DISCUSSION OF

10-K Disclosure Repetition and Managerial Reporting Incentives
What Have We Learned and Where Do We Go with Textual Research?

Travis A. Dyer
Mark H. Lang
The University of North Carolina at Chapel Hill

Lorien Stice-Lawrence
University of Southern California

ABSTRACT: We discuss the evolution of research on textual attributes in accounting and suggest potential directions for future research using Cazier and Pfeiffer (2017) to illustrate the challenges and opportunities in the literature. We categorize the literature based on whether a given textual attribute is assumed to be “discretionary” versus “nondiscretionary” and whether it is assumed to be “helpful” versus “harmful.” The literature to date has been largely descriptive, with limited focus on causality, and we argue that there are substantial opportunities for research that better isolate the discretionary components of textual disclosure, as well as identifying specific contexts in which a given attribute is more likely to be helpful than harmful.

Keywords: textual analysis; disclosure effectiveness; narrative disclosure.

JEL Classifications: D8; M4.

INTRODUCTION

Cazier and Pfeiffer (2017; hereafter, CP) document associations between measures of redundancy in 10-K text, firm performance, and litigation risk, which they argue reflect incentives by managers to use redundancy to obfuscate poor performance, especially in contexts with high levels of litigation risk. They also present evidence suggesting that repetition may be linked to slower price discovery for Form 10-Ks in some settings.

Our charge in this discussion was to present (1) the related literature more generally, (2) how this paper fits in, and (3) opportunities for future research. As a result, we provide a broad perspective on the literature, using CP as an example to illustrate specific challenges and opportunities for future research. In the interest of parsimony, we focus our discussion on papers evaluating textual attributes of Form 10-Ks and related text, and cite specific papers as examples rather than attempting a thorough review of the literature.

CHARACTERIZATION OF THE TEXTUAL ANALYSIS LITERATURE

Textual analysis research is relatively new to the accounting literature. While the advent of modern financial accounting research is often tied to Ball and Brown (1968), textual analysis in accounting lags by roughly 40 years, coming to prominence only within the last decade, beginning with research such as Li (2008). In fact, one might argue that Ball and Brown (1968) were to accounting numbers what Li (2008) was to the associated text. The two papers are analogous in several ways. First,
both were driven by the availability of large, electronically accessible datasets—CRSP and Compustat in the case of Ball and Brown (1968) and the SEC’s EDGAR database of corporate filings in the case of Li (2008). Second, neither study was truly the “first” in the sense that both had been preceded by other, typically smaller-sample, studies, but both were careful, large-sample analyses that anchored substantial literatures. Third, both were relatively descriptive in providing initial evidence of correlations between measures derived from accounting information and variables of interest to a range of constituencies.

Following Li (2008), the literature has grown quickly and broadened substantially in at least two ways. First, the range of textual measures has increased dramatically from the Gunning Fog Index and length in Li (2008) to include, among others, stickiness, repetition, file size, cosine similarity, specificity, salience, and a variety of word-count-related measures. In addition, the range of text considered has expanded to include such diverse financial documents as analyst reports (Huang, Lehavy, Zang, and Zheng 2017), conference calls (Allee and DeAngelis 2015), press releases (Guay, Samuels, and Taylor 2016), news media (Kothari, Li, and Short 2009), and social media (Chen, De, Hu, and Hwang 2014).

Reflecting its early stage, the literature remains largely descriptive with limited emphasis on causality. In addition, authors have generated a large number of new measures and results without, for the most part, integrating them with findings from earlier studies. As a result, it is difficult to draw strong, causal conclusions or policy prescriptions. Further, because the studies often do not build on each other, there does not yet appear to be a cohesive consensus on the causes and consequences of the textual attributes.

HOW DOES THE CP PAPER FIT INTO THE EXISTING LITERATURE?

