see that, and it does tell a different story because you're not seeing -- the challenge that we all face, and we've talked about it prior, is that with equity accounting, we're taking out revenue dollars in backlog and they're flowing in on the equity side in the P&L, so you don't see the 1:1 correlation between change in backlog and what's going into the P&L. And believe me, that's why we spent a lot of time last year describing how to be -- how to think about it in terms of the duration of the project, the normal S curve, so that as you could look at what is really happening in the backlog, that you could get a sense of what was coming out on Ichthys. And we've stayed away from commenting specifically on any one project from a revenue side or a backlog adjustment, and so that's why we continue to refer folks who have questions -- because it is a gap and how you're looking at your model -- to think more analytically of the normal distribution and percent complete times the backlog that we booked.	0.0112	0.0168

3	Okay, I hope that answers that question and I'll take the strategic inventory. As Olaf already pointed out, it's in two areas. One is natural products, I'll come to that a bit later, and the other one is sunscreen products. The sunscreen filters, for those who are with us since a longer time, this is a normal effect and it's a seasonal effect which occurs each year. But the other one, which is even higher this time, is natural strategic inventories. Just to name some areas, it's onion, carrot, acerola, vanilla and vegetables, mainly the Diana business. So it clearly shows that obviously, with the Diana acquisition, we did the best acquisition in the market as the strong trend towards natural solutions obviously is a strong supporter of our strong organic growth, and the high inventory shows our confidence in that business. Okay?	0.0115	0.1282
3	Yes, I mean I think the Abbott folks have plan, its aggressive for this fiscal year consistent with our goals of double-digit growth. And I think expectations are that they're going to hit their numbers and maybe do better than what they planned. We are clearly working with them on a quarterly basis where at the end of every quarter Rick is going to have a meeting with the Abbott folks to review what worked and what didn't work, and what changes if any we'd need to make. And I'm optimistic that this Abbott thing is going to work. We're keeping close to Abbott Centre.	0.0092	0.0247
4	No we were not getting any ounces from Lower West at the end of Q2, most of it was coming through lower grade material, I mentioned the transition some open-pit to under ground. So you're slowing down the open-pit at the expense of increasing the development level at the under ground and portion of that was also result of the inflationary pressures in the country. Ludovico, I don't know if you wanted to supplement that with anything else.	0.0154	0.0335
4	Yes. Just to take the second piece first, the structural piece is about 15% of the total approximately $55 million we talked about. The structural piece, which is head count, that does have annual savings related than what we realized in 2020. As far as the regional split of the cost savings, for the most part, we're about 75% of that's going to be in the Americas, and then about 20% in Europe, and the other 5% in APMEA. Rough split. We do get a little bit more savings in the Americas versus Europe.	0.0187	0.2047
4	Vinay, I mean, I would say, if not all of it, then most of it is actually model change, because the whole point of our version 11 is telling you, for the same exposure, the same dollars at risk, you're likely to lose more money. That's the message of it in a nutshell. So, naturally, you should expect everyone's loss expectations to go up, if they are using RMS 11 and they were previously using 10 and they are maintaining their exposures, if it's not going up it's because of reduced exposure.	0.0122	0.0084
5	No fully metal changes again that's very much - the ground if you're look at timing of revenues and finally it's all about the course of the year. So, same methodology though.	0.0321	0.0027
5	Well, you know I'll let our results speak for the question of how all our agencies are performing. And obviously with these kinds of results we're not relying on one particular agency. And as you know well we're there with - from the beginning, this has been a journey in terms ore repositioning our people and our organizations. And I said this publically and privately, I'd never been more comfortable with a leadership of our organizations throughout all of our agencies and I am right now. Our most recent repositioning involve putting Mullen and low together to form the MullenLowe Group and I am very pleased with how that's come out of the box in terms of new business wins and servicing at our existing clients. So when you combine three global networks of MullenLowe, McCann and FCB, on top of all the independent agencies we have in the U.S. you know some of them are not performing where we would like them to be and of course MullenLowe has a ways to go before they are achieving levels of McCann or SCV and frankly that's the opportunity we have at IPG are media offerings to best in class, they are recognize obviously UM in terms of its agencies, the year awards and its positive results of new business. Weber Shandwick and Golin being the top of the charts in terms of winning and gaining market share and certainly Weber winning all the recognition awards, our sports marketing and our experiential, so I am very pleased with all of our offerings that doesn't need to say we can't refine that in terms of adding additional talent and go to market strategies and acquisitions. And given our financial position, we are in a position to do that. So I am feeling - where we told the team is - you know we are fighting any football, you know this is still a journey, we still have margin improvement that's available to us and that's what we're focusing on, but in the meantime, I have no doubt that if there is an opportunity for us to pitch business, we will put together a world-class opportunity in team in place and we will continue to focus on our existing client base, which is among the best in the world. So other than that I am pretty - I am feeling pretty comfortable about where we stand.	0.0212	0.4620
5	Jeff, in addition to that probably the - follow on question is the synergy question that a lot of people ask. As we continue our integration planning we are increasingly confidents that we will meet were feed our synergy targets that potential to actually realize them a lot earlier than planned.	0.0210	0.4927

