

Exploring the Use of Contextual Analysis of Written Communications to Identify Fraud Risk: Analysis and Comparison of Enron and Facebook Internal Communications

Subash Adhikari

Southeastern Louisiana University, USA

G. Stevenson Smith

Independent

[Doi: 10.60154/jaepp.2024.v25n1p15](https://doi.org/10.60154/jaepp.2024.v25n1p15)

Submitted: 20 August 2023

Accepted: 29 February 2024

Published: 15 March 2024

Copyright 2024 Author(s)

Under Creative Commons CC-BY 4.0

OPEN ACCESS

Cite As:

Adhikari S. & Smith G.S. (2024). *Exploring the Use of Contextual Analysis of Written Communications to Identify Fraud Risk: Analysis and Comparison of Enron and Facebook Internal Communications*. Journal of Accounting, Ethics & Public Policy, JAEPP, 25 (1), 15. <https://doi.org/10.60154/jaepp.2024.v25n1p15>

Abstract

Written texts can reveal the personality traits of the authors of those documents. Beyond the differences in writing styles of individuals, the choice of words and use of vocabulary can be different for individuals depending on their personality traits. In this study, we use Linguistic Inquiry and Word Count (LIWC2015) software to analyze the content of emails from a sample of Enron and Facebook executives. We compare the attributes revealed in these emails at both the individual and company levels. The results of the analysis indicate that there are significant differences in the personality traits of individuals in the two companies. Context analysis can offer valuable insights into the traits of business leaders and serve as a useful tool for identifying potential fraud risks.

Keywords: Linguistic analysis; LIWC2015; Enron; Facebook

I. Introduction

Analysis of written texts and the meta data in those texts can help identify underlying character traits of their author. The linguistic analysis of such texts provides insights into the writer's personality. Such character traits can be evaluated for an individual or for a group of individuals. Analysis of the written communication of a group may be indicative of the organizational

culture existing within that group. In this study, we analyze the written communication (emails) of the executives at Enron and Facebook using Linguistic Inquiry and Word Count (LIWC2015)¹.

At the conclusion of the Enron trial, the Enron emails were released by the Federal Energy Regulatory Commission (FERC) in 2003. The Enron dataset consists of emails sent by more than 150 senior executives. The British Parliament released emails of Facebook executives in 2018 because of Mark Zuckerberg's refusal to appear before a parliamentary committee. Those emails included two hundred and fifty pages of private Facebook emails sent between 2012 and 2015.

The choice of words in written language has been associated with personality styles (Pennebaker and Graybeal, 2001). Several other studies have examined the relationship between personality and written language (Boyd and Pennebaker, 2015, Boyd and Pennebaker, 2017; Boyd & Schwartz, 2021; Chung and Pennebaker, 2018; Stachl et al., 2020). Personality traits such as Machiavellianism, narcissism, and psychopathy (collectively referred to as "dark triad") have been associated with fraud risk (Furnham, et al., 2013; Paulhus and Williams 2002).

In this study, we analyze and identify the differences in personality traits of executives at Enron and Facebook by analyzing the words used in their emails. Enron and Facebook serve as proxies for a "financially fraudulent company" and a typical "non-financially fraudulent company", respectively. We collected and analyzed emails from selected executives in both companies to identify their character traits based on the language used in their correspondence. The differences in such personality traits have the potential to help identify executives exhibiting a predisposition toward fraudulent conduct.

This paper is organized into six sections. Section two identifies similarities and differences in the background of Enron and Facebook. We discuss prior studies in Section three and describe the data collection procedure in section four. Section five provides details and results of analysis. Section six discusses the limitations of the paper and Section seven concludes the paper.

¹ LIWC (2015) was created at the University of Texas (Pennebaker et al., 2015) and has been widely used in research studies since the 1990s (for example, Robinson, et al., 2014; Ritter, et al., 2014; Underwood, et al., 2012; Holtgraves, 2011; and Tumasjan, et al., 2010). LIWC2015 application uses a pre-defined set of words (LWIC dictionary) composed of almost 6,400 words, word stems, and select emoticons. Each word in the text being analyzed that match with the LWIC dictionary is assigned to a category. One word can fall under multiple categories. For example, the word cried is part of five-word categories: sadness, negative emotion, overall affect, verbs, and past focus.

II. Comparisons between Enron and Facebook

Enron operated in the energy sector while Facebook is in the technology industry, specifically in social media. Enron's 2000 10-k annual report identifies business operations in transportation and distribution, wholesale services, retail energy services, broadband services, and others as business segments. The largest source of revenue for Enron was from wholesale services which involved the marketing and delivery of physical commodities and financial and risk management services. Enron's Internet online trading platform resulted in \$2.5 billion a day in trading volume in 2001. On the other hand, Facebook disclosed in its 2018 10k annual report that it operates as a single entity that generates substantially all of its revenue on its Internet platforms from advertising revenues.

Though there are obvious differences between the two companies, many commonalities exist. For example, Enron was founded by a charismatic leader, Ken Lay, who led the company into a new and innovative business strategy using an online commodity trading platform. Facebook was founded by Mark Zuckerberg, another charismatic leader. Both companies experienced initial tremendous growth in their business operations. Facebook's business model is entirely reliant on the Internet, and so was Enron's online trading platform (Enron Online) where billions of commodities were traded. To expand its online footprint in 2000, Enron was working on a contract agreement with Blockbuster to provide online music, games, and movies, which was a streaming breakthrough for that time. Enron was regarded to be one of the most innovative companies at the time. It received the "America's Most Innovative Company" award from *Fortune* for six consecutive years between 1996 and 2001. Facebook, with its new mobile ad formats including full-screen, 360-degree videos, and interactive ads has also been an innovative company that essentially changed the way people interact with each other. Another similarity is that both companies were leaders in their respective businesses. Enron was the number one seller of natural gas in North America, and Facebook is the number one social media company. Consequently, the management of both companies needed to deal with controversy, strong competition, and threats against their business models.

Enron and Facebook have both been criticized and there are (were) controversies surrounding the companies. However, a major difference between the two companies is that Enron's executives were found guilty of a financial conspiracy, whereas Facebook's executives have never been charged with any financial fraud. At Enron, several executives including CEO Kenneth Lay, CFO Andrew Fastow, and former CEO Jeffrey Skilling were convicted and sentenced for fraud and conspiracy. In this study, Enron and its executives represent a financially fraudulent company. Although a lawsuit was filed against Facebook for alleged privacy violations in 2018, the case has not been

settled.² Yet no one is alleging that Facebook, and its executives are guilty of financial fraud. Therefore, it is reasonable to consider Facebook as representative of a typical “non- financially-fraudulent-company”. In this study, a comparison is made between the personality traits exhibited by executives at Enron and Facebook. These two corporations represent “financially-fraudulent-company” and “non- financially-fraudulent-company”, respectively, for this study.

