COMMUNICATION &LANGUAGE at work
Issue no. 3 | February 15th 2014
Social Media and/in Business Communication

The power of language in corporate financial reports

Darina M. Slattery

(pp. 53-63)

http://ojs.statsbiblioteket.dk/index.php/claw/article/view/16555

| Bridging Theory and Practice |
http://ojs.statsbiblioteket.dk/index.php/claw

The power of language in corporate financial reports

Darina M. Slattery
Lecturer
School of Languages, Literature, Culture and Communication, University of Limerick

Abstract

Financial information is extremely valuable to investors and other interested parties. This information, which can be qualitative or quantitative in nature, can be analyzed and subsequently used to try to predict future share prices and/or determine market sentiment. Financial writers need to bear this in mind when writing reports, as their message(s) could be interpreted in unexpected ways and this could cause undesirable market reactions. In this article, I provide an overview of some studies that examined the writing style and tone of financial reports. I also provide an overview of some studies that examined the use of positive and negative words in financial reports. I conclude with reference to some recent studies that involved the automatic analysis and classification of financial content. Whilst the success of automated tools has been limited, to a certain extent, tools are being used increasingly to assist with the daunting task of interpreting complicated and lengthy financial documents. Once these tools improve, it will not be so easy for financial writers to disguise bad news in the midst of good news.

Introduction

There is a general consensus that there are three different types of financial information: information that is available in past stock prices, information that is available to all the public, and information that is both available to the public and available privately to insiders (Fama 1970; Haugen 1990; Hellstrom and Holmstrom 1998; Elton et al. 2003). There are two main lines of arguments about the possible impact that information can have on the value of financial instruments. Advocates of the efficient markets hypothesis (EMH) believe that the price of a financial instrument properly reflects all available information immediately (Fama 1970). If security prices respond to all available information quickly, then the market is deemed *efficient* and no excess profits or returns can be made. An excess return is the return from an investment that *exceeds* some pre-determined benchmark or index (e.g. the Standard & Poors' 500 index) which has a similar level of risk. The other line of argument is made by fundamental and technical analysts who argue that the market is *inefficient* because information

disseminates slowly through the market and prices under- or over-react to the information (Haugen 1990). By *under-reaction*, we mean that the average return following a good news announcement (e.g. company X has expanded its product line) is *higher* than the average return following a bad news announcement (e.g. company X is being sued for litigation). An *over-reaction* arises when the average return following a series of good news announcements is *lower* than the average return following a series of bad news announcements.

This article provides a review of some literature on fundamental content analysis. Fundamental content analysis involves evaluating the value of a financial instrument using quantitative and qualitative content derived from company financial statements, news stories, analysts' reports, and discussion forums. Frequently used online news sources include Reuters, the Wall Street Journal, and Yahoo! Finance. Other sources of fundamental information include online financial analysis tools such as Dow Jones News Analytics and online databases such as the Securities Exchange Commission's EDGAR system. Technical analysis, which involves evaluating time series patterns and trends relating to *previous prices* of a financial instrument and the volume of trading, with a view to predicting future prices and volumes, is beyond the scope of this article.

The Language of Financial Reports and News

Numerous researchers have analyzed the language in annual and interim reports and news articles, for various reasons, including:

- narrative analysis (e.g. examining the phrases used in the narratives of reports to determine the quality of the reports)
    - see Beattie et al. 2004
- quantitative and qualitative analysis of content (e.g. examining reports at word, sentence or paragraph level to see if quantitative and qualitative data are used for different purposes)
    - see Back et al. 2001; Kloptchenko et al. 2004
- writing style and tone analysis (e.g. examining reports at word, sentence or paragraph level to identify changes in writing style or tone made by different types of companies, or after certain company events occur)
    - see Kloptchenko et al. 2004; Feldman et al. 2008; Loughran et al. 2008
- thematic analysis (e.g. coding reports on a sentence-by-sentence basis to identify themes and therefore examine corporate communication strategies)
    - see Kohut and Segars 1992
- effectiveness analysis (e.g. examining the effectiveness of reports in terms of their credibility, efficacy, commitment and responsibility)
    - see Segars and Kohut 2001

