Linguistic errors and investment decisions : the case of ICO white papers

Reference:
Thewissen Jennifer, Thewissen James, Torsin Wouter, Arslan-Ayaydin Özgür.- Linguistic errors and investment decisions : the case of ICO white papers
The european journal of finance - ISSN 1466-4364 - (2022), p. 1-43
Full text (Publisher's DOI): https://doi.org/10.1080/1351847X.2022.2075780
To cite this reference: https://hdl.handle.net/10067/1884270151162165141

Linguistic errors and investment decisions:

The case of ICO white papers

Jennifer Thewissen
University of Antwerp
Université Saint-Louis
jennifer.thewissen@uantwerpen.be

James Thewissen
Université catholique de Louvain
james.thewissen@uclouvain.be

Wouter Torsin
HEC Management School, University of Liège
wtorsin©uliege.be

Özgür Arslan-Ayaydin
University of Illinois at Chicago
orslan@uic.edu

Linguistic errors and investment decisions:

The case of ICO white papers

ABSTRACT

Drawing on language expectancy theory, we predict that linguistic errors in ICO white papers negatively impact investors' willingness to financially contribute to ICO projects. We manually annotate a sample of 546 ICO white papers according to 13 different error subcategories related to spelling and grammar. The error-annotated data are subsequently submitted to regression analyses which confirm that linguistic errors discourage potential investments in ICOs. Specifically, our analyses reveal the presence of "high penalty" vs. "low penalty" errors which result in higher vs. lower financial investment losses for the ICOs. The negative impact of language errors is stronger when ICO white papers are (1) written in native English-speaking countries and (2) from countries without cryptocurrency regulation. Results from an experiment confirm that this relationship is not driven by the entrepreneur- or investor-specific characteristics. Overall, we highlight that the reader identifies linguistic errors as a major 'red flag' that ultimately affects financial decision-making.

Keywords: Initial coin offerings, white papers, linguistic errors

JEL Codes: G32, F30, G41, M20, O16.

"Good grammar is credibility, especially on the internet. In blog posts, on Facebook statuses, in e-mails, and on company websites, your words are all you have. They are a projection of you in your physical absence. And, for better or worse, people judge you if you can't tell the difference between their, there, and they're."

- *Kyle Wiens (2012)*
Harvard Business Review

1. Introduction

This paper is at the crossroads between the finance literature and English linguistics and addresses a hitherto unexplored topic, namely the relationship between formal and grammatical errors[1] (henceforth, linguistic errors) found in business communication and their impact on investors' behavior. A growing body of research indicates that narratives and the associated use of language help leverage resources by conveying a comprehensible identity to a firm (Lounsbury and Glynn 2001; Martens et al. 2007). Narrative linguistic attributes such as style, tone and readability are shown to have a tangible impact on subsequent investment decisions (Henry 2008; Boudt and Thewissen 2019; Chen et al. 2020; De Amicis et al. 2020; Gao et al. 2021). However, there is little evidence concerning the impact of the linguistic errors found in business communication on the ability to raise capital. Yet, studies in linguistics, psychology and communication highlight the association between linguistic errors and their negative impact on the reader's perception of the quality of the information provided (Kreiner et al. 2002) and the author's ability, skill and cognitive intelligence (Varnhagen 2000). Accurate language use, on the other hand, has been shown to give an impression of professionalism, eloquence, credibility and capacity for higher-order thought (Varnhagen 2000; Figueredo and Varnhagen 2005). This study builds on language expectancy theory (LET) and examines whether the presence of linguistic errors in the Initial Coin Offerings' (ICO) prospectus, or white paper,

[1] Formal errors include spelling errors, but also duplicated words (e.g., **the the*), missing words (e.g., *this the main asset* where the verb *is* is missing) and morphological errors *(* unuseful* instead of *useless*). Grammatical errors are operationalised as subject-verb agreement problems (e.g. *it *present* instead of *it presents*). Section 4.3 and Table 2 outline the full error categorization system used.

serves as a signal of quality, and subsequently affects investment decisions during the ICO fundraising process.

ICOs are predominantly ventures at a very early stage (Fisch 2019; Allen et al. 2021), during which an issuing party sells digital assets ('tokens') to investors over a pre-defined period with the objective of raising funds to develop a project. The owners of the tokens can use these to obtain the entrepreneurs' services at a later stage, or trade the tokens on secondary markets. Crypto-tokens are liquid and face few obstacles in international transactions, which explains their popularity and their market value of $31 billion between 2016 and 2019 (Howell et al. 2020). Typically, an ICO process starts with the publication of its white paper. Available on the project's website, this document promotes the project to investors and provides information on the project's value proposition, technical features, team, background and objectives (Cerchiello et al. 2019). Aside from the details provided in white papers, information for investors is scarce as they cannot rely on detailed due diligence, which is common in venture capital transactions. Unlike crowdfunding, there is no mediating platform with an inherent incentive to weed out bad actors (Agrawal et al. 2014), nor is there a regulatory body such as the Securities Exchange Commission assessing the quality of the information provided (Howell et al. 2020). This lack of reliable and credible information is amplified by the fact that there is little direct access to issuers, who represent predominantly early-stage ventures without proven track records and developed products (Fisch et al. 2021). This asymmetry of information causes heightened uncertainty about the underlying quality of the entrepreneurial project, team and in turn, about the reliability of the information provided in the white paper (Blaseg 2018). Altogether, this *New Digital Wild West* is highly risky, opaque and unregulated, which renders the identification of high-quality ICOs a challenging task for both investors and regulators (Robinson 2018).

White papers matter to investors. A survey conducted by Fisch et al. (2021) shows that white papers are read by a large proportion of investors. Specifically, 31.5% of the survey respondents indicate that they not only read the document in detail, but also try to understand everything, while 49.1% read the white paper and try to understand its general content. Given the centrality of white papers as a source of information in the ICO process, past research mainly looks for signals of ICO quality in the linguistic cues contained in the narratives of ICO white papers. For example, Zhang et al. (2019), Samieifar and Baur (2020), Feng et al. (2019) and Dittmar and Wu (2019) study white papers and document that the textual style characteristics, such as the readability, number of words and tone of the narrative, help identify successful ICOs. Altogether, prior research concludes that written/linguistic attributes contain credible signals that investors pay attention to and that influence the investment chances of particular ventures. However, in the context of token sales, signalling is very cost-efficient and easy to imitate. Given the fierce competition for growth capital and the fact that token offerings are often designed in such a way that ventures can raise funds for a specific project only once, there persists an incentive among entrepreneurs to send positively biased signals to increase the expected funding amount by exaggerating statements in the white paper (Momtaz 2020b). This flexible feature of the disclosure content therefore calls for a more extensive and robust set of dimensions along which to analyse narratives (Hoberg and Lewis 2017). In this paper, we answer this call and focus on linguistic errors as a signal that, by definition, is unlikely to be voluntarily introduced in the narrative of the financial communication to investors.

Drawing from LET (Burgoon and Miller 1985), our primary prediction is that linguistic errors will negatively impact investors' incentive to contribute to an ICO project. LET dictates that if language use violates what is expected as appropriate communication behavior, this will impact the persuasiveness of the message and might cause a change in attitude opposite to that which the message aims to achieve. Given that linguistic errors suggest a lack of skill, legitimacy, eloquence and credibility (Varnhagen 2000; Figueredo and Varnhagen 2005), it is possible that they violate investors' decisions about what constitutes a well-crafted and persuasive communication attempt, thereby leading to a reduction in the amount raised during the fundraising period. Alternatively, there may be no significant impact on investor behavior for several reasons. In particular, the cross-border and decentralized nature of ICOs makes the identification of the origin of investors and their mastery of English difficult to establish. As a result, we cannot determine the extent to which the average ICO investor understands the lexical and semantical complexities of the English language. Moreover, not every linguistic error may impact investor behavior in the same way, as, for example, the omission of a hyphen might be negligible compared to wrong verb conjugation.[2] Finally, given the highly international and uncertain ICO context, many factors of ICO quality co-exist and it may be that other factors of quality subdue the signal sent by linguistic errors. For instance, Shrestha et al. (2021) find that the information value of heuristic signals is mitigated by the presence of ICO regulation. Against this backdrop, whether and when linguistic errors determine decisions to invest in an ICO requires an empirical investigation.

We manually error annotate a sample of 546 ICO white papers for formal and grammatical errors. All the papers are written in English, and issued between 2015 and 2020. About 82.8% of the white papers have at least one linguistic error. The average number of errors among the white papers is slightly over eight, with one single paper including more than 100 errors. Because each type of error is unlikely to have the same value-relevance to investors, we decided against treating errors as an undifferentiated set and carried out a fine-grained error analysis which distinguishes between 13 error categories (Bestgen and Granger 2011).[3] The most frequent error types in the ICO texts were found to be: (1) missing letters (e.g. **investr*

[2] Interestingly, prior research in linguistics and psychology provides little indication on which type of error is expected to lead to a violation of the reader's expectations regarding the author's ability.

[3] Specifically, we adjust the Bestgen and Granger (2011) taxonomy and include a 13th category that relates to grammatical errors (GVN). The Appendix A2 provides detailed examples from ICO white paper for each error category we consider in this paper.

instead of *investor*), (2) subject- verb agreement errors ('the company **sell* the product', instead of 'the company *sells* the product'), (3) word split errors (e.g. **every one* instead of *everyone*), and (4) errors involving the doubling of consonants (e.g. **paralel* instead of *parallel*). The sum of the 13 distinct errors is subsequently divided by the number of words in each document to obtain a percentage-based measure of linguistic inaccuracy.

After controlling for other linguistic attributes of white papers including tone, length and readability, a series of ICO-specific characteristics, continent dummies, and time fixed effects, we find a negative and economically significant relationship between linguistic errors and the total amount raised by the entrepreneurs. In monetary values, relative to the average amount raised during an ICO ($13 mil.), a one standard deviation increase in errors reduces the total dollar amount raised by 36% (a monetary loss of $4.68 million). Interestingly, when the impact of individual error types is considered, we find that not all error types have the same impact on investors. Three error types can be qualified as "high-penalty" in the sense that they have a significantly negative impact on investment: (i) words with missing letters, (ii) grammatical errors, and (iii) multiple formal errors in the same word. Other error types such as the misuse of an apostrophe or the incorrect splitting of a word, do not appear to influence investors' decisions significantly. We also find some evidence that the marginal effect of an additional error decreases with an increasing amount of inaccuracies in the white paper. Our results are robust to alternative measures of ICO success, sample-selection biases and a variety of different additional controls. We conclude that investors perceive the writing quality of white papers as a signal of the project's quality and subsequently incorporate this dimension of financial communication into their decision-making process.

We conduct a series of sensitivity tests to increase our understanding of the underlying dynamics between investors' behavior and linguistic errors. First, we examine the mitigating impact of ICO regulation on the information value of linguistic errors. Prior literature has long since acknowledged that regulation is instrumental in establishing the credibility of economic transactions by providing a basis for the project's legitimacy to investors (Sutinen and Kuperan 1999; Chelli et al. 2014), which, in turn, increases investor trust (Aldrich and Fiol 1994). Such increase in trust reduces the need for investors to lean on signals when making their investment decisions (Shrestha et al. 2021; Van Dijk and Zeelenberg 2003; Chaiken and Maheswaran 1994). We therefore expect the influence of linguistic errors to be mitigated with the presence of ICO regulation and strong institutions. Based on the classification of ICO regulation proposed by Shrestha et al. (2021) and the proxy for institutional strength proposed by Li and Zahra (2012), we find that linguistic errors in the white papers of ICOs from unregulated countries or from countries with weak institutions have a stronger negative impact on ICO success. Second, we show that the influence of linguistic errors is mitigated for ICOs established in a non-native English-speaking country. This result adds to prior research showing that people have a higher tolerance for errors made by non-native English speakers, and perceive these errors as less bothersome than those made by native speakers (Wolfe et al. 2016).

A potential concern with our results is that unobserved characteristics correlated with linguistic errors (such as the team's competence or investors' knowledge of the English language) explain investors' decision to contribute to an ICO (e.g., Younkin and Kuppuswamy 2018; Mohammadi and Shafi 2018). We therefore correct for this potential endogeneity between linguistic errors and the amount raised during the ICO and conduct three tests to exclude the possibility that alternative variables explain the impact of linguistic accuracy on ICO success. We first follow Momtaz and Fisch (2020) and employ a restricted control function regression (rCF). We then exploit an instrumental variable approach using the use of brackets in the text as an instrument. Finally, we conduct an experiment on Amazon Mechanical Turk (mTurk) where we develop a fictive ICO project pitch with 32 different manipulations, varying in levels of linguistic errors, ranging from zero to six. Each test confirms the presence of a negative and highly significant association between linguistic errors in white papers and investors' willingness to invest in the ICO.

This study contributes to prior literature in several ways. First, our findings shed light on the impact of qualitative information on investors' decision making and adds to prior work that examines the influence of disclosures' tone and readability on investors' decisions (Henry 2008; Loughran and McDonald 2016; Henry et al. 2021). Whereas prior research mostly focuses on the tone or readability of corporate disclosures, there is only anecdotal evidence on alternatives disclosure measures that capture the linguistic aspects of managers' business communications, such as linguistic errors. For instance, Mollick (2014) investigates the dynamics of crowdfunding projects and defines a dummy variable that indicates whether a project's description contains a linguistic error. Gao et al. (2021) further examine the impact of the readability of the description of P2P lending initiatives on investors' likelihood to finance the project. The authors evoke the use of linguistic errors to define their aggregated readability score, but do not directly examine the influence of linguistic errors on investors' behavior. In this paper, we are interested in quantifying the direct monetary impact of linguistic errors on investors' behavior. As such, we go beyond its use as a control or composite variable and study the value-relevance of linguistic errors in a financial context. It must also be noted that our focus on ICOs significantly contrasts with the P2P crowdfunding setting used by Gao et al (2021), where the platform provides guarantees to investors and, therefore, takes the investment risk (Huang et al. 2021a).[4] On the contrary, in the context of ICOs, participants are not lenders, but investors in tradable securities who bear the risk of their decisions. As a result, linguistic errors are likely to constitute a more relevant signal of quality for ICO investors than for lenders on P2P platforms.

[4] In fact, Huang et al. (2021a) discuss that P2P individual lenders do not observe the loan listings, nor make loan decisions.

Second, this paper contributes to our understanding of the fast developing FinTech market in which ICOs have become a major component (e.g., Chod and Lyandres 2020; Corbet et al. 2019). Given their highly uncertain informational and regulatory context, ICOs of high-quality are difficult to identify, which led prior research to primarily focus on signals of ICO quality. Signals such as the entrepreneur's connectedness and loyalty (Benedetti and Kostovetsky 2021; Momtaz 2020a), assessment of ICO team's confidence through their pictures (Huang et al. 2021b), name change (Akyildirim et al. 2020) or venture-specific features (e.g. team size and country-of-origin) have been linked with the amount raised during the ICO (Amsden and Schweizer 2018; Shrestha et al. 2021). Other signals, such as the code quality, as well as website characteristics and the degree of social media presence, have been shown to influence ICO success (Rhue 2018). Prior research also finds that white papers with greater transparency, readability, positivity and technical details encourage investors and signal the venture's future performance (Howell et al. 2020; Zhang et al. 2019; Fisch 2019). Our study complements these results by zooming in on a rather underexplored feature, namely the impact of linguistic errors. Relative to other identified signals, analysing the impact of linguistic errors provides several advantages for market participants and academics. Compared to the white paper's tone (Dittmar and Wu 2019) or readability (Zhang et al. 2019) that can be manipulated by entrepreneurs, linguistic errors are, by nature, involuntary. This suggests that linguistic errors are less likely to be related to a strategic manipulation of the information. In fact, the underlying issue with the signals identified in prior literature is that the ICO market is highly inefficient because investors cannot see through the *cheap talk* by ICO firms (Momtaz 2020b). As a result, it is essential to identify signals of ICO quality that are less likely to be opportunistically set by entrepreneurs. In addition, linguistic errors rely on a set of recognized pre-defined rules and are therefore less susceptible to subjectivity than other measures of textual attributes, such as tone and readability that remain highly debated linguistic attributes.[5] Linguistic errors are therefore more straightforward to identify and may serve as a more robust signal of project quality.

Our third contribution is theoretical and proposes the application of LET to the analysis of financial communication. We argue that linguistic persuasion theories such as LET constitute an appropriate framework to improve the efficiency of financial communication, and thereby positively impact investment engagement. LET highlights the construct of "expectancies", that is to say expectations about the linguistic properties of language in given contexts. Violations of these expectations will affect the persuasiveness of the message. In financial contexts, potential investors (explicitly or implicitly) rely on a series of deeply-rooted expectations and norms regarding how business projects should be efficiently presented. For instance, investors may expect the tone to be positive and engaging, but they also expect written financial documents to have a high degree of readability, thereby reducing the processing load for the reader. In this paper we theorize that entrepreneurs are expected to display low levels or an absence of linguistic errors when they pitch their ventures in ICO white papers. We conclude that an explicit awareness of LET helps shape, hone and refine the persuasiveness of financial communication and better understand firms' ability to raise funds.

This paper is further organized as follows. Section 2 provides a review of the relevant literature and Section 3 develops our hypotheses. Section 4 provides an overview of the data collection and methodology, while Section 5 summarizes the results. Section 6 provides additional tests, while robustness tests are discussed in Section 7. Section 8 tackles endogeneity and sample selection concerns, while Section 9 concludes.

[5] See Loughran and McDonald (2016) and Boudt and Thewissen (2019) for empirical evidence on the subjectivity of readability or tone measures.

2. Literature review

2.1 Prior literature on linguistic errors

While new ventures have been using communication to effectively convince funders about their potential and legitimacy (Cornelissen and Clarke 2010; Lounsbury and Glynn 2001), it is only recently that the finance and entrepreneurship literature has started emphasizing the importance of language in fund-raising (Halbinger and Reichtein 2015; Martens et al. 2007). For instance, Parhankangas and Ehrlich (2014) conduct a content analysis of application materials submitted by entrepreneurs to potential business angels and examine how linguistic attributes affect their likelihood of securing funding. They conclude that business angels prefer investment proposals characterized by a moderate use of positive tone, moderate levels of promotion of innovation, supplication and blasting of competition, and high levels of opinion conformity. Other studies also show how the style of crowdfunding pitches influences the success of fundraising campaigns. For instance, Larrimore et al. (2011) find that, in the context of peer-to-peer lending requests, the use of extended narratives, concrete descriptions and quantitative words that are likely related to one's financial situation has a positive association with funding success. Furthermore, Gao et al. (2021) show that the readability of a description text on the Prosper peer-to-peer platform is positively appreciated by the lenders. Overall, this literature illustrates how individuals can optimize their persuasiveness by monitoring their usage of language in online environments.

In addition to the tone, style and readability of a disclosure, linguistic errors are likely to play a foundational role in effective written business communication and have a profound impact on readers' perceptions (Figueredo and Varnhagen 2005; Praise and Meenakshi 2015). Linguistic errors have been shown to be significantly related to negative impressions regarding the author's ability, precision, and even intelligence levels (Varnhagen 2000); a conclusion that applies across a number of other fields. For instance, in journalism, media texts that include such linguistic errors tend to be perceived as being lower in quality, credibility and informativeness (Appelman and Schmierbach 2017; Formentin, et al. 2021). Seamon (2001) specifically notes that news consumers lose confidence in the news media outlets when they encounter such errors.

In Second Language Acquisition research, formal and grammatical errors tend to be associated with lower to intermediate levels of learner linguistic proficiency (Thewissen 2015). In organization sciences, several studies point out the negative correlation between linguistic errors in professional resumés and subsequent recruitment decisions. For instance, Martin-Lacroux and Lacroux (2016) find that formal errors in application forms had as much of a negative impact on recruiters as a lack of professional experience.

In the finance literature, however, the analysis of the impact of linguistic errors on investors' behavior is more limited. Mollick (2014) studies the dynamics of crowdfunding projects and defines a dummy control variable that indicates whether a project's description contains a linguistic error. They find that the presence of linguistic errors decreases the likelihood of success of the crowdfunding project. Gao et al. (2021) investigate the impact of the readability of the description of P2P lending initiatives on investors' likelihood to finance the project. The authors rely on the use of linguistic errors to define their aggregated readability score, but, as mentioned before, do not directly examine the influence of linguistic errors on investors' behavior.

Overall, while linguistic errors negatively affect the perceptions of journalists, head-hunters, (non)academic readers, as well as teachers, a key question that remains to be answered is to what extent they constitute a signal of quality and influence investors' decision-making in a highly asymmetric informational environment such as ICOs. Adding to the anecdotal evidence of prior literature, this paper draws upon LET to examine whether the presence of writing errors have significant real effects on the amount raised during ICOs.