This table presents in Panel A and B three randomly sampled remarks drawn from each quintile of the non-zero region of the TCS^{IDX} and TCS^{EMB} distribution, respectively. In addition, three sentences are drawn from the large set of remarks with a score of 0. Occurrences of wordlist entries are colour-coded in grey and reported alongside the sampling quintile, the remark's score, and the main TCS score which is computed over the respective remark instead of the entire transcript.

OA Tab. 8 Transformation Wordlists

Panel A: Transformation Wordlist Derived from Textbook Indices

Entry	w	rf	Entry	w	rf	Entry	w	rf
growth	3	0.7319	company	1	0.7189	utilization	1	0.1218
target	3	0.4263	people	1	0.6646	excess	1	0.1204
metric	3	0.1602	plan	1	0.6422	uncertainty	1	0.1156
leadership	3	0.0886	revenue	1	0.5845	synergy	1	0.1116
stakeholder	3	0.0215	process	1	0.5596	monitor	1	0.1099
merger_acquisition	3	0.0039	benefit	1	0.5142	innovation	1	0.1036
cost	2	0.7370	deal	1	0.5009	finance	1	0.1034
focus	2	0.6505	share	1	0.4892	split	1	0.0985
customer	2	0.6175	activity	1	0.4461	free_cash_flow	1	0.0961
value	2	0.4820	strategy	1	0.4372	inflation	1	0.0953
cash	2	0.3998	volume	1	0.4366	cost_structure	1	0.0867
acquisition	2	0.3972	trend	1	0.4313	insurance	1	0.0865
asset	2	0.3779	buy	1	0.4294	sale_force	1	0.0849
play	2	0.3775	demand	1	0.4175	engineering	1	0.0844
management	2	0.3372	set	1	0.4128	consolidation	1	0.0842
control	2	0.2640	model	1	0.3915	acceleration	1	0.0827
cash_flow	2	0.2555	return	1	0.3873	clarity	1	0.0809
success	2	0.2331	support	1	0.3858	advertising	1	0.0785
market_share	2	0.1881	technology	1	0.3592	franchise	1	0.0751
bank	2	0.1586	contract	1	0.3419	phone	1	0.0727
commitment	2	0.1466	follow	1	0.3407	communication	1	0.0723
revenue_growth	2	0.1164	approach	1	0.3222	act	1	0.0719
business_model	2	0.1083	report	1	0.3197	owner	1	0.0718
firm	2	0.1059	job	1	0.3122	selling	1	0.0712
value_proposition	2	0.0656	decline	1	0.3007	implementation	1	0.0711
competitive_advantage	2	0.0411	state	1	0.2893	training	1	0.0699
customer_need	2	0.0409	inventory	1	0.2886	knowledge	1	0.0660
election	2	0.0312	transaction	1	0.2666	disruption	1	0.0659
diversity	2	0.0265	brand	1	0.2485	wind	1	0.0629
value_creation	2	0.0223	supply	1	0.2397	core_business	1	0.0619
digital_transformation	2	0.0098	debt	1	0.2376	overtime	1	0.0603
market_development	2	0.0088	wait	1	0.2312	maturity	1	0.0598
accountability	2	0.0072	purchase	1	0.2275	taxis	1	0.0568
return_asset	2	0.0065	option	1	0.2199	customer_want	1	0.0556
climate_change	2	0.0035	shareholder	1	0.2188	consumption	1	0.0555
digital_technology	2	0.0034	credit	1	0.2149	truck	1	0.0538
adaptation	2	0.0033	tax	1	0.2103	belief	1	0.0524
monopoly	2	0.0027	distribution	1	0.2042	complexity	1	0.0520
restructuring_plan	2	0.0026	gross_margin	1	0.2003	commerce	1	0.0513
ethic	2	0.0013	access	1	0.1959	turnaround	1	0.0492
deregulation	2	0.0012	competition	1	0.1894	transformation	1	0.0467
executive_compensation	2	0.0008	power	1	0.1705	culture	1	0.0454
firm_size	2	0.0002	story	1	0.1704	automation	1	0.0411
business_model_innovation	2	0.0002	offering	1	0.1702	section	1	0.0411
fall_angel	2	0.0001	assumption	1	0.1634	differentiation	1	0.0409
top_fee	2	0.0001	profit	1	0.1543	crisis	1	0.0403
creative_destruction	2	0.0001	priority	1	0.1459	sale_growth	1	0.0402
freemium_business_model	2	0.0000	ratio	1	0.1343	business_unit	1	0.0394
management_walk	2	0.0000	premium	1	0.1318	create_value	1	0.0394
change	1	0.7464	productivity	1	0.1260	capital_structure	1	0.0392

