AIS Transactions on Human-Computer Interaction

Volume 14 | Issue 2 Article 5

6-2022

# Understanding the Message and Formulation of Fake Online Reviews: A Language-production Model Perspective

Boran Wang
*The University of Sydney*, bwan3230@uni.sydney.edu.au

Kevin K.Y. Kuan
*The University of Sydney*, kevin.kuan@sydney.edu.au



## Recommended Citation

Wang, B., & Kuan, K. K. (2022). Understanding the Message and Formulation of Fake Online Reviews: A Language-production Model Perspective. *AIS Transactions on Human-Computer Interaction, 14*(2), 207-229. https://doi.org/10.17705/1thci.00167
DOI: 10.17705/1thci.00167

Transactions on Human-Computer Interaction

Volume 14 Issue 2

6-2022

# Understanding the Message and Formulation of Fake Online Reviews: A Language-production Model Perspective

Boran Wang

*School of Electrical and Information Engineering, The University of Sydney, bwan3230@uni.sydney.edu.au*

Kevin K.Y. Kuan

*School of Computer Science, The University of Sydney, kevin.kuan@sydney.edu.au*



**Recommended Citation**

Wang, B., & Kuan, K. K. Y. (2022). Understanding the message and formulation of fake online reviews: A language-production model perspective. *AIS Transactions on Human-Computer Interaction*, *14*(2), pp. 207-229.

DOI: 10.17705/1thci.00167

Available at http://aisel.aisnet.org/thci/vol14/iss2/5

Transactions on Human-Computer Interaction

Research Paper DOI: 10.17705/1thci.00167 ISSN: 1944-3900

# Understanding the Message and Formulation of Fake Online Reviews: A Language-production Model Perspective

**Boran Wang**
School of Electrical and Information Engineering
The University of Sydney

**Kevin K.Y. Kuan**
School of Computer Science
The University of Sydney

**Abstract:**

Consumers have become ever more reliant on online reviews. Therefore, fake reviews have also become increasingly rampant and eroded online review platforms' credibility. Previous literature suggests that particular linguistic styles can manifest in fake reviews with reference to the varying stages of the language-production process. Drawing on the language-production model as our theoretical foundation, we examine the psycholinguistic styles of fake online reviews at the message and formulation level. We performed a computational linguistic analysis on 66,940 reviews from Yelp. Our results suggest that fake reviews align more with deceptive writing in terms of the message-level variables such as length and psychological (affective, cognitive, social, and perceptual) cues. Interestingly, we found that they align less with deceptive writing in terms of the formulation-level variables such as readability, pronouns, and part-of-speech tags, which may be due to the fake review writers' conscious attempt to follow the language styles that genuine reviews adopt.

**Keywords:** Fake Information, Fake Online Reviews, Language-production Model, Review Message, Review Formulation, Linguistic Characteristics.

Shuk Ying (Susanna) Ho was the accepting senior editor for this paper.

# 1 Introduction

Web technologies that incorporate social platforms have become an increasingly prominent information-transmission medium in the contemporary era. These online platforms provide a means for individual users to directly share their own user-generated content (UGC) with a large audience quickly with minimal effort and without external controls (Fontanarava et al., 2017). While these social platforms provide easier access to content that other users generate and higher information-transmission rates, the fact that they lack external scrutiny and accountability also means that misinformation can spread more easily throughout them (Fontanarava et al., 2017).

User-generated online reviews exemplify this trend. Along with developments in e-commerce, businesses today depend more on digital word of mouth (WoM) than ever before (Chevalier & Mayzlin, 2006), which has increased consumer reviews' importance. Against such a background, users have become ever more reliant on reviews in that they use them as an important decision-making basis when shopping for products and services online (Chen et al., 2008; Cheung et al., 2009). Therefore, reviews can shape consumers' attitudes towards the business and, hence, affect their purchase decisions (Dellarocas & Narayan, 2007; Gretzel & Yoo, 2008).

However, due to online reviews' importance, people and businesses sometimes generate fake reviews (Ren & Ji, 2017). Moreover, while the fact that one can easily access online review platforms has increased the transparency of the services and products that businesses provide, it has also posed a threat to online reviews' credibility based on the fact that users face few (if any) restrictions in posting them (Yoo & Gretzel, 2009) and no one generally holds them accountable for the content they post, which results in unreliable and even fake reviews to abuse the system (Mukherjee et al., 2013b). Fake reviews refer to deceptive reviews that one produces to intentionally mislead consumers in their decision-making process (Zhang et al., 2016). They generally lack information based on one's actual experience in using a product or service (Zhang et al., 2016) or they conflict/do not concur with such an experience.