2.2 Language Expectancy Theory

LET provides a relevant framework to explore how linguistic errors may elicit a negative investor reaction. LET is a formalized message-based model of persuasion about message strategies and attitude of behavioral change, which specifically addresses expectations in language patterns (Burgoon and Miller 1985; Averbeck 2010; Averbeck and Miller 2014). LET assumes that language is as a rule-governed mechanism where one person either follows or violates expectations. In particular, LET focuses on how message features (e.g., intensity, sentiment, word choices, style) violate the reader's expectations concerning the appropriate communication.

Whether intentional or unintentional, violations of the receiver's expectations have implications for the persuasiveness of the message (Burgoon and Miller 1985). Using language that negatively violates expectations (e.g. being aggressive where deemed inappropriate), results in a negative appraisal of the message and the source, which may prompt a behavior in the opposite direction than intended by the source.[6] Extensive research on LET focuses on various expectations regarding language use, such as aggression, opinionated language, fear appeals, language intensity (Burgoon and Miller 1985) and lexical complexity (Averbeck and Miller 2014). Yet, this type of relationship dynamic hardly exists in situations where entrepreneurs are in quest for funding. Based on prior evidence on the impact of linguistic errors on the reader's impression of the writer (Kreiner et al. 2002; Praise and Meenakshi 2015; Varnhagen 2000), we extend LET by examining whether linguistic errors violate investors' decisions in an emerging category of new ventures that persistently suffer from information and regulatory uncertainty, initial coin offerings.

[6] For instance, Burgoon et al. (1988) identify that when female individuals, who are presumed to use less aggressive language choices in persuasive communication, use more instrumental verbal aggression, they were seen as negatively violating expectations. In turn, this results in lower ratings of persuasiveness and source credibility. At the same time, Burgoon et al. (1988) show that male individuals could either use aggressive or unaggressive verbal strategies and still be persuasive.

2.3 Initial Coin Offerings

ICOs are a relatively novel means for early ventures to raise financing (Colombo et al. 2021). Depending on the business model and the underlying purpose of the venture, different types of tokens can be issued. Generally, a distinction is made between utility tokens and security tokens. Utility tokens are arguably the most common form, and are issued with the purpose to be traded later on for a future service. For instance, an ICO has the purpose of creating a mobile application to which access can only be granted in exchange for the correct token. During the coin offering, investors interested in the app, may purchase tokens for access later on and often at a discount or with a purchasing bonus. Once a venture sold a sufficient amount of tokens, it can become listed on token platforms in which investors can also trade these tokens for other forms of cryptocurrency. This type of financing has been particularly popular in the earlier days of digital offerings, because of the large flexibility given the lack of proper regulation (see Momtaz 2020b). Security tokens (also referred to as Security Token Offerings or STOs) approximate a more traditional investment product, in which the value of the token is inherently tied to the value of the company, and which are subject to traditional securities law (Lambert et al. 2021; Momtaz 2021).[7] Ultimately, these offerings allow investors to receive financial rewards such as interest or dividends from the issuing company

Regardless of the nature of the token, parallels can be drawn between ICOs and (i) initial public offerings (IPOs), where firms sell a part of their equity to the public in a stock market and (ii) the crowdfunding market, where entrepreneurs raise money from a heterogeneous set of investors through online platforms. ICOs share key features with both markets. First, ICOs and IPOs are very similar with regard to the concept of listing. In spite of the utility-security nature of the token, investors that no longer wish to contribute to the firms can easily sell their

[7] Note that, another form of utility-token based ICOs are Initial Exchange Offerings. Herein, crypto trading platforms issue tokens directly on their exchanges (against a fee) and allows investors to immediately purchase and trade these tokens with other interested investors.

token on an exchange (Huang et al. 2020). Second, while IPOs are an interesting means for raising capital for established private firms with large underlying assets, ICOs and crowdfunding are particularly interesting for cash-constrained start-up ventures to attract financing from a pool of online investors (Philippi et al. 2021). In the same way as utility tokens can be traded later on for a specific service, crowdfunding backers receive a future product or service in return for their past pledge. Third, in all three forms of financing, disclosure plays a fundamental role in attracting investors. Crowdfunding websites (e.g. Kickstarter) require ventures to fill in information about the project and require the disclosure of potential risks and challenges. Similarly, IPOs issue their financial data and future prospects in a prospectus. The predominant way of communicating the concept and the idea of the coin offering to potential investors takes place through the ICO white paper. The white paper contains the necessary information describing the firm's business model and is the only document available for investors to base their decision-making on. Fisch et al. (2021)'s survey confirms that white papers in general are read by the investors, which is in sharp contrast with investors' limited interest in the disclosures by publicly listed firms.[8]

However, relative to IPOs and crowdfunding, there are several reasons why ICOs constitute an advantageous setting to test the impact of linguistic errors on investors' behavior. Whereas IPOs are subject to securities litigation and crowdfunding campaigns have intermediating platforms that can weed out bad actors (Agrawal et al. 2014), ICOs – and particularly utility token-based ICOs – offer start-up companies a decentralized money supply alternative without a specific regulatory framework (Corbet et al. 2019), which increases investors' risk of losing their money to a potential scam or fraud (Amsden and Schweizer 2018).[9] This absence of regulations also results in a lack of any standardized disclosure format

[8] Loughran and McDonald (2017) identify that the average publicly listed 10-K report is only requested on average about 28 times by investors immediately after its filing at the SEC, putting the extent to which traditional capital market investors care about direct channels of corporate communication to question.

[9] ICOs and cryptocurrencies are very often associated with scams and fraud. For instance, Benedetti and Kostovetsky (2021) indicate that less than half of the ICOs remain active for more than 120 days after they have finished issuing tokens to the

with regard to white papers. In certain instances, a white paper is a one-page document containing basic financial information related to the token sale, while, in other cases, white papers exceed 70 pages. Moreover, Momtaz (2020b) finds that entrepreneurs tend to manipulate white paper content to attract investors by exaggerating the positive features of the project. This form of moral hazard in signaling is particularly pronounced in the ICO market, where mostly uninformed retail investors invest in start-up firms solely based on the white paper, without any investor protection (Momtaz 2020b; Butticè et al. 2021). Given ICOs' lack of regulation and their high potential for fraudulent intent (Fisch et al. 2021), it is likely that investors will be more sensitive to the presence of flaws such as linguistic errors. Errors are, by definition, unlikely to be a strategic choice or exaggeration of the entrepreneur, and are therefore *involuntarily* introduced in the narrative.

3. Hypotheses development

3.1 The impact of language errors on investors' behavior

The ICO market is highly unregulated and suffers from pervasive information asymmetry. Investors have access to relatively little information and base their monetary decisions on white papers. This leads to the expectation that white papers need to be well-crafted documents (that is, void of linguistic errors). Drawing on LET, linguistic errors may reduce the quality of a project, subsequently decreasing the persuasiveness of financial communication and elicit resistance among investors. Accordingly, we expect that the reader identifies linguistic errors as a significant 'red flag' leading to less funding from investors.

On the other hand, it may be that linguistic errors elicit no reaction from the reader at all. Despite being carefully read by investors (Fisch et al. 2021), the discretionary and largely heterogeneous nature of white papers raises the question of whether they contain credible

public. One recent famous example is the case of 'OneCoin', where about $4.9 billion dollars of investments went missing (Marson 2020). We also refer the readers to https://deadcoins.com/, where over 1,000 failed cryptocurrencies and scams are listed based on user input.

information. Theory argues that the disclosure of information leads to liquid and efficient markets (Grossman and Hart 1980; Healy and Palepu 2001), and a central prediction in economics is that such incentives would lead entrepreneurs to provide relevant information to investors voluntarily. However, as often observed in traditional markets, without oversight, in non-ideal conditions, firms are disinclined to make adequate disclosures (Beyer et al. 2010). In fact, Adhami et al. (2018) observe no significant relationship between the availability of white papers on the probability of reaching the stated funding goal, while Momtaz (2020b) observes only a marginal relationship between the number of words in white papers and the time-to-funding and finds that white paper length is unrelated to the amount raised, the time-to-listing, the first-day return, or the first-month token-price volatility. In the presence of incentives to exaggerate positive information such as in the case of ICOs (Momtaz 2020b), the scepticism that surrounds the credibility of narratives means that rational users are more likely to discount information (El-Haj et al. 2019), which may reduce the value-relevance of linguistic cues in white papers. If voluntary white papers are interpreted as being boilerplate disclosures with no real incremental value, investors are unlikely to consider linguistic errors as a signal of ICO quality.[10]

Furthermore, factors of ICO quality co-exist and it may be that other determinants of quality subdue the value of linguistic errors as a key signal of quality. For instance, Shrestha et al. (2021) find that the value of signals such as the country of origin in explaining ICO success is mitigated by the presence of ICO regulation. In addition, the cross-border, international and decentralized nature of ICOs makes the identification of the origin of the crowd of investors and, therefore, their mastery of the English language difficult to establish. This means that we cannot establish the extent to which ICO investors comprehend the lexical and semantical complexities of English, which potentially diminishes the influence of linguistic errors on

[10] Departing from the context of ICO white papers, Park et al. (2010) explore the influence of misspellings in electronic meetings and find that neither participants' satisfaction, or their productivity suffers from linguistic errors.

investors' behavior. Against this backdrop, whether linguistic errors determine investors' decision to invest in an ICO remains an empirical question. We pose the following hypotheses.

Hypothesis 1A – Linguistic errors are not associated with the dollar amount raised during ICO campaigns.

Hypothesis 1B – Linguistic errors are negatively associated with the dollar amount raised during ICO campaigns.

3.2 Cross-sectional analyses

The impact of linguistic errors on the amount raised is likely to vary cross-sectionally with geography of ICOs (see Huang et al. 2020; Shrestha et al. 2021). We predict that linguistic errors should have the greatest impact on ICOs from countries that are (i) unregulated regarding ICOs, (ii) weak in terms of institutional strength and (iii) English-speaking.

3.2.1 ICOs from unregulated countries

The topic of ICO regulation is contentious. Not only are ICOs loosely regulated, but they also enable ventures to gather large amounts of funding, while bypassing the costs relating to compliance and intermediaries (Zetzsche et al. 2017). Conversely, tokens often have little intrinsic value and do not lead to any legal entitlement, which significantly increases the investment risk due to misconduct (Fisch 2019; SEC 2017; Momtaz 2020b). This investment risk is particularly high for non-professional investors who do not have the knowledge and the resources to perform careful due diligence. As a result, authorities have eyed the rise of ICOs rather cautiously, which reflects in the preliminary and fairly inconsistent rules across countries (Nestarcova 2018). While China is taking the decision to ban ICOs, Swiss authorities have defined extensive regulations and guidelines regarding ICOs, curtailing some ambiguities. Other countries, on the other hand, fail to provide any regulation or guidelines.

Hypothesis 1 relies on the notion that investors use linguistic errors to assess the quality of an ICO. However, in addition to this cue, countries are also able to directly ensure investor protection by defining regulations on cryptocurrencies and ICOs. Prior literature has long since acknowledged that regulation is instrumental to the credibility of economic transactions by providing legitimacy to investors (Sutinen and Kuperan 1999; Chelli et al. 2014), which, in turn, leads to investor trust (Aldrich and Fiol 1994). Such increased trust reduces the extent to which investors lean on signals when making their funding decisions (Van Dijk and Zeelenberg 2003; Chaiken and Maheswaran 1994). We therefore predict the influence of signals such as linguistic errors to vary in accordance with the development of ICO regulations. That is, we expect that investors interpret pro-active regulatory steps taken by the authorities concerning ICOs as a source of legitimacy and will be less reliant on signals such as linguistic errors. Our second hypothesis is formulated as follows:

Hypothesis 2 – The negative relationship between linguistic errors and the dollar amount raised during ICO campaigns is stronger for ICOs registered in unregulated countries.

3.2.2 ICOs from countries with weak investor protection

Extending on Hypothesis 2, we further investigate the notion of geographical differences in explaining the relationship between linguistic errors and funding outcome. In particular, we examine whether the institutional strength of the issuing country influences the signal sent by linguistic errors. Perceptions concerning institutional strength are arguably of particular importance in the ICO setting, in which investors make resource allocation decisions based on relatively little verifiable information and nearly no personal contact with the entrepreneur. Institutions have been shown to play a crucial role in determining the reputation of firms in a country (Brammer and Jackson, 2012; Newburry, 2012). As argued by North (1991), "institutions exist to reduce uncertainty in the world", and play a crucial role in increasing trust, which serves as a basis for economic transactions (Shrestha et al. 2021;

Bachmann and Inkpen, 2011; Welter, 2012; Williamson, 1993; Zucker, 1986). Based on North (1991), Shrestha et al. (2021) find that ICOs originating in countries with stronger institutions obtain more funding during the ICO and that this effect is stronger for countries with regulation surrounding ICOs. They conclude that investors rely on preconceptions about the country of origin as a heuristic for the ICO project's unobserved trustworthiness. Under this heightened trust, we therefore expect investors to be less concerned about linguistic errors if the ICO is established in a country with strong institutional strength. We formulate our third hypothesis as follows:

Hypothesis 3 – The negative relationship between linguistic errors and the dollar amount raised during ICO campaigns is stronger for ICOs established in countries with weak institutional strength.

3.2.3 ICOs from English-speaking countries

Given that non-native writers of English are less linguistically articulate than native writers (Planken 2005), a higher tolerance is expected for the linguistic errors made by the non-native writers. Rubin and Williams-James (1997) find that when English instructors are given (fabricated) information about student identities, they are more lenient with essays that are attributed to non-native English writers compared to native English writers. Similarly, Wolfe et al. (2016) report that business people have a higher tolerance for the errors made by non-native English speakers, and perceive these errors as less bothersome than those made by native speakers. This evidence indicates that the reader tends to hold lower expectations regarding the mastery of English by non-native speakers. Following LET, we therefore expect the presence of linguistic errors in white papers of ICOs established in an English-based country to have a more negative impact than those in white papers of ICOs written in non-English-speaking countries.[11]

[11] An ICO stemming from a non-native English-speaking country might still have international English-speaking founders and advisory board members who are not native to the country from where the coin offering is launched. As such, our country-

Hypothesis 4 – The negative relationship between linguistic errors and the dollar amount raised during ICO campaigns is stronger for ICOs established in English-speaking countries.

4. Variable definitions and research design

This section presents the data selection process, variables, methods and descriptive statistics of our sample. The definition of the variables is presented in Table 1.

< Insert Table 1 about here. >

4.1 Sample selection

Collecting ICO data for empirical analyses is a challenging task as ICOs allow ventures to circumvent intermediaries and rely on decentralization. Ventures can directly provide the relevant information on their websites alone, and therefore, a centralized repository with details of all ICOs does not exist. However, given the growing interest in ICO investments, third-party ICO-tracking websites have emerged, which offer detailed information on considerably large pools of ICOs. As our primary source of data, we use the prominent ICO-listing website, ICOBench.com, which is widely used in the literature (for examples of application, see Momtaz 2020b; Fisch 2019; Amsden and Schweizer 2018; Howell et al. 2020). To verify and obtain additional information, we rely on sources such as ICOdrops.com and ICOsbull.com, and also

based measure may capture the investors' expectancy of English proficiency in a conservative manner. We however prefer to adhere to this country-level measure rather than focus on the presence of native founders and board members for two reasons. First, although we believe that investors attach great importance to members of a specific ICO, the level of their involvement is not necessarily verifiable. For example, Benebit, a fraudulent ICO, allegedly used fake board members to appear more legitimate, in which victims even claimed they used fake passports as a proof of identity and that the board-members' images were plucked from a boys' school website (see https://news.bitcoin.com/benebit-ico-runner-2-7-million-investor-funds/). Second, the presence of a legitimate native English-speaking team advisor is not necessarily proof that this person has been involved in the writing/proofreading of the ICO white paper. Advisors are known to join an ICO team because they are paid to simply promote the crypto asset to the community at large, not to effectively contribute to the ICO. For instance, the research team of Cointelligence analysed the Cremit ICO and detected many red flags. To their surprise, Vladimir Nikitin, who is considered a top expert in evaluating and detecting flawed ICOs, was part of the advisory team, raising the question on the effectiveness and involvement of advisory board members (see https://www.cointelligence.com/content/ico-expert-corruption/). Phillip Nunn, another listed advisor on the team was contacted and he mentioned not having any knowledge about his involvement in the project whatsoever. In turn, a country-based measure of non-native English speaking can be considered as a conservative way of the average investors' interpretation of their English proficiency. According to our measure, an ICO stemming from a native English-speaking country is more proficient in English, whereas an ICO stemming from a non-native English-speaking country is expected to be less proficient. If there are native English team members in the latter category, and if this would cause investors to expect higher levels of English proficiency, we would expect to find little statistical difference between the two groups. Therefore, the reported effect in this paper would always be an underestimation of the true economic effect and should not result in an overestimation.

the issuing firms' webpage. We obtain information on a total of 3,900 ICOs launched between August 2015 and June 2020. After requiring the availability of all of our control variables, we obtain a final potential sample of 1,340 ICOs.[12]

We next collect all available white papers. We gather white papers for little under half of the ICOs, which is similar to the percentage retrieved by Momtaz (2020a). It should be noted that, although we do not find white papers for the remaining half, it does not *ipso facto* imply that these projects did not have a white paper during their coin offering. Since we collected the white papers after the ICO, we find that several ICOs no longer have a working link to their white papers, which may be the result of changes in their websites. In such cases, manual hand-collection was attempted. For our final sample, we filter out all the ICOs where the white papers are available, and for which the PDF creation date took place prior to the ICO ending. As such, we keep about 612 projects. Out of these projects, 66 were either not written in English, the file was corrupt and could not be opened, or were not convertible to a *.txt* file (that is, these files were image-based). Ultimately, our final sample comprises 546 distinct ICO projects issued from 74 different countries.[13] Our sample size is comparable to that of Fisch and Momtaz (2020) and Huang et al. (2021b), who analyse 565 and 515 ICOs, respectively.

4.2 Research design

To test Hypothesis 1, we estimate the following cross-sectional regression:

$$AMOUNT\_RAISED = \alpha + \beta \cdot LINACC + \rho \cdot Controls + \varepsilon, \quad (1)$$

[12] Note that, the most crucial variable here is the dollar amount raised as it captures investors' collective behavior. It is this variable that is missing most often on the tracking websites.

[13] It is important to note that our sample consists of ICOs with all available control variables and those having a white paper. While some ICOs may not have had a white paper during the ICO, many other projects remove their white papers from their websites after the completion of the ICO. Furthermore, some white papers are in languages other than English, and we also find some white papers in non-readable files. Given the lack of a consistent factor that adequately describes the selection process and also satisfies the exclusion restriction criteria for our main analyses, we resort to sub-sample analyses as recommended by Puhani (2000) and do not apply a selection model. However, to avoid our results to be driven by this choice, we repeat our main Hypothesis using a two-step Heckman selection model, for which we report the second stage in robustness tests (see Table 8).

where ε are robust standard errors clustered at the country-level.[14] $AMOUNT\_RAISED$ is the natural logarithm magnitude of the success of the ICO and measures the amount collected during the ICO. This proxy for ICO success has been used in several studies in prior research (e.g., Fisch 2019; Adhami et al. 2018), and captures the investors' collective behavior regarding the ICO. $LINACC$ is the linguistic inaccuracy of the ICO white paper, measured as the ratio of the total number of formal and grammatical errors to the total number of words in the white paper (see also Section 4.3). If according to Hypothesis 1A, linguistic errors have no impact on the amount raised, we expect β to be statistically insignificant. Alternatively, if following Hypothesis 1B, the success of the fundraising decreases with the presence of linguistic errors in the ICO white paper, we expect β to be significantly negative. $Controls$ represents a set of determinants that explain the success of the ICO (see Section 4.4), including year fixed effects, and continent and crypto-industry (category) dummies as defined by ICOBench.[15]

Hypothesis 2 through 4 respectively examine how ICO regulation, institutional strength, and the country-of-origin affect the relationship between linguistic errors and the amount raised during the ICO. We test the following model by introducing an interaction term in Equation (1) between $LINACC$ and the variable of interest ($FACTOR$):

$$AMOUNT\_RAISED = \alpha + \gamma \cdot LINACC \cdot FACTOR + \beta \cdot LINACC + \mu \cdot FACTOR + \rho \cdot Controls + \varepsilon, \qquad (2)$$

where ε are robust standard errors clustered by country. $Controls$ is the same set of control variables and fixed effects as in Equation (1). $FACTOR$ either represents the absence of crypto-specific regulation ($UNREGULATED$), a country's institutional strength ($INST$) or whether the ICO is located in countries that are native English speaking (NES). $UNREGULATED$ is a dummy variable equalling one if the ICO had no active ICO regulation during the coin offering

[14] Our main results continue to hold if we (i) do not cluster, (ii) adopt a two-way clustering (country and year), or (iii) use the Huber White standard error without clustering. These results are not reported, but available upon request.
[15] Note that an ICO can be active in multiple categories at the same time, such as "Health" and "Education".

period; zero otherwise. To identify such countries, we follow the methodology proposed by Shrestha et al. (2021).[16] $INST$ is a continuous variable that estimates the strength of institutions of a country. To construct the institution variable, we take the first principal component of six Worldwide Governance Indicators. These are; (i) control of corruption, (ii) rule of law, (iii) government effectiveness, (iv) regulatory quality, (v) political stability, and (vi) voice and accountability (see Li and Zahra 2012; Shrestha et al. 2020). NES is a dummy variable that is equal to one if (one of) the official language(s) of the country where the ICO is established is English; zero otherwise. If the impact of language errors is more pronounced for ICOs conducted in countries without regulations or for ICOs registered in NES countries, we expect a negative coefficient loading on the interaction variable (γ).