- performance and readability analysis (e.g. examining the use/non-use of strong, clear, and concise writing in reports, to convey a message)
    - see Subranamiam et al. 1993; Loughran and McDonald 2011
- positive and negative word analysis (e.g. examining words and/or the technical characteristics of reports to classify a report as having positive or negative news)
    - see Hildebrandt and Snyder 1981; Thomas 1997
- market sentiment analysis (e.g. examining how frequently certain words or phrases appear in reports and using these frequencies to plot a time series of the market sentiment)
    - see Gillam et al. 2002; Ahmad et al. 2003; Ahmad et al. 2005; Devitt and Ahmad 2007; Daly et al. 2009
- summary generation (e.g. examining the selection of words used in reports to produce report summaries automatically)
    - see de Oliveira et al. 2002

Some of these studies examined single words, others examined phrases, and others examined more complex technical characteristics such as the number of syllables per word or the number of passive constructions. As the literature in this area is vast, this article provides an overview of some studies on writing style and tone analysis as well as positive and negative word analysis, in financial reports. These studies should be of interest to students of professional communication who may be responsible for corporate communication when they enter industry. These studies should also be of interest to professionals already working in the financial services industry, because the language of financial reports can be analyzed in a variety of ways and it can result in a different market reaction, depending on *how* it is analyzed. Also, these professionals have a legal obligation to ensure they do not intentionally deceive the public by using language inappropriately.

Writing Style and Tone Analysis

In this section we will briefly outline three studies that involved the examination of writing style and tone in financial reports. The first study, by Kloptchenko et al. (2004), examined text and financial ratios to identify the likely future performance. The second study, by Feldman et al. (2008), examined the frequencies of positive and negative words to identify tone changes. The third study, by Loughran et al. (2008), examined ethics-related terms to see if certain types of firms tended to use these terms more than others.

Kloptchenko et al. (2004) analyzed text and financial ratios in quarterly reports downloaded from the Web, with a view to identifying indications to likely future financial performance. Using seven quarterly reports for three telecommunication companies during 2000-2001, they found that qualitative and quantitative data seem to represent different things; text tends to give hints about future performance through the

use of optimistic language (e.g. *increase*, *share growth*, and *strong demand*), whereas financial ratios tend to refer to past performance.

They also found that the writing style and tone in a company's financial report tends to change before a major company event; the tone tends to represent the future performance more than the past or current performance. However, some limitations of their study include a very small data set (only three companies were used), a limited vocabulary set (they were all telecommunications companies), and a significant usage of proprietary names, all of which may have skewed the results to some extent.

Feldman et al. (2008) examined changes in tone in the Management Discussion and Analysis (MD&A) section of corporate reports, to see if this section adds any incremental information to that already provided by preliminary earnings surprises, accruals, and operating cash flows. By counting the frequencies of positive and negative words, they found that tone changes in MD&As yield excess average returns (i.e. returns above some pre-determined benchmark) and that the returns tend to drift for longer periods that extend beyond the subsequent quarter's preliminary earnings announcements. In addition, they found the change in tone was incrementally more informative when firms were small and analyst following was weak.