III. Previous studies of Enron’s dataset

Prior studies have examined the Enron email dataset to identify various aspects of linguistic characteristics in the emails. Diesner, et al. (2005) used network analysis to investigate the structure, behavior and idiosyncrasies of communication networks in Enron. The authors report that previously disconnected employees engage in mutual communication during a crisis period. The authors also report that messaging interactions bypassed formal chains of communication during the crisis period. Kang, et al (2010) studied the “content-actor” relationships in emails using NetLens. The analysis identified paired content and individual relationships. Aven (2015) used the Enron emails to examine the mechanisms that individuals use to remain undetected when sharing information. The authors found that the members of corrupt projects communicate less and have fewer reciprocal relations than the non-corrupt project members.

Network analysis of the emails was used by Fragale, et al. (2012). They used linguistic markers to determine signs of reverence within the content of the emails. Mohammad (2012) analyzed Enron emails to examine the differences in emotions in emails. The author reports differences among genders in the use of emotional words. Salami (2014) studied the Enron email corpus to investigate the simplicity of email communication. The author found that the readability of emails change with change in the hierarchical direction of communication, hierarchical distance between communicators, and the size of the audience. Past researchers have used the Enron corpus to determine personal network relationships among the executives and staff (Gahrooei and Paynabar 2018; Fox, et al. 2016; McCallum, et al. 2007; Kurihara, et al., 2006). Other researchers have used a linguistic analysis of email to study employee threat risk and forensic authorship (Brown, et al., 2013; Wright 2014). Previous studies are largely concerned with identifying characteristics of the linguistic networks in the emails and not with identifying the traits associated with fraudulent behavior.

IV. Data

² At the time of the writing of the article, Facebook has set aside \$5B for possible fines related to the Cambridge Analytica Data case.

For the purpose of this study, private emails sent by executives from Enron and Facebook were analyzed using LIWC2015 linguistic software. Both sets of emails were released as a result of legal cases involving the companies. The Enron scandal became public knowledge in October 2001 and Enron's emails were released in 2002.³ The date of Enron emails used in the study range from Jan 27, 2000 to Jan 15, 2002 and the date of Facebook emails ranged from Oct 7, 2012 to March 30, 2015. Facebook emails were released in the UK by a parliamentary committee at the end of 2018. Although Enron emails have been used in prior studies, as noted, we are not aware of Facebook emails used in research of this nature.

The Enron data, made public by FERC in 2003, includes more than 500,000 emails sent by 158 Enron executives during a span of over three and a half years, from 1999 to 2002. These emails were exchanged at a time when many at Enron were aware of the company's underlying problems even though they had not become public knowledge. The database includes 92% of Enron staff emails.⁴ None of the emails included attachments. The Enron executives' emails in the dataset were sent by senior managers thirty-five of whom were indicted and charged with fraud and conspiracy for their involvement in the fraud. In this study, we analyze the emails of the top six Enron executives. They are (1) Kenneth Lay, founder of Enron; (2) Jeffrey Skilling, CEO; (3) Andrew Fastow, CFO; (4) Kenneth Rice, CEO of broadband services; (5) Mark Koenig Head of Investor Relations; and (6) Ben Glisan, Treasurer. We chose these executives because of their significant role and responsibilities in determining Enron's financial results.

Before analyzing the Enron emails, the dataset was converted into a corpus⁵ for analysis. Previous studies have also developed an Enron corpus for analysis by eliminating extraneous data in the emails. The email headers and duplicate emails were manually removed before analysis.

The Facebook emails are available in several online sources (for example <https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Note-by-Chair-and-selected-documents->

³ There is no exact date for when the fraudulent activities at Enron actually began. The public became aware of the fraud in October 2001.

⁴ See: *Information Released in the Enron Investigation* <https://www.ferc.gov>.

⁵ A dataset is the raw data that is collected. A corpus is the dataset that has been cleaned up. As in the case of emails, headers and dates have been manually removed and typos corrected. We corrected typos to understand the intended communication in the email. We also suspect that some typos may have been created during the digital transformation of the emails from one source to other. We took this additional step to remove typos, but LIWC2015 converts all text files to lower case before processing them therefore grammar, capitalization, and sentence structure are not required to be corrected. Finally, it should be noted that an annotated corpus is created from the corpus by grammatically separating the words in the corpus into different categories such as positive or negative connotations.

ordered-from-Six4Three.pdf). Two hundred and fifty pages of private Facebook emails, sent between 2012 and 2015, were released by the U.K. Parliament. Originally, they were collected as evidence in a California lawsuit against Facebook by Six4Three, an app development company (Roose 2018). Six4Three alleged that Facebook was restricting their access to user data on the Facebook platform. The U.K.'s Digital, Culture Media, and Sports Committee seized these documents as part of a larger investigation into Facebook's practices in the U.K. Eventually the emails were put online in 2018. The Committee had been investigating Facebook's role with Cambridge Analytica (CA). CA had harvested private information from the Facebook platform and used it for political purposes. Facebook's emails were released because Mark Zuckerberg would not testify before the Parliamentary Committee reviewing Facebook's practices.

Executives and staff emails were included in the release. The following top executives' emails in the exchanges were chosen for analysis: (1) Mark Zuckerberg, CEO; (2) Konstantinos Papamiltidas, Director of Platform Partnerships; (3) Sam Lessin, VP of Product Development Related to Partnerships; (4) Simon Cross Product Manager, Plugin Apps; and (5) Sachin Monga, Platform Team, Social Media Interactions.⁶ These executives were chosen as they were the highest-ranking executives involved in the email exchanges. Although these executives did not have the same job functions as those executives at Enron, their decision-making oversight in the organizational hierarchy matched those of the executives in Enron. Similar to the Enron dataset, we cleaned up the Facebook emails by removing headers and other peripheral elements, focusing solely on the email texts for our analysis.

V. Analysis

The context analysis was used to study the Enron and Facebook corpus.⁷ Context analysis categorizes large amounts of textual material or corpus to identify word patterns that lead to the identification of the writer's personality traits (Cohn et al., 2004; Pennebaker 2011; Pennebaker et al., 2003; Pollach 2012). Once the word patterns are categorized based on lexical categorizations used by a mapping tool for automated context analysis, an annotated corpus is created. LIWC2015's categorizations were used for that purpose. The tool maps a text file with counts on dictionary-determined categories as well as the different senses that can be assigned or tagged to a

⁶ The emails were collected from a discussion thread used by the top executives at Facebook to discuss their platform business practices.