4.3 Measuring linguistic errors: A corpus annotation method

The formal and grammatical errors are manually annotated following the linguistic method known as computer-aided error analysis (Dagneaux et al. 1998) or error annotation (Thewissen 2015, 2021). This method proposes a systematic way of detecting, flagging and correcting errors in a corpus of linguistic data. To do this, the ICO white paper PDFs are converted into a *.txt* format ("saved as text") and opened by using Microsoft Word. Because the presence of errors that are flagged by a spellchecker have the largest impact on the reader's perceptions (Figueredo and Varnhagen 2007), we rely on the well-known Microsoft Word spellchecker to objectively identify linguistic errors.

Each type of error is unlikely to have the same value-relevance to investors. We therefore draw inspiration from the error taxonomy of Bestgen and Granger (2011) and carry out a fine-grained error analysis distinguishing between 13 error categories. Each time a formal or grammatical error is detected, we manually insert a corresponding error tag in front of the error; and the correction is inserted following the error between dollar signs, as shown in the

[16] Note that ICOs from countries that legally banned crypto-trading and coin offerings are not included in our sample.

example column in Table 2. This is done for all ICO white papers and means that formal and grammatical errors are not just treated as one broad, undifferentiated category, but are further subdivided to allow for refined analyses into whether certain error types have a stronger impact on investor behavior than others.[17] To define $LINACC$, we take the sum of each subcategory k and divide by the total number of words in the ICO white paper:

$$LINACC = \frac{\sum_{k=1}^{13} FS_k}{TW}, \tag{3}$$

where FS_k refers to each error subcategory k and TW to the total number of words in the ICO white paper. Table 2 describes the error taxonomy applied to identify linguistic errors.

< Insert Table 2 about here. >

4.4 Control variables

To obtain unbiased estimates of the effect of linguistic errors on the funding amount achieved during the coin offering period, we include a number of control variables. Given the novelty of the ICO literature, the list of control variables largely varies across studies, and a consensus in this regard has yet to emerge (Shrestha et al. 2021). Nonetheless, we rely on Fisch (2019), Adhami et al. (2018), Huang et al. (2020, 2021b), and Amsden and Schweizer (2018) in selecting a list of prominent ICO-level variables that are available on ICO tracking websites.

To account for various linguistic attributes in the white paper, we include the text readability ($READABILITY$), proxied by the Flesch reading ease score (e.g., Kincaid et al. 1975). This measure quantifies the difficulty in reading a document and is derived from a linear combination of average words per sentence and average syllables per word. A higher score

[17] Another reason why we do not solely rely on the spellchecker software for our measure of linguistic inaccuracy and perform error tagging is that much of the field-specific lingo applied in the crypto-currency universe would be considered as improper English, but are not necessarily linguistically inaccurate. For instance, many words such as 'tokenization' and 'democratization' do not occur in the dictionary and are consequently flagged by a spellchecker. They however have a semantically valid meaning within the crypto-universe and are written correctly. Moreover, specific to the field of cryptocurrencies, words may be intentionally misspelled. For instance, "*nsure*" is the name of a token and should not be confused with an erroneous version of the word "*ensure*". Overall, blindly relying on an automated spellchecking software would result in a noisy interpretation of linguistic accuracy.

represents a greater ease in understanding the text. We also control for the tone in white paper text ($TONE$) using a bag-of-words approach. To identify positive and negative words in the white paper, we rely on the Loughran and McDonald (2011) dictionaries. We then take the spread in the proportion of positive and negative words in the white paper and divide by the total number of words. Beyond including the logarithm of the total number of words ($TOTALWORDS$), we follow Loughran and McDonald (2016) and include the file size, measured as the logarithm of the document size in bytes of the original pdf ($FILESIZE$).

Apart from the white paper textual attributes, we control for various project characteristics. We include a group of three dummy variables indicating if the issued tokens are built on the Ethereum platform ($ETHEREUM$), if the issuer details a minimum amount that needs to be purchased during the coin offering period ($MINIMUM$) and whether the project implements the whitelist or know-your-customer guidelines ($WHITE\_KYC$). Implementing a Whitelist and Know Your Customer policy in the ICO process is an indication of regulatory compliance. As dealing with cryptocurrencies essentially allows anonymity to buyers, these compliances help ensure the identity of the buyers and mitigate the potential for illicit activities. However, there is little evidence on whether these compliances affect ICO success (Amsden and Schweizer 2018).

Moreover, we include a number of ICO-specific attributes that are regularly used in the empirical literature. Specifically, our models include the dollar price for a token (ICO_PRICE) and dummy variables for whether a hard cap or soft cap was specified ($HARD\_CAP$ and $SOFT\_CAP$). Soft caps indicate the minimum amount targeted to be raised and are found to favourably influence ICO's success (Amsden and Schweizer 2018; Howell et al. 2020). Hard caps indicate the maximum amount that the firm intends to raise. Apart from guiding investors regarding the project feasibility after funding, an upper limit is often established to maintain scarcity and preserve the value of the issued tokens. Furthermore, we distinguish between ICO utility and security token issues (see Block et al. 2020) and include the dummy variable $UTIL/SEC$, which equals one if the ICO issues a utility token.

Prior studies also document the significance of community engagement via the Telegram, Github, Facebook and Twitter social media on ICOs' success (Howell et al. 2020; Sharma and Zhu 2020). We therefore control for this by creating indicator variables for the presence of a Facebook page, Twitter account, Telegram account and a Github page ($FACEBOOK$, $TWITTER$, $TELEGRAM$, and $GITHUB$). We also control for the total logarithm of the number of team members involved in the ICO (N_O_TEAM). A simple headcount of the team could indicate the scope of the project and its capacity to handle the ICO process and the various tasks to successfully materialize the project. In fact, Amsden and Schweizer (2018) and Cerchiello et al. (2018) find a significant positive relationship between the success of ICOs and their number of team members.

In addition, it is likely that investors base their investment decision on a variety of unobservable characteristics pertaining to the project or the team, which are difficult to quantify. To reduce the impact of such omitted variables, we include expert ratings which would capture an overall quality of the project prior to the culmination of the ICO. ICOBench.com rates listed ICOs using a combination of standardized algorithm and independent experts' evaluation. The ratings incorporate various factors, such as the trustworthiness of the team, quality of the product, venture's social media presence and business strategy, and a short legal review. We therefore include the aggregated rating that received by a group of experts from ICOBench (ICO_RATING), alongside the logarithm of the number of experts that rated the project ($N\_O\_EXPERTS$).

Finally, prior literature has acknowledged the importance of geography in driving ICO success (Shrestha et al. 2020; Huang et al. 2020, 2021b). We therefore control for relevant country-specific attributes. First, we include a dummy variable indicating if the project is based in a country considered to be a tax haven (*TAX_HAVEN*). This variable indicates whether the specified ICO country is a tax haven based on a list of 52 tax havens prepared by Hines (2010). Second, we introduce several country-level variables that capture the country-level economic development. These are: (i) the logarithm of the GDP per capita (*GDP*) as well as (ii) the logarithm of the population (*POPULATION*) of the country and year in which the ICO was issued, alongside (iii) the financial development index (*FDI*), which measures the efficiency of financial services meeting business needs as per World Bank (2016) (see Huang et al. 2020). Next, we draw inspiration from Huang et al. (2021b) and include a location dummy (*LOCATION*) for whether the ICO was issued in one of the five countries with the highest total fundraising amount in our sample (which in our case are United States, United Kingdom, Singapore, Switzerland, and Estonia). Second, we follow Shrestha et al. (2021) and, instead of adding country-fixed effects as some countries only have one ICO in our sample, we control for continent fixed effects.

4.5 Descriptive statistics

Table 3 presents the summary statistics of our final sample. We find a large heterogeneity in linguistic errors ($LINACC$) across our sample of ICO white papers. The average white paper includes little over eight linguistic errors (either formal or grammatical). Yet, as indicated by its large standard deviation, this number varies strongly across our sample from zero to 136. When expressed as a ratio relative the total number of words in each document, we find that 0.122% of all words in a white paper are erroneous on average. Overall about 17.2% of all white papers in our sample are error-free. When investigating specific error types, we find that the type of linguistic errors that occurs the most is incorrect splitting/hyphenating of words (*FS_SPLIT*), such as **far reaching* instead of *far-reaching* or **real time* instead of *real-time*. In fact, this component seems to be the largest error category and accounts for about 60% of all errors. Other more frequent errors are the omission of a letter within a word (*FS_MIS*), such as

**refered* instead of *referred* or **choses* instead of *chooses*, the erroneous addition of a letter, such as *distribuition instead of *distribution* (*FS_RED*) and errors pertaining to verb agreement (*FS_GVN*), e.g. *they *amounts*. The least frequent formal errors are morphological errors (*FS_MORPH*) such as *the * trustful* instead of *trustworthy*, the omission of a word or article in the sentence (*FS_MISSING*) such as **nerve-racking* instead of *nerve-wracking*, the erroneous doubling of a letter (*FS_DOUB21*), such as **pllatform* instead of *platform*, and the erroneous duplication of two distinct words (*FS_DUP*), for instance *"Nicola has has experience in [...]"*. Appendix A1 provides an example of each error category found in white papers.

Concerning the control variables, we find that an ICO in our sample raises on average $13 million. We find that 25.4% of the ICOs are issued in countries where English is the native language (*NES*), while 43.3% are from countries with limited ICO regulation (*UNREGULATED*). We further find that 86.5% of the tokens are based on the Ethereum blockchain (*ETHEREUM*). The average token sells for about 1.27 dollars (*ICO_PRICE*) during the coin offering, and about 89.2% (63.9%) of the ICOs have specified a hard (soft) target (*HARD_CAP* / *SOFT_CAP*). About 41% have specified a minimum amount that needs to be purchased (*MINIMUM*), and we observe that the number of team members varies vastly between one and 53 with a median number of 14 (*N_O_TEAM*). With regard to the outsider monitoring of the ICO, it appears that the average ICO obtains a rating of 3.382 out of a possible five (*ICO_RATING*), which stems from an average of about six experts (*N_O_EXPERTS*). Note that there exists a large variation across the number of experts reviewing the projects, which can range from zero to about 96. Moreover, about 70% of the ICOs in our sample employ a whitelist / Know Your Customer policy (*WHITE_KYC*), in which some identifying information about the ICO's customers is being recorded. 98.2% (92.1%) of the ICOs in our sample have a Twitter (Facebook) account and 90.1% (63%) have a Telegram account (Github repository). Finally, we find that the average document tone (*TONE*) of the white papers is positive, the average number of words (*TOTALWORDS*) is about 7,000 and we report a large variation in terms of readability (*READABILITY*), as measured by the Flesch reading ease score. The top five countries (*LOCATION*) cover about 48.5% of our total sample and 33.6% of the ICOs are located in a tax haven country. We find a large standard deviation for our country-level variables (*GDP*, *POPULATION*, and *FDI*), which reflects the large geographical diversity in our dataset.

In Table 4, we compare the linguistic accuracy of ICO white papers between countries where ICOs are regulated, those with higher than-media and lower-than-media institutional strength, and countries where English is one of the official languages. While differences in linguistic errors between regulated (institutional strong) vs. unregulated (institutional weaker) countries are virtually negligible, we observe that linguistic errors are far more present in countries where English is not one of the official languages. In fact, the average number of linguistic errors in ICOs from native English countries equals 5.51 per white paper, compared to 9.32 errors in the Non-English group. The difference is statistically significant at a 99% confidence level (t-test = 3.97).

< Insert Tables 3 and 4 about here >

In Table 5, we report the correlation table of the dependent, independent and control variables. We find that linguistic errors and the amount raised during the ICO are negatively correlated with a coefficient equal to -0.233, which is significant at a 99% confidence level. This result provides initial evidence supporting Hypothesis 1A. On the other hand, we find no statistically significant correlation between the linguistic error variable ($LINACC$) and whether the ICO is from a country with ICO regulation ($UNREGULATED$), and a weak negative correlation between *LINACC* and *INST*. However, we find a positive and significant correlation of 0.251 between *LINACC* and the dummy variable *NES*, which indicates that the white papers of ICOs issued in countries where English is officially one of the native languages tend to be more linguistically accurate. Altogether, our variables are sufficiently correlated, which makes multivariate regressions an appropriate statistical analysis tool.

< Insert Table 5 about here >

5. Empirical results

5.1 The impact of linguistic errors on the amount raised during an ICO

In Table 6, Model (1) reports the results of Equation (1) estimated without the control variables. Consistent with Hypothesis 1B, we find a statistically significant and negative coefficient for the $LINACC$ variable, indicating that a higher level of linguistic inaccuracy is associated with a lower amount raised during the coin offering period. In Model (2) of Table 6, we add the control variables to exclude a potential alternative explanation for our findings, which substantially increases the Adj. R2 from 2.5 to 16.7%. After controlling for the ICO's rating, the white paper's tone and readability, as well as for other ICO-specific characteristics (e.g., ICO_RATING, $TWITTER$, TAX_HAVEN), we find that our results are not driven by other linguistic features contained in the white paper or by the experts' rating of the ICO's quality. Although the estimated coefficient of $LINACC$ is substantially smaller, it remains highly significant at a 95% confidence level. This effect is also economically sizeable. In fact, an increase in the ratio of total errors from Q25 to Q75 corresponds to a decrease in the raised amount of 18.2%.[18] Relative to the average amount raised in our sample ($13.002 million), this drop amounts to a $2.37 million dollar loss in potential revenue for the ICO.

In terms of LET, this result indicates that investors do not expect ICO white papers to include deviations from the linguistic norm and perceive errors as a negative signal about the

[18] More precisely, this is a drop of -18.232% and is calculated as the difference between the Q75 value of our ratio of total spelling errors (0.00173) minus the Q25 value of the ratio of total spelling errors (0.00018), multiplied by the regression coefficient (-117.625). An alternative explanation is that a one standard deviation increase in the total ratio of inaccuracies reduces the total dollar amount raised by about 36% (which translates to a total monetary loss of $4.68 million relative to the average ICO in our sample).

quality of the project. In turn, investors decide against contributing financially to the project, which results in a significantly lower amount raised during the campaign.

< Insert Table 6 about here >

Intuitively, as the number of errors in the white paper increases, we expect the marginal impact of errors to decrease. That is, the interdependence between linguistic errors and the amount raised during an ICO is likely to be non-linear. To test this non-linear relationship, we include in Equation (1) the variable $LINACC^2$ and report these findings in Model 3. We find that, while $LINACC$ remains significantly negative, the coefficient loading on $LINACC^2$ is positive and significant at a 99% confidence level. These coefficients suggest the presence of a curvilinear relationship, confirming that linguistic errors matter less as the total number of errors increase in the white paper. However, if we graphically illustrate the linear and non-linear relationship in Figure 1 for reasonable levels of linguistic errors ($LINACC$) in our sample ([0; .076]), we reach more refined conclusions. The full black line represents the linear impact of Model (1), while the red dashed line highlights the overall effect of linguistic errors when we include the quadratic term in Model (3). We see that the impact is, for reasonable values of $LINACC$, negative in both specifications, but that the red dashed curve is less steep. This indicates that, as $LINACC$ increases, there is a decreasing effect of linguistic errors on the amount raised. The fact that we do not observe a minimum in this range of linguistic errors does not exclude the presence of a curvilinear relationship. It only indicates that errors consistently have a marginally decreasing and negative impact on the amount raised during the ICO, without ever reaching an optimum within a reasonable range of linguistic errors.[19]

< Insert Figure 1 about here >

We further decompose the total measure of linguistic errors into its different components. Interestingly, we find that not all error types result in an investor penalty. Specific

[19] The minimum optimum value lies substantially outside traditional levels of linguistic errors, with 230 errors per ICO white paper, on average.

error types, which we refer to as ‘high-penalty errors’, lead to higher financial losses. High-penalty errors include (1) missing letters, (2) subject-verb agreement errors, and (3) words with multiple errors. Low-penalty errors, on the other hand, did not result in financial losses. These include errors on the use of the apostrophe, the splitting of words, or full missing words in a sentence. One explanation for the lower impact of these error types might be that the readers may not necessarily have spotted them as errors involving apostrophes or the splitting of words are indeed subtler errors that even native-speakers of English are likely to make themselves. In other words, the impact of linguistic errors on ICO success significantly depends on the type of error.

Looking at the control variables, we find that both the size of the team and the rating the ICO obtained lead to higher levels of funding. This can be interpreted from the angle of credibility, in that larger teams and those rated more highly by experts are arguably perceived as being more trustworthy. Pertaining to the white paper, we further show that textual length is significantly associated with success during the coin offering process and find some evidence that more readable white paper reach higher funding, which is in line with the findings of Zhang et al. (2019).

5.2 Factors influencing the effect of linguistic inaccuracies on the amount raised during an ICO

In the previous section, we find that linguistic errors induce investors to diminish their contribution to ICO projects. However, the magnitude of this impact is likely to vary cross-sectionally with ICO-specific characteristics. We next examine the mitigating impact of ICO-related regulation on the value of linguistic errors, and whether investors are less tolerant of linguistic inaccuracies for ICOs that originate from countries where English is (one of) the native languages.

Hypothesis 2 to 4 focus on country-specific moderating factors and argue that the impact of linguistic errors on the amount raised during the ICO is likely to be mitigated in countries that are regulated, with strong investor protection or non-native English-speaking. In particular, Hypothesis 2 predicts that ICO regulation mitigates the effect of language errors on ICO success. The results of Equation (2) are reported in Model 1 of Table 7. The coefficient for the interaction variable $LINACC \cdot UNREGULATED$ is negative and significant at the 95% confidence level. This result suggests that investors of ICOs located in countries without ICO regulation tend to attach more value to signals that can be inferred from white papers, which is in line with the argument of Shrestha et al. (2021) that ICO investors rely more on heuristics to make their investment decision in an unregulated setting.

< Insert Table 7 about here. >

Hypothesis 3 further argues that ICOs from countries with a stronger institutional background infer higher trust among investors and are therefore less likely to be penalized for linguistic errors in white papers. In line with our hypothesis, Model 2 of Table 7 reports a positive and significant coefficient on the interaction variable $LINACC \cdot INST$, which indicates that investors tend to allocate less importance to linguistic errors of ICOs located in countries with a higher institutional strength. Finally, Hypothesis 4 predicts that ICOs originating from a Native English-speaking country are penalized more for linguistic errors than ICOs stemming from a non-Native English-speaking country. Consistent with our Hypothesis, Model 2 of Table 7 reports a negative coefficient loading on the interaction variable $LINACC \cdot NES$, which is significant at a 90% confidence level. This result suggests that the penalty for linguistic errors is stronger for ICOs from countries where English is a native language.

6. Additional analysis – ICOs from sin industries and the position of linguistic errors in the white paper

In this section, we conduct additional analyses to highlight the role played by social norms and the position of linguistic errors in white papers.

6.1 The role of linguistic errors in sin industries

Social norms are a significant driving force of individual behavior (Kübler 2001) and are likely to play a significant role in investment decisions within the ICO market. We therefore question whether ICOs that violate social norms (i.e., ICOs in sin industries) are perceived differently. Prior evidence in the corporate finance literature argues that firms in sin industries (i.e. tobacco, gambling, and alcohol) tend to receive a high degree of scrutiny because of the nature of the products they sell. In fact, a sin firm is a "company under siege from lawyers, politicians and public opinion" (Edgecliffe-Johnson 2001). To compensate this heightened scrutiny, Kim and Venkattachalam (2011) show that managers of sin stocks tend to provide financial information of higher quality to attract a wider investment and analyst base. Particularly, they find that sin stocks have accruals that better predict future cash flows and recognize losses in a timelier fashion relative to a variety of control groups of non-sin firms.