The CP paper is relatively representative of the current state of textual analysis in accounting and serves as a useful starting point to characterize the existing literature and potential directions for future research. One way to characterize existing research is as fitting loosely into one of four cells in a 2 x 2 matrix, as presented in Figure 1. On the horizontal axis, papers differ based on whether the textual attribute of interest is assumed to be primarily “discretionary” or “nondiscretionary.” For example, if the complexity of text in the 10-K increases after a merger, then do the authors assume that the increased complexity primarily reflects a strategic choice on the part of managers to increase opacity, or does it reflect a change in the firm’s economic circumstances that naturally leads to longer sentences with more complex words? The distinction is important because the interpretation of the results, as well as the associated policy implications, hinges on the extent to which the textual attributes are assumed to reflect managerial discretion.

While we use the discretionary/nondiscretionary split to organize our discussion, we recognize that it is often difficult to disentangle the two in practice. What we have in mind is the following thought exercise. Imagine a machine that takes the economic events for the period along with the filing requirements and generates text much like the news-writing “bots” in journalism. If events such as a complex merger or a new FASB pronouncement would naturally result in a document that is longer and less readable, we would refer to that as “nondiscretionary.” If, on the other hand, managers with particular incentives use their discretion to decrease the readability of the document and increase opacity, then we would refer to that as “discretionary.” A good example of the distinction is the argument in Li (2008) that managers reporting losses choose

FIGURE 1

A Characterization of Textual Analysis Research

<table>
<thead>
<tr>
<th>Discretionary</th>
<th>Nondiscretionary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Helpful</strong></td>
<td></td>
</tr>
<tr>
<td>Few papers</td>
<td>Lang &amp; Stice-Lawrence (2015)</td>
</tr>
<tr>
<td><strong>Harmful</strong></td>
<td></td>
</tr>
<tr>
<td>Cazier &amp; Pfeiffer (2017)</td>
<td>Few papers</td>
</tr>
</tbody>
</table>

1 As we discuss in more detail below, empirically operationalizing the distinction is challenging. Partially mitigating that difficulty is the fact that, in many studies such as CP, the analysis is estimated with firm-fixed effects, making identification somewhat easier.
Discussion of: 10-K Disclosure Repetition and Managerial Reporting Incentives

Discussion

longer 10-Ks to obfuscate (“discretionary”), as contrasted with the Bloomfield (2008) discussion of Li (2008), which argues that managers reporting losses are likely to be explaining more complex underlying economics that necessitate lengthier discussions (“nondiscretionary”).

The vertical axis in Figure 1, on the other hand, is based on whether the characteristics being considered are assumed to be “helpful” or “harmful” in a particular context. For example, in Li (2008) the underlying presumption is that increased 10-K length and complex language reduce information content (“harmful”). An alternative would be to argue that increased 10-K length and complex technical discussions are designed to increase the information content (“helpful”), which would fundamentally change the interpretation of the results.

Examples of papers are provided in each cell of the matrix in Figure 1 to help illustrate the categories. CP and Li (2008) fit in the lower left-hand (“discretionary/harmful”) cell as examples of papers that assume that the textual attribute being observed is “discretionary” from the standpoint of the manager and the manager is using that discretion to obfuscate (“harmful”). Arguably, most of the literature occupies that cell—managers are assumed to have substantial discretion over reporting and to use that discretion to reduce the informativeness of the resulting text.

At the other extreme is the upper right-hand (“nondiscretionary/helpful”) cell characterized by research such as Lang and Stice-Lawrence (2015). In that study, the authors examine trends in financial reporting globally, particularly focusing on IFRS adoption as a plausibly exogenous shock, and textual constructs such as harmonization and convergence of text. They provide evidence that trends in worldwide reporting requirements (“nondiscretionary”) have resulted in improvements in informative attributes of annual report text such as harmonization and convergence (“helpful”). The authors conclude that changes in global reporting requirements over time have likely increased the usefulness of textual reporting.

The lower right-hand (“nondiscretionary/harmful”) cell is illustrated by Dyer, Lang, and Stice-Lawrence (2017). The authors focus on regulatory changes in the U.S. (“nondiscretionary”) that have been associated with increases in textual attributes that prior literature suggests reduce informativeness of disclosure (“harmful”), including (among others) length, boilerplate, redundancy, stickiness, and Fog. They conclude that major regulatory changes in the U.S. appear to have resulted in attributes of text that increase informational processing costs for investors.