⁷ There is a difference between the meaning of "content" and "context." In linguistic applications, content is what is written whereas context is concerned with the meta data structure behind the written text.

word, such as sadness or anxiety, based on an algorithm.⁸ The algorithms in the tool have been independently rated by linguists and tested for reliability (Tausczik and Pennebaker 2010). Li and Chignell (2010) followed this approach to show how personality types could be identified with the words written by blog writers. Argamon and Levitan (2005) and Oberlander and Gill (2006) also used this approach and found that word usage could identify the trait of extroversion in writers. Martínez-Huertas, et al. (2022) reported that LIWC variables had a significant relationship with big five personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism) in their sample. Pennebaker and King (1999) and Pennebaker (2011) have written extensively about the relationship between the personality traits associated with writing style. Pennebaker and King (1999) state:

Thus, language use may be thought of as an arena in which the impact of the person is unavoidable (p. 1309).

Other authors have adopted this methodology with business datasets. Demers and Vega (2008) used this methodology to evaluate earnings reports. Abrahamson and Park (1994) used context analysis to identify traits found in the letters sent to stockholders. Smith (2017) used this methodology to identify characteristics of accountants in literature. Ingram and Frazier (1983) used context analysis to investigate environmental disclosures from a sample of firms.⁹

The Background for the Analysis.

There are many well-recognized categorization tools for context classification. Within these categorization tools, there are word dictionaries with meanings assigned to word usages. One such system, in use since the 1990s, is the linguistic tool LIWC2015. LIWC2015 was developed by coders with linguistic backgrounds (Pennebaker, et al., 2015). LIWC2015 has been widely used to identify traits in various documents from texts to tweets

⁸In an *annotated* corpus that is grammatically tagged, words have been assigned word labels. Examples of categorization tools that can be used to develop an annotated corpus are the General Inquirer, Diction 7.0, KH Coder, Minnesota Contextual Context Analysis, and Whissell's Dictionary of Affect in Language. Some of these are freely available online. Several of these tools only provide limited results and others have not been maintained such as the General Inquirer. Although the algorithm used among tools may differ, the results from the LIWC2015 have been widely used and tested in research studies. For that reason, the LIWC 2015 was chosen over other categorization tools.

⁹A limitation of context analysis is that it is difficult to establish a causal relationship to the word categories. In our paper, although suggestions are made as to the reason for aggregated trends in scores, we are not attempting to identify a specific cause of these traits.

(Robinson, et al. 2014; Ritter, et al., 2014; Underwood, et al., 2012; Holtgraves, 2011; and Tumasjan, et al., 2010).

LIWC2015 allows for the linguistic patterning of text to identify traits associated with a writer. It is composed of two basic parts: (1) dictionaries and (2) algorithms to process text. Essentially, LIWC2015 collects metadata from within a text. The LIWC2015 dictionary contains nearly 6,400 words, word stems, and emoticons used to process text. LIWC2015 captures eighty-six percent of the words people use in writing and speech (Pennebaker, et al., 2015). Independent judges were initially used to build word lists in the dictionaries. Later, other groups of judges continued to rate the goodness of fit for the 81 categories in LIWC2015 as well as the words in the dictionary. In LIWC2015, summary variables are included in the categories. These variables are: (1) analytical thinking; (2) clout; (3) authenticity; and (4) emotional tone. These variables are the only non-transparent section of the ratings as shown in Table 1. Although words are grouped into these four categories, no word examples are disclosed in LIWC2015 due to commercial development agreements. Before being released, the LIWC2015 was tested by measuring each category for reliability and validity scores based on real-world text sets (Pennebaker, et al., 2015).

LIWC2015 processes text files sequentially. It finds each word and then determines if it is in its dictionary. Next, it includes that word in its count and uses those word totals to calculate a percentage function, i.e., as a percentage of the total words in the document. Words not counted are not part of the total word count. It then uses a scaled algorithm to determine a psycholinguistic category as its output. Word counts and word percentages are summed by category. Before counts are made, the word roots are lemmatized by reducing word stems by removing suffixes such as “ly” “ion” or “ing” or “ize” and “izes.” It should also be noted that one word may fit in several category traits. In those cases, the word might contribute to such characteristics as both “sadness” and “negative emotion”.

Once we obtain the LIWC2015 scores, we perform the two-tailed t-tests to compare the scores throughout this study. This is because we do not make directional predictions when comparing the scores. Due to the nature of the scores, we deem t-tests to be the appropriate method to examine whether the differences in scores are statistically significant.

The Sample

The analysis of the dataset begins with a selection of the trait categories. The selection was based on those traits that would be most revealing regarding possible fraudulent behavior. For example, the power and overconfidence associated with narcissistic behavior was important for the

analysis. Each email was checked for correct spelling, and typos¹⁰ and converted into the proper file format before being processed with the LIWC 2015. The emails from Enron and Facebook were processed using the nineteen traits in Table 1. The main divisions in Table 1 are Linguistic and Psychological Processes. Linguistic Processes identified by LIWC2015 are language descriptions such as counts of verbs, nouns, adverbs, and prepositions for example. These are described as “function words.” From a linguistic viewpoint, functional words show how people view and interact with the world around them. Table 1 lists the Summary Variables for Linguistic Processes under this category as well as those for Numbers and Negations such as “no.” Psychological Processes and their subclassifications are used to identify the emotional context in the email samples.

Table 1. The LIWC Traits Selected for the Context Analysis of the Enron and Facebook Datasets

Traits	Description	Word Examples (total number of words in category)
Part I: Linguistic Processes¹¹		
Analytic Thinking (see Pennebaker et al., 2014)	High score: formal, logical, hierarchical thinking; Low score; informal, personal, narrative thinking	Summary Variable: No words listed
Clout (See Kacewicz et al., 2013)	High score: “high expertise” and confident; Low score: more tentative or humble; the score can indicate confidence and level of leadership.	Summary Variable: No words listed

Table 1 Continued

Authentic (See Newman et al., 2003)	High score: honest, personal, and disclosing text; Low score: more guarded and distanced in disclosure	Summary Variable: No words listed
Tone (Emotional tone) (See: Cohn et al., 2004)	High score (>50) positive or upbeat emotion; Low score (<50) more anxiety, sadness,	Summary Variables: No words listed

¹⁰ We correct typos before running analysis because LIWC2015 will drop words that are not in dictionary and would not count these words in the analysis. Keeping typos would drop those words from the analysis and make the results less reliable.