We investigate whether such heightened scrutiny also exists for ICOs in sin industries. We define "sin ICOs" (*SIN*) as projects that engage in the development of activities related to tobacco, gambling and alcohol by reading the white paper. We next test whether linguistic errors in white papers of ICOs from sin industries tend face a larger penalty compared to non-sin ICOs. Our sample contains 6.1% ICOs in sin industries. In line with Kim and Venkattachalam (2011) who find a superior reporting quality for firms in sin industries, the significantly negative correlation factor between $LINACC$ and SIN in Table 5 shows that ICOs in sin industries tend to contain less linguistic errors. Despite this higher quality, in Panel A of Table 8, we find that the coefficient loading on the interaction variable $LINACC \cdot SIN$ is negative and significant at a 99% confidence level. This result is in line with our expectations and indicates that ICOs from sin industries are penalized substantially more for linguistic inaccuracies.

< Insert Table 8 about here. >

6.2 The position of linguistic errors in the white paper

So far, our main analyses assume that all errors in the white paper hold an equal negative impact irrespective of their position in the document. Yet, Boudt and Thewissen (2019) show that the position of words in a text has informative value for investors. The authors rely on the serial position effect and argue that readers recall information better when it is presented last (recency effect) and to a lesser extent when it is presented first (primacy effect) in a vector of words, compared to the middle (Baddeley and Hitch 1977; Glanzer and Cunitz 1966; Roediger and Crowder 1976). To test whether the primacy and recency effects are present, we split each white paper into four parts, each containing the same number of words. We then measure the linguistic errors in each bin (*LINACC_Q1*, *LINACC_Q2*, *LINACC_Q3*, and *LINACC_Q4*). Univariate t-statistics reported in Appendix A2 confirm that there are no significant differences between the occurrences of linguistic errors across the different quartiles. However, Panel A of Table 7 confirms that the errors towards the end of the document tend to have the strongest impact on funding outcome. Overall, similar to Boudt and Thewissen (2019), this result suggests that the position of these errors in the text has a significant impact of investors' perception of the quality of the project and that an error towards the end of the document is better recalled by investors, who penalize ICOs' fundraising outcome by 25% for a one standard deviation increase.[20]

[20] This finding is calculated by multiplying the standard deviation of *LINACC_Q4* (0.0017) with the coefficient reported in Panel A of Table 7 (-146.781), and amounts to 24.9%.

7. Robustness tests

Our main finding that linguistic errors reduce the amount raised during an ICO aligns with our Hypotheses. However, we bear in mind that the exact quantification of this effect depends on the measurement of our dependent variable and is contingent upon sample selection biases. Therefore, in this section, we conduct several tests to ensure the robustness of our results.

7.1 Likelihood of success

Our main analyses investigate the impact of linguistic errors on the amount raised during the ICO. To show that our results do not depend on the definition of the dependent variable, we test our hypothesis using alternative dependent variables. Following Amsden and Schweizer (2018) and Howell et al. (2020), we rely on the identification of successful ICOs as tokens that are traded on a secondary exchange. We first use a binary variable ($TRADED$) indicating whether ICO token is eventually traded on the most prominent exchange (i.e. Coinmarketcap.com). Given that issuing tradable tokens is a common objective of issuers, regardless of the token type, *TRADED* acts as a consistent and unbiased indicator of ICO success. Moreover, identifying successful ICOs as those listed on a popular exchange allows us to incorporate the expert due diligence of a third party to identify successful ICOs (Howell et al. 2020). We run Equation (1) again under a Probit specification, where *TRADED* is the dependent variable. Our results are reported in Panel C of Table 8. These results also confirm that projects with higher linguistic errors achieve lower funding. We also define the variable *HARD_CAP_ACHIEVED* that takes the value of '1' if the amount raised exceeded the hard cap that was pre-specified during the ICO, zero otherwise. Since not all the ICOs have had a predefined hard cap, we continue with a limited sample of 485 ICOs. We find a negative and significant association between *LINACC* and *HARD_CAP_ACHIEVED*, which highlights that ICOs with more linguistic errors in their white paper have a lower likelihood of achieving the pre-defined target amount.

7.2 Sample selection bias

Our analyses may be exposed to selection bias because we were able to find the white paper for half of the ICO projects for which all data was available. This might be a significant concern for our empirical analyses as it might indicate selectivity in our data selection. The source of this non-disclosure can be multiple, such as a fraudulent behavior, a limited development of the project, or a non-working link to an existing white paper that could not be found manually. Another concern is that there might be a limited overlap between the different data sources that we use in this study. Ignoring this component in our data selection might bias our analyses.

The selection model introduced by Heckman (1979) provides a potentially useful tool in this situation, since it allows to both test and correct for potential biases created by unobservable characteristics. We therefore follow Momtaz (2020b) and conduct a Heckman (1979) two-stage regression. The first step (unreported) involves estimating a probit model for the likelihood that a project publishes a white paper, given some project quality characteristics.[21] The second step uses the density and the distribution functions from the first stage to compute Inverse Mills Ratio (*IMR*), which is subsequently included in the subsequent analyses to control for sample selectivity. Model (5) in Panel D of Table 8 shows that our main results remain qualitatively similar using a two-step framework.

7.3 The moderating impact of linguistic features

Thus far, the paper demonstrated that linguistic inaccuracies is negatively associated with funding outcomes. Yet, the extent to which linguistic inaccuracies are perceived as a poor signal of the project's underlying quality is likely to be contingent on other features of the white paper disclosure. To investigate this notion, we repeat our main analysis using a split sample analysis

[21] There exists no clear consensus concerning the control variables to include in the first-stage regression. We use our set of control variables, apart from the different white paper attributes.

based on other linguistic features: (i) the level of technical emphasis (Fisch 2019), (ii) the readability (Zhang et al. 2019), and (iii) the tone of the white paper (Dittmar and Wu 2019.

In order to estimate the technical emphasis of white papers, we borrow the composed word list as per Liu et al. (2021). They exploit a basic glossary blockchain and crypto-specific words as a benchmark to their machine learning based version (the word list includes distinct terms such as *blockchain*, *proof of concept*, *proof of work*, *distributed ledger*, *DDOS*, etc.). We find that the average white paper mentions such terms about 429.5 times (sd 299.8). In Panel A of Table A4, we perform a split sample analysis with the aim of understanding how linguistic inaccuracies manifest themselves over technical vs. non-technical whitepapers. We find a small, yet statistically significant difference in that linguistic inaccuracies occur more frequently in papers with a higher technical emphasis. When splitting the sample based on readability and white paper tone, we find no such differences between the groups. In Panel B, we next perform separate analyses based on a median split of the three distinct linguistic attributes. We find that linguistic inaccuracies play a significant role in explaining the amount raised in the subsamples of higher technical emphasis, lower levels of readability, and lower levels of disclosure tone. Overall, these findings are consistent with more complex, opaque, and less promotional white papers facing a larger penalty for linguistic inaccuracies.

8. Endogeneity concerns

A potential concern with our results is that we treat the variable for linguistic errors as exogeneous. An alternative explanation is that unobserved effects that are correlated with linguistic errors (such as the entrepreneurial characteristics) may explain investors' decision to invest (e.g., Younkin and Kuppuswamy 2018; Mohammadi and Shafi 2018). In fact, it may be that investors are sensitive to unobservable signals and are not necessarily affected by the confounding presence of linguistic errors. In addition, although our theoretical development examines investors' collective behaviour regarding linguistic errors, we are unable to control for individual investor characteristics due to the proprietary and decentralized nature of ICOs. As such, our results may be driven by investor features, such as the investor's mastery of the English language. Therefore, obtaining causal evidence is challenging as investors are likely to base their financial decision-making on a variety of attributes that are not readily quantifiable, but that might be correlated with the number of linguistic errors. In turn, our current economic interpretation may either be a spurious statistical artefact (that is, our significance is a false positive), or driven by an omitted variable bias. Several methodological approaches exist to account for this endogeneity. In line with prior studies (e.g. Colombo and Grilli 2010), we employ a (i) the restricted control function (rCF), (ii) instrumental variable and (iii) an experimental design approach.

8.1 Restricted control function (rCF)

We build on seminal work by Heckman (1978) and revisit our analyses using a two-step restricted control function regression. The rCF approach mitigates the possibility of endogeneity possibly generated by spurious correlations or reverse causality, leading to the "experimental average treatment effect" (Heckman 1990; Colombo and Grilli 2010) of linguistic errors on the amount raised during the ICO. The benefit of the rCF approach over matching methods is that, rather than assuming that the conditioning set of relevant control variables is sufficiently complete, the rCF method models omitted variables (Heckman and Navarro-Lozano 2004). Another advantage is that the generalized residual obtain from the first stage can be regarded as an explicit test for endogeneity (Colombo and Grilli 2010).

In the first step of the restricted control function analysis, we estimate the probability that white papers contain at least one linguistic error, based on the same control variables as those employed in Equation (1). Using this model, we calculate the generalized residual ($GENRES$)

and insert this residual as an additional control variable in a second step regression, where we estimate the impact of linguistic errors on the total amount raised. We report the results of the second step in Model (6) of Table 8 (Panel D), and find that the relationship between *LINACC* and A*MOUNT_RAISED* remains significantly negative.

8.2 Instrumental variable approach

We next use the instrumental variable approach to ensure that unobserved characteristics that may simultaneously explain the linguistic errors and the amount raised (such as the entrepreneur's skill) are not driving our results. A proper instrument exclusively affects the outcome variable (*AMOUNT_RAISED*) through its impact on the independent variable (*LINACC*). As such, it needs to fit both the relevance (that is, the instrument should sufficiently be correlated with *LINACC)* and exclusion criteria (no direct correlation between the instrument and *AMOUNT_RAISED*).

We define our instrument as the number of brackets – that is: '(', ')', '[', and ']' – in the white paper relative to the number of non-alphanumeric characters (*BRACKET*).[22] Brackets are generally used to insert explanations, side-notes, or comments into a sentence, which allows the author to write the text in a more condensed manner, rather than spreading the statement over multiple sentences. The correct use of brackets therefore requires more effort from the author in the structuring of the information. As such, we expect that increased attention should diminish the occurrence of linguistic errors. Conversely, we do not expect the use of brackets to explain the amount raised. That is, these characters themselves contain no information value to explain ICOs' performance.

Our correlation analysis confirms the intuition. The correlation between our instrument, *BRACKET*, and the independent variable, *LINACC*, is significantly negative ($r(544) = -0.235$,

[22] Non-alphanumeric characters are comprised of the following: ! " # $ % & ' ( ) * + , - . / : ; < = > ? @ [\ ] ^ _ \ { | } ~).

p < 0.01). At the same time, there is no significant correlation between *BRACKET* and *AMOUNT_RAISED* ($r(544) = -0.032$, p = 0.459). Given that the instrument appears to fit both the restriction and exclusion criteria, we proceed by estimating the impact of *BRACKET* on *LINACC* as follows:

$$LINACC = \alpha + \beta \cdot BRACKET + \rho \cdot Controls + \varepsilon, \quad (4)$$

where *Controls* represent the same set of control variables as in Equation (1) and ε are robust standard errors clustered at the country-level. We report these result in Model 7 of Panel E in Table 8, in which we omit to report the estimates for the control variables for parsimony. The inclusion of the instrument, significantly improves the baseline specification, with an increase in adjusted R^2 from 0.16 to 0.246. The first-stage F-statistic for testing the hypothesis that the instruments are unrelated to the endogenous regressor suggests that our instrument is sufficiently strong.

We next obtain the fitted values from Equation (4) (*LINACC_ESTIMATED*) and repeat our main analysis using these fitted values. The results are reported in Model 8 of Panel E of Table 8. We find a statistically significant and negative relationship between *LINACC_ESTIMATED* and *AMOUNT_RAISED*. Moreover, the results of the Wu-Hausman test suggests that OLS is an equally consistent method to the IV analysis, which increases our confidence in the inferences of our initial analyses.

8.3 Experimental design

To complement the rCF and instrumental variable approaches, we conduct an experiment to further mitigate our concerns regarding omitted variables. The benefit of an experimental method is that it allows us to exclusively examine the impact of varying levels of linguistic errors, while keeping the ICO-project, the content of the narrative and the entrepreneur's characteristics constant. In addition, we record an array of investor-specific characteristics that we use as additional control variables.

8.3.1 *The experiment and summary statistics*

We develop a fictive short ICO white paper and provide 32 different manipulations with mistakes (EX_ERRORS) ranging from zero to six.[23,24] The white paper describes the ICO and, at the bottom of the document, provides a biography of the author, which is the same across all versions of the white paper (and in which we never introduce linguistic errors). We administered these pitches via Amazon's Mechanical Turk (mTurk), which has been used as a source of survey for experimental data in past research (see, e.g., Anglin et al. 2017; Allison et al. 2017; Chua 2013). Participants were paid $50 cents to complete the study. We received responses from 600 individuals.[25] To assess the validity of these responses, we employed several screening devices including timers, directed answers, and content checks (Allison et al. 2017). After removing incomplete and erroneous responses, we are left with a sample of 346 responses to one randomly selected scenario/manipulation (i.e. one response pertaining to each ICO pitch).

We collect a large array of information on the participants. We create a dummy variable as to whether the participants have a bachelor's degree or higher ($EX\_EDUCATION$), and ask whether they consider themselves as being financial risk taking ($EX\_FINRISK$) based on a five-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (5) to the question "Do you take financial risks?". We further include a dummy variable taking the value of '1' if the investor is male (EX_GENDER), control for the investor's age (EX_AGE), and add a dummy variable as to whether the participant has considered investing, or has invested in cryptocurrencies before ($EX\_PRIORINVEST$). Furthermore, we ask each investor about their perceptions of the entrepreneur's ability (EX_AUTHOR) by rating the author on a scale of 0 to 100. We request each participant to invest a dollar amount between zero and USD 1,000 to this

[23] Appendix A3 provides two examples of the ICO white papers used in the experiment.
[24] Given that across all ICO pitches the total number of words is the same, we do not scale the errors by the fraction of total words and instead take the logarithm of 1 + the number of errors.
[25] The average time to complete our survey was 4 minutes and 5 seconds.

project (EX_INVEST) and additionally request the participant to specify their overall willingness to invest by basing on a five-point Likert scale ($EX\_WILLINGNESS$) to the question "Would you be willing to invest in this project?". This analysis allows us to test individual investor's financial decision-making as a function of the number of linguistic errors, while keeping the author's information fixed. We display two examples of ICO projects used in the experiment in Appendix A3.

Summary statistics on the participants are provided in Panel A of Table 9. The median participant is a male of about 31 years old, and has at least obtained a Bachelor's degree. Moreover, about 67% of the participants are native English-speaking, are financially risk taking ("Probably yes" or "Definitely yes" on the answer to the question "Do you take financial risks?"), while 48.5% have either considered investing or have invested in cryptocurrencies before. Regarding the ICO pitch, the participants rated the ICO's author as competent (a mean score of 71.62 out of 100). The median participant would be willing to invest in the ICO ("Somewhat agree" to the question "Would you be willing to invest in this project?") and, when translated to a monetary amount, the average participant would be willing to invest USD 450.6 out of USD 1,000.

< Insert Table 9 about here. >

8.3.2 Results of the experiment

We regress our set of control variables and the number of linguistic errors (EX_ERRORS) on the total amount invested (EX_INVEST) in Model (1) of Panel B of Table 9. We find a negative and significant relationship between the number of errors in the investment project and the dollar amount invested. In Model (2) of Panel B, we find similar negative impact of linguistic errors on the participants' willingness to invest ($EX\_WILLINGNESS$). Overall, these findings of our experimental design lend further validation to our main result by showing that linguistic errors are associated with a lower overall investment commitment by investors. Regarding our control variables, the sign and significance of our control variables are consistent throughout all of the four models (except for EX_NATIVE, which is not significant in Model (2). It is interesting to note that participants are more willing to invest in ICOs if they rate the author's ability to be higher, if they are less averse to risk, and if they have a higher education level. Moreover, we find a strong and significant association between the willingness to invest as well as the dollar amount invested if the participant has invested in cryptocurrencies prior to the experiment.

9. Conclusion

This study investigates whether linguistic errors affect the financial outcome in an ICO context. Drawing upon LET and applying it to the context of ICO white papers, we argue that linguistic errors influence investors' decision about what is a well-crafted and persuasive communication attempt, especially if entrepreneurs are looking to attract external funds. By analysing the linguistic errors in 546 ICO white papers, we find that errors constitute a significant penalty for fundraising success and that different error types have varying degrees of impact on ICO funding. In particular, we identify the high-penalty errors – mainly words with multiple errors and errors pertaining to grammar – which were found to have a particularly 'off-putting' effect on the reader. Shedding light on underlying mechanisms, our cross-sectional analyses find more pronounced effects of linguistic errors for ICO white papers from countries with restricted ICO legislation, lower institutional strength and written in English-speaking regions. We also find that the impact of linguistic errors on the amount raised is stronger for ICOs in sin industries and for errors towards the end of the ICO white paper. Finally, an experiment-based study conducted in a simulated ICO context and tests based on restricted control functions further suggest the presence of a causal relationship between the types of linguistic errors and investors' decision to invest in the ICO project.

This paper contributes to a burgeoning field of research on initial coin offerings, and particularly papers demonstrating the importance of white paper disclosure features (Zhang et al. 2019; Fisch 2019; Dittmar and Wu 2019). Our paper demonstrates that market participants pay close attention to white paper content (see also Fisch et al. 2021) and penalize those with more inaccuracies. As such, our paper provides new evidence on the importance of language in entrepreneurial ventures, and are relevant for policymakers and regulators. That is, regulating entities may benefit from understanding that white papers matter to investors, and that such documents are completely unregulated and unstructured, unlike disclosures in the public firm setting. In addition, coin offerings established by entrepreneurs with a lower knowledge of English may face funding difficulties and may make it more difficult for people with lower levels of education or from migrant backgrounds to attain similar opportunities.

Despite using several tests to mitigate endogeneity concerns and issues regarding our data construction process, it is important to note that we still cannot exclude a certain bias in our sample related to missing observations or variables. However, this concern is shared by all ICO studies (see Momtaz 2020a for a discussion) and requires future research to have access to a larger sample and more reliable data as the standardization of ICO information increases. Nonetheless, this study increases our understanding on the importance of using accurate language in financial communication and that, in addition to characteristics such as tone or readability, it plays an active role in securing or losing investment funds. Following this result, future research would be interested in identifying other signals of ICO quality. For instance, the impact of the entrepreneur's gender on investment decisions would be a worthy research avenue to pursue. LET argues that women have a much more restricted bandwidth to communicate relative to their language use because of their lower perceived credibility (Burgoon and Miller 1985). Therefore, when women rely on more aggressive strategies (threats, intense language, fear appeals), they are negatively evaluated and the persuasiveness of the message decreases.

In addition, future research could zoom in specifically on the identification of the additional signals of quality included in the ICO white paper, such as the white paper similarity with preceding projects and assess whether similar projects tend to be perceived negatively by investors.

References

Adhami, S., Giudici, G., & Martinazzi, S. (2018). Why do businesses go crypto? An empirical analysis of initial coin offerings. Journal of Economics and Business, 100, 64–75.

Agrawal, A., Catalini, C., & Goldfarb, A. (2014). Some Simple Economics of Crowdfunding. Innovation Policy and the Economy, 14, 63-97.

Akyildirim, E., Corbet, S., Sensoy, A., & Yarovaya, L. (2020). The impact of blockchain related name changes on corporate performance. Journal of Corporate Finance, 65, 101759.

Aldrich, H. E., & Fiol, C. M. (1994). Fools Rush in? The Institutional Context of Industry Creation. The Academy of Management Review, 19(4), 645.

Allen, F., Gu, X., & Jagtiani, J. (2021). A Survey of Fintech Research and Policy Discussion. Review of Corporate Finance 1(3-4), 259-339.

Allison, T. H., Davis, B. C., Webb, J. W., & Short, J. C. (2017). Persuasion in crowdfunding: An elaboration likelihood model of crowdfunding performance. *Journal of Business Venturing*, *32*(6), 707-725.

Amsden, R., & Schweizer, D. (2018). Are blockchain crowdsales the new 'Gold Rush'? Success determinants of Initial Coin Offerings. Working Paper, available at SSRN.

Anglin, A. H., Wolfe, M. T., Short, J.C., McKenny, A.F. & Pidduck, R.J. (2018). Narcissistic rhetoric and crowdfunding performance: A social role theory perspective. Journal of Business Venturing, 33(6), 780-812,

Appelman, A., & Schmierbach, M. (2018). Make no mistake? Exploring cognitive and perceptual effects of grammatical errors in news articles. Journalism & Mass Communication Quarterly, 95(4), 930-947.