Fake reviews often arise due to competition. Businesses nowadays exploit such loopholes to produce fake positive reviews to advertise for themselves or undermine their competitors' reputation by posting fake negative reviews (Ott et al., 2013; Wang et al., 2011). Other than the businesses, consumers and potential consumers can also produce both positive and negative unwarranted reviews to misinform others about a product or service for either economic or personal reasons (e.g., a personal connection with the business) (Hunt, 2015). This misinformation not only misguides potential consumers when making decisions but pollutes the review platform environment and undermine users' trust in platforms, which reduces their usefulness as a whole (Ho & Richardson, 2013; Luca & Zervas, 2016).

Past studies have investigated economic incentives for creating fake reviews and found that they have become increasingly prevalent and sometimes even provided as a commercial service (Luca & Zervas, 2016). Ott et al. (2012) examined reviews from six popular online review platforms (Expedia, Hotels.com, Orbitz, Priceline, TripAdvisor, and Yelp) and found that, although at different growth rate, fake reviews have become a growing problem across them all. Fake review positing as a commercial service does not constitute a new phenomenon either. Back in 2011, *The New York Times* reported that one could purchase a fake five-star Yelp review on Amazon Mechanical Turk for just 25 U.S. cents (Segal, 2011). These fake reviews have drastically undermined online review platforms' transparency and credibility. However, while research has well established the psychology behind lie detection, one cannot easily detect fake review over the Internet since the medium lacks commonly used cues, such as fidgeting and other body language (Luca & Zervas, 2016; Wu et al., 2010).

While detecting online fake reviews represents a challenging task (Mukherjee et al., 2013b), in an effort to minimize users' exposure to fake reviews, many online review platforms have taken appropriate countermeasures and created automated review-filtering systems. Such platforms typically keep their review-filtering algorithms a secret, but they generally create them through supervised learning techniques using linguistic features and behavioral patterns as input (Mukherjee et al., 2013b). The proprietary arrangement restricts general users from understanding fake reviews' characteristics, which leaves potential consumers to identify fake reviews only according to their own intuitive measurements, which might lack accuracy in comparison. Moreover, although previous studies have considered the relationship between some psycholinguistic characteristics and fake reviews (Banerjee & Chua, 2014a; Banerjee et al., 2015), we currently lack empirical evidence about whether psycholinguistic characteristics can adequately distinguish fake reviews from genuine ones in conjunction with a theoretical framework.

Therefore, in this research, we address the following research question:

> **RQ**: What psycholinguistic characteristics differentiate fake and genuine online reviews and how do they reflect the differences in reviews' content and formulation.

In particular, we identify psycholinguistic measures that have the power to predict fake reviews and explain these measures' theoretical validity based on the language-production model (Levelt, 1993) perspective. This study constitutes among the first attempts to investigate psycholinguistic characteristics under the language-production model. We believe that, by fully analyzing and categorizing these variables, the results will further explain how and why fake reviews demonstrate certain psycholinguistic characteristics and, thus, contribute to a better, automated response to the rampant fake review problem.

# 2 Theoretical Background

Previous studies suggest that one can establish credibility through different linguistics characteristics (Banerjee & Chua, 2014b; Pennebaker, 2011). Research has studied this idea extensively in lie detection. Therefore, we draw on the relevant literature on deception and lie detection to study these characteristics.

Deception and lie detection have been a popular study area for centuries. Philosophers have debated deception's moral implications, while social psychological researchers have examined different cues to deception (Hartwig & Bond, 2014). Deception detection refers to a technique that people widely use either consciously or unconsciously in both normal and specific contexts (e.g., forensics) (Granhag & Strömwall, 2004). While typical everyday lies are quite trivial, lies in certain situations can have significant ramifications, which explains why researchers have focused on investigating deception and its cues and produced a substantial body of work on the topic (Hartwig & Bond, 2014). From previous studies, we identified three main paths to lie detection: 1) by observing how someone behaves (body movements, gaze, pitch, etc.), 2) by analyzing the speech content, or 3) by measuring the speaker's physiological indicators (Vrij et al., 2000). While once a popular approach to detect lies through measuring physiological indicators, polygraphs have waned in popularity as more recent studies have questioned its accuracy (Vrij, 2008). In more recent times, researchers have suggested that, for the best results, lie detection in verbal communications should exploit both verbal (the speech content) and non-verbal behaviors (e.g., tones, eye movements, and body language) (Vrij et al., 2000). However, in non-real-time written communication contexts, one can only analyze the exchanged text. Nonetheless, a growing body of research suggests that one can still reveal deception and other underlying thoughts via counting and categorizing different word types in written texts (Newman et al., 2003; Zhou & Sung, 2008).