In the final, upper left-hand (“discretionary/helpful”) cell are papers that consider textual attributes chosen by managers (“discretionary”) that are designed to improve the informativeness of textual disclosure (“helpful”). Interestingly, this cell appears to have received the least attention in the literature. While not fitting neatly, a paper that illustrates the potential direction for research in this cell is Guay et al. (2016). It does not fit neatly into the cell because the portion of the paper related directly to 10-K text examines poor readability (“harmful”) caused by exogenous events such as regulatory changes and economic events within the firm (“nondiscretionary”). However, the main goal of the paper is to examine whether managers faced with reporting 10-K text that is naturally characterized by low readability (i.e., complex and lengthy text) respond by increasing voluntary disclosure, such as management earnings forecasts, to mitigate the unhelpful attributes of the 10-K text. An alternative that would have fit neatly into the “discretionary/helpful” cell would have been to examine whether managers respond to “harmful” attributes associated with “nondiscretionary” (e.g., SEC- or FASB-mandated) disclosure with discretionary text that is designed to help in clarifying the potentially confusing mandated disclosure (e.g., clear, concise and specific disclosure in sections of the 10-K over which they have more discretion, such as Management’s Discussion and Analysis).

While the preceding classification is admittedly crude, we would argue that authors of papers in the existing literature implicitly make assumptions that place their research into one of the four categories. Further, as we will argue below, a lack of specificity in terms of what is being assumed on the discretionary/nondiscretionary and helpful/harmful dimensions muddies the interpretation of the results and limits the contribution of the existing literature. In addition, the splits in Figure 1 help to directly identify opportunities for future research. For example, as noted above, most of the existing research focuses on managers using discretion to create opacity. There is much less research that speaks directly to contexts in which managers choose textual attributes (or associated tabular or graphical representations) to intentionally provide clarity in otherwise complicated contexts, which must also be important in practice. Similarly, there is relatively little research on determinants and consequences of nondiscretionary textual attributes, although in practice it seems likely that much of the textual content of the 10-K is in response to specific reporting requirements coupled with underlying economics.

Opportunities for Future Research

Discretionary versus Nondiscretionary

A major challenge in interpreting the existing literature on textual characteristics is in disentangling discretionary from nondiscretionary disclosure attributes. While authors typically have a maintained assumption about whether the attributes they observe are primarily discretionary or nondiscretionary, that split is often not fully developed empirically. For example, CP
implicitly assume that an observed correlation between repetition in 10-K text (particularly if it includes specific references to accounting standards) and variables such as firm performance reflects an attempt on the part of management to increase opacity (i.e., managers of poorly performing firms use textual redundancy to make it more difficult for investors to infer the firm’s underlying economics). However, analogizing from Bloomfield (2008), poor performance may be inherently more difficult to explain and may thus benefit from additional repetition. For example, managers who are trying to report transparently may remind readers of the accounting rules for impairments every time they discuss losses caused by asset write-downs, which could increase the informativeness of the resulting disclosure as well as the measured redundancy.

The notion that underlying economic factors are important determinants of the extent of observed repetition in practice seems clear from the fact that many of the control variables in the regression in Cazier and Pfeiffer (2017, Table 4) have significant explanatory power, suggesting that economic factors play a substantial role in explaining repetition. Relatedly, it is unclear whether many of the control variables are better thought of as reflecting economic variation that influences the nondiscretionary component of textual attributes (which is implicit in including them among the control variables) or as variables influencing incentives for discretionary disclosure attributes (suggesting that they should be primary variables of interest). For example, variables such as research and development, intangibles, special items, and volatility could as easily reflect underlying economic attributes that influence managers’ incentives to provide more informative disclosure.