Averbeck, J.M. (2010). Irony and Language Expectancy Theory: Evaluations of Expectancy Violation Outcomes. Communication Studies, 61, 356-372.

Averbeck, J.M., & Miller, C. (2014) Expanding Language Expectancy Theory: The Suasory Effects of Lexical Complexity and Syntactic Complexity on Effective Message Design, Communication Studies, 65(1), 72-95.

Baddeley, A. D., & Hitch, G. J. (1977). Recency re-examined. Attention and Performance VI, 647-667.

Bachmann, R., & Inkpen, A. C. (2011). Understanding institutional-based trust building processes in inter-organizational relationships. *Organization studies*, *32*(2), 281-301.

Benedetti, H., & Kostovetsky, L. (2021). Digital tulips? Returns to investors in initial coin offerings. Journal of Corporate Finance, 66, 101786.

Bestgen, Y. & Granger, S. (2011). Categorising Spelling Errors to Assess L2 Writing. Int. J. Continuing Engineering Education and Life-Long Learning, 21, 235-252.

Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. Journal of Accounting and Economics, 50(2-3), 296–343.

Blaseg, D. (2018). Dynamics of Voluntary Disclosure in the Unregulated Market for Initial Coin Offerings. Working paper.

Block, J.H., Groh, A., Hornuf, L., Vanacker, T. & Vismara, S. (2020). The entrepreneurial fnance markets of the future: A comparison of crowdfunding and initial coin oferings. Small Business Economics, 57, 865–882

Boudt, K., & Thewissen, J. (2019). Jockeying for position in CEO letters: Impression management and sentiment analytics. Financial Management, 48(1), 77-115.

Burgoon, M., & Miller, G.R. (1985). An expectancy interpretation of language and persuasion. In H. Giles & R. Clair (Eds.) The social and psychological contexts of language (pp. 199–229). London: Lawrence Erlbaum Associates.

Burgoon, J. K., Newton, D. A., Walther, J. B., & Baesler, E. J. (1988). Nonverbal expectancy violations and conversational involvement. Journal of Nonverbal Behavior, 13(2), 97–119.

Burgoon, J. K., Bonito, J. A., Ramirez, A. Jr., Dunbar, N. E., Kam, K., & Fischer, J. (2002). Testing the Interactivity Principle: Effects of Mediation, Propinquity, and Verbal and Nonverbal Modalities. Interpersonal Interaction, Journal of Communication, 52(3), 657–677.

Butticè, V., Collewaert, V., Stroe, S., Vanacker, T., Vismara, S., & Walthoff-Borm, X. (2021). Equity Crowdfunders' Human Capital and Signal Set Formation: Evidence From Eye Tracking. Entrepreneurship Theory and Practice, 104225872110268.

Cerchiello, P., Tasca, P., & Toma, A. M. (2019). ICO Success Drivers: A Textual and Statistical Analysis. The Journal of Alternative Investments, 21(4), 13–25.

Chua, R.Y.J. (2013). The Costs of Ambient Cultural Disharmony: Indirect Intercultural Conflicts in Social Environment Undermine Creativity. Academy of Management Journal, **56**, 1545–1577.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. Journal of Personality and Social Psychology, 66(3), 460–473.

Chen S., Huang B. & Shaban M. (2020). Naïve or sophisticated? Information disclosure and investment decisions in peer to peer lending. Journal of Corporate Finance, forthcoming.

Chelli, M., Durocher, S., & Richard, J. (2014). France's new economic regulations: insights from institutional legitimacy theory. Accounting, Auditing & Accountability Journal, 27(2), 283-316.

Chod, J. & Lyandres, E. (2020). A Theory of ICOs: Diversification, Agency, and Information Asymmetry. Management Science, forthcoming.

Colombo, M., Fisch, C., Momtaz, P., & Vismara, S. (2021). The CEO Beauty Premium: Founder CEO Attractiveness and Firm Valuation in Initial Coin Offerings. Strategic Entrepreneurship Journal, forthcoming.

Colombo M., & Grilli, L. (2010). On growth drivers of high-tech startups: exploring the role of founders' human capital and venture capital. Journal of Business Venturing, 25(6), 610-626.

Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. International Review of Financial Analysis, 62, 182-199.

Cornelissen, J. P. & Clarke, J. S. (2010). Imagining and Rationalizing Opportunities: Inductive Reasoning and the Creation and Justification of New Ventures. Academy of Management Review, 35, 539–557.

Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. System, 26(2), 163–174.

De Amicis C., Falconieri S. & Tastan M. (2020). Sentiment analysis and gender differences in earnings conference calls. Journal of Corporate Finance, forthcoming.

Dittmar, R. F., & Wu, D. A. (2019). Initial coin offerings hyped and dehyped: an empirical examination. Working paper, available at SSRN.

Edgecliffe-Johnson, A. (2001). The Challenge of Making a Profit in Hostile Territory. Financial Times, Online, accessed April 20, 2020.

El-Haj, M., Rayson, P., Walker, M., Young, S., & Simaki, V. (2019). In search of meaning: Lessons, resources and next steps for computational analysis of financial discourse. Journal of Business Finance & Accounting, 46, 265-306.

Feng, C., Li N., Wong M. H., & Zhang M. (2019). Initial coin offerings, blockchain technology, and white paper disclosures. Working Paper, available at SSRN.

Figueredo, L., & Varnhagen, C. K. (2005). Didn't You Run the Spell Checker? Effects of Type of Spelling Error and Use of a Spell Checker on Perceptions of the Author. Reading Psychology, 26(4-5), 441–458.

Fisch, C. (2019). Initial coin offerings (ICOs) to finance new ventures. Journal of Business Venturing, 34, 1–22.

Fisch, C., Masiak, C., Vismara, S., & Block, J. (2021). Motives and profiles of ICO investors. Journal of Business Research, 125, 564-576.

Fisch, C., & Momtaz, P. P. (2020). Institutional investors and post-ICO performance: an empirical analysis of investor returns in initial coin offerings (ICOs). Journal of Corporate Finance, 64, 101679.

Formentin, M., Hettinga, K. & Appelman, A. (2021). Two Wrongs Don't Make a Right: Journalists' Perceptions and Usage of Press Releases. Corporate Reputation Review, 24, 65–75.

Gao, Q., Lin, M., & Sias, R. W. (2021). Words matter: The role of texts in online credit markets. Journal of Financial and Quantitative Analysis (Forthcoming).

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. Journal of verbal learning and verbal behavior, 5(4), 351-360.

Grossman, S. J., & Hart, O. D. (1980). Takeover Bids, The Free-Rider Problem, and the Theory of the Corporation. The Bell Journal of Economics, 11(1), 42.

Halbinger, M. A., & Reichstein, T. (2015). Entrepreneurs' social skills. In Academy of Management Proceedings (Vol. 2015, No. 1, p. 14550). Briarcliff Manor, NY 10510: Academy of Management.

Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. Journal of Accounting and Economics, 31(1-3), 405–440.

Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica: Journal of the Econometric Society, 47(1), 153-161.

Heckman, J., & Navarro-Lozano, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics*, *86*(1), 30-57.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? Journal of Business Communications, 45(4), 363-407.

Hines, J.R. (2010). Treasure islands. The Journal of Economic Perspectives, 24(4), 103-126.

Hoberg, G., & Lewis, C. (2017). Do fraudulent firms produce abnormal disclosure? Journal of Corporate Finance, 43, 58–85.

Howell, S., Niessner, M., & Yermack, D. (2020). Initial Coin Offerings: Financing Growth with Cryptocurrency Token Sales. The Review of Financial Studies, 33(9), 3925–3974.

Huang, Y., Li, X., & Wang, C. (2021a). What does peer-to-peer lending evidence say about the Risk-Taking Channel of monetary policy? Journal of Corporate Finance, 66, 101845.

Huang, W., Meoli, M., & Vismara, S. (2020). The geography of initial coin offerings. Small Business Economics, 101007

Huang, W., Vismara, S., & Wei, X. (2021b). Confidence and capital raising. Journal of Corporate Finance, 101900.

Kim, I., & Venkatachalam, M. (2011). Are Sin Stocks Paying the Price for Accounting Sins? Journal of Accounting, Auditing & Finance, 26(2), 415–442.

Kincaid, J., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel.

Kreiner, D. S., Schnakenberg, S. D., Green, A. G., Costello, M. J., & McClin, A. F. (2002). Effects of spelling errors on the perception of writers. The Journal of General Psychology, 129(1), 5–17.

Kübler, D. (2001). On the regulation of social norms. Journal of Law, Economics, and Organization, 17(2), 449-476.

Lambert, T., Liebau, D. & Roosenboom, P. (2021). Security token offerings. Small Business Economics, Forthcoming.

Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., & Gorski, S. (2011). Peer to Peer Lending: The Relationship Between Language Features, Trustworthiness, and Persuasion Success. Journal of Applied Communication Research, 39(1), 19–37.

Li, Y., & Zahra, S. A. (2012). Formal institutions, culture, and venture capital activity: A cross-country analysis. Journal of Business Venturing, 27(1), 95–111.

Liu, Y., Sheng, J., and W. Wang (2021). Technology and Cryptocurrency Valuation: Evidence from Machine Learning. Working Paper available at SSRN 3577208.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35-65.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. Journal of Accounting Research, 54(4), 1187-1230.

Loughran, T., & McDonald, B. (2017). The Use of EDGAR Filings by Investors. Journal of Behavioral Finance, 18(2), 231–248.

Lounsbury, M., & Glynn, M. A. (2001). Cultural entrepreneurship: stories, legitimacy, and the acquisition of resources. Strategic Management Journal, 22(6-7), 545–564.

Marson, J. (2020). OneCoin Took In Billions. Then Its Leader Vanished, The Wall Street Journal, August 27th, 2020

Martens, M. L., Jennings, J. E., & Jennings, P. D. (2007). Do the Stories They tell get them the Money They Need? The Role of Entrepreneurial Narratives in Resource Acquisition. Academy of Management Journal, 50(5), 1107–1132.

Martin-Lacroux, C., & Lacroux, A. (2017). Do employers forgive applicants' bad spelling in résumés? Business and Professional Communication Quarterly, 80(3), 321-335.

Mohammadi, A. & Shafi, K. (2018). Gender differences in the contribution patterns of equity-crowdfunding investors. Small Business Economics, 50, 275–287.

Momtaz P. P. (2020a). Initial coin offerings, asymmetric information, and loyal CEOs. Small Business Economics, Forthcoming.

Momtaz, P. P. (2020b). Entrepreneurial finance and moral hazard: Evidence from token offerings. Journal of Business Venturing, 106001.

Momtaz, P. P. (2021). Security tokens. Working Paper.

Nestarcova, D. (2018). A Critical Appraisal of Initial Coin Offerings: Lifting the "Digital Token's Veil". Brill Research Perspectives in International Banking and Securities Law, 3(2-3), 1-171.

Parhankangas, A., & Ehrlich, M. (2014). How entrepreneurs seduce business angels: An impression management approach. Journal of Business Venturing, 29(4), 543-564.

Park, M., Aiken, M., Lindblom, T., & Vanjani, M. (2010). Spelling and grammatical errors in electronic meetings. Issues in Information Systems 11 (1), 384-391.

Philippi, S., Schuhmacher M., & Bastian, N. (2021). Attracting Investors in Initial Coin Offerings: The Relevance of Specific Technological Capabilities for Fundraising Success. Review of Corporate Finance 1(3-4), 455-485.

Planken, B. (2005). Managing rapport in lingua franca sales negotiations: A comparison of professional and aspiring negotiators. English for Specific Purposes, 24(4), 381–400.

Praise, S., & Meenakshi, K. (2014). Importance of grammar in communication. International Journal of Research Studies in Language Learning, 4(1), 97-101.

Puhani, P. (2000). The Heckman Correction for Sample Selection and Its Critique. Journal of Economic Surveys, 14(1), 53–68.

Rhue, L. (2018). Trust is all you need: an empirical exploration of initial coin offerings (ICOs) and ICO reputation scores. Working paper, Available at SSRN.

Robinson II, R. (2018). The New Digital Wild West: Regulating the Explosion of Initial Coin Offerings. Tennessee Law Review, 85(4), 897-960.

Roediger, H. L., & Crowder, R. G. (1976). A serial position effect in recall of United States presidents. Bulletin of the Psychonomic Society, 8(4), 275-278.

Rubin, D. L., & Williams-James, M. (1997). The impact of writer nationality on mainstream teachers' judgments of composition quality. Journal of Second Language Writing, 6(2), 139–154.

Seamon, M. (2001). Assessing the need for change in J-School grammar curricula. Journalism & Mass Communication Educator 55(4), 60–69.

Security Exchange Comission. (2017). Statement on Cryptocurrencies and Initial Coin Offerings. Online accessed at: https://www.sec.gov/news/public-statement/statement-clayton-2017-12-11.

Sharma, Z. & Zhu, Y. (2020). Platform building in initial coin offering market: Empirical evidence. Pacific-Basin Finance Journal, 61.

Shrestha, P., Arslan-Ayaydin, Ö, Thewissen J., & Torsin, W. (2021). Institutions, regulations and initial coin offerings: An international perspective. International Review of Economics and Finance, 72, 102–120.

Sutinen, J. G., & Kuperan, K. (1999). A socio-economic theory of regulatory compliance. International Journal of Social Economics, 26(1-3), 174–193.

Thewissen, J. (2013). Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus. Modern Language Journal, 97, 77-101.

Thewissen, J. (2015). Accuracy across proficiency levels: A learner corpus approach. Corpora and Language in Use. Presses Universitaires de Louvain: http://www.i6doc.com/fr/livre/?GCOI=28001100217910

Thewissen, J. (2021). Accuracy. In N. Tracy-Ventura, & M. Paquot (Eds.) Handbook of SLA and Corpora. Routledge. https://www.routledge.com/The-Routledge-Handbook-of-Second-Language-Acquisition-and-Corpora/Tracy-Ventura-Paquot/p/book/9780815352877

van Dijk, E. & Zeelenberg, M. (2003), The discounting of ambiguous information in economic decision making. Journal of Behavioral Decision Making, 16, 341-352.

Varnhagen, C. K. (2000). Shoot the messenger and disregard the message? Children's attitudes toward spelling. Reading Psychology, 21(2), 115–128.

Welter, F. (2012). All you need is trust? A critical review of the trust and entrepreneurship literature. International Small Business Journal 30 (3), 193-212.

Williamson, O. E. (1993). Calculativeness, trust, and economic organization. The Journal of Law and Economics 36 (1, Part 2), 453-486.

Wiens, K. (2012). I won’t hire people who use poor grammar. Here’s why. Harvard Business Review, 20.

Wolfe, J., Shanmugaraj, N., & Sipe, J. (2016). Grammatical Versus Pragmatic Error. Business and Professional Communication Quarterly, 79(4), 397–415.

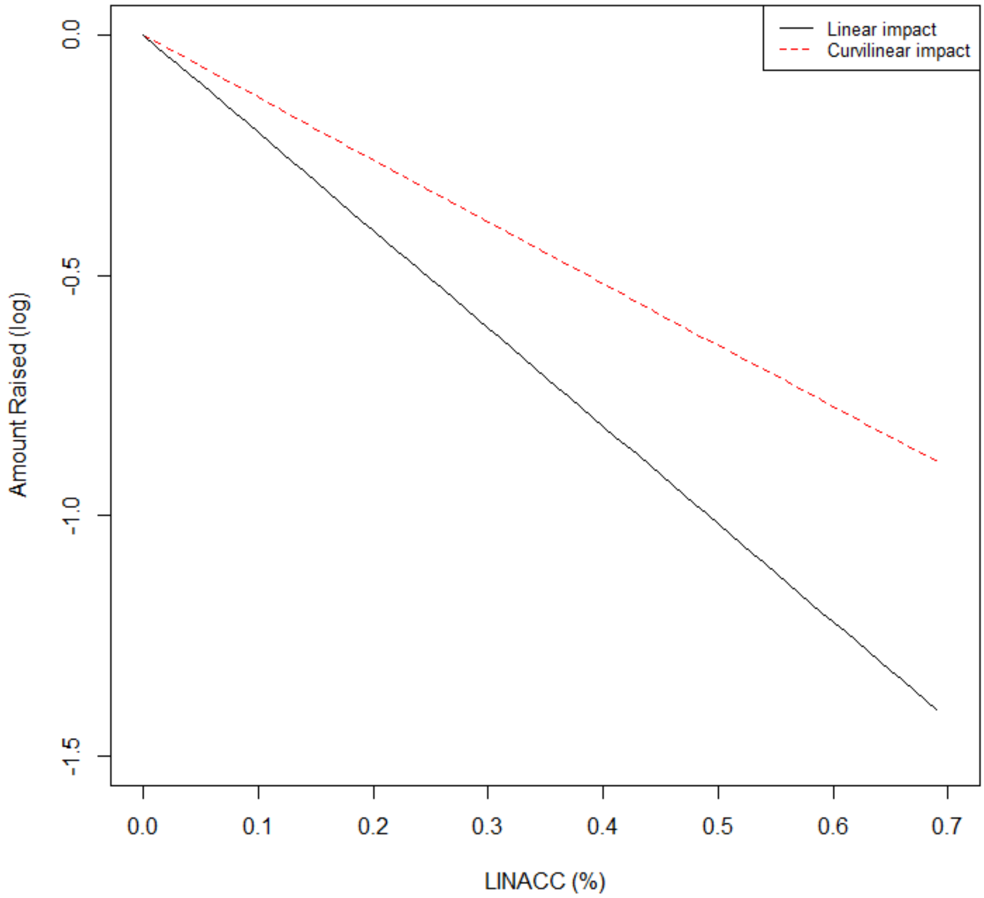
Younkin P. & Kuppuswamy V. (2018). The Colorblind Crowd? Founder Race and Performance in Crowdfunding. Management Science, 64(7), 2973-3468.

Zetzsche, D. A., Buckley, R. P., Arner, D. W., & Föhr, L. (2017). The ICO Gold Rush: It's a scam, it's a bubble, it's a super challenge for regulators. University of Luxembourg Law Working Paper, (11), 17-83.

Zhang, S., Aerts, W., Lu, L., & Pan, H. (2019). Readability of token white paper and ICO first-day return. Economics Letters, 180, 58-61.

Zucker, L. G. (1986). Production of trust: Institutional sources of economic structure, 1840-1920. Research in Organizational Behavior, 8, 53—111

Figure 1 – The linear and curvilinear impact of linguistic errors in ICO white papers on the amount raised during an ICO



This figure reports the marginal impact of linguistic errors on the amount raised during an ICO. The black line reports the linear impact, while the red dashed line reports investors' curvilinear relationship with linguistic errors.

Table 1 – Variable Definitions

Variable	Definition
Dependent Variables	
AMOUNT_RAISED	The logarithm of the US dollar amount raised during the coin-offering period in US dollars.
TRADED	Indicates whether the token is eventually traded on a dominant crypto currency exchange (Coinmarketcap).
HARD_CAP_ACHIEVED	Indicates whether the dollar amount raised exceeded the hard cap specified.
Independent Variables	
FS	The total number of linguistic errors from the categorization scheme (FS_k) of Bestgen and Granger (2011) found in the ICO white paper. $$FS = \sum_{k=1}^{13} FS_k$$ Refer to Appendix A1 for examples on the k categories.
LINACC	The ratio of total formal (FS) errors present in the ICO white paper. We refer to Appendix A1 for a more detailed breakdown of the 13(k) distinct linguistic inaccuracies. $$LINACC = \frac{FS}{TW}$$
NES	Dummy variable if the ICO's home country is located in a country where English is considered an official language, see: http://globed.eu/wp-content/uploads/2019/11/English_official_language.pdf.
UNREGULATED	Indicator whether the ICO is located in a country without any regulations pertaining to cryptocurrencies and coin offerings at the time of the coin offering per Shrestha et al. (2021).
INST	A country's institutional strength following the methodology adopted by Li and Zahra (2012) and constructed as the first principal component of six Worldwide Governance Indicators. These are: (i) control of corruption, (ii) rule of law, (iii) government effectiveness, (iv) regulatory quality, (v) political stability, and (vi) voice and accountability.
Control Variables	
TAX_HAVEN	Indicates whether the country is located in a tax haven (Hines 2010).
ETHEREUM	Indicates whether the project blockchain is built on the Ethereum platform.
ICO_PRICE	The USD price of one token issued during the coin offering.
ICO_RATING	The ICOBench rating that the coin obtained from external experts during the coin offering.
N_O_EXPERTS	Logarithm of 1 + the number of experts that rated the ICO.
MINIMUM	Indicates whether a minimum investment amount is specified.