¹¹ Linguistic processes are related to the grammatical characteristics of sentence structure. First person and non-first-person descriptions. Counts of verbs, adverbs, prepositions, and pronouns would be a measure of linguistic processes. In Table 1, only the summary variables of these traits are noted along with number and Negations word counts. No word examples are provided for summary variables as required under the contract agreement for developing this section of LIWC2015.

	and hostility. A score of around 50 indicates ambivalence.	
Negate (negations)	More inhibition	No, not, never (62)
Numbers	Numerical orientation of the document	Second, thousand (36)
Part II: Psychological (Emotional) Processes¹²		
<i>Cognitive Processes</i>		
Posemo (Positive Emotion)	More Optimistic; More extravert	Love, nice, sweet (620)
Negemo (Negative Emotion)	More neurotic	Hurt, ugly, nasty (744)
Anx (Anxiety)	More neurotic	Worried, fearful (116)
Anger	More neurotic	Hate, kill, annoyed (230)
Sadness	More neurotic	Crying, grief, sad (136)
Insight	Suggest learning or understanding	Think, know (259)
Tentat (Tentative)	Less self-assurance	Maybe, perhaps (178)
<i>Drivers and Needs</i>		
Power	More control	Superior, bully (518)
Reward	Compensation	Take, prize, benefit (120)
Risk	Hazard or threat	Danger, doubt (103)
<i>Time Orientation</i>		
Focuspresent (Present focus)	Current focus	Today, is, now (424)
Focusfuture (Future focus)	More goal oriented	May, will, soon (97)
<i>Personal Concerns</i>		
Money	More personal oriented	Audit, cash, owe (226)

Enron and Facebook Traits.

We first analyze the emails of Enron executives. To get the LIWC score for individual executives, we merge all emails of each Enron executive into one text document for each executive and then analyze the texts using LIWC2015. The results of the analysis are presented in Table 2.

¹²Tausczik and Pennebaker (2010) state “emotion words people use provide important psychological cues to their thought processes, emotional states, intentions, and motivations” (p.37). Table 1 uses a list of emotional traits identified by Pennebaker and associates as pertinent in the context analysis of data.

Table 2. LIWC Traits and their LIWC percentage scores for the top six Enron executives convicted in the Enron trial. In the table, the highest score in each row is in *Italics* and the lowest score in **bold**. *

Executive	Andy Fastow, CFO	Ben Glisan, Treasurer	Kenneth Lay, Founder	Mark Koenig, Investor Relations	Kenneth Rice, CEO Broadband	Jeffery Skilling, CEO
Word Count	467	433	10502	736	518	2582
Analytic Thinking	52.35	90.84	84.16	78.18	60.62	<i>91.66</i>
Clout	41.53	69.43	<i>89.20</i>	70.20	74.43	87.97
Authentic	32.06	8.34	24.02	<i>32.09</i>	28.36	29.32
Tone	29.35	98.25	82.28	81.93	98.35	88.28
Negations	<i>3.21</i>	0.92	0.56	0.95	<i>0.77</i>	0.39
Number	1.93	1.85	3.20	2.85	<i>4.63</i>	2.56
Positive Emotion	3.00	5.54	3.69	3.53	<i>5.79</i>	4.22
Negative Emotion	<i>2.78</i>	0.00	0.53	0.41	0.19	0.54
Anxiety	0.00	0.00	<i>0.06</i>	0.00	0.00	0.00
Anger	<i>0.43</i>	0.00	0.08	0.00	0.19	0.00
Sadness	<i>1.71</i>	0.00	0.17	0.14	0.00	0.39
Insight	<i>2.78</i>	2.08	2.79	2.58	<i>3.28</i>	2.52
Tentative	<i>3.85</i>	2.54	1.82	1.49	3.47	1.55
Power	2.36	2.77	3.88	1.90	1.54	3.83
Reward	1.28	0.23	1.21	1.22	<i>2.32</i>	1.74
Risk	1.28	0.46	0.66	0.41	<i>1.35</i>	0.23
Present Focus	<i>15.20</i>	9.70	8.21	11.01	11.39	9.26
Future Focus	1.50	0.23	<i>2.81</i>	1.77	1.54	1.98
Money	2.57	<i>6.93</i>	5.31	2.45	1.54	0.97

*This percentage score is computed by merging all emails of the listed executives and analyzing them as a single text. For example, all emails of Fastow are merged into a single text document and then analyzed in LIWC system as one text document.

Table 2 presents the LIWC2015 score of nineteen traits for the Enron executives selected for the analysis. The word counts represent the total number of words analyzed by LIWC2015. As expected, Kenneth Lay and Jeffery Skilling have the highest number of words in the analysis as they are the Founder and CEO and CEO of the companies respectively. Many communications in a company come from the CEO to the employees on a regular basis. Next, the table presents the scores of the executives on various attributes. We highlight the largest scores in each trait in *italics* and the lowest score in **bold**. The scores represent the percentage of the words in each category used by each of the executives. For example, Jeffery Skilling has the largest analytic score (91.66) and Andy Fastow has the lowest analytic score (52.35). This means that 91.66% of the words used by Jeffery Skilling are categorized as analytical, while 52.3% of the words used by Jeffery Skilling are categorized as analytical.

It should be noted that there is no benchmark (or norm) for scores in each category for a normal person. The difference in the scores can be interpreted as Jeffery Skilling using more analytical language than Andy Fastow. However, we cannot conclude that Andy Fastow is a person with low analytic skills.

As the comparative results in Table 2 are reviewed, the highest and lowest percentage scores in each row are highlighted in *Italics and bold*, respectively. The first trend that becomes apparent are the scores for Andrew Fastow, Enron's CFO. Mr. Fastow shows the highest scores for Negations (inhibition), Negative Emotions, Anger, Sadness, Tentativeness, and Present Focus. His lowest scores are for Analytic, Clout, Tone, and Positive Emotions. His psychological traits display less self-assurance, and more neurotic issues such as Anger. He is not confident about his leadership (Clout), has more anxiety and hostility (Tone), and faces more inhibition than the other executives. The causality of these personality traits cannot be identified, yet it serves as a marker of concern for the CFO of one of America's largest corporations to exhibit such traits. As the CFO, he was dealing with Enron's financial issues on a daily basis, and should have been cognizant of the imminent fiscal collapse.

In Table 2, Ken Lay's traits show high levels of Clout and Power as would be expected, and a focus more on the future and less on the present than the other executives. Most of the executives were not very concerned with any threat issues such as Anxiety or Anger (0.0 scores) which is an indication of their high level of confidence in themselves. The most notable trait exhibited by Jeffrey Skilling was his Analytical approach to problems, which is a measure of intelligence. He also had a very low (.39) score for the Negation trait which is a measure of inhibitions such as shyness. Other traits such as Sad and Negative emotion exhibited the same tendencies. His low concern for risk suggests a high level of certainty.

We also analyze the emails of Facebook executives. To get the LIWC score for individual executives, we merge all emails of each Facebook executive into one text document for each executive and then analyze the text using LIWC2015. The results of the analysis are presented in Table 3. The comparative results for the five Facebook executives shown in Table 3 are for the same set of traits as in Table 2. The Facebook executives in Table 3 were involved with product development. This product development was directed at settling how Facebook was going to allow app developers and media companies, such as Netflix, to interact and have access to the Facebook platform. This development group was responsible for setting policies as to pricing and full or limited access to Facebook's data. The email exchanges are an extensive collection of their views about these issues.