Loughran et al. (2008) examined ethics-related terms in Form 10-K reports between 1994 and 2006, to see if they could identify types of firms that had a tendency to use these terms. Form 10-Ks are more commonly referred to as the annual reports. These reports are frequently reviewed by the general public as they provide a good overview of the company's overall performance during the year. Loughran et al. found that, in the pre-regulatory period up to 2002 when the use of ethics-related terms in reports was voluntary, such terms only appeared in 8% of the reports. They also found that terms specifically related to a 'code of conduct' appeared less than 1% of the time. In the post-regulatory period, when reporting regulations were much stricter, code-related terms appeared in almost 60% of 10-Ks, as firms were legally obliged to discuss their code. Loughran et al. then focused their attention on reports disclosed during the pre-regulation period to see if any firms that used ethics-related terms had been identified as 'sin stocks', were involved in class action lawsuits, or had received poor corporate governance scores. Sin stocks typically refer to public companies involved in industries that sell alcohol, tobacco, or gaming products. They found that these 'problematic' firms were more likely to use ethics-related terms than other firms, probably to appeal to investors who were concerned about deception, but the proportional differences between these 'problematic' firms and other firms were even more pronounced in the pre-regulatory period. They proposed that further study could investigate the link between long-term stock performance and the use of ethics-related terms in 10-Ks, not just for sin stocks but also for poor-governance firms. Whilst they did not report specific results regarding returns, they did report that the sin stocks in their dataset performed "relatively well", unlike the poor-governance firms they studied (p.18).

Positive and Negative News Analysis

In this section we will outline two studies which examined how financial language was used differently to convey positive and negative news in financial reports. The first study, by Hildebrandt and Snyder (1981), examined the frequency of occurrences of positive and negative words, to see if companies used more positive than negative words, even in years when they were not performing well financially. The second study, by Thomas (1997), involved a more in-depth analysis. It examined verb usage and thematic structures, to try to determine if these were used differently when positive or negative news was being conveyed.

Hildebrandt and Snyder (1981) applied the 'Pollyanna Hypothesis' to the writing of annual reports. In the context of business communication, the Pollyanna Hypothesis states that people tend to use positive words more frequently and directly than negative words (Hildebrandt and Snyder 1981).

They proposed three related hypotheses:

- "Positive words occur more frequently in annual letters to stockholders regardless of a financially good or bad year.
- Negative words occur less frequently in a good year than a bad year.
- German respondents will parallel American respondents when viewing positive and negative words in isolation" (*ibid*, pp.5-6).

They viewed twelve annual letters to stockholders in 1975 (a bad year) and twelve letters in 1977 (a good year). The companies were selected from the Dow Jones Industrials. They used a list of 356 positive and negative antonym pairs and translated these into German. The entire list was given to 100 Graduate Business students, who identified the preferred (positive) and non-preferred (negative) words; this was done to make sure there was agreement. They then read the 24 annual letters, recorded occurrences of each of the 356 words, and then placed each occurrence into one of three classifications (positive statement, negative statement, or neutral statement), depending on its contextual usage. With regard to the first and second hypotheses, they found that there were significantly more occurrences of positive words than neutral or negative words, regardless of year i.e. 68.9% of words were positive in context in 1975 (a bad year) and 79.5% of words were positive in context in 1977 (a good year). We would expect this to be the case as companies will obviously prefer to use positive language than negative language, whenever possible. These results were found to be statistically significant using t-tests. They proved the third hypothesis by showing the words to the American and German reviewers in isolation (i.e. not in context) and found that over half of the preferred or positive words selected had 90% agreement in both languages. It is important to note that when these words were used in context, the meaning often changed quite significantly. For example, whilst *increased* is a preferred word and

decreased a non-preferred word, in the following statement, *increased* has a negative connotation: "The OPEC nations, because of their consolidated position, were able to *increase* crude oil prices about $1 per barrel in October" (*ibid*, p.9). This example demonstrates how difficult it is to analyze financial texts by merely looking at words in isolation of one another. If a potential investor or a student of communication performs a simple word count of the 'positive' and 'negative' words in a financial report, it is quite likely that he/she will not get a true picture of the message in the report. Even if the message is negative (from an investment point-of-view), there can still be more 'positive' than 'negative' words, because report writers often disguise bad news in the midst of good news. In this next example, which relates to a reduction in profits (bad news), there are at least two positive words (profits and profitability), as opposed one negative word (declined): "Even though *profits* have *declined* since the first quarter, we expect to return to *profitability* in the next quarter". The decline in profits is a fact, whereas the return to profitability is merely an *expectation*.