N_O_TEAM	Logarithm of 1 + the number of members in the team behind the ICO.
HARD_CAP	Indicates whether a soft cap is specified.
SOFT_CAP	Indicates whether a hard cap is specified.
UTIL/SEC	Indicates whether the ICO issues a utility token.
WHITE_KYC	Indicates whether the ICO implements Whitelisting and Know Your Customer (KYC) compliances.
FACEBOOK	Indicates whether the ICO has a link to a Facebook account.
GITHUB	Indicates whether the ICO has a link to a Github account.
TWITTER	Indicates whether the ICO has a link to a Twitter account.
TELEGRAM	Indicates whether the ICO has a link to a Telegram account.
FILESIZE	The size of the original white paper PDF file (in MB).
TOTAL_WORDS	The number of words in the ICO white paper
READABILITY	Flesch Reading Ease Readability score of the white paper text (Kincaid et al. 1975).
TONE	The number of positive minus negative words in white paper text, divided by the total number of words, based on the dictionaries provided by (Loughran and McDonald 2011).
GDP	Natural logarithm of the GDP per capita (in USD) at the end of the year that the ICO has been issued.
POPULATION	Natural logarithm of a country's population level at the end of year that the ICO has been issued
FDI	Based on eight survey questions, the Financial Development Index measures the efficiency of financial services meeting business needs and the availability of financing through local equity markets and the trustworthiness and confidence of banking systems. Source: World Bank (2016).
LOCATION	Dummy variable taking the value of '1' if the initial coin offering stems from the top five countries in terms of total amount raised (these are: USA, UK, Estonia, Switzerland, and Singapore) (as per Huang et al. 2021b)
Year	Dummy variable taking the value of '1' if the initial coin offering period ended in year t. Zero otherwise
Industry	Dummy variable taking the value of '1' if the ICO belonged to a specific category as defined by ICOBench, zero otherwise.
Continent	Indicates in which continent the country of the issuing ICO belongs.
SIN	Indicates whether the ICO is active in the tobacco, gambling, or alcohol industry.
BRACKET	The number of brackets: '(', ')', '[', and ']', divided by the number of non-alphanumeric characters in the white paper.
Variables from the mTurk experiment	
EX_ERRORS	Logarithm of one plus the number of linguistic errors in the ICO white paper pitch.
EX_INVEST	The dollar amount invested by the participant (ranges from 0 to USD 1,000).

EX_WILLINGESS	The participant's willingness to invest in the ICO based on a five-point Likert scale, as a response to the question "Would you be willing to invest in this project?". The answers of the participants range from "strongly disagree" (1) to "strongly agree" (5).
EX_AUTHOR	The participant's evaluations of the entrepreneur behind the ICO project. The participant can rate the entrepreneur between 0 and 100.
EX_PRIORINVEST	Dummy variable taking the value of '1' if the participant has considered investing, or has invested in cryptocurrencies before, zero otherwise.
EX_FINRISK	The participant's level of financial risk taking based on a five-point Likert scale, as a response to the question "Do you take financial risks?". The answers of the participants range from "definitely no" (1) to "definitely yes" (5).
EX_AGE	The age of the participant in years.
EX_GENDER	Dummy variable taking the value of '1' if the participant is male, zero otherwise.
EX_NATIVE	Dummy variable taking the value of '1' if the participant is native English speaking, zero otherwise.
EX_EDUCATION	Dummy variable taking the value of '1' if the participant has obtained a Bachelor's degree or higher, zero otherwise.

Table 2 – Error Taxonomy Used in the ICO Corpus (Based on Bestgen and Granger 2011)

Error tag/code	Explanation of error type	*Examples*
FS_mis	Omission of a letter	1. *(FS_mis) complelty $completely$* 2. *(FS_mis) wether $whether$* 3. *(FS_mis) mecanisms $mechanisms$*
FS_red	Addition of a letter	1. *(FS_red) assignement $assignment$* 2. *(FS_red) eighthееn $eighteen$* 3. *(FS_red) develope $develop$*
FS_doub12	Single letter instead of double letter	1. *Many (FS_doub12) adicts $addicts$* 2. *The accident (FS_doub12) ocurred $occurred$ in the morning*
FS_doub21	Double letter instead of single letter	1. *Many (FS_doub21) appartments $apartments$ are sold* 2. *The text is too (FS_doub21) detailled $detailed$*
FS_XY	Substitution of one letter	1. *He is very (FS_XY) dependend $dependent$* 2. *This happened throughout their (FS_XY) lifes $lives$*
FS_swap	Interchange of two adjacent letters	1. *(FS_swap) concieved $conceived$* 2. *(FS_swap) birht $birth$*
FS_apost	Error involving an apostrophe	1. *A leopard never changes (FS_apost) it's its spots.* 2. *Let's sort out the (FS_apost) childrens' $children's$ clothes.*
FS_split	Word segmentation – erroneous splitting or joining of words	1. *(FS_split) airpollution $air pollution$* 2. *(FS_split) free-time $free time$* 3. *(FS_split) eventhough $even though$*
FS_morph	Wrong prefix (e.g. un-, in-, im-, dis-, etc.) or suffix (e.g. -ness, -ment, -ity, etc.) Wrong verb inflection (e.g. **hitted* instead of *hit*)	1. *(FS_morph) unpolite $impolite$* 2. *(FS_morph) politic $political$* 3. *(FS_morph) paintors $painters$* 4. *(FS_morph) hitted hit* 5. *(FS_morph) swum $swam$*
FS_mult	Multiple (2 or more) errors of the same type or different types on the same word.	1. *(FS_mis) drasticly $drastically$* 2. *(FS_mult) weter $whether$* 3. *(FS_mult) eventough $even though$* 4. *(FS_mult) pollitic $political$*
FS_Missing	Full missing word	1. *The company (FS_Missing) 0 is one of a kind.*
FS_Dup	Duplicated word	1. *The company is (FS_Dup) is 0 one of a kind.*
FS_GVN	Grammar verb number (subject-verb agreement errors)	2. *Our project (FS_GVN) show $shows$ that…*

Table 3 – Summary Statistics

	mean	sd	min	Q25	median	Q75	max
Panel A – Linguistic accuracy (count)							
FS	8.368	13.069	0	1	4	10	136
FS_MIS	0.677	1.762	0	0	0	1	26
FS_RED	0.572	1.144	0	0	0	1	9
FS_DOUB12	0.105	0.496	0	0	0	0	6
FS_DOUB21	0.078	0.306	0	0	0	0	2
FS_XY	0.379	1.218	0	0	0	0	20
FS_SWAP	0.134	0.528	0	0	0	0	4
FS_APOST	0.327	0.908	0	0	0	0	8
FS_SPLIT	5.076	10.871	0	0	1	6	130
FS_MORPH	0.011	0.134	0	0	0	0	2
FS_MULT	0.238	0.81	0	0	0	0	9
FS_GVN	0.623	2.503	0	0	0	0	50
FS_MISSING	0.032	0.484	0	0	0	0	11
FS_DUP	0.117	0.445	0	0	0	0	5
Panel B – Linguistic errors (Ratio) x 100							
LINACC	0.122	0.306	0	0.018	0.066	0.173	0.731
LINACC_MIS	0.011	0.026	0	0	0	0.015	0.115
LINACC_RED	0.009	0.007	0	0	0	0.012	0.115
LINACC_DOUB12	0.001	0.006	0	0	0	0	0.041
LINACC_DOUB21	0.001	0.004	0	0	0	0	0.024
LINACC_XY	0.006	0.027	0	0	0	0	0.088
LINACC_SWAP	0.002	0.010	0	0	0	0	0.047
LINACC_APOST	0.005	0.131	0	0	0	0	0.074
LINACC_SPLIT	0.069	0.184	0	0	0.015	0.095	0.567
LINACC_MORPH	0	0	0	0	0	0	0
LINACC_MULT	0.004	0.018	0	0	0	0	0.082
LINACC_GVN	0.008	0.019	0	0	0	0	0.123
LINACC_MISSING	0	0.001	0	0	0	0	0.009
LINACC_DUP	0.002	0.006	0	0	0	0	0.036
Panel C – Dependent and control variables							
*AMOUNT_RAISED** (in mil. USD)	13.002	51.401	0.001	1.486	4.591	13.726	1000
TAX_HAVEN	0.336	0.473	0	0	0	1	1
ETHEREUM	0.865	0.342	0	1	1	1	1
ICO_PRICE (in USD)	1.273	10.951	0	0.04	0.1	0.456	230.77
ICO_RATING	3.382	0.624	1.6	2.9	3.4	3.9	4.7
*N_O_EXPERTS**	5.924	10.201	0	0	2	7	97
MINIMUM	0.408	0.492	0	0	0	1	1
*N_O_TEAM**	15.339	8.161	1	9	14	20	53
HARD_CAP	0.892	0.311	0	1	1	1	1
SOFT_CAP	0.632	0.483	0	0	1	1	1
WHITE_KYC	0.69	0.463	0	0	1	1	1
UTIL/SEC	0.978	0.147	0	1	1	1	1
FACEBOOK	0.921	0.271	0	1	1	1	1
GITHUB	0.63	0.483	0	0	1	1	1
TWITTER	0.982	0.133	0	1	1	1	1
TELEGRAM	0.901	0.299	0	1	1	1	1
*FILESIZE** (in MB)	3.979	6.683	0.174	1.116	2.423	4.778	125.688
*TOTALWORDS**	7706.449	4108.284	209	5209.5	6910	9723	34081
TONE	0.002	0.006	-0.016	-0.001	0.002	0.006	0.016
READABILITY	45.109	7.336	26.318	40.835	44.808	48.988	67.322
*POPULATION** (in mil.)	89.628	224.575	0.031	2.872	8.372	65.637	1378.665
*GDP**	38701.4	26225.88	707.9	17403.8	42325.3	55645.6	167313.3
FDI	0.65	0.232	0.09	0.46	0.73	0.89	0.95
NES	0.254	0.436	0	0	0	1	1
LOCATION	0.485	0.5	0	0	0	1	1
UNREGULATED	0.433	0.494	0	0	0	1	1
INST	2.436	1.717	-3.029	1.998	3.008	3.706	4.448
SIN	0.061	0.24	0	0	0	0	1
BRACKET	0.004	0.003	0.001	0.003	0.004	0.004	0.015

This table provides the summary statistics of the linguistic inaccuracy dimensions expressed in absolute numbers (Panel A) and scaled by the total number of words in the text (Panel B), as well as the main independent and dependent variables (Panel C).* indicates that logarithmic values are used in regressions, but are expressed in their regular form for presentation purposes. Definitions of the variables are provided in Table 1.

Table 4 – Univariate Comparison: NES and Regulation

	Panel A - *UNREGULATED*			Panel B - *INST*			Panel C - *NES*		
	Regulated	Unregulated	T-Test	Low INST	High INST	T-Test	Non-NES	NES	T-Test
FS	8.679	8.141	0.469	8.494	7.198	1.357	9.341	5.518	3.970***
FS_MIS	0.756	0.619	0.903	0.722	0.623	0.640	0.717	0.56	0.756
FS_RED	0.62	0.538	0.817	0.635	0.481	1.641*	0.654	0.333	3.752***
FS_DOUB12	0.188	0.044	1.813*	0.161	0.036	3.154***	0.121	0.057	1.775*
FS_DOUB21	0.103	0.059	1.607	0.087	0.069	0.684	0.08	0.071	0.300
FS_XY	0.483	0.303	1.558	0.448	0.263	1.883*	0.438	0.206	2.666***
FS_SWAP	0.098	0.159	-1.406	0.127	0.142	-0.749	0.153	0.078	1.755*
FS_APOST	0.393	0.278	1.444	0.334	0.312	0.290	0.317	0.355	-0.414
FS_SPLIT	5.034	5.106	-0.077	4.766	4.502	0.349	5.649	3.397	2.797***
FS_MORPH	0.004	0.016	-1.104	0.007	0.016	-0.769	0.015	0	1.903*
FS_MULT	0.282	0.206	1.038	0.294	0.166	1.917*	0.298	0.064	4.621***
FS_GVN	0.603	0.638	-0.176	0.729	0.474	1.274	0.736	0.291	2.761***
FS_MISSING	0.009	0.05	-1.158	0.047	0.016	0.796	0.041	0.007	1.201
FS_DUP	0.107	0.125	-0.458	0.137	0.097	1.072	0.123	0.099	0.672

This table reports univariate comparisons in terms of linguistic accuracy and its distinct dimensions along three dimensions: whether the ICO is present in a country with regulations concerning ICOs as per Shrestha et al. (2021) (Panel A), whether the ICO is issued in a country with lower than (or equal to) median vs. higher than median institutional strength (Panel B), and whether the ICO's home country is native English speaking (Panel C). Definitions of the different linguistic inaccuracies are reported in Table 1. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on a Welch two-sample t-test. Definitions of the variables are provided in Table 1

Table 5 – Correlation Table

Panel A – Correlation table of the main dependent, independent, and control variables

	LINACC	*AMOUNT RAISED*	*TAX_ HAVEN*	*ETHEREUM*	*ICO_ PRICE*	*ICO_ RATING*	*N_O_ EXPERTS*	*MINIMUM*	*N_O_ TEAM*	*HARD_ CAP*	*SOFT_ CAP*	*WHITE_ KYC*	*UTIL/ SEC*
AMOUNT_RAISED	-0.163***												
TAX_HAVEN	0.016	0.016											
ETHEREUM	0.005	0.004	0.046										
ICO_PRICE	0.003	0.003	-0.049	-0.006									
ICO_RATING	-0.095**	-0.095**	0.033	-0.009	-0.055								
N_O_EXPERTS	-0.055	-0.055	0.037	0.064	-0.044	0.464***							
MINIMUM	0.049	0.049	0.055	0.006	-0.038	0.182***	0.049						
N_O_TEAM	-0.074*	-0.073*	0.032	0.061	-0.015	0.375***	0.286***	0.025					
HARD_CAP	-0.117***	-0.117***	0.001	0.014	-0.051	0.190***	0.081*	0.100**	0.102**				
SOFT_CAP	0.018	0.017	0.059	0.015	0.009	0.207***	0.075*	0.199***	0.126***	0.384***			
WHITE_KYC	0.125***	0.125***	0.155***	0.008	-0.018	0.283***	0.086**	0.366***	0.111***	0.192***	0.353***		
UTIL/SEC	0.024	-0.049	0.001	0.158***	-0.051	-0.003	0.022	-0.028	-0.007	0.028	0.042	0.034	
FACEBOOK	-0.015	-0.018	-0.074*	-0.038	0.026	0.322***	0.128***	0.148***	0.108**	0.112***	0.121***	0.192***	0.052
GITHUB	-0.064	-0.002	0.046	0.002	0.003	0.434***	0.255***	0.156***	0.157***	0.190***	0.151***	0.188***	-0.063
TWITTER	-0.037	0.002	0.067	-0.053	0.006	0.174***	0.068	0.002	0.083**	-0.003	-0.019	0.160***	-0.019
TELEGRAM	0.069	0.031	0.146***	0.009	-0.002	0.283***	0.133***	0.152***	0.176***	0.001	0.147***	0.170***	-0.009
FILESIZE	0.009	0.021	0.056	0.037	-0.031	0.078*	0.034	0.101**	0.025	0.100**	0.009	0.036	-0.015
TOTALWORDS	-0.096	-0.122***	0.072*	0.006	-0.015	0.269***	0.181***	0.071*	0.262***	0.083*	0.073*	0.095**	-0.027
TONE	0.088**	0.088**	0.018	-0.079*	0.019	-0.059	0.014	-0.023	0.022	0.022	0.026	-0.037	0.079*
READABILITY	0.160***	0.160***	-0.151***	-0.031	0.027	-0.077*	-0.008	0.059	-0.125***	-0.021	-0.010	0.136***	0.035
POPULATION	-0.031	-0.039	-0.473***	-0.075*	0.020	-0.153***	-0.091**	-0.101**	-0.052***	-0.100**	-0.092**	-0.208***	-0.015
GDP	-0.149***	0.143***	0.280***	0.115***	-0.038	0.0491	0.001	-0.019	0.021	0.019	-0.059	0.067	-0.075*
FDI	-0.157***	0.134***	0.148***	0.038	0.011	-0.056***	-0.055***	-0.041	-0.013	-0.058	-0.062	-0.035	-0.110***
LOCATION	-0.040	0.057	0.098**	-0.017	-0.056	0.066	-0.016	-0.128***	-0.033	-0.086**	-0.109**	-0.098**	0.045
NES	0.251**	0.015	-0.239***	0.061	0.053	-0.096**	-0.067	-0.063	-0.103**	-0.023	-0.112***	0.033	0.005
INST	-0.082*	0.101**	0.311***	0.079*	-0.018	0.036	-0.027	0.038	0.028	0.034	-0.019	0.107**	-0.071*
UNREGULATED	-0.005	0.022	0.128***	0.340***	-0.037	0.069	-0.019	-0.000	-0.026	-0.004	-0.092**	-0.034	0.002
SIN	-0.171***	0.000	-0.038	-0.030	-0.019	-0.054	-0.042	-0.013	0.034	-0.012	-0.007	-0.055	0.025

	FACE BOOK	*GITHUB*	*TWITTER*	*TELEGRAM*	*FILE SIZE*	*TOTAL WORDS*	*TONE*	*READ ABILITY*	*POPUL.*	*GDP*	*FDI*	*LOCATION*	*NES*	*INST*	*UN REGULATED*
FACEBOOK															
GITHUB	0.134***														
TWITTER	0.148***	0.149***													
TELEGRAM	0.183***	0.183***	0.136***												
FILESIZE	0.022	0.022	0.058	0.056											
TOTALWORDS	0.095**	0.096**	0.070*	0.107**	0.089*										
TONE	-0.005	-0.005	-0.019	0.021	0.013	-0.121***									
READABILITY	-0.056	-0.056	-0.045	-0.031	0.015	-0.131***	0.081*								
POPULATION	-0.021	0.016	0.003	-0.162***	-0.059	-0.057	0.023	0.080*							
GDP	-0.019	0.035	0.002	0.057	-0.021	0.064	-0.109***	-0.151***	-0.126***						
FDI	-0.031	0.034	0.032	0.011	-0.025	0.033	-0.051	-0.115***	0.310***	0.744***					
LOCATION	-0.015	-0.072*	-0.047	-0.052	-0.012	-0.000	-0.025	-0.024	0.275***	0.347***	0.293***				
NES	-0.015	-0.016	0.017	-0.069	-0.022	-0.094**	-0.028	0.069	0.250***	0.379***	0.451***	0.536***			
INST	-0.015	0.017	0.007	0.046	-0.027	0.061	-0.067	-0.107***	-0.146***	0.805***	0.562***	0.293***	0.358***		
UNREGULATED	-0.042	-0.043	-0.006	-0.076*	-0.000	-0.019	-0.037	-0.037	-0.159***	-0.470***	-0.396***	-0.416***	-0.340***	-0.038	
SIN	-0.036	-0.037	-0.022	-0.016	0.048	0.024	0.117***	-0.005	0.009	-0.046	-0.059	-0.028	0.075	0.051	0.051

Panel B – Correlation of distinct linguistic inaccuracy dimensions (Ratio of linguistic errors x 100)

LINACC	*MIS*	*RED*	*DOUB12*	*DOUB21*	*XY*	*SWAP*	*APOST*	*SPLIT*	*MORPH*	*MULT*	*GVN*	*MISSING*
RED	0.107**											
DOUB12	0.183***	0.047										
DOUB21	0.171***	0.038	0.184***									
XY	0.124***	0.103**	0.132***	0.105**								
SWAP	0.155***	0.118***	-0.012	0.025	0.070*							
APOST	0.142***	0.155***	0.149***	0.149***	0.049	-0.004						
SPLIT	0.073*	0.133**	0.049	0.059	0.134***	0.027	0.118***					
MORPH	0.068	-0.016	-0.017	-0.021	0.118***	0.082*	0.060	-0.035				
MULT	0.083*	0.161***	0.041	0.057	0.401***	0.010	0.071*	0.286***	-0.023			
GVN	0.031	0.114***	0.019	0.009	0.057	0.018	0.039	0.232***	-0.020	0.206***		
MISSING	0.014	0.083**	-0.014	-0.017	0.001	-0.017	0.111***	-0.027	-0.005	0.132***	0.016	
DUP	0.051	-0.022	0.026	0.105**	0.018	0.026	-0.032	-0.086**	0.039	0.033	-0.006	0.075*

This table presents two sets of correlation matrices: one for ICO-level variables (Panel A) and one for the distinct components of linguistic inaccuracy (Panel B). Definitions of the variables are provided in Table 1. The table shows Pearson correlation coefficients with significance levels of 10 percent, 5 percent, and 1 percent denoted with *, ** and ***, respectively.