Table 3. LIWC Traits and LIWC percentage scores for Facebook's executives whose emails were released in the UK and had more than 400 words in their email messages. In the table, the highest score in each row is in *italics* and the lowest score in **bold**. *¹³

Executive Traits	K. Papamildidas, Director of Platform Partnerships	Mark Zuckerberg, CEO	Sam Lessin, VP of Product Development related to Partnerships	Simon Cross Product Manager, Plugin Apps	Sachin Monga Platform Team, Social Media Interactions
Word Count	4134	1774	7442	694	433
Analytic Thinking	<i>77.49</i>	53.81	66.87	77.46	61.54
Clout	88.34	83.81	77.41	<i>93.65</i>	69.43
Authentic	17.46	<i>36.20</i>	24.73	6.92	14.11
Tone	<i>92.70</i>	74.11	89.57	55.66	86.05
Negations	1.04	<i>2.20</i>	1.53	0.58	2.08
Number	1.31	1.52	1.24	1.15	<i>2.31</i>
Positive Emotion	<i>4.43</i>	3.61	4.42	2.45	3.70
Negative Emotion	0.22	<i>1.01</i>	0.60	0.86	0.23
Anxiety	0.10	0.39	0.13	<i>0.43</i>	0.23
Anger	0.00	0.28	0.12	<i>0.43</i>	0.00
Sadness	0.07	0.11	<i>0.17</i>	0.14	0.00
Insight	2.56	<i>3.27</i>	3.18	0.86	2.77
Tentative	3.29	<i>4.68</i>	3.29	2.59	2.54
Power	1.89	1.47	2.31	2.88	<i>3.46</i>
Reward	1.28	2.03	2.02	<i>2.16</i>	1.62
Risk	0.31	<i>0.45</i>	0.28	0.43	0.23
Present Focus	11.13	13.25	12.34	<i>14.55</i>	11.32
Future Focus	<i>1.96</i>	1.58	1.06	1.73	1.62
Money	1.26	3.78	3.13	0.58	2.08

*This percentage score is computed by merging all emails of the listed executives and analyzing them as a single text. For example, all emails of Papamildidas are merged into a single text document and then analyzed in LIWC system as one text document.

As reported in Table 3, Mark Zuckerberg, CEO of Facebook, exhibited strong traits of Authenticity, Risk, and Money as compared with the other executives. High authenticity reflects honesty as noted in Table 1. Yet, his concern with risk and financial issues likely affects his other personality traits. This is apparent as he has the highest scores in Negations, Negative Emotion, and Tentative compared with the other executives. Inhibition, negativity, and hesitation in decision making are part of these traits (as seen in Table 1). Such traits tend toward hesitation in decision making. These personality traits are

¹³ The LIWC scores for data with less than 400 words is considered unreliable.

not exhibited by Lay or Skilling at Enron. Lay's and Skilling's email messages exhibited Power and Clout as seen in Table 2.

Two employees at Enron and Facebook had job functions that were directly related to product management. At Enron, Kenneth Rice was CEO of broadband services, a separate product, and at Facebook, Simon Cross was a product manager dealing with apps allowed on its platform. When the traits of these two executives are compared, there are marked differences. Rice's traits exhibit positive emotions (Tone 98.35; Positive emotion 5.79; and Insight 3.28). When comparing Simon Cross' traits (Tone 55.66; Positive emotion 2.45; and Insight 0.86) to those of Rice, it is apparent that Rice has lower scores. Although there are differences in the job functions among product managers, they are generally responsible for the success or failure of a product line. Simon Cross further demonstrates a fundamental difference in traits between Facebook's and Enron's C-suite. Except for Andy Fastow, most executives at Enron displayed a robust self-assurance similar to Jeffrey Skilling's traits.

Several authors have identified a correlation between narcissistic behavior and fraudulent tendencies (Johnson, et al. 2013; Rijsenbilt and Commandeur 2013; Blickle, et al., 2006). Characteristics of narcissism include the need for power, dominance and an unrealistic belief in themselves. Narcissistic personalities tend to use manipulation and risky behavior to achieve their goals.

Schrand and Zechman (2012) and Chen (2010) also found a correlation between fraudulent behavior and overconfidence. Corporate leaders usually have a high level of confidence in themselves. Yet, Skilling was overly confident in his abilities with the initial success of Enron and his recognition in the nation as an icon of a successful business leader. Personality traits such as concerns with power, domination, and display of overconfidence are correlated with fraudulent behaviors. In Table 3, Jeffrey Skilling (3.83) and Ken Lay (3.88) both exhibit traits associated with Power as compared with the other executives at Enron and Facebook's Mark Zuckerberg (1.47). The traits at Enron were defined by dominance and overriding confidence.

Table 4. A Comparison of the Average of all LIWC percentage scores for Skilling and Zuckerberg and a determination of their statistical difference using t-tests. *

Traits	Average Score for Skilling Emails	Average Score for Zuckerberg Emails	P-Value (two Tail)
Word Count	2582	1774	n/a
Analytic Thinking	77.53	56.95	0.0392
Clout	83.42	85.27	0.7630
Authentic	29.77	35.02	0.5275
Tone	76.04	65.38	0.4489
Negations	0.23	1.79	0.0266
Number	2.46	0.91	0.0536

Positive Emotion	5.86	3.00	0.0312
Negative Emotion	0.60	0.79	0.6854
Anxiety	0.00	0.37	0.1129
Anger	0.00	0.23	0.0887
Sadness	0.43	0.09	0.1544
Insight	1.61	2.92	0.0532
Tentative	1.29	4.94	0.0001
Power	4.50	1.20	0.0023
Reward	1.66	1.98	0.6194
Risk	0.57	0.40	0.7181
Present Focus	9.75	13.05	0.0235
Future Focus	1.68	1.33	0.4418
Money	1.54	4.72	0.1995

*The scores reported are the average score of individual emails sent by Skilling and Zuckerberg. They are then tested for the statistical significance of the differences in the average of the scores.

Table 4 presents comparison of the traits for the CEOs of Enron and Facebook. For the purpose of this comparison, we first analyze each email of the executives separately. Then we compute the average score of the traits for the executives. We then perform t-test to examine whether the average score of the executives is significantly different¹⁴. We highlight the attributes that are significantly different (having p-value less than 0.05) in bold.