Thomas (1997) examined the differences between good news and bad news communicated in management letters at the start of annual reports, with a view to identifying management's motivations and priorities. To avoid differing company styles, she gathered the annual reports for one company over a five-year period (1984-1988). She chose this particular period as the company in question changed from being a profitable to an unprofitable one during that period. Her main research goal was to confirm or question the concept of the 'Pollyanna hypothesis' in the context of annual reports, as discussed in Hildebrandt and Snyder (1981). She wanted to see if language usage was different in profitable and unprofitable years. The findings from this research are useful to communication students and financial industry professionals because they highlight, once again, the challenges faced by the public when trying to analyze the meaning contained within financial reports.

Unlike Hildebrandt and Snyder, who examined the frequency of occurrences of positive and negative words in reports, Thomas performed a more in-depth analysis of the language by examining transitivity structures and thematic structures. Transitivity "describes a clause according to the kind of verb used, the participants, and the circumstances" (*ibid*, p.53). With regards transitivity structures, she first looked at the number of passive verb constructions in each clause in the management letters. She found there was an increase in the number of passive constructions from 10% in 1984 (a profitable year) to 20% in 1988 (a loss-making year). Thomas suggested that this makes sense, as messengers tend to distance themselves from negative messages whenever possible. Distancing is a feature of deception and can be examined by investigating linguistic styles (Newman et al. 2003).

Thomas also found that verbs of 'being' in the first and last paragraphs doubled in the five-year period from 33% to 66%, giving a much more objective or factual impression that should not be questioned. Typically, the first paragraph prepares the reader for the

message that will follow and the last paragraph reminds the reader of the key points, so these are both important paragraphs. There was also an increase in the number of non-human participants acting as agents (examples cited include *opportunities*, *fiscal 1988*, and *machine tool market*). Focusing again on the first and last paragraphs, Thomas found that there was a shift away from the writer of the message; instead, nonhuman circumstances were causing these factual situations, so they could not be questioned. Nonhuman participants only appeared in 25% of clauses in 1984 (a profitable year) but they appeared in 87% and 73% in 1987 and 1988 respectively (these were both loss-making years).

"Thematic structure in systemic theory is the part of the clause that serves as a point of departure for the message - what the message is about" (Thomas 1997, p.56). When Thomas examined thematic structures, she found that there were two types of themes: (1) the personal pronoun *we* and (2) a variety of inanimate nominal groups, such as the nonhuman agents mentioned previously. Even though the usage did not progressively increase or decrease each year, she found that overall usage of the pronoun *we* decreased from 75% in 1984 to 27% in 1988 and the usage of inanimate nominal groups increased from 25% in 1984 to 73% in 1988. These thematic findings are largely in line with the transitivity findings outlined earlier, suggesting that management distanced itself from the poor performance in the latter years.

Thomas also briefly examined condensations; one example of a condensation in Business English is *transition*, as in *transition year*, which could mean things are improving if linked with recent profits, or disimproving if linked with losses or declining profits. Another example is *gradually improving market* – does this mean that the situation was poor in previous times and now it is improving or that the situation is not improving as well as had been hoped? Does *profitability* mean that they (only) have sufficient funds to pay a dividend or that they are making a healthy profit? In one of the 1987 reports, she identified the following sentence in the first paragraph: "*Cross & Tracker had a difficult year in fiscal 1987 as markets proved weaker than expected and pricing pressures and other factors combined to severely erode margins.*" (*ibid*, p.63). This sentence, which contains several contractions (*markets proved weaker*, *pricing pressures*, *other factors*, and *erode margins*), demonstrates that once again, management clearly abdicated responsibility for the dire financial situation. Business students should bear these issues in mind because they may one day be responsible for corporate communication and they will be legally obliged to ensure they do not intentionally deceive the investing public. The public also needs to be aware of these issues, because they may misinterpret corporate messages and consequently make poor financial decisions.