Going forward, a significant challenge (and opportunity) for the literature will be in developing better approaches for separating discretionary from nondiscretionary textual attributes. In some sense, the situation is analogous to the research on “abnormal accruals” that began with simple expectation models, such as those in Jones (1991), and evolved into a substantial literature attempting to formally model discretionary and nondiscretionary accruals. Analogizing to the Jones (1991) model, the starting point would be to identify settings with comparable underlying economics but differing incentives. For example, Jones (1991) examines the effects of incentives to manage earnings around import relief investigations and uses the “normal” relation between accruals, changes in revenues, and property, plant and equipment to estimate the nondiscretionary portion. The subsequent literature has refined the model by thoughtfully identifying other factors that naturally drive the nondiscretionary components of accruals. One could imagine a very similar approach to formally identifying the likely nondiscretionary components of textual attributes.

To be effective, it would be important for such an approach to identify settings in which other economic factors are likely to be similar, but incentives are likely to differ. Earnings performance, as in CP, seems like a relatively poor choice of incentive variable because it is difficult to imagine contexts in which reported earnings capture incentives but not underlying economics. However, it may be possible to estimate the relation between incentives created by poor performance and disclosure attributes by using earnings discontinuities around, for example, zero to identify contexts in which incentives change sharply relative to underlying economics.

More generally, authors could use exogenous variation to help identify the causal effect. CP take a step in that direction by using litigation exposure from Kim and Skinner (2012) as a proxy for litigation-based incentives for redundancy. However, their measure, which uses economic variables to identify litigation risk (growth, return variability, turnover, etc.), could as easily be capturing differences in underlying economics. A more promising approach (which CP mention briefly) would be, instead, to identify incentives associated with litigation using shocks to the external likelihood of lawsuits based on changes in the litigation environment caused by, for example, precedent from recent court rulings or changes in the mix of local judges, which would permit a difference-in-differences design. Barring that, it would be possible to exploit cross-sectional variation in incentives across managers based on, for example, sensitivity of compensation to share price (if that is the mechanism the authors have in mind). However, that also highlights the importance of identifying the mechanism through which the authors believe the effect obtains (i.e., is the goal to temporarily keep bad news from being reflected in share price or is it to make it more difficult to attribute the observed poor performance to specific actions by the manager?).

Closely related, it would be helpful to explore the process by which the 10-K is prepared to understand the extent to which given textual attributes are likely to be discretionary versus nondiscretionary. For instance, Cazier and Pfeiffer (2017, Appendix A) provide examples of redundancy in financial statement footnotes. What would be the mechanism by which top management in a large corporation could influence redundancy in specific footnote text as a means to create opacity and obscure poor performance? Perhaps it would be possible to use the section in which the text appears, coupled with knowledge about the process by which the 10-K is prepared, to identify text that is more likely to be under the direct control of management. For example, McMullin (2016) uses information about auditor identities to infer that much of the variation one sees in footnote text is more likely the result of the specific auditor rather than a direct choice by management.

**Differentiating between Helpful and Harmful Textual Attributes**

The second dimension highlighted in Figure 1 is the split between whether a given set of textual attributes is assumed to be helpful versus harmful. Again, this is an area in which the existing literature is inconclusive, providing opportunities for future...
research. While authors often attempt to avoid taking an explicit stand on the issue, it seems central to motivating the literature. For example, CP are motivated with quotes suggesting that redundancy is harmful. However, it is easy to imagine situations in which repetition would be informative. More generally, for an attribute to be of interest to a broad constituency, it seems important to provide evidence linking that attribute to the informativeness of the resulting text.

A significant challenge in assessing whether a given textual attribute is likely to enhance the information content of text is that it will often depend both on the financial information user in question, as well as on the content of the text. For example, retail investors, lenders, and institutional investors all have different needs and capabilities. Further, the 10-K is a lengthy document including many types of information about differing aspects of firm fundamentals, some of which may have been preempted by previous disclosure, making it very difficult to disentangle the extent to which the manner in which a disclosure is written affects market participants beyond the actual content of the disclosure itself. This, combined with the endogeneity issues inherent in any firm choice, makes it very difficult to convincingly link textual attributes to outcomes in the context of the 10-K.