The literature suggests that linguistic features can provide hints that facilitate deception detection (Banerjee & Chua, 2014b; Pennebaker, 2011). However, different features have varying strengths in relation to their ability to detect lies: some attributes indicate what a text expresses prima facie, but other methods to measurement text's psychological details may provide even more information on deception due to their hidden nature (Vrij et al., 2011).

## 2.1 Language-production Model

To understand how fake review writers fail to establish perceived credibility through the language they choose in their reviews, we draw on literature on language production to examine the effect that different psycholinguistic variables have on review credibility.

Studies in language production decompose the spoken or written language-production process into various sequential stages. These studies were used for various practical purposes, such as to detect lies and deception (Bott & Williams, 2018; Lane et al., 2006). By understanding the process that human brains undergo when formulating a message, one can better examine the relationships between the linguistic variables and information credibility and potentially reveal the underlying ground for such relationships. Hence, we use language-production models to understand fake reviews in a more structured and systematic manner.

Among the different language-production models, we draw on Levelt's (1993) classic work as our theoretical foundation. Researchers consider Levelt's (1993) language-production model a good framework to examine at which levels of the production process the lying effect occurs (Bott & Williams, 2018). According to the model, speech production goes through three essential levels: message, formulation, and articulation (Levelt, 1993). While the stages go through a sequential procedure to some extent (e.g., Levelt (1993) explained that conceptual planning at the message level needs to exist to some extent for formulation to occur), from a temporal perspective, the stages can run in parallel to some extent. However, from a component perspective, Levelt (1993) treats each stage as an autonomous processor that produces certain outputs at its end.

To understand Levelt's (1993) language-production model and its applicability to the present study, we discuss each level in the model. At the message level, speakers conceive the intention to communicate and select the relevant information to achieve it (Levelt, 1993). At this level, speakers must produce an overall plan to achieve their goals and subgoals (e.g., choose a method and select information to realize each goal/subgoal) (Levelt, 1993). As Levelt (1993) explains, the way in which speakers conceive intention drives their later formulation and articulation choices to effectively carry through the intention until they finally deliver the message. To put it simply, this level includes selecting the topics, concepts, and ideas but still remains as thoughts awaiting formulation and later articulation. Subsequently, at the formulation level, speakers process the conceptual ideas into linguistic sentences, which includes syntax planning (selecting words and sentence structures) and articulation planning (e.g., selecting tones and pitches) (Levelt, 1993). Syntax planning takes place before articulation planning. Syntax planning involves retrieving lemmas from speakers' lexicon and creating grammatical relations to reflect conceptual ideas in the message (Levelt, 1993). Syntax planning results in a surface structure, which then becomes the basis that speakers use for articulation planning in which they decide on the tune and rhyme (Levelt, 1993). At the end of the formulation level, speakers formulate textual expressions, choose words, and have prepared the message they will articulate. Finally, in the articulation level, speakers then execute the articulation plan and produce an audible output (verbally pronounces the words) through a sequence of neuromuscular movements[1].

Levelt (1993) treats each level as individual "autonomous processors" that takes an input and produce certain outputs. Considering each level in this way allows one to consider level subsets on their own merits without compromising the language-production model's efficacy as a whole. Thus, given online reviews' non-verbal nature, we focus on the first two levels in Levelt's language-production model to understand fake online reviews' characteristics at both the message level, which captures how speakers form intention and select information, and at the formulation level, which captures how speakers transform conceptual information into linguistic structures, while leaving out the irrelevant phonetic planning process at the formulation level.

Figure 1 illustrates the research model, which comprises three variable categories: message level, formulation level, and control variables. The message level includes review length and the psychological cues, while the formulation level includes review readability and the linguistic style variables.

[1] The articulation stage is sometimes referred to as 'Articulatory Stage' (Garrett, 1975) or 'Phonological Encoding Stage' (Bock & Levelt, 1994).