Conclusions

Nowadays, there is a huge volume of financial content available to investors and the general public. Sources of financial content include financial news sites (e.g. Reuters), online databases of corporate reports (e.g. the US EDGAR system), and online discussion forums. Some methods of examining financial content include analyzing the writing style and tone or analyzing positive and negative words. In this article, I provided readers with an overview of just some of the studies in these areas.

This article should be of interest to students of business communication who may be trying to interpret complicated and lengthy financial documents as part of their studies. In addition, these students may one day work as professionals in the financial services industry and be responsible for corporate communication. If they do so, they will be legally obliged to ensure they do not intentionally deceive the public through their communications. The general public also needs to be aware that simple analyses of positive and negative words may not be sufficient to interpret complex reports and that various strategies have been adopted by report writers in the past, to disguise bad news in the midst of good news.

In the last decade, research has focused increasingly on the *automatic* analysis and classification of financial content. The content features analyzed have included single words, keyword records and phrases, as well as ratios and variables. The research goal has typically been to improve share price predictability (see for example Tetlock et al. 2008, Loughran and McDonald 2011, and Slattery and Sutcliffe 2012). However, the complexity of language, coupled with the avoidance strategies frequently adopted by financial writers, make the successful analysis and classification of such content extremely difficult.

References

Abarbanell, J. S. and V. L. Bernard (1992). "Test of analysts' overreaction/ underreaction to earnings information as an explanation for anomalous stock price behavior." The Journal of Finance 47(3): 1181-1207.

Ahmad, K., L. Gillam and D. Cheng (2005). Textual and quantitative analysis: towards a new, e-mediated social science. 1st International Conference on e-Social Science, Manchester, National Centre for e-Social Science.

Ahmad, K., L. Gillam, D. Cheng, T. Taskaya-Temizel, S. Ahmad, P. Manomaisupat, H. Trablousi and A. Hippisley (2003). The mood of the (financial) markets: In a corpus of words and of pictures Corpus Linguistics 2003 Conference, Lancaster University, University Centre for Computer Corpus Research on Language.

Back, B., J. Toivonen, H. Vanharanta and A. Visa (2001). "Comparing numerical data and text information from annual reports using self-organizing maps." International Journal of Accounting Information Systems 2(4): 249-269.

Beattie, V., B. McInnes and S. Fearnley (2004). "A methodology for analysing and evaluating narratives in annual reports: a comprehensive descriptive profile and metrics for disclosure quality attributes." Accounting Forum 28(3): 205-236.

Daly, N., C. Kearney and K. Ahmad (2009). Correlating Market Movements with Consumer Confidence and Sentiments: a Longitudinal Study. Text Mining Services 2009. G. Heyer. Leipzig, Germany, Leipzeiger Beitrage zur Informatik.

de Oliveira, P. C. F., K. Ahmad and L. Gillam (2002). A financial news summarisation system based on lexical cohesion. Making Money in the Financial Services Industry, a Workshop at the Terminology and Knowledge Engineering Conference (TKE 2002), Nancy, France.

Devitt, A. and K. Ahmad (2007). Sentiment Polarity Identification in Financial News: A Cohesion-based Approach. 45th Annual Meeting of the Association of Computational Linguistics (ACL 2007), Prague, Czech Republic, Association for Computational Linguistics.

Elton, E. J., M. J. Gruber, S. J. Brown and W. N. Goetzmann (2003). Modern portfolio theory and investment analysis. New York, John Wiley & Sons, Inc.

Fama, E. F. (1970). "Efficient capital markets: A review of theory and empirical work." The Journal of Finance 25(2): 383-417.