As seen in table 4, there are significant differences in the attributes of the two executives such as Analytic, Negotiations, Positive Emotion, Tentative, Power, and Present Focus. Zuckerberg exhibited more uncertainty (Negation, Tentative) in his messaging than Skilling. Skilling showed more of power orientation in his messages. A strong Power trait indicates that Skilling, as CEO, used a management style that exercised higher levels of direct control over his subordinates. When an executive uses a high level of control, he is confident in his decisions, a trait Skilling exhibited. This difference in traits between Skilling and Zuckerberg is particularly apparent in the scores for Tentative. A higher percentage of the words used by Zuckerberg (4.94) are associated with tentativeness. Thus, Zuckerberg may be more open in making decisions. He exhibits less control and may be more receptive to the perspectives of his team in decision-making. Skilling is not tentative in his messaging (1.29). Although the percentage scores show there is not much of a comparative difference between the two executives in future focus, there is

¹⁴ To make the comparison, each executives' score on each email message was separately averaged for the nineteen characteristics. For example, all emails of Zuckerberg are analyzed separately to get the score for each attributes. Then we compute the average score of each attribute for Zuckerberg. We do the same for Skilling and perform t-test to compare the average score. This approach is used for the results presented in Tables 5, 6, 7, and 8.

a significant difference in their focus on the present. Zuckerberg is much more focused on the present than Skilling, who may have other executive interests.

Table 5 presents the comparison of the traits depicted in the emails of Enron and Facebook executives. The average score presented in the table is computed by separately analyzing each email of the executives in the respective companies. After that, we compute the average score in each trait based on the scores for each email. Traits that are significantly different (p-value less than 0.05) are presented in bold.

Table 5. A Comparison of the Average of all LIWC percentage scores for all Enron and Facebook messages and a determination of their statistical difference using t-tests.

Traits	Average Score for Enron Emails	Average Score for Facebook Emails	P-Value (two Tail)
Word Count	15238	14109	n/a
Analytic Thinking	74.46	68.07	0.0761
Clout	72.16	80.03	0.0231
Authentic	30.52	23.35	0.0656
Tone	72.73	80.87	0.0558
Negations	1.00	0.98	0.9326
Number	2.10	1.31	0.0522
Positive Emotion	5.13	5.27	0.8688
Negative Emotion	0.54	0.26	0.0221
Anxiety	0.04	0.10	0.2039
Anger	0.05	0.08	0.6449
Sadness	0.24	0.05	0.0153
Insight	2.15	2.46	0.3156
Tentative	2.02	3.79	0.0016
Power	3.04	1.91	0.0152
Reward	1.16	1.77	0.1058
Risk	0.54	0.22	0.0231
Present Focus	11.05	12.97	0.0387
Future Focus	1.77	1.87	0.7555
Money	2.89	1.25	0.0001

*The scores reported are the average scores of the individual emails sent by Enron and Facebook executives. They are then tested for the statistical significance of differences in the average of the scores.

As seen in Table 5, there is a significant difference in the Negative Emotion, Sadness, Tentative, Power, Risk, and Money traits between Enron and Facebook emails. Taken together, it paints a picture of Enron's executives as being more neurotic, power/dominant and risk orientated, along with having a direct concern about themselves. Facebook executives exhibit significantly less of these traits as compared with Enron. In addition, Facebook executives exhibit a lower level of self-assurance in their messaging compared with the high level of confidence found in Enron's emails. The Enron executives involved in the fraud exhibit a strong sense of self-assurance, confidence in

their beliefs, and a display of power, which are not observed among Facebook executives. The combination and interaction of these traits led to the fraudulent financial scheme that defrauded investors. After the extent of fraud was revealed, an example of Skilling's completed confidence in himself led to his 2002 testimony in front of Congress.¹⁵

In Table 2, Fastow seems to be an outlier compared to the other Enron executives, specifically for Clout, Negative Emotion, Sadness, and Tentative. To examine whether Fastow is driving the results in Table 5, we run the analysis without Fastow included. The results of the additional analysis are qualitatively similar to that in Table 5 except for the "Clout" trait. "Clout" is no longer significant when Fastow is excluded from the analysis suggesting that Fastow is driving the difference.

The Enron Executives

So far, our discussion and analysis have focused on the differences between Enron and Facebook executives. We next look at the potential differences between the Enron executives. In this section, we compare and identify differences in the traits of Fastow and Lay and Fastow and Skilling. This comparison has been presented in Tables 6 and 7 respectively. These executives played key roles in shaping Enron's strategic plans and were involved in the fraud, making it valuable to compare their trait patterns.

Fastow's traits have so far shown a stronger negative orientation. As CFO, he was involved in keeping losses off Enron's books. Therefore, comparing his attributes with the other two executives has the potential to reveal some differences in their personality traits. As seen in Tables 6 and 7, Lay and Skilling have more clout than Fastow. Therefore, we can expect the power in the relationship to be on the side of Lay and Skilling. One trait that a CFO should possess is "Authentic" which measures honesty and full disclosure in messaging and relationships. Consistent with this expectation, Fastow shows a significant difference on this trait with Lay (p-value 0.0263) and a higher orientation toward this trait compared to Skilling but not statistically significant.

Table 6. A Comparison of the Average of all LIWC percentage scores for Fastow and Lay and a determination of their statistical difference using t-tests. *

Traits	Average Score for Fastow Emails	Average Score for Lay Emails	P-Value (two Tail)
Word Count	467	10502	n/a

¹⁵"Skilling is the only member of Enron's senior management to testify before Congress, appearing before a House subcommittee earlier this month". (April Witt. February 27, 2002. Skilling's Testimony Met With Skepticism. The Washington Post. Accessed April 8, 2019. https://www.washingtonpost.com/archive/politics/2002/02/27/skillings-testimony-met-with-skepticism/b65cc993-4017-4735-a49c-012e027e61d2/?noredirect=on&utm_term=.be783f4d6694)

Analytic Thinking	51.29	80.69	0.0000
Clout	43.72	83.65	0.0002
Authentic	46.66	24.15	0.0263
Tone	42.54	74.09	0.0196
Negations	3.63	0.61	0.0028
Number	1.32	2.14	0.2451
Positive Emotion	2.43	5.54	0.1168
Negative Emotion	1.54	0.57	0.1970
Anxiety	0.00	0.13	0.0242
Anger	0.26	0.04	0.2359
Sadness	0.79	0.17	0.2598
Insight	1.65	2.36	0.4091
Tentative	4.83	1.54	0.0497
Power	1.64	3.72	0.0330
Reward	1.70	1.06	0.6070
Risk	0.71	0.57	0.7057
Present Focus	16.55	10.36	0.0200
Future Focus	1.04	2.30	0.0094
Money	3.22	3.97	0.5644

*The percentage scores are the average score of individual emails sent by Fastow and Lay. They are then tested for the statistical significance of differences in the average of the scores.