CP attempt to link repetition to informativeness by assessing the association between the level of repetition and the speed of price discovery over a relatively long return window following the release of the 10-K. However, it is unclear how much can be learned from that analysis. First, to the extent that repetition reflects underlying economics, it may be the underlying economics rather than the repetition that affect the speed of price discovery. Further, even taking the results at face value, it is unclear why a relatively small amount of repetition would result in a delay of up to 20 days in price discovery. Finally, it is difficult to reconcile the results in CP with those of Li (2017), which provide evidence that repetition can increase the overall informativeness of the annual report.

The reality is that disclosure is nuanced, and repetition (and other textual attributes) may be beneficial in some contexts and detrimental in others. Further, drawing inferences based on the 10-K seems particularly difficult given the length and wide range of content. That being said, this would seem to be an important context that could benefit from the creative use of multiple methods in more focused settings to attempt to reach consistent answers.

For example, laboratory studies offer the opportunity to manipulate specific textual attributes while holding constant the underlying economics. Of course, while laboratory studies have the benefit of internal validity, archival studies potentially have greater external validity. One approach to dealing with this issue is to combine archival and laboratory studies as in Bonsall, Leone, Miller, and Remnekamp (2017). A particularly clever feature of that study is their use of a laboratory setting to validate their measure of readability, which they then apply to prospectuses in an archival setting.

Going forward, it seems important to focus on settings in which the effect of a given textual attribute and the underlying mechanism can be more clearly identified. This could be an area particularly well suited to laboratory studies (perhaps combined with focused archival analyses) since it is difficult to imagine how one could tease out causal effects and mechanisms in a purely archival setting. Similarly, a laboratory setting provides a potential method for disentangling the effects of textual attributes on different user groups. For example, the effect of an attribute such as repetition could be compared between relatively sophisticated and unsophisticated users. Similarly, subjects could be compared making equity investing decisions relative to, say, lending decisions.

A related issue is whether the 10-K is a useful starting point to understand textual attributes, given its length, the extent to which it is driven by regulatory requirements, and the fact that much of the information it includes has likely been preempted by other disclosure. As noted by Bonsall et al. (2017), a more fruitful starting point might be to assess the effects of textual attributes on shorter, more-focused, disclosure that is more likely to convey new information to the market. Once the effects of textual attributes are better understood in tighter settings, the move to a broader document such as the 10-K would seem more interpretable.

A tightly controlled setting might also be useful in helping to winnow and synthesize the wide range of textual attributes developed in the literature. One of the factors that limits the progress of the literature is the scattershot approach to date. As a practical matter, it seems very unlikely that all of the measures that have been identified are independent or incrementally important. Further, there may be significant interactions among constructs (e.g., repetition may be more beneficial when readability is low). There is significant scope for studies that empirically synthesize the various attributes into, for example, factors and that identify when specific factors are more or less important in practice. Again, a controlled setting such as a laboratory study or focused archival research (along with, perhaps, insight from related disciplines such as linguistics) could be well suited to such an analysis.

2 CP estimate that the repetition in a typical 10-K ranges from 200 words to 350 words over the sample period, with a double-spaced page constituting about 300 words. It is unclear how one page of verbatim repetition in a lengthy 10-K would meaningfully delay price discovery for the marginal investor.
Finally, techniques such as latent Dirichlet allocation (LDA) could be exploited to identify topical content within a
document as a means for understanding the interaction between content and textual attributes. It might well be the case, for
example, that multi-syllabic words are useful in discussing financial instruments but not in describing the firm’s results for the
period. As discussed in more detail in Dyer et al. (2017), it is possible to use LDA to link textual attributes to specific content at
the paragraph level within a lengthy document such as a 10-K, potentially facilitating such an analysis.

CONCLUSION

Cazier and Pfeiffer (2017) present an interesting set of results on the determinants and outcomes of repetition in corporate
10-Ks. In this discussion we have attempted to use their research to highlight many important topical areas for which strong
inference from the existing literature is still lacking. Overall, we believe that the literature on textual attributes of disclosure is
still in its infancy, providing significant opportunities for future research.

REFERENCES


doi:10.2308/jfr-51912


2491047


Kothari, S. P., X. Li, and J. E. Short. 2009. The effect of disclosures by management, analysts, and business press on cost of capital, return


doi:10.1016/j.jacceco.2008.02.003


University Bloomington.