Panel B: Transformation Wordlist Derived from Word Embedding Models

Wordlist entry	w	Wordlist entry	w	Wordlist entry	w
reorganizational	0.8829	rationalization_modernization	0.5518	simplify	0.5136
reorganization	0.8489	de_layering	0.5513	centralization	0.5129
transformation	0.8055	delayer	0.5511	reallocate_resource	0.5127

restructure	0.7870	restructuring_charge	0.5496	evolution_revolution	0.5111
organizational_change	0.7802	rewiring	0.5495	redesign	0.5102
restructuring	0.7646	streamline_simplify	0.5485	decentralization	0.5093
reorganize	0.7528	integration	0.5464	reformation	0.5093
organizational_realignment	0.7317	reintegration	0.5433	transition	0.5087
change_management	0.7159	digitization	0.5415	platforming	0.5071
realignment	0.7038	reposition	0.5404	rescaling	0.5062
restructuration	0.6965	leadership_team	0.5377	eliminate_redundant	0.5058
cultural_transformation	0.6932	branch_rationalization	0.5370	duplication	0.5040
organizational_structure	0.6896	reallocation_resource	0.5348	initiative	0.5035
reorganizing	0.6866	rebranding	0.5339	reinvigoration	0.5032
organizationally	0.6865	merger	0.5337	sap_implementation	0.5023
organizational_redesign	0.6787	org	0.5335	eliminate_redundancy	0.5017
transform	0.6730	functionalize	0.5329	reorientation	0.5011
organizational_alignment	0.6721	consolidation	0.5310	reassignment	0.5010
organizational_effectiveness	0.6678	functional_excellence	0.5277	automation_digitization	0.4998
operationalization	0.6460	digitally_transform	0.5273	streamline_supply_chain	0.4974
streamlining	0.6395	administrative_function	0.5271	internationalization	0.4943
organizational	0.6357	reengineere	0.5266	replatforming	0.4930
simplification	0.6348	delayere	0.5263	evolutionary_revolutionary	0.4927
refocus	0.6320	executional_excellence	0.5258	retool	0.4924
reshape	0.6274	recapitalization	0.5256	digitize	0.4919
organizational_setup	0.6273	downsizing	0.5250	talented_executive	0.4913
realign	0.6187	resizing	0.5250	automation_digitalization	0.4913
delayering	0.6152	regionalization	0.5246	transformative_medicine	0.4903
transformational	0.6124	reconceptualization	0.5243	reconsolidation	0.4902
digital_transformation	0.6029	institutionalization	0.5236	retooling	0.4896
reengineering	0.6024	rightsize	0.5235	decentralize	0.4883
headcount_reduction	0.5984	operational_effectiveness	0.5235	reprioritization	0.4874
refocusing	0.5976	org_structure	0.5218	heavy_lifting	0.4873
streamline	0.5962	organize	0.5205	revitalize	0.4872
reorg	0.5950	simplify_streamline	0.5203	mutualization	0.4871
reshaping	0.5891	structural_reform	0.5197	focusing	0.4866
rightsizing	0.5890	coststructure	0.5194	rethinking	0.4864
reinvention	0.5885	separation	0.5189	seasoned_executive	0.4862
restructuring_et_cetera	0.5871	digitalization	0.5188	globalization	0.4861
professionalization	0.5851	reconfiguration	0.5187	senior_management	0.4859
organization	0.5850	operational_excellence	0.5187	erp_implementation	0.4843
streamline_operation	0.5731	span_layer	0.5183	embark	0.4837
cost_cutting	0.5730	simplifying	0.5177	rewire	0.4834
organization_et_cetera	0.5713	consolidating	0.5175	leadership	0.4827
rationalization	0.5604	professionalize	0.5173	senior_leadership	0.4818
simplification_modernization	0.5594	distraction	0.5171	verticalization	0.4816
repositioning	0.5592	resize	0.5160	chief_operating	0.4804
transformative	0.5573	undertake	0.5157	merge	0.4804
realigning	0.5545	revitalization	0.5147	revamping	0.4797
reinvent	0.5534	restructure_balance_sheet	0.5137	rethink	0.4796