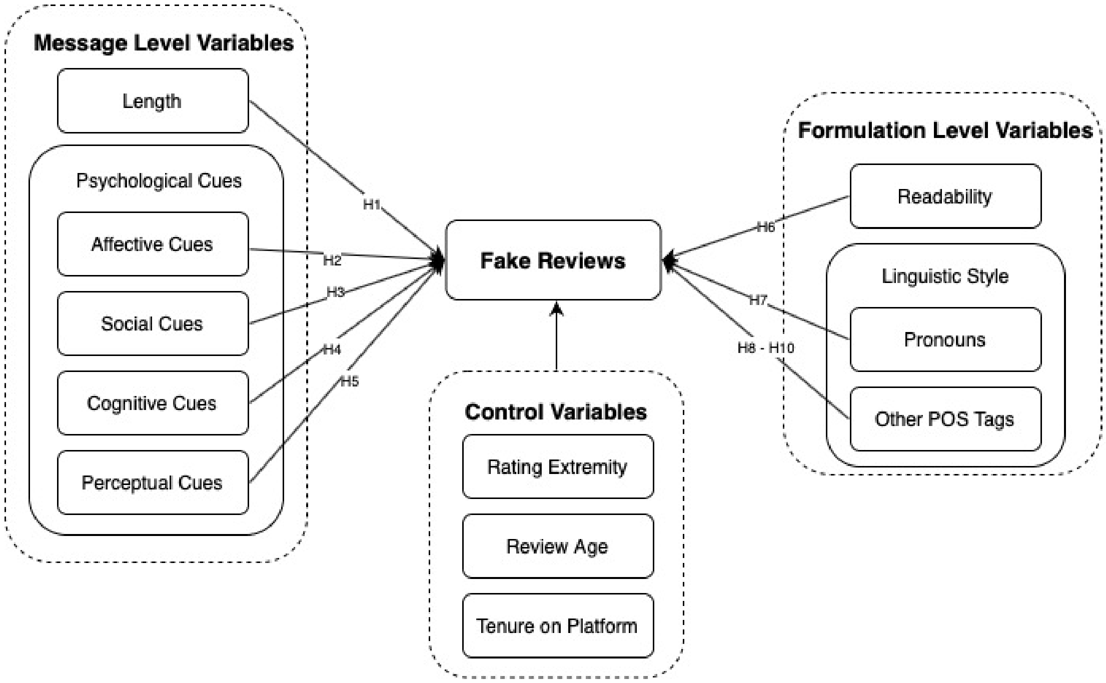


**Figure 1. Research Model**

## 2.2 Message-level Variables

At the message level, speakers confirm the intention to lie or to achieve specific purposes (Bott & Williams, 2018) and then decide on the relevant content and its scope (Levelt, 1993). In the online review context, during this process, an author would make important decision about what intentions they want to achieve through their review (intention formation), such as to persuade readers or describe something, to create a particular mood or psychological effect, to share a little or lot of information (review length), and/or to share particular information (information selection). However, at this stage, the information remains conceptual until authors process it at the formulation level and execute it (Levelt, 1993). Although one cannot directly measure the information related to this stage due to its conceptual nature (i.e., it still exists in their mind without any linguistic or grammatical structures), one can certainly trace the reviewer's intention and information choices from the final reviews that they produce based on Levelt's (1993) argument that message-level decisions contribute and, to a certain degree, drive the later production stages. Thus, reviewers carry the choices they make at the message level through to the final written review. Therefore, in order to fully investigate their mindset at the message level, one can consider the content a review includes (information selection) the message's intention (intention formation). Thus, by manipulating review length, which reflects how much content a review contains, and psychological cues, which reflect a review's intentions and nature, one can reveal reviewers' underlying thoughts.

### 2.2.1 Length

Passage length, which can potentially reflect how much information a text embeds, constitutes a prima facie linguistic element to detect deception. It serves to indicate text's richness, which reflects how much information it conveys. Accordingly, this element has its roots in the decisions an author makes about what information to select and, thus, belongs to the message level.