Feldman, R., S. Govindaraj, J. Livnat and B. Segal (2008). The Incremental Information Content of Tone Change in Management Discussion and Analysis. Available at: http://ssrn.com/abstract=1126962.

Gillam, L., K. Ahmad, S. Ahmad, M. Casey, D. Cheng, T. Taskaya Temizel, P. C. F. de Oliveira and P. Manomaisupat (2002). Economic news and stock market correlation: a study of the UK market. Making Money in the Financial Services Industry, a Workshop at the Terminology and Knowledge Engineering Conference (TKE 2002), Nancy, France.

Haugen, R. A. (1990). Modern investment theory. New Jersey, Prentice Hall, Englewood Cliffs.

Hellstrom, T. and K. Holmstrom (1998). Predicting the stock market, Mälardalen University, Sweden.

Hildebrandt, H. W. and R. D. Snyder (1981). "The pollyanna hypothesis in business writing: Initial results, suggestions for research." The Journal of Business Communication 18(1): 5-15.

Kloptchenko, A., T. Eklund, J. Karlsson, B. Back, H. Vanharanta and A. Visa (2004). "Combining data and text mining techniques for analyzing financial reports." Intelligent Systems in Accounting, Finance and Management 12(1): 29-41.

Kohut, G. F. and A. H. Segars (1992). "The president's letter to stockholders: An examination of corporate communication strategy." The Journal of Business Communication 29(1): 7-21.

Loughran, T. and B. McDonald (2011). Measuring Readability in Financial Disclosures. Summer Finance Seminar Series, University of Notre Dame.

Loughran, T. and B. McDonald (2011). "When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks." The Journal of Finance 66(1): 35-65.

Loughran, T., B. McDonald and H. Yun (2008). "A Wolf in Sheep's Clothing: The Use of Ethics-Related Terms in 10-K Reports." Journal of Business Ethics 89(1): 39-49.

Newman, M. L., J. W. Pennebaker, D. S. Berry and J. M. Richards (2003). "Lying words: predicting deception from linguistic styles." Personality and Social Psychology Bulletin 29(5): 665-675.

Segars, A. H. and G. F. Kohut (2001). "Strategic communication through the world wide web: An empirical model of effectiveness in the CEO's letter to shareholders." Journal of Management Studies 38(4): 535-556.

Slattery, D. M. and R. F. E. Sutcliffe (2012). Automatic Analysis of Financial Event Phrases and Keywords in Form 8-K Disclosures. In proceedings of the 10th Terminology and Knowledge Engineering Conference (TKE 2012), Madrid, Spain.

Subranamiam, R., R. G. Insley and R. D. Blackwell (1993). "Performance and readability: A comparison of annual reports of profitable and unprofitable corporations." The Journal of Business Communication 30(1): 49-61.

Tetlock, P. C., M. Saar-Tsechansky and S. Macskassy (2008). "More than words: quantifying language to measure firms' fundamentals." The Journal of Finance 63(3): 1437-1467.

Thomas, J. (1997). "Discourse in the marketplace: The making of meaning in annual reports." The Journal of Business Communication 34(1): 47-66.

Author

Darina M. Slattery
Lecturer

Darina M. Slattery holds a PhD in Computer Science, an MSc in Software Localization, and a BBS with French Degree. She has been teaching technical communication, instructional design, and e-learning design and development at the University of Limerick since 2003 and has published her research in refereed journals and international conference proceedings. Her teaching and research interests include e-learning design and development, technical communication theory, financial content analysis, virtual teams and information design. In the past, she has been Program Director for two successful graduate programs—the MA in Technical Communication and E-Learning and the Graduate Certificate in Technical Writing. She is currently Section Leader for the Technical Communication Section at the University of Limerick.

Contact:

Darina.Slattery@ul.ie

Office LC1-014

School of Languages, Literature, Culture and Communication

University of Limerick

Limerick

Ireland

Web: http://www.ul.ie/techcomm and http://www.staff.ul.ie/slatteryd