This table presents in Panel A and B the top 150 wordlist entries for the textbook index-based and embedding model-based transformation vocabulary, respectively. Panel A is sorted by the number of books which list the entry in their index (w) as well as the relative document frequency (rf), i.e., share of transcripts in which they occur at least once. Panel B sorted by the entries' cosine similarity with the average seed embedding (w). In both cases, the weighting terms reweight word counts based on the presumed representativeness of each token for our focal construct.

OA Tab. 9 Out-of-Sample Performance of Alternative Measurment Approaches

Performance Metric	Contextual Classifier	Wordlist from Textbook Indices	Wordlist from Word Embedding Model
ROC-AUC	0.8983	–	–
Accuracy	0.8360	0.4440	0.7997
Precision	0.5152	0.2209	0.3664
Recall	0.7907	0.8837	0.2050
F1	0.6239	0.3535	0.2629

This table presents selected metrics showcasing the classification performance of our contextual classification model (column (1)) vis-a-vis the two alternative wordlist approaches (columns (2) and (3)). Metrics are computed on 500 sentences that are used for training our human coders but are unknown to the model (i.e., out-of-sample). *ROC-AUC* (receiver operating characteristic under the curve) is a prominent measure of model performance in the binary classification case (e.g., $T = 0$ and $T = 1$). It trades-off the amount of true positive and false positive predictions at all possible values of $P(T|s)$. A value of 1 (0.5) reflects a perfect (uniformative) classifier. *Accuracy* gives the percentage of correctly classified sentences (true positives plus true negatives divded by the sample size). *Precision* gives the ratio of true positives to the sum of true and false positives (i.e., the proportion of correctly predicted transformation statements). *Recall* gives the ratio of true positives to the sum of true positives and false negatives (i.e., proportion of identified transformation statements). *F1 score* gives the harmonic mean of precision and recall. The calculcation of the ROC-AUC is based on the $P(T|s)$. In contrast, we assume $T = 1$ ($T = 1$) if $P(T|s) \geq 0.5$ ($P(T|s) < 0.5$) for the latter four metrics. Because the wordlist approaches do not return classification probabilities by design, we cannot report any statistic for the ROC-AUC.

OA Tab. 10 Comparative Validation Tests of Transformational Change Scores

Panel A: Convergent validity (composite metrics)

	SV^{4D}	SV^{6D}	ED^{BF}	ED^{ATF}
TC	0.0132***	0.0080***	0.0075***	0.0034***
	(4.04)	(2.67)	(8.14)	(4.66)
TC^{IDX}	0.0116***	0.0073**	0.0041***	0.0016**
	(3.32)	(2.21)	(4.29)	(2.23)
TC^{EMB}	0.0066**	0.0066**	0.0045***	0.0015**
	(2.51)	(2.43)	(5.51)	(2.54)

Panel B: Concurrent validity (stand-alone metrics)