On the one hand, deceivers can intentionally write longer and more elaborated passages to appear more comprehensive if they perceive persuasive effort as beneficial in the particular circumstance (Burgoon et al., 2003). In fact, in examining deception in an instant messaging context, Hancock et al. (2007) found that the deceivers use more words overall (Hancock et al., 2007), while Anderson and Simester (2013) found they use more explanations to appear more persuasive. On the other hand, an overwhelming number of psychological experiments have shown a positive correlation between text length and informational credibility as having more material to write about reflects a reviewers' actual first-person experience (Pennebaker, 2011). Since individuals often produce fake reviews without actual experience with a product and/or service, difficulties in information gathering presents a practical barrier in writing fake reviews (Banerjee & Chua, 2014c). Accordingly, this barrier may lead to shorter reviews. Indeed, studies have found a positive association between a review's length and its overall influence (Baek et al., 2012). For instance, Baek et al. (2012) and Newman et al. (2003) found longer reviews with both positive and negative information to have a higher utility to readers compared to biased reviews that contain only one side of an argument (Baek et al., 2012; Newman et al., 2003). Following this approach, we can expect credible reviews to exceed fake reviews in length due to the increased likelihood that they will cover both the pros and cons of an experience in their arguments. Furthermore, since genuine reviewers have had an actual experience with a product/service and have more likely experienced both its pros and cons, one can expect them to have an advantage in gathering the informative details and persuasive arguments that online reviews require and a better ability to produce longer persuasive arguments. Thus, one can expect genuine reviews to exceed fake reviews in overall length. Accordingly, we hypothesize that:

**H1**: Genuine reviews exceed fake reviews in length.

### 2.2.2 Psychological Cues

Psychological cues include several important brain function cues that reflect subtle activities or intentions that reside in someone's mind. Psychological cues constitute a significant linguistic dimension for credibility since liars often pay significant attention to the content they deliver, and they may divulge their state of mind unconsciously through the way they express themselves (DePaulo et al., 2003; Ekman & Friesen, 1969).

**Affective cues**: affective cues refer to positive and negative sentiment words (Pennebaker et al., 2015) that can leave an impression and affect readers' minds. In the online review context, positive and negative sentiment words (e.g., love, nice, hurt, ugly, and nasty; Pennebaker et al., 2015) that can affect readers' minds reflect these cues (Tian et al., 2021). As such, affective cues can reflect the writer's intention to affect readers in a specific way (e.g., by associating the writer's experience with a product/service with negative sentiment).

Yoo and Gretzel (2009) found a positive correlation between affect intensity and perceived credibility and, more specifically, high affect intensity to relate to negative credibility. Moreover, in a study on deception in text-based asynchronous computer-mediated communication, Zhou et al. (2004) proposed that deceivers might strategically articulate language via using more affective language—a logical proposal since studies in distinguishing memory have found that people remember sentimental words better than other words (Hu et al., 2017).

Given that financial incentives generally drive actors to post deceptive reviews (e.g., to boost their business or slander competition) (Malbon, 2013; Maurer & Schaich, 2011), reviewers may be enticed to include more affective cues in an attempt to manipulate the readers' moods and their purchase decisions (Ho & Lim, 2018). On the contrary, since reviewers mostly write genuine reviews to recommend a product/service or offer their critical judgments about it, they should not be as sentimentally loaded as fake reviews. Thus, we hypothesize that:

**H2:** Fake reviews have more affective cues than genuine reviews.

**Social cues**: social cues refer to social connections and structures, such as references to family and friends (Pennebaker et al., 2015). These cues can reflect the reviewer's intent to make social connections with the readers. Thus, social cues belong to the message level since people form intentions and confirm a review's nature at that level.

Psychologists have proposed that reviewers use this device to form a closer connection with the audience to build trust (Tausczik & Pennebaker, 2010). Social cues can indicate interpersonal concerns (Proyer & Brauer, 2018). People lower in individuality tend to focus more on their peers and social relationships (Burke & Dollinger, 2005). For instance, reviewers may refer to their audience as "pal" or "friend" to establish a personal connection through the review. Moreover, social cues may reflect writers' intention to exhibit their social authority (Gal & Woolard, 1995), a means to make the reviews appear more credible. We can further explain the use of social cues with trust formation theory. Researchers have discovered that emotional connection represents one of two major approaches to build trust in online communities. Thus, reviewers, through leveraging their social capital to facilitate information adoption, can create a sense of belonging and, hence, achieve relational closeness (Fan & Lederman, 2018). Readers will then expect that they can rely on the reviewer's words (Chau et al., 2011; Ho & Chau, 2013). In this case, we expect that genuine review writers would make not as much effort to form a closer connection compared to fake review writers. Thus, we hypothesize that:

**H3:** Fake reviews have more social cues than genuine reviews.

**Cognitive cues**: cognitive cues emphasize the writer's cognitive load. In other words, they reflect how much effort an author put into drawing a logical conclusion based off the writer's observations (Pennebaker et al., 2015). Cognitive cues reflect the thinking processes that relate to whether a reviewer sources information from objective logical reasoning or subjective actual experience. Accordingly, cognitive cues belong to the message level.