Table 6 – Linguistic Accuracy and ICO Amount Raised

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
(Intercept)	15.380***	10.932***	10.934***	11.121***	10.644***	10.697***	10.673***	10.716***
	(0.094)	(1.992)	(2.002)	(2.690)	(2.001)	(2.009)	(2.256)	(2.054)
LINACC	-203.372**	-117.625**	-131.019**					
	(52.768)	(59.889)	(60.081)					
LINACC^2			372.821***					
			(132.891)					
LINACC_MIS				-527.113*				
				(323.019)				
LINACC_RED					-437.670			
					(366.914)			
LINACC_DOUB12						693.132		
						(1001.721)		
LINACC_DOUB21							24.281	
							(1745.113)	
LINACC_XY								-590.292
								(598.297)
ICO_RATING		0.408***	0.398***	0.412***	0.437***	0.442***	0.439***	0.427***
		(0.154)	(0.154)	(0.152)	(0.158)	(0.157)	(0.167)	(0.157)
TAX_HAVEN		0.049	0.059	0.022	0.045	0.036	0.031	0.034
		(0.182)	(0.182)	(0.212)	(0.186)	(0.186)	(0.214)	(0.194)
ETHEREUM		-0.198	-0.206	-0.177	-0.172	-0.169	-0.172	-0.188
		(0.208)	(0.209)	(0.208)	(0.201)	(0.203)	(0.209)	(0.208)
ICO_PRICE		-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
		(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.006)	(0.004)
N_O_EXPERTS		0.106	0.108	0.114	0.112	0.115	0.116	0.114
		(0.072)	(0.073)	(0.080)	(0.072)	(0.072)	(0.081)	(0.075)
MINIMUM		-0.159	-0.173	-0.164	-0.136	-0.134	-0.131	-0.140
		(0.155)	(0.155)	(0.154)	(0.155)	(0.154)	(0.153)	(0.156)
N_O_TEAM		0.422***	0.412***	0.437***	0.450***	0.456***	0.451***	0.444***
		(0.153)	(0.153)	(0.148)	(0.153)	(0.152)	(0.148)	(0.157)
HARD_CAP		0.096	0.146	0.108	0.094	0.103	0.107	0.089
		(0.268)	(0.269)	(0.251)	(0.269)	(0.271)	(0.252)	(0.267)
SOFT_CAP		-0.395**	-0.408**	-0.373**	-0.380**	-0.385**	-0.379**	-0.377**
		(0.153)	(0.154)	(0.166)	(0.152)	(0.154)	(0.167)	(0.159)
UTIL/SEC		-0.212	-0.215	-0.191	-0.194	-0.216	-0.211	-0.219
		(0.307)	(0.308)	(0.477)	(0.315)	(0.318)	(0.478)	(0.319)
WHITE_KYC		-0.020	-0.033	-0.081	-0.074	-0.087	-0.091	-0.094
		(0.224)	(0.224)	(0.247)	(0.229)	(0.229)	(0.248)	(0.230)
FACEBOOK		-0.200	-0.209	-0.226	-0.231	-0.242	-0.239	-0.227
		(0.257)	(0.259)	(0.284)	(0.258)	(0.258)	(0.286)	(0.263)
TWITTER		-0.090	-0.078	-0.089	-0.003	-0.100	-0.079	-0.069
		(0.556)	(0.561)	(0.547)	(0.543)	(0.519)	(0.557)	(0.551)
TELEGRAM		-0.128	-0.127	-0.116	-0.114	-0.140	-0.122	-0.157
		(0.238)	(0.237)	(0.258)	(0.241)	(0.237)	(0.259)	(0.244)
GITHUB		-0.235	-0.229	-0.248	-0.237	-0.238	-0.243	-0.237
		(0.159)	(0.159)	(0.162)	(0.157)	(0.158)	(0.162)	(0.152)
FILESIZE		0.047	0.038	0.047	0.046	0.032	0.037	0.041
		(0.070)	(0.070)	(0.072)	(0.071)	(0.071)	(0.072)	(0.072)
TOTALWORDS		0.209*	0.209*	0.203*	0.209*	0.225*	0.226*	0.217*
		(0.119)	(0.119)	(0.106)	(0.121)	(0.119)	(0.106)	(0.120)
TONE		-4.378	-4.997	-5.939	-5.678	-5.904	-6.056	-6.306
		(11.565)	(1.162)	(11.971)	(11.584)	(11.555)	(12.001)	(11.562)
READABILITY		-0.015	-0.016*	-0.014	-0.015	-0.016*	-0.016	-0.015
		(0.010)	(0.009)	(0.010)	(0.010)	(0.009)	(0.011)	(0.010)
POPULATION		-0.037	-0.036	-0.038	-0.031	-0.034	-0.035	-0.034
		(0.038)	(0.039)	(0.036)	(0.035)	(0.035)	(0.043)	(0.035)
GDP		0.064	0.059	0.058	0.067	0.066	0.066	0.062
		(0.123)	(0.122)	(0.141)	(0.119)	(0.119)	(0.141)	(0.119)
FDI		0.436	0.432	0.549	0.515	0.589	0.575	0.533
		(0.561)	(0.561)	(0.599)	(0.517)	(0.548)	(0.601)	(0.538)
LOCATION		0.124	0.113	0.144	0.183	0.119	0.114	0.135
		(0.173)	(0.172)	(0.187)	(0.171)	(0.173)	(0.193)	(0.176)
Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Dummies	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Continent Dummies	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.025	0.167	0.168	0.164	0.162	0.160	0.160	0.162
Num. Obs.	546	546	546	546	546	546	546	546

This table reports the results on Hypothesis 1 on the relationship between linguistic accuracy and the amount raised during the ICO period (Model 1-2). Model 3 introduces a quadratic variable. Models 4 – 16 report a breakdown for the independent linguistic inaccuracies that collectively comprise *LINACC*. Variable definitions are reported in Table 1. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on a two-sided t-test. Robust standard errors, clustered at the country level, are reported between parentheses.

Table 6 (continued) – Linguistic Accuracy and ICO Amount Raised

	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14	Model 15	Model 16
(Intercept)	10.650***	10.730***	10.684***	10.681***	10.912***	10.441***	10.657***	10.730***
	(2.012)	(2.045)	(2.019)	(2.013)	(1.995)	(2.007)	(2.016)	(2.018)
LINACC_SWAP	537.681							
	(1178.721)							
LINACC_APOST		596.564						
		(522.581)						
LINACC_SPLIT			15.235					
			(75.166)					
LINACC_MORPH				-1041.059				
				(1462.482)				
LINACC_MULT					-1615.412**			
					(795.601)			
LINACC_GVN						-1187.021***		
						(372.012)		
LINACC_MISSING							4543.121	
							(6649.983)	
LINACC_DUP								-708.234
								(1109.110)
ICO_RATING	0.443***	0.425***	0.438***	0.438***	0.389**	0.420***	0.444***	0.435***
	(0.156)	(0.158)	(0.158)	(0.158)	(0.152)	(0.152)	(0.157)	(0.157)
TAX_HAVEN	0.039	0.021	0.035	0.030	0.016	0.026	0.033	0.034
	(0.186)	(0.192)	(0.186)	(0.189)	(0.184)	(0.184)	(0.184)	(0.184)
ETHEREUM	-0.170	-0.165	-0.176	-0.170	-0.149	-0.163	-0.165	-0.186
	(0.205)	(0.208)	(0.205)	(0.203)	(0.203)	(0.207)	(0.202)	(0.202)
ICO_PRICE	-0.003	-0.002	-0.002	-0.003	-0.003	-0.003	-0.003	-0.003
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
N_O_EXPERTS	0.113	0.117	0.115	0.116	0.241	0.098	0.119*	0.118*
	(0.071)	(0.074)	(0.072)	(0.072)	(0.072)	(0.072)	(0.071)	(0.072)
MINIMUM	-0.131	-0.139	-0.132	-0.131	-0.155	-0.149	-0.126	-0.131
	(0.155)	(0.155)	(0.155)	(0.155)	(0.154)	(0.153)	(0.155)	(0.154)
N_O_TEAM	0.452***	0.452***	0.449***	0.452***	0.431***	0.418***	0.446***	0.453***
	(0.152)	(0.156)	(0.154)	(0.153)	(0.149)	(0.149)	(0.153)	(0.153)
HARD_CAP	0.108	0.112	0.106	0.106	0.123	0.065	0.109	0.106
	(0.270)	(0.275)	(0.271)	(0.271)	(0.265)	(0.271)	(0.212)	(0.269)
SOFT_CAP	-0.387**	-0.393**	-0.381**	-0.380**	-0.395**	-0.403**	-0.380**	-0.374**
	(0.155)	(0.158)	(0.154)	(0.152)	(0.154)	(0.156)	(0.153)	(0.153)
UTIL/SEC	-0.195	-0.184	-0.213	-0.212	-0.213	-0.268	-0.217	-0.196
	(0.309)	(0.324)	(0.316)	(0.318)	(0.317)	(0.306)	(0.317)	(0.317)
WHITE_KYC	-0.096	-0.081	-0.079	-0.090	-0.061	-0.021	-0.088	-0.097
	(0.227)	(0.229)	(0.234)	(0.228)	(0.224)	(0.238)	(0.229)	(0.230)
FACEBOOK	-0.242	-0.244	-0.237	-0.236	-0.240	-0.225	-0.252	-0.248
	(0.256)	(0.261)	(0.257)	(0.258)	(0.259)	(0.253)	(0.258)	(0.255)
TWITTER	-0.073	-0.068	-0.081	-0.081	-0.108	-0.097	-0.070	-0.084
	(0.539)	(0.527)	(0.541)	(0.539)	(0.589)	(0.565)	(0.539)	(0.542)
TELEGRAM	0.126	0.109	0.124	0.131	0.125	0.111	0.107	0.129
	(0.240)	(0.244)	(0.241)	(0.245)	(0.138)	(0.238)	(0.234)	(0.241)
GITHUB	-0.246	-0.244	-0.244	-0.241	-0.236	-0.190	-0.247	-0.239
	(0.157)	(0.153)	(0.158)	(0.158)	(0.159)	(0.157)	(0.158)	(0.158)
FILESIZE	0.036	0.039	0.037	0.036	0.041	0.047	0.036	0.036
	(0.070)	(0.071)	(0.071)	(0.070)	(0.069)	(0.069)	(0.070)	(0.070)
TOTALWORDS	0.235**	0.224*	0.225*	0.226*	0.224*	0.223*	0.226*	0.225*
	(0.112)	(0.121)	(0.119)	(0.119)	(0.116)	(0.115)	(0.119)	(0.118)
TONE	-5.968	-5.874	-5.897	-6.233	-4.273	-6.933	-6.343	-5.859
	(11.578)	(11.565)	(11.621)	(11.615)	(11.657)	(11.462)	(11.573)	(11.571)
READABILITY	-0.016*	-0.015	-0.016	-0.016	-0.015	-0.015	-0.015	-0.015
	(0.009)	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)
POPULATION	-0.035	-0.035	-0.035	-0.036	-0.033	-0.030	-0.035	-0.036
	(0.035)	(0.035)	(0.035)	(0.035)	(0.034)	(0.035)	(0.035)	(0.035)
GDP	0.058	0.064	0.066	0.066	0.059	0.083	0.065	0.063
	(0.121)	(0.119)	(0.119)	(0.119)	(0.123)	(0.121)	(0.119)	(0.118)
FDI	0.604	0.557	0.569	0.574	0.522	0.443	0.575	0.579
	(0.555)	(0.537)	(0.550)	(0.549)	(0.559)	(0.554)	(0.549)	(0.547)
LOCATION	0.172	0.164	0.138	0.194	0.183	0.176	0.179	0.165
	(0.179)	(0.185)	(0.169)	(0.174)	(0.171)	(0.168)	(0.182)	(0.180)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Continent Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.160	0.162	0.160	0.160	0.172	0.177	0.161	0.161
Num. Obs.	546	546	546	546	546	546	546	546

Table 7 – Cross-sectional Variation: Regulation, Institutional strength, and Native English countries

	Model 1	Model 2	Model 3
(Intercept)	12.353***	11.837***	11.374***
	(2.308)	(2.095)	(2.094)
LINACC	-221.214***	-242.173***	-72.695
	(75.921)	(85.329)	(70.167)
UNREGULATED	0.132		
	(0.175)		
LINACC x *UNREGULATED*	-203.112**		
	(99.314)		
INST		-0.002	
		(0.006)	
LINACC x *INST*		62.750**	
		(29.369)	
NES			-0.287
			(0.232)
LINACC x *NES*			-192.913*
			(115.862)
ICO_RATING	0.385**	0.422***	0.395**
	(0.166)	(0.151)	(0.154)
TAX_HAVEN	0.038	0.037	0.002
	(0.215)	(0.176)	(0.183)
ETHEREUM	-0.194	-0.205	-0.203
	(0.207)	(0.207)	(0.206)
ICO_PRICE	-0.003	-0.000	-0.004
	(0.004)	(0.002)	(0.004)
N_O_EXPERTS	0.120	0.112*	0.107
	(0.080)	(0.067)	(0.073)
MINIMUM	-0.161	-0.175	-0.159
	(0.152)	(0.154)	(0.156)
N_O_TEAM	0.419***	0.430***	0.411***
	(0.147)	(0.154)	(0.153)
HARD_CAP	0.055	0.117	0.131
	(0.250)	(0.260)	(0.271)
SOFT_CAP	-0.368**	-0.390**	-0.409***
	(0.165)	(0.151)	(0.153)
WHITE_KYC	-0.278	-0.007	-0.021
	(0.251)	(0.218)	(0.258)
UTIL/SEC	-0.283	-0.232	-0.252
	(0.473)	(0.300)	(0.307)
FACEBOOK	-0.211	-0.211	-0.190
	(0.282)	(0.261)	(0.252)
TWITTER	-0.105	-0.060	-0.079
	(0.549)	(0.052)	(0.055)
TELEGRAM	0.095	0.143	0.112
	(0.257)	(0.223)	(0.237)
GITHUB	-0.216	-0.251	-0.221
	(0.161)	(0.161)	(0.160)
FILESIZE	0.049	0.051	0.046
	(0.071)	(0.071)	(0.071)
TOTALWORDS	0.227**	0.216*	0.213*
	(0.105)	(0.113)	(0.121)
TONE	-8.544	-5.561	-3.375
	(11.982)	(11.913)	(11433)
READABILITY	-0.016	-0.014	-0.015
	(0.010)	(0.010)	(0.010)
POPULATION	-0.053	-0.029	-0.047
	(0.044)	(0.034)	(0.037)
GDP	0.055	0.037	0.048
	(0.148)	(0.163)	(0.126)
FDI	0.582	0.509	0.471
	(0.601)	(0.575)	(0.558)
LOCATION	0.145	0.100	0.088
	(0.167)	(0.164)	(0.194)
Year FE	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes
Continent Dummies	Yes	Yes	Yes
Adj. R^2	0.178	0.174	0.169
Num. Obs.	546	546	546

This table reports the results pertaining to Hypothesis 2, and 3 on the relationship between linguistic accuracy and the amount raised during the ICO period, and the moderating factor of (i) the ICO home country being regulated as per Shrestha et al. (2021) (Model 1), (ii) the institutional strength of the country in which the ICO is issued, and (iii) whether the ICO stems from a country where English is a native language (Model 3). Variable Definitions are reported in Table 1. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on a two-sided t-test. Robust standard errors, clustered at the country level, are reported between parentheses.

Table 8 – Additional Analyses

Panel A – Sin Industry

Dependent variable	Model 1 *log AMOUNT_RAISED*
LINACC	-91.866*
	(51.279)
LINACC x *SIN*	-549.762***
	(11.376)
SIN	0.642
	(0.548)
Controls	Yes
Year FE	Yes
Industry Dummies	Yes
Continent Dummies	Yes
Adj. R^2	0.173
Num. Obs.	546

Panel B – Position of Errors in the text

Dependent variable	Model 2 *log AMOUNT_RAISED*
LINACC_Q1	42.517
	(55.674)
LINACC_Q2	-52.295
	(32.079)
LINACC_Q3	28.632
	(52.808)
LINACC_Q4	-146.781***
	(72.999)
Controls	Yes
Year FE	Yes
Industry Dummies	Yes
Continent Dummies	Yes
Adj. R^2	0.179
Num. Obs.	546

Panel C – Ex-post listing and hard cap achieved

Dependent variable	Model 3 *TRADED*	Model 4 *HARD_CAP_ACHIEVED*
LINACC	-300.637**	-260.023**
	(149.951)	(129.812)
Controls		Yes
Year FE		Yes
Industry Dummies	Yes	Yes
Continent Dummies	Yes	Yes
McFadden. R^2	0.279	0.328
Num. Obs.	546	485

Panel D – Heckman (2nd step) and Restricted Control Function (2nd step)

Dependent variable	Model 5 log *AMOUNT_RAISED*	Model 6 *log AMOUNT_RAISED*
LINACC	-78.489*	-161.351**
	(40.127)	(74.914)
IMR	Yes**	No
GENRES	No	Yes***
Controls	Yes	Yes
Year FE	Yes	Yes
Industry Dummies	Yes	Yes
Continent Dummies	Yes	Yes
Adj. R^2	0.182	0.204
Num. Obs.	546	546

Panel E – Instrumental Variable Analysis

Dependent variable	Model 7 (Step 1) *LINACC*	Model 8 (Step 2) *log AMOUNT_RAISED*
BRACKET	-0.121***	
	(0.020)	
LINACC_ESTIMATED		-308.902*
		(170.703)
Controls	Yes	Yes
Year FE	Yes	Yes
Industry Dummies	Yes	Yes
Continent Dummies	Yes	Yes
F-statistic (Weak Instrument)	12.084***	
Wu-Hausman statistic		1.525
Adj. R^2	(0.160†) 0.246	0.147
Num. Obs.	546	546

This table reports additional tests pertaining to the relationship between linguistic accuracy and the funding outcome of ICOs. In Panel A, we study the moderating factor of an ICO being active in a "sin" industry (*SIN*). In Panel B, we divide the white papers into quartiles and measure the fraction of linguistic errors per quartile. In Panel C, we regress the linguistic accuracy on the ex-post measure *TRADED*, which takes the value of '1' if the ICO is listed on a secondary platform (Coinmarketcap), and zero otherwise. We also regress *LINACC* on *HARD_CAP_ACHIEVED*, which takes the value of '1' if the amount raised exceeded the pre-defined ICO hard cap. In Panel D, we conduct tests to mitigate endogeneity issues. We report in Model 5 the second stage of a two-step selection procedure, in which the presence of our ICOs is weighted against a large sample of 1,340 ICOs, using all non-white paper related control variables. The calculated Inverse Mills Ratio (*IMR*) is included in the second step. In Model 6, we report the results of our restricted control function, where we include the generalized residuals (*GENRES*) to our main Equation (4). Panel E provides the results of the instrumental variable analysis. The first step (Model 7) regresses the number of brackets in the text, divided by the total number of words, against our independent variable (*LINACC*). † represents the adjusted R^2 when excluding the instrumental variable. The fitted values of the first step (*LINACC_ESTIMATED*) are then regressed against the amount raised in the second step (Model 8). Variable Definitions are reported in Table 1. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on a two-sided t-test. Robust standard errors, clustered at the country level, are reported between parentheses.