	PPE	R&D	SG&A	Financial Leverage	Employees	Revenue Conc.	Change in Goodwill	Acquisition Cost
TC	0.0242***	0.0114**	0.0137***	0.0041***	0.0218***	0.0035**	0.1330***	0.0071***
	(3.00)	(1.97)	(4.64)	(6.30)	(9.94)	(2.17)	(7.99)	(12.88)
TC^{IDX}	-0.0022	0.0007	0.0041	0.0033***	0.0040*	0.0033*	0.0334**	0.0031***
	(-0.24)	(0.11)	(1.16)	(4.84)	(1.75)	(1.91)	(2.15)	(5.46)
TC^{EMB}	-0.0047	-0.0053	-0.0002	0.0019***	0.0042**	-0.0023	0.0325***	-0.0001
	(-0.72)	(-1.20)	(-0.07)	(3.41)	(2.20)	(-1.50)	(2.73)	(-0.23)

(1) Property, Plant & Equipment (PPE), Research & Development (R&D), Sales, General & Administrative (SG&A), Financial Leverage (LEV), Employees,

Panel C: Predictive validity

	PPE	R&D	SG&A	Financial Leverage	Change in Goodwill
TC_{-1}	0.0068***	0.0064***	0.0042***	0.0015***	0.0164***
	(7.43)	(2.91)	(4.33)	(6.78)	(10.93)
TC_{-2}	0.0021**	0.0003	0.0010	0.0005**	0.0040***
	(2.13)	(0.15)	(1.12)	(2.18)	(2.69)
TC_{-3}	-0.0003	0.0028	-0.0007	0.0001	-0.0005
	(-0.33)	(1.09)	(-0.72)	(0.51)	(-0.32)
TC_{-4}	0.0001	-0.0001	-0.0010	-0.0005**	-0.0024*
	(0.16)	(-0.02)	(-1.13)	(-2.38)	(-1.68)
TC_{-1}^{IDX}	0.0038***	0.0016	0.0014	0.0011***	0.0088***
	(4.24)	(0.68)	(1.43)	(5.16)	(5.84)
TC_{-2}^{IDX}	-0.0008	0.0024	0.0003	0.0003	0.0037**
	(-0.84)	(1.07)	(0.35)	(1.29)	(2.48)
TC_{-3}^{IDX}	-0.0008	0.0024	0.0004	0.0003	0.0011
	(-0.86)	(1.05)	(0.37)	(1.26)	(0.72)
TC_{-4}^{IDX}	-0.0013	-0.0001	0.0005	-0.0003	-0.0008
	(-1.50)	(-0.04)	(0.55)	(-1.63)	(-0.57)
TC_{-1}^{EMB}	0.0033***	0.0025	0.0024***	0.0003*	0.0036***
	(3.89)	(1.34)	(2.83)	(1.87)	(2.95)
TC_{-2}^{EMB}	-0.0001	-0.0002	-0.0005	0.0002	0.0007
	(-0.10)	(-0.08)	(-0.68)	(1.00)	(0.57)
TC_{-3}^{EMB}	0.0002	0.0016	0.0001	0.0001	-0.0005
	(0.30)	(0.88)	(0.09)	(0.41)	(-0.41)
TC_{-4}^{EMB}	0.0004	-0.0003	0.0003	0.0001	0.0007
	(0.45)	(-0.18)	(0.47)	(0.41)	(0.61)

This table presents the results of our comparative validation tests. We rerun the regression specifications introduced in section 5.4 of the paper and report the coefficients (t statistics in parentheses) for TC, TC^{IDX}, and TC^{EMB} as main independent variable, respectively. Panel A, B, and C list the results for the convergent validity tests (i.e., associations of the TC scores with the related composite proxies), concurrent validity tests (i.e., associations of the TC scores with absolute changes in the stand-alone metrics), and predictive validity test (i.e., associations of the double-lagged TC scores with absolute changes in the quarterly stand-alone metrics). TC scores are standardizes to ensure comparability. Each regression specification includes year- and firm-fixed effects and clusters standard errors per firm. Statistical significance levels are defined as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Alammar, J. (2018a). The Illustrated Transformer. https://jalammar.github.io/illustrated-transformer/.

Alammar, J. (2018b). The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning). http://jalammar.github.io/illustrated-bert/.

Allee, K. D., & DeAngelis, M. D. (2015). The Structure of Voluntary Disclosure Narratives: Evidence from Tone Dispersion. *Journal of Accounting Research, 53*, 241–274. doi: https://doi.org/10.1111/1475-679X.12072.