The act of deceiving requires one's brain to process, maintain, and deliver a fake story in a convincing way, which consumes cognitive resources (Newman et al., 2003). Generally, individuals who author credible reviews will unconsciously signal more processes through the diction they choose (Tausczik & Pennebaker, 2010). Hence, we can expect credible reviews to use more cognitive cues and, thus, to evidence cognitive processes to a greater extent than fake reviews. As a result, we hypothesize that:

**H4:** Fake reviews have fewer cognitive cues than genuine reviews.

**Perceptual cues**: perceptual cues describe someone's first-person sensory experience, which includes their descriptions in relation to what the person sees, hears, and feels (Pennebaker et al., 2015). Since choosing information, which includes perceptual experiences, belongs to the message level, perceptual cues belong to the message level.

All these perceptual cues rely highly on the actual physical experience, especially aural and tactile cues; thus, one cannot easily fabricate them in deceptive situations (Lin, 2004). Although deceivers intend to be as descriptive and broad as possible when attempting to fabricate experiences (Tausczik & Pennebaker, 2010), they can find it difficult to pay much attention to forge the nuanced perceptual cues since they already need to use a high proportion of their cognitive resources to maintain a fabricated story (Vrij et al., 2000). Even if some sophisticated fake reviewers pay conscious attention on adding perceptive cues in their content, their perceptive description would not contain as much detail as genuine reviews since they did not have a real first-person physical experience with the product/service in question (Hauch et al., 2015). Therefore, we would expect genuine reviews to include more perceptual details compared to deceptive reviews and hypothesize that:

**H5**: Fake reviews have fewer perceptual cues than genuine reviews.

## 2.3 Formulation-level Variables

At the formulation level, reviewers transform the information they chose and their intentions into syntactic and linguistic expressions that they can (but do not yet) communicate (Garrett, 1975; Levelt, 1993). During this stage, reviewers make important choices on the words to use and formulate their ideas into grammatical sentences (Levelt, 1993). By the end of the stage, they would have a word-to-word draft in their minds that they will later type out (Levelt, 1993). Therefore, from tracing a review's words and syntactic features, one can reveal a reviewer's mindset during the formulation level. In drafting a review word by word, authors follow two important steps: word choice and grammatical sentence formation. Thus, through examining a review's readability (which reflects the syntactic expressions the writer chose) and its linguistic styles (which reflects linguistic grammatical decisions), one may ascertain the reviewer's mental state at the formulation level.

### 2.3.1 Readability

Readability refers to sentence complexity and word choice. Review readability depends on the way in which one expresses linguistic sentences rather than the conceptual ideas in them; thus, it belongs to the formulation level.

Previous studies have found a negative correlation between sentence complexity and deception due to the cognitive load necessary to maintain a fake story that contradicts or does not result from a real experience, and such a process requires psychological effort (Tausczik & Pennebaker, 2010). Other studies have suggested that deception is associated not only with lower sentence complexity but also lower language diversity (Burgoon et al., 2003; Zhou et al., 2004). Hence, deceivers have less mental capacity to produce more logically flowing complex texts compared to truth tellers. Interestingly, studies that have adopted a different perspective have made a similar observation in that deceivers use more conjunction words for more complex, less readable sentence (Banerjee & Chua, 2014b). Considering these arguments, the contextual background, and the reasons why deceivers create fake online reviews, we can expect that deceptive writers either attempt to use larger words (words with more letters) and longer sentences to intentionally ensure readers perceive their reviews as more sophisticated and, in turn, more credible (Levy, 2017; Yoo & Gretzel, 2009). Genuine reviews may instead have higher readability (less complicated sentences and simpler word choices) since their authors write them primarily to convey their thoughts about a service/product rather than to manipulate others to perceive them as credible. Thus, we hypothesize that:

**H6:** Fake reviews have poorer readability than genuine reviews.

### 2.3.2 Linguistic Style

Linguistic style captures the grammatical details that a text embeds. Linguistic style focuses on the more nuanced details in text that readers rarely pay attention to. These details can indicate subtle ways in which genuine and non-genuine reviewers differ in the mind.