Table 9 – Experiment on Linguistic Inaccuracies and Individuals' Decision to Invest

Panel A – Summary statistics

	mean	sd	min	Q25	median	Q75	max
EX_ERRORS+	3.035	1.799	0	2	3	4	6
EX_INVEST	450.6	272.512	0	200	500	651	1,000
EX_WILLINGNESS	3.532	1.157	1	3	4	4	5
EX_AUTHOR	71.62	19.637	0	61	75	85	100
EX_PRIORINVEST	0.485	0.501	0	0	0	1	1
EX_FINRISK	3.772	1.317	0	3	4	5	5
EX_AGE	32.25	8.815	18	25	31	36	74
EX_GENDER	0.624	0.485	0	0	1	1	1
EX_NATIVE	0.673	0.469	0	0	1	1	1
EX_EDUCATION	0.791	0.406	0	1	1	1	1

Panel B – Multivariate analysis

	Model 1	Model 2
Dependent variable	*EX_INVEST*	*EX_WILLINGNESS*
(Intercept)	-58.275	1.201***
	(63.382)	(0.278)
EX_ERRORS	-38.096*	-0.195**
	(20.975)	(0.092)
EX_AUTHOR	4.129***	0.012***
	(0.614)	(0.003)
EX_PRIORINVEST	161.866***	0.518***
	(25.977)	(0.114)
EX_FINRISK	20.581*	0.221**
	(10.581)	(0.046)
EX_AGE	-0.909	0.005
	(1.295)	(0.005)
EX_GENDER	-43.494	0.088
	(29.423)	(0.011)
EX_NATIVE	50.284*	0.042
	(29.513)	(0.129)
EX_EDUCATION	154.929***	0.509***
	(34.450)	(0.151)
Adj. R^2	0.385	0.863
Num. Obs.	346	346

This table reports the results of an experiment in which participants are asked to read an ICO pitch and express their willingness to invest as well as the dollar amount invested, while the number of linguistic errors vary per pitch. Panel A reports the summary statistics of the sample. + indicates that a logarithm is used in the multivariate regression, but that untransformed values are reported for interpretation purposes. Panel B reports the results of multivariate regressions between the number of linguistic errors (*EX_ERRORS*) and the amount invested by the participants (*EX_INVEST*) (Model 1) and the investors' willingness to invest (*EX_WILLINGNESS*). Variable Definitions are reported in Table 1. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on a two-sided t-test. Robust standard errors are reported between parentheses.

Appendix

A1 – Examples of the linguistic error categories

In the appendix below, we provide some further examples on the different categories of linguistic errors encountered in the different ICO white papers.

FS_Mis

Linguistic errors regarding the omission of a letter are included in this category. Below, an excerpt of the Patron White paper can be found, in which the word “Bord” should refer to “Board”. This is a clear omission of the letter ‘a’. Patron sold about $40,000,000 worth of tokens (PAT token) and provides a platform for social media influencers to buy and sell data (for more information, see: https://icobench.com/ico/patron).

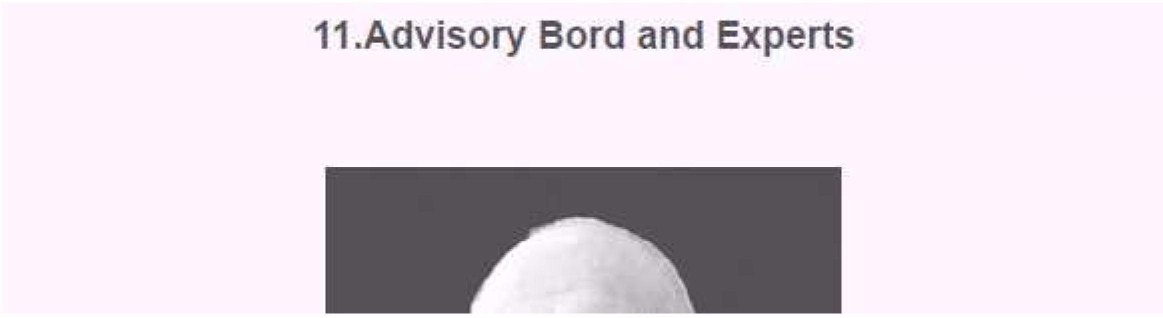


Figure A.1: Excerpt from the Patron White paper

FS_Red

Linguistic errors pertaining to the erroneous addition of a letter are included in this category. Below, an excerpt of the CFun White paper can be found, in which they mention that “The ICO will offier”, in which “offier” erroneously includes an additional ‘i’. CFun sold about $15,000,000 worth of tokens (PAT token) and aims to exploit blockchain technologies to properly determine the extent to which users are collaborates/co-authors/co-owners of projects (for more info: https://icobench.com/ico/cfun).


5.3 ICO Details

The ICO will offier a total amount of 500 million CFun. 1 CFun is equal to 0.03 USD. The soft cap of the ICO is set at 4.5 million USD, and hard cap is set at 15 million USD.

Figure A.2: Excerpt from the CFun White paper

FS_Doub12

This category contains linguistic errors pertaining to the use of a single letter instead of double letters. Below, an excerpt of the Cajutel-Sarl White paper, an ICO from Switzerland, can be found. They mention that “… which will bring additional busines to the country”. Herein, the word “busines” should be written as “business” as an additional ‘s’ should have been added ad the end of the word. Cajutel sold about $152,598 worth of tokens (CAJUTEL token) and aims to build a cost-effective broadband access network (for more info, see : https://icobench.com/ico/cfun).

diversification. Latest development have confirmed that the country is well on its path. After several years of economic downturn and political instability, in 1997, Guinea-Bissau entered the CFA franc monetary system, bringing about some internal monetary stability.
Guinea-Bissau is a member of the Organization for the Harmonization of Business Law in Africa (OHADA).[39]
Guinea-Bissau has a natural resources of oil but it is currently not being harvested but development to do so is in progress which will bring additional busines to the country.
The currency is the west African CFA which is tied to the Euro (655.957 CFA = 1€)

Figure A.3: Excerpt from the Cajutel-Sarl White paper

FS_Doub21

This category contains linguistic errors pertaining to the use of double letters instead of a single letter. Below, an excerpt of the Pickcio-chain White paper can be found. They mention that "… and to determine if incomming data is trustworthy". Herein, the word "incomming" should be written as "incoming", in that an additional 'm' has been included. Ultimately, Pikcio-chain issued about $2,269,958 worth of tokens (PKC token) during their ICO and its purpose was to build a data-exchange platform. However, the Pikcio-chain disbanded in early 2020 (for more info, see: https://icobench.com/ico/pikciochain and https://neonewstoday.com/general/pikcioag-announces-it-is-shutting-down-after-two-years-of-operation-switcheo-delists-pkx/).

Trust Capital Index

From a security perspective, each separate party that validates the same piece, or collection of data, will add an entry into the ledger for that data and will in effect create a sort of reputation of trustworthiness for the data and the data owner. This reputation system is called the Trust Capital Index, or TCI in short. It is a way for parties to judge the reliability of users and certain data on the system and to determine if incomming data is trustworthy.

Figure A.4: Excerpt from the Pikcio-chain White paper

FS_XY

This category contains linguistic errors pertaining to the substitution of a letter. Below, an excerpt of the MeeTip White paper can be found. They erroneously use "Messasing service" to refer to "Messaging service". Herein, a simple typo is made in that the letter 'g' was erroneously replaced by the letter 's'. Ultimately, MeeTip issued about $3,226,597 worth of tokens (MTIP token) during their ICO and its purpose was to be used on OutMySphere app, to purchase drinks and tip waiters (for more info, see: https://icobench.com/ico/meetip).

Messasing service allows you to chat with people around you and your friends. You will also have the ability to instantly send MeeTip to them. To protect privacy, end-to-end message encryption will be integrated. With this, only you and the person you are communicating with can read what is sent, and nobody in between, not even us.

Figure A.5: Excerpt from the MeeTip White paper

FS_Swap

This category contains linguistic errors pertaining to the swapping of two adjacent letters. Below, an excerpt of the UService White paper can be found. They erroneously use "variuos" instead of "various". Herein, the letters 'u' and 'o' simply switched places. Ultimately, MeeTip issued about $23,829,663 worth of tokens (UST token) during their ICO and its purpose is to connect car industry specialists (For more info, see: https://icobench.com/ico/uservice).

> According to **LeasePlan** company research car owning cost in variuos countries varies from **364 euro to 708 euro** per month.

Figure A.6: Excerpt from the UService White paper

FS_Apost

This category contains linguistic errors pertaining to the substation of a letter. Below, an excerpt of the Golden Fleece White paper can be found. They erroneously use "it's" instead of "its" (note that this excerpt also contains a grammatical error ("increase" instead of "increases") and a swapped letter ("at" instead of "as")). Ultimately, Golden Fleece issued about $500,000 worth of tokens (GFL token) during their ICO and its purpose was to build an ICO mining company in Georgia (for more info, see: https://icobench.com/ico/golden-fleece).

> Golden Fleece (GFL) White Paper (V3)
>
> 3.3 Golden Fleece Mining Operations
>
> Golden Fleece revenue increase over years at it's capacity increases,

Figure A.7: Excerpt from the Golden Fleece White paper

FS_Split

This category contains linguistic errors pertaining to the segmentation of words. Below, an excerpt of the Biofactorycoin White paper can be found. Herein, they write "ice-cream" instead of "ice cream", in which an erroneous hyphen is being used. Ultimately, Biofactorycoin issued about $39,223,918 worth of tokens (BFC token) during their ICO and its purpose was to sell biological dairy products directly from the factory using cryptocurrencies as payment (for more info, see: https://icobench.com/ico/biofactorycoin).

> Our team is ready to accept this challenge and to substitute solitary milk with a line of the ecological milk production (organic milk, yogurt with increased quantity of proteins and vegan ice-cream). We know how to produce ecological and useful for health production using contemporary technologies. Investing to our BFC token crypto currency you will contribute to not only the innovations of nowadays food market but will directly take a part in the increasing of the food industry and contribute to healthy life style.

Figure A.8: Excerpt from the Biofactorycoin White paper

FS_Morph

This category contains linguistic errors pertaining to the wrong use(s) of pre- and suffixes. Below, an excerpt from the Tokenstars ACE White paper can be found. Herein, they write "cause" instead of "because", although "'cause" is often used in spoken language, there is clearly a prefix "be" missing. Ultimately, TokenStars ACE issued about $5,000,000 worth of tokens (ACE token) during their ICO and its purpose is to introduce blockchain based rewards to young sports talent and ensure that young talent receives payment (for more info, see: https://icobench.com/ico/tokenstars-ace).

> One of the most remarkable conflicts between an agency and a player was Ivan Lendl's case, the winner of seven Grand Slam tournaments and undisputable No.1 in the ATP ranking. Ivan was dissatisfied with the work of agency, cause at the top of his career he experienced lack of attention from his agents. Later Lendl accused Proserv of exploiting his image for 'packaging' with other players at less favorable terms.

Figure A.9: Excerpt from the Tokenstars Ace White paper

FS_Mult

This category contains linguistic errors pertaining to the occurrence of multiple forms of errors in the same word. Below, an excerpt from the BitRewards White paper can be found. Herein, they write "Websitesof" instead of "Website", in which (i) erroneous segmentation is being used: "Websitesof" → "Websites of" and (ii) we believe that they refer to a single website, rather than multiple websites (alternatively, they may refer to trading websites - i.e. "Websites of trading platforms" -, in which case there are still multiple errors since at least one word is missing). Ultimately, BitRewards issued about $6,583,400 worth of tokens (Bit token) during their ICO and its purpose is to introduce blockchain based rewards for online shopping (for more info, see: https://icobench.com/ico/bitrewards).

> Regulatory authorities in many countries carefully study the enterprises and operations related to cryptocurrencies. In this regard, regulatory measures, investigations or actions may affect the activities of BitRewards Network and even limit or prevent its development in the future. Any person who undertakes to purchase BIT must be aware of BitRewards Network business model. This white paper and the Terms and Conditions available at the Websitesof may change or need to be modified due to

Figure A.10: Excerpt from the BitRewards White paper

FS_Missing

This category contains linguistic errors pertaining to the omission of a complete word. Below, an excerpt from the PlanEx White paper can be found. Herein, they write "hundreds of thousands dollars" instead of "hundreds of thousands of dollars", in which the word "of" was omitted. Ultimately, Planex issued about $3,700,000 worth of tokens (Bit token) during their ICO and its purpose is to provide a platform to exchange cryptocurrencies for fiat currencies (for more info, see: https://icobench.com/ico/planex).

> Placing assets of a company on one classical stock exchange costs hundreds of thousands dollars and usually it is possible only on terms of the country on which the company is registered. As a result, only big companies are able to afford emission of securities.

Figure A.11: Excerpt from the PlanEx White paper

FS_Dup

This category contains linguistic errors pertaining to the erroneous inclusion or repetition of a complete word. Below, an excerpt from the Monster Byte White paper can be found. Herein, they write "… the the shuffled deck", in which the word "the" was repeated. Ultimately, Monster Byte issued about $1,000,000 worth of tokens (MBI token) during their ICO and its purpose is to provide blockchain related bankroll services for online gaming companies (for more info, see: https://icobench.com/ico/monster-byte).

> ○ At a high level, Provably Fair Shuffling is a technique that Monster Byte leverages which allows the user to reshuffle the deck before it is dealt. The system first shuffles the deck once, and then presents the user with a hash of the the shuffled deck before any wagers are made. Then, whenever the user places a wager, he can optionally provide a client_seed. The system then uses the client_seed provided to

Figure A.12: Excerpt from the Monster Byte White paper

FS_GVN

This category contains linguistic errors pertaining wrongful conjugation or tense use of a verb. Below, an excerpt from the Lambda White paper can be found. Herein, they write "We estimates that", in which the conjugation should clearly be "we estimate". Ultimately, Lambda issued about $15,000,000 worth of tokens (LAMB token) during their ICO and its purpose is to provide a large-scale data storage infrastructure (for more info, see: https://icobench.com/ico/lambda-1).

> **To a certain degree, validators are analogous to mining pools in the bitcoin system.**
> We estimates that hundreds of validator nodes will be contained in the system.

Figure A.13: Excerpt from the Lambda White paper

Table A2 – Univariate differences between quartile linguistic inaccuracies

(a) *vs.* *(b)*	Mean (a)	Mean (b)	t-test
LINACC_Q1 vs. *LINACC_Q2*	0.110	0.113	-1.560
LINACC_Q1 vs. *LINACC_Q3*	0.110	0.112	-0.212
LINACC_Q1 vs. *LINACC_Q4*	0.110	0.106	0.41
LINACC_Q2 vs. *LINACC_Q3*	0.113	0.112	1.39
LINACC_Q2 vs. *LINACC_Q4*	0.113	0.106	1.02
LINACC_Q3 vs. *LINACC_Q4*	0.112	0.106	0.625

This table reports univariate differences between the occurrences of linguistic errors in the different quartiles of the white paper. Mean values are reported, multiplied by 100 for presentation purposes. Variable Definitions are reported in Table 1.

A3 – Two versions of the mTurk experiment ICO pitch

This appendix provides two examples of the ICO white paper to the respondents on mTurk. The first example contains no linguistic errors, while the second version of the white paper contains six errors, which are underlined.

No linguistic errors

About TATT:
TATT is not just aiming to challenge incumbent banks, it is a collaborative banking platform that aims to dramatically improve the personal economies of our customers. Customers today not only face low interests rate, but a multitude of banking scandals have resulted in a loss of trust. We are currently building a new and innovative financial platform that suits the needs of our globalizing world. One that goes beyond dollars and euros, but incorporates all other currencies and even allows cryptocurrencies to be deposited. A more important feature is that revenues are shared with customers based on each of their particular contributions.

TATT has been under development since November 2018. The current technology, including its back-end architecture, big data, blockchain, and front-end modules, is already serving 5,000 beta users, and is running on cloud infrastructure. TATT's collaborative model is built around its native token, the TATT Coin. TATT intends to issue the TATT token as a Virtual Financial Asset ("VFA") under the Virtual Financial Assets Act in Malta (a member state of the European Union), which is the first nation to enact a regulatory framework for virtual assets that leverage distributed ledger technologies.

In order to meet TATT's capital needs for the following 18 months, we wish to collect $20 million. TATT intends to issue and distribute the TATT Coin by undertaking an Initial Virtual Financial Assets Offering ("IVFAO") in accordance to the VFA Act, once it has satisfied all the applicable requirements prescribed by the VFA Act, including but not limited to the registration of the whitepaper with the MFSA.

About the TATT CEO: ***Brad Johnson***
Mr. Johnson is dedicated to the development of enterprises and startups, particularly in the banking and retail industry. He has over 30+ years of Private International, Commercial and Public Development banking experience. He currently serves as the Chairman of Fidelity Bank and acted as its former CEO, prior to starting TATT. In the past, Mr. Johnson worked for many internationally renowned banks in Asia, Europe and the U.S.

Beyond his direct experience in Banking, he has been appointed to advisory boards in several major regional corporations in a multitude of sectors such as manufacturing, agriculture, international trade, education, and the non-profit sector. Mr. Johnson obtained a B.S. from the University of Michigan and is a graduate of the University of Florida, where he earned a B.S. in Economics and a master's degree in Applied Economics.

Six linguistic errors (underlined)

About TATT:
TATT is not just aiming to challenge incumbent banks, it is a *collabrative* banking platform that aim to dramatically improve the personal economies of our customers. Customers today not only face low interests rate, but a *multute* of *bankin* scandals *has* resulted in a loss of trust. We are currently building a new and *inovativ* financial platform that suits the needs of our globalizing world. One that goes beyond dollars and euros, but incorporates all other currencies and even allows cryptocurrencies to be deposited. A more important feature is that revenues are shared with customers based on each of their particular contributions.

TATT has been under development since November 2018. The current technology, including its back-end architecture, big data, blockchain, and front-end modules, is already serving 5,000 beta users, and is running on cloud infrastructure. TATT's collaborative model is built around its native token, the TATT Coin. TATT intends to issue the TATT token as a Virtual Financial Asset ("VFA") under the Virtual Financial Assets Act in Malta (a member state of the European Union), which is the first nation to enact a regulatory framework for virtual assets that leverage distributed ledger technologies.

In order to meet TATT's capital needs for the following 18 months, we wish to collect $20 million. TATT intends to issue and distribute the TATT Coin by undertaking an Initial Virtual Financial Assets Offering ("IVFAO") in accordance to the VFA Act, once it has satisfied all the applicable requirements prescribed by the VFA Act, including but not limited to the registration of the whitepaper with the MFSA.

About the TATT CEO: *Brad Johnson*
Mr. Johnson is dedicated to the development of enterprises and startups, particularly in the banking and retail industry. He has over 30+ years of Private International, Commercial and Public Development banking experience. He currently serves as the Chairman of Fidelity Bank and acted as its former CEO, prior to starting TATT. In the past, Mr. Johnson worked for many internationally renowned banks in Asia, Europe and the U.S.

Beyond his direct experience in Banking, he has been appointed to advisory boards in several major regional corporations in a multitude of sectors such as manufacturing, agriculture, international trade, education, and the non-profit sector. Mr. Johnson obtained a B.S. from the University of Michigan and is a graduate of the University of Florida, where he earned a B.S. in Economics and a master's degree in Applied Economics.

Table A4 – The moderating impact of other linguistic features

Panel A – Univariate Comparison

	TECHNICAL TERMS ≤ median (*TECHNICAL TERMS*)	*TECHNICAL TERMS* > median (*TECHNICAL TERMS*
LINACC (mult. *x* 100)	0.100	0.159
	t-stat = -1.947*	
	READABILITY ≤ median (*READABILITY*)	*READABILITY* > median (*READABILITY*)
LINACC (mult. *x* 100)	0.156	0.108
	t-stat = 1.453	
	TONE ≤ median (*TONE*)	*TONE* > median (*TONE*)
LINACC (mult. *x* 100)	0.153	0.105
	t-stat = 1.592	

Panel B – Regression Analyses

Dependent variable	*log AMOUNT_RAISED*	

Panel B.1 – Technical Terms

	Model 1 *TECHNICAL TERMS* ≤ median (*TECHNICAL TERMS*)	Model 2 *TECHNICAL TERMS* > median (*TECHNICAL TERMS*)
LINACC	-37.133	-142.168*
	(85.460)	(84.241)
Controls	Yes	Yes
Year FE	Yes	Yes
Industry Dummies	Yes	Yes
Continent Dummies	Yes	Yes
Adj. R^2	0.085	0.237
Num. Obs.	273	273

Panel B.2 – Readability

	Model 3 *READABILITY* ≤ median (*READABILITY*)	Model 4 *READABILITY* > median (*READABILITY*)
LINACC	-107.387	-119.491*
	(291.173)	(74.281)
Controls	Yes	Yes
Year FE	Yes	Yes
Industry Dummies	Yes	Yes
Continent Dummies	Yes	Yes
Adj. R^2	0.148	0.189
Num. Obs.	245	301

Panel B.3 – Tone

	Model 5 *TONE* ≤ median (*TONE*)	Model 6 *TONE* > median (*TONE*)
LINACC	-155.280*	-76.143
	(89.751)	(75.854)
Controls	Yes	Yes
Year FE	Yes	Yes
Industry Dummies	Yes	Yes
Continent Dummies	Yes	Yes
Adj. R^2	0.148	0.184
Num. Obs.	273	273

This table reports the relationship between linguistic inaccuracies and the dollar amount raised, contingent on the white papers' (i) technical emphasis, (ii) readability, and (iii) tone. Panel A reports the results of univariate differences between the white papers. Mean values are reported, multiplied by 100 for presentation purposes. Panel B reports the results of regression analyses. Variable Definitions are reported in Table 1. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on a two-sided t-test. Robust standard errors are reported between parentheses.