Artstein, R., & Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics.

Bentley, J. W. (2021). Improving the Statistical Power and Reliability of Research Using Amazon Mechanical Turk. *Accounting Horizons, 35*, 45–62. doi: https://doi.org/10.2308/HORIZONS-18-052.

Bernstein, M. (2022). Labeling and Crowdsourcing. https://datacentricai.org/labeling-and-crowdsourcing/. Accessed 28 February 2022.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics, 5*, 135–146. doi: https://doi.org/10.1162/tacl_a_00051.

Brown, S. V., & Tucker, J. W. (2011). Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications. *Journal of Accounting Research, 49*, 309–346. doi: https://doi.org/10.1111/j.1475-679X.2010.00396.x.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. http://arxiv.org/pdf/1810.04805v2.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebel, J. (2020). Hedging Climate Change News. *The Review of Financial Studies, 33*, 1184–1216. doi: https://doi.org/10.1093/rfs/hhz072.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as Data. *Journal of Economic Literature, 57*, 535–574. doi: https://doi.org/10.1257/jel.20181020.

Hoberg, G., & Phillips, G. (2016). Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*. doi: https://doi.org/10.3386/w15991.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016a). *FastText.zip: Compressing text classification models*. http://arxiv.org/pdf/1612.03651v1.

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016b). *Bag of Tricks for Efficient Text Classification*. http://arxiv.org/pdf/1607.01759v3.

Kelly, B., Papanikolaou, D., Seru, A., & Taddy, M. (2018). *Measuring Technological Innovation over the Long Run*. Cambridge, MA.

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology*. Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE.

Lee, J. (2016). Can Investors Detect Managers' Lack of Spontaneity? Adherence to Predetermined Scripts during Earnings Conference Calls. *The Accounting Review, 91*(1), 229–250.

Li, K., Liu, X., Mai, F., & Zhang, T. (2021a). The Role of Corporate Culture in Bad Times: Evidence from the COVID-19 Pandemic. *Journal of Financial and Quantitative Analysis*, 1–68. doi: https://doi.org/10.1017/S0022109021000326.

Li, K., Mai, F., Shen, R., & Yan, X. (2021b). Measuring Corporate Culture Using Machine Learning. *The Review of Financial Studies, 34*, 3265–3315. doi: https://doi.org/10.1093/rfs/hhaa079.

Liu, Y., Ott, M., Goyal, N., Du Jingfei, Joshi, M., Chen, D., et al. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. http://arxiv.org/pdf/1907.11692v1.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). *Efficient Estimation of Word Representations in Vector Space*. http://arxiv.org/pdf/1301.3781v3.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013b). Distributed Representations of Words and Phrases and their Compositionality.

Mikolov, T., Yih, W., & Zweig, G. (2013c). Linguistic Regularities in Continuous Space Word Representations.

Muller, M., Wolf, C. T., Andres, J., Desmond, M., Joshi, N. N., Ashktorab, Z., et al. (2021). Designing Ground Truth and the Social Life of Labels, 1–16. doi: https://doi.org/10.1145/3411764.3445402.

Peterson, K., Schmardebeck, R., & Wilks, T. J. (2015). The Earnings Quality and Information Processing Effects of Accounting Consistency. *The Accounting Review, 90*, 2483–2514. doi: https://doi.org/10.2308/accr-51048.

Rong, X. (2014). *word2vec Parameter Learning Explained*. http://arxiv.org/pdf/1411.2738v4.

Rush, A. M., Nguyen, V., & Klein, G. (2018). The Annotated Transformer. http://nlp.seas.harvard.edu/2018/04/03/attention.html. Accessed 25 March 2021.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. http://arxiv.org/pdf/1910.01108v4.

Sheng, V. S., Provost, F., & Ipeirotis, P. G. (2008). Proceedings of the 14th ACMKDD International Conference on Knowledge Discovery & Data Mining: Las Vegas, NV, USA, August 24 - 27, 2008.

Song, S. (2021). The Informational Value of Segment Data Disaggregated by Underlying Industry: Evidence from the Textual Features of Business Descriptions. *The Accounting Review, 96*, 361–396. doi: https://doi.org/10.2308/TAR-2017-0572.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is All you Need.