**Pronouns**: pronouns refer to subject referencing words that substitute for nouns or noun phrases. Since choosing pronouns belongs to the transformation from a conceptual idea to linguistically expressible sentences, pronouns belong to the formulation level.

Although the way in which people use pronouns may vary depending on the context, in situations that involve personal experience, psychologists have found that deceivers use less first-person pronouns due to feeling guilt (Pennebaker, 2011). Furthermore, deceivers frequently use other pronouns to dissociate themselves from the falsehoods they convey (DePaulo et al., 2003; Ickes et al., 1986).

People use first-person pronouns to describe or evaluate an event through a more private and subjective viewpoint. Unlike the singular “I”, which implies that a writer takes responsibilities fully and personally, the plural “we” illustrates a more collective picture with shared responsibilities, which provides a means for writers to dissociate themselves by hiding in a larger group (Little & Skillicorn, 2008; Schmid, 2014). In this way, reviews that avoid sole responsibility may lead readers to doubt their authenticity. However, people can certainly experience an event in a group setting; thus, reviews that use first-person plural pronouns in such a situation would not compromise credibility. In the consumer review context, due to the possibility that individuals experience a product or service in a group, readers should not attribute reviews that use “we” as deceptive but rather as indicating actual personal experience. Accordingly, we treat first-person pronouns as a whole (i.e., we include both singular and plural pronouns).

As we discuss above, other pronouns allow writers to dissociate themselves from falsehoods, and studies in online reviews have found that false posts use more other pronouns because writers exploit them as a mean to avoid lying outright in contrast to deceiving others in their own first-person voice (Moon et al., 2021; Ott et al., 2011). Moreover, from a different perspective, according to simplified thinking theories, liars tend to subconsciously simplify matters and substitute other persons’ or subjects’ names with second-person, third-person, and impersonal pronouns (Lin et al., 2017). Hence, we hypothesize that:

**H7**: Fake reviews have fewer first-person pronouns than genuine reviews.

**H8**: Fake reviews have more other pronouns than genuine reviews.

**Other parts-of-speech (POS) tags:** POS tags reflect sentence structures through separating words into their respective grammatical categories; thus, they can reflect the sentence structure from a microscopic viewpoint. Similar to pronouns, other POS tags reflect the words one chooses and, hence, belong to the formulation level.

Previous machine learning (ML) studies have found that authentic reviewers are more likely to use more adjectives, articles, and prepositions in their sentences but fewer verbs and adverbs (Banerjee et al., 2015; Ott et al., 2011). However, these studies did not explain the underlying reasons for such observations. Other more theoretical studies have shone light onto the potential reasons for such observations. For instance, Tausczik and Pennebaker (2010) suggested that articles (e.g., a, an, the) signal an upcoming concrete noun, and concrete nouns indicate upcoming concrete information regarding a specific topic. Similarly, adjectives provide descriptive details and knowledge about an event that we can expect reviewers to have gathered from a realistic experience. Writers generally use prepositions (e.g., to, which, above) as linking words between different phrases and to form a connection between nouns; in this way, they can reflect the writer’s knowledge level.

Verbs and adverbs represent two other important POS tags. Verbs convey core action in a passage, while adverbs provide descriptive details on the verbs (Tausczik & Pennebaker, 2010). Since verbs appear frequently in daily life, one can use them to fabricate a story with minimal effort (Liu, 2020). Furthermore, it also makes logical sense that credible authors will use verbs to make more evaluate statements when reconstructing their experience rather than focus on mere actions (Rayson et al., 2002). As such, we hypothesize that:

**H9:** Fake reviews have fewer adjectives, articles, and prepositions than genuine reviews.

**H10:** Fake reviews have more verbs and adverbs than genuine reviews.

# 3 Methodology

## 3.1 Data Collection and Preparation

It is difficult to obtain real-world labeled review dataset in this field of study (Zhang et al., 2016). Some studies have used participants or Amazon Mechanical Turk to manually generate fake reviews (Jindal & Liu, 2007; Lim et al., 2010; Ott et al., 2011). However, others have criticized using manually generated experimental reviews because they exhibit different word distributions compared to real-life reviews, especially given that research has found that real-world spammers make more effort in making their reviews credible (Mukherjee et al., 2013b). Therefore, in this study, we chose a real-world commercial dataset since it better represents actual fake and genuine reviewers' mental model and expertise.