A closer look at tables 6 and 7 indicates that Fastow exhibits a higher prevalence of negative traits compared to both Lay and Skilling. He has less Clout than Skilling and Lay. Fastow has exhibited little self-assurance, along with many inhibitions (Table 2). As CFO, he appears to have a compliant role to the other two executives. There are indications that when a relationship such as this one exists between a group of executives, the less powerful executive is strongly under the influence of the more powerful executives. It appears that both Lay and Skilling had control and power over Fastow.

Table 7. A Comparison of the Average of all LIWC percentage scores for Fastow and Skilling and a determination of their statistical difference using P-Value tests of significance. *

Traits	Average Score for Fastow Emails	Average Score for Skilling Emails	P-Value (two Tail)
Word Count	467	2528	n/a
Analytic Thinking	51.29	77.53	0.0021
Clout	43.72	83.42	0.0002
Authentic	46.66	29.77	0.1403
Tone	42.54	76.04	0.0249
Negations	3.63	0.23	0.0008
Number	1.32	2.46	0.1574
Positive Emotion	2.43	5.86	0.0276
Negative Emotion	1.54	0.60	0.2304
Anxiety	0.00	0.00	-
Anger	0.26	0.00	0.1574
Sadness	0.79	0.43	0.5371

Insight	1.65	1.61	0.9657
Tentative	4.83	1.29	0.0357
Power	1.64	4.50	0.0196
Reward	1.70	1.66	0.9808
Risk	0.71	0.57	0.7878
Present Focus	16.55	9.75	0.0037
Future Focus	1.04	1.68	0.2558
Money	3.22	1.54	0.2558

*The percentage scores reported are the average score of individual emails sent by Fastow and Skilling. They were then tested for the statistical significance of differences in the average of the scores.

VI. Limitations of the study

One weakness of context analysis is that positive words are used in phrases that might have negative connotations. For example, the word “cool” as related to a person might have the meaning that a person is “really hip” or that they are standoffish. Research related to determining the connotation of a word based on the phrase in which the word appears is still under development (Wilson et al., 2009). The corpus used in LIWC2015 attempts to overcome this problem by considering the phrase not just the word in isolation. Further, there are cultural differences and use of organization-specific slang in different organizations that may cause subtle differences in the language used in each organization that may not be captured by LIWC2015. Another limitation of context analysis is coding bias due to the subjectivity among the various linguists performing the coding. LIWC2015 has been consistently used for coding traits since the 1990s, and it is a widely recognized research tool used for context analysis. Therefore, coding bias is not considered an issue.

Another aspect of context analysis that limits its utility is its inability to establish causality. Consequently, the traits identified cannot be associated with specific causes. Hence, we refrain from speculating on the causality of the identified traits in the datasets.

Most of the Enron emails analyzed in this study were composed in the year 2001. It is possible that the linguistic characteristics revealed in these emails are amplified because the executives were trying to contain the consequences of the negative publicity and scrutiny. Researchers may examine whether and how linguistic characteristics change when someone is under scrutiny by examining the linguistic characteristics of written communication pre- and- post- public knowledge of financial misconduct and investigation.

Conclusion

It has been shown that there are important traits associated with Skilling and Zuckerberg, especially, those that are significantly different. As leaders of the companies, they exhibit personal traits that steer the companies in different directions. Skilling's traits reflect those of someone who was highly motivated by Power who had high levels of confidence in his actions and who wanted control over his employees. Although confidence is important for a leader of a company, overconfidence can lead to myopic decisions made without consultation. The traits associated with Skilling are those of a narcissistic personality and may be linked with fraudulent behavior. Zuckerberg appears to be a leader who is interested in consultative relationships. The traits identified in his emails show that he is hesitant in making decisions without consultation, and his management style is more participative than those executives who controlled Enron.

It is also apparent that Andy Fastow did not have the Power or Clout to change the operational strategies toward which Enron was being directed. Lay's and Skilling's Power and Clout trait levels were much stronger than Fastow's. This disparity may have implications as to the source of the fraudulent activities at Enron.

Future researchers should consider the application of other linguistic tools for evaluating fraudulent traits in other organizations with the purpose of developing a more extensive dataset and list of "fraudulent" traits. The results in the paper indicate that the "red flag" traits of fraud, such as narcissism, exist and can be identified using LIWC2015. If there is a suspicion of corporate fraud within an organization, context analysis can serve as a tool to explore further into the suspicion. Further, auditors who are concerned about fraud within their auditee could analyze the available written communication of the executives to evaluate potential fraudulent traits. Context analysis can also be used as a preemptive tool by forensic fraud investigators. It will not identify the fraud, but through the analysis of C-suite writing samples, it can determine if concerns about potential fraud within an organization should be amplified.

Context analysis can be a tool for forensic profiling. However, intracompany communications are not publicly available, and accessing them internally can be challenging. In this age of digital tracking and surveillance, people are increasingly concerned and suspicious of any measure to monitor their activities and communication. Therefore, accessing the communication of individuals within an organization for the purpose of contextual analysis can pose significant ethical and legal challenges. Consideration and research can be directed towards using publicly available communications such as earnings calls and other public communications of the executives.

Context analysis can be used to identify traits in numerous documents. Some studies based on financial reports have been described in the paper. Yet, there are other areas for such analysis. For example, editorials in accounting

journals, corporate conference calls, review notes, and reports in the financial press can provide a rich source of information about the underlying attributes of the writer. Such an analysis can help identify biases of the writer that may not be initially apparent.

Conflict of Interest: The authors reported no conflict of interest.

Data Availability: All data are included in the content of the paper.

Funding Statement: The authors did not obtain any funding for this research.

References:

1. Abrahamson, E., & Park, C. (1994). Concealment of negative organizational outcomes: An agency theory perspective. *Academy of Management Journal*, 37(5), 1302-1334.
2. Argamon, S., & Levitan, S. (2005, June). Measuring the usefulness of function words for authorship attribution. *Proceedings of the Joint Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing*, 1-3.
3. Aven, B. L. (2015). The paradox of corrupt networks: An analysis of organizational crime at Enron. *Organization Science*, 26(4), 980-996.
4. Blickle, G., Schlegel, A., Fassbender, P., & Klein, U. (2006). Some personality correlates of business white-collar crime. *Applied Psychology*, 55(2), 220-233.
5. Boyd, R. L., & Pennebaker, J. W. (2015). Did Shakespeare write Double Falsehood? Identifying individuals by creating psychological signatures with text analysis. *Psychological science*, 26(5), 570-582.
6. Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: A new approach to personality in a digital world. *Current opinion in behavioral sciences*, 18, 63-68.
7. Boyd, R. L., & Schwartz, H. A. (2021). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*, 40(1), 21-41.
8. Brown, C. R., Watkins, A., & Greitzer, F. L. (2013, January). Predicting insider threat risks through linguistic analysis of electronic communication. *46th Hawaii International Conference on System Sciences* 1849-1858.
9. Chen, S. (2010). The role of ethical leadership versus institutional constraints: A simulation study of financial misreporting by CEOs. *Journal of Business Ethics*, 93, 33-52.

10. Chung, C. K., & Pennebaker, J. W. (2018). What do we know when we LIWC a person? Text analysis as an assessment tool for traits, personal concerns and life stories. *The Sage handbook of personality and individual differences*, 341-360.
11. Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, 15(10), 687-693.
12. Demers, E. & Vega, C. (2008). Soft information in earnings announcements: news or noise?, available at:
<https://www.federalreserve.gov/econres/ifdp/soft-information-in-earnings-announcements-news-or-noise.htm>
13. Diesner, J., Frantz, T. L., & Carley, K. M. (2005). Communication networks from the Enron email corpus "It's always about the people. Enron is no different". *Computational & Mathematical Organization Theory*, 11, 201-228.
14. Enron email dataset, <http://www.cs.cmu.edu/~enron>. Accessed February 5, 2019.
15. Facebook email dataset,
<https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Note-by-Chair-and-selected-documents-ordered-from-Six4Three.pdf>. Retrieved: February 16, 2019.
16. Fox, E. W., Short, M. B., Schoenberg, F. P., Coronges, K. D., & Bertozzi, A. L. (2016). Modeling e-mail networks and inferring leadership using self-exciting point processes. *Journal of the American Statistical Association*, 111(514), 564-584.
17. Fragale, A. R., Sumanth, J. J., Tiedens, L. Z., & Northcraft, G. B. (2012). Appeasing equals: Lateral deference in organizational communication. *Administrative Science Quarterly*, 57(3), 373-406.
18. Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of personality: A 10 year review. *Social and personality psychology compass*, 7(3), 199-216.
19. Gahrooei, M. R., & Paynabar, K. (2018). Change detection in a dynamic stream of attributed networks. *Journal of Quality Technology*, 50(4), 418-430.
20. Hales, S. and Rabey, G. (2011) The frontline manager: fronting up to organizational change. *Industrial and Commercial Training*, (43)6, 368-376.
21. Holtgraves, T. (2011). Text messaging, personality, and the social context. *Journal of Research in Personality*, 45(1), 92-99.
22. Ingram, R., & Frazier, K. (1983). Narrative disclosures in annual reports. *Journal of Business Research*, 11(1), 49-60.

23. Johnson, E. N., Kuhn, J. R., Apostolou, B. A., & Hassell, J. M. (2013). Auditor perceptions of client narcissism as a fraud attitude risk factor. *Auditing: A Journal of Practice & Theory*, 32(1), 203-219.
24. Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33(2), 125-143.
25. Kang, H., Plaisant, C., Elsayed, T., & Oard, D. W. (2010). Making sense of archived e-mail: Exploring the Enron collection with NetLens. *Journal of the American Society for Information Science and Technology*, 61(4), 723-744.
26. Kurihara, K., Y. Kameya, & T. Sato (2006). A frequency-based stochastic blockmodel. *Workshop on Information Based Induction Sciences*.
27. Li, J. & M. Chignell (2010). Birds of a feather: How personality influences blog writing and reading. *International Journal of Human-Computer Studies* 68(9), 589-602.
28. Martínez-Huertas, J. Á., Moreno, J. D., Olmos, R., Martínez-Mingo, A., & Jorge-Botana, G. (2022). A Failed Cross-Validation Study on the Relationship between LIWC Linguistic Indicators and Personality: Exemplifying the Lack of Generalizability of Exploratory Studies. *Psych*, 4(4), 803-815.
29. McCallum, A., Wang, X., & Corrada-Emmanuel, A. (2007). Topic and role discovery in social networks with experiments on enron and academic email. *Journal of artificial intelligence research*, 30, 249-272.
30. Mohammad, S. M. (2012). From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53(4), 730-741.
31. Newman, M., J. Pennebaker, D. Berry, & J. Richards (2003). Lying words: Predicting deception from linguistic styles. *Personality and Psychology Bulletin* 29(5), 665-675.
32. Oberlander, J., & Gill, A. J. (2006). Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discourse processes*, 42(3), 239-270.
33. Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of research in personality*, 36(6), 556-563.
34. Pennebaker, J. W. (2011). The secret life of pronouns. *New Scientist*, 211(2828), 42-45.
35. Pennebaker, J. W., & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science*, 10(3), 90-93.

36. Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6), 1296.
37. Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.
38. Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PloS one*, 9(12), e115844.
39. Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, 54(1), 547-577.
40. Pollach, I. (2012). Taming textual data: The contribution of corpus linguistics to computer-aided text analysis. *Organizational research methods*, 15(2), 263-287.
41. Rijsenbilt, A., & Commandeur, H. (2013). Narcissus enters the courtroom: CEO narcissism and fraud. *Journal of business ethics*, 117, 413-429.
42. Ritter, R. S., Preston, J. L., & Hernandez, I. (2014). Happy tweets: Christians are happier, more socially connected, and less analytical than atheists on Twitter. *Social Psychological and Personality Science*, 5(2), 243-249.
43. Robinson, M. D., Boyd, R. L., & Fetterman, A. K. (2014). An emotional signature of political ideology: Evidence from two linguistic content-coding studies. *Personality and Individual Differences*, 71, 98-102.
44. Roose, K. (2018). Facebook Emails Show Its Real Mission: Making Money and Crushing Competition. *The New York Times*, <https://www.nytimes.com/2018/12/05/technology/facebook-emails-privacy-data.html> accessed on, 6.
45. Salami, A. (2014). Factors Affecting Effective Organizational Communication?. In *Academy of Management Proceedings* (Vol. 2014, No. 1, p. 15746). Briarcliff Manor, NY 10510: Academy of Management.
46. Schrand, C. M., & Zechman, S. L. (2012). Executive overconfidence and the slippery slope to financial misreporting. *Journal of Accounting and Economics*, 53(1-2), 311-329.
47. Smith, G. (2017). The accountant: A character in literature. *Meditari Accountancy Research* 25(1), 2-27.
48. Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24-54.

49. Tumasjan, A., Sprenger, T., Sandner, P., & Welp, I. (2010, May). Predicting elections with twitter: What 140 characters reveal about political sentiment. *Proceedings of the International AAAI Conference on Web and Social Media*, 4(1) 178-185.
50. Underwood, M. K., Rosen, L. H., More, D., Ehrenreich, S. E., & Gentsch, J. K. (2012). The BlackBerry project: capturing the content of adolescents' text messaging. *Developmental Psychology*, 48(2), 295.
51. Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3), 399-433.
52. Wright, D. (2014). Stylistics versus Statistics: A corpus linguistic approach to combining techniques in forensic authorship analysis using Enron emails (*Doctoral dissertation*, University of Leeds).