Yelp, as one of the largest crowd-sourced review platforms, holds an enormous review database for a wide range of businesses (Luca, 2016). More importantly, Yelp has developed its own review-filtering algorithm to flag fake reviews (Zhang et al., 2016). Although Yelp does not make its algorithm public, it displays the fake reviews in a different section on their platform for the public (Mukherjee et al., 2013b). Previous research has proven Yelp's fake review-filtering algorithm to be reliable and accurate to a great extent (Mukherjee et al., 2013b; Weise, 2011) and closely correlated to spamming behaviors (Mukherjee et al., 2013a). Thus, we find it rational to treat Yelp's review-filtering results as ground truth for our purposes in this study.

To control the subject variation across different businesses categories, we included only restaurant reviews for our analyses. The dataset contained 66,940 reviews on 98 restaurants in Chicago posted during the period from October, 2004, to September, 2012. Among the reviews, Yelp's automatic filtering algorithm flagged 8,303 as fake and, thus, considered the remaining 58,637 reviews as genuine (unfiltered).

We measured reviews' linguistic characteristics (except readability) using Linguistic Inquiry and Word Count (LIWC) 2015. Pennebaker et al. (2015) developed the LIWC 2015 dictionary, which comprises nearly 6,400 words, through a rigorous process over many years. They have extensively examined and verified its internal reliability and external validity, and hundreds of studies have found LIWC categories to be valid in various psychological domains (Pennebaker et al., 2015; Tausczik & Pennebaker, 2010). Researchers have also used the software to examine online consumer behaviors and online reviews (e.g., Geng et al., 2020; Hancock et al., 2007; Karami & Zhou, 2015; Ott et al., 2011; Park, 2019; Plotkina et al., 2020; Yang et al., 2015).

## 3.2 Data Analysis

To examine the differences in psycholinguistic features between fake and genuine reviews, we used the following three logistic regression models.

**Model 1:** *Filtered Flag ~ Rating Extremity + Review Age + Tenure on Platform*

**Model 2:** *Filtered Flag ~ Rating Extremity + Review Age + Tenure on Platform + Length + Affective Cues + Social Cues + Cognitive Cues + Perceptual Cues*

**Model 3:** *Filtered Flag ~ Rating Extremity + Review Age + Tenure on Platform + Length + Affective Cues + Social Cues + Cognitive Cues + Perceptual Cues + Readability + i & we + other pronouns + adj & article & prep + adverb & verb*

The dependent variable, *Filtered Flag*, is a binary variable, where "1" indicates a fake (filtered) review and "0" represents a genuine (regular and unfiltered) review. We categorized the linguistic independent variables into two levels: message level and formulation level. The message level variables included the following: 1)

*Length*, which we measured based on word count (WC), constitutes how many words a review contained in total; 2) *Affective Cues*, which include positive and negative sentiment words (e.g., happy, love, sweet, hurt, ugly, worried, hate, and grief); 3) *Social Cues*, which include family, friends, and gender references (e.g., buddy, neighbor, mate, uncle, boy, mom, and dad); 4) *Cognitive Cues*, which include insights, causation, discrepancy, tentative, certainty, and differentiation words (e.g., cause, know, ought, because, should, perhaps, else, and always); and 5) *Perceptual Cues*, which include visual, aural, and touch words (e.g., look, heard, feeling, listen, and touch) (Pennebaker et al., 2015).

The formulation level included five variables under three subcategories: 1) *Readability*, which we measured with the Coleman-Liau Index (CLI), for which a lower value indicates a more readable passage; 2) *I + we* sums the "I" and "we" measurements that LIWC provides to overview the first-person pronouns; 3) *other pronouns* refer to pronouns excluding "I" and "we" according to LIWC (e.g., second-person, third-person, and impersonal pronouns); 4) *adj+article+prep* sums the adjective, article, and preposition measurements from LIWC; and 5) *adverb+verb* sums adverb and verb measurements from LIWC.

Moreover, we included three control variables: 1) *Rating Extremity*, which refers the absolute difference in stars between the reviewer's star rating on the business and the existing business rating; 2) *Review Age*, which refers to the number of days from the review date to the data-collection date (we used this variable to capture the effect that a review's age had on its perceived credibility); and 3) *Tenure on Platform,* which refers to the number of days between a review's date and the date the reviewer posted their first review. Tables 1 and 2 provide the descriptive statistics for each variable.

Given the unbalanced data, models built based on such data would gravitate towards classifying fake reviews class into the genuine reviews class, which would compromise the model's quality. We used the synthetic minority over-sampling technique (SMOTE), among the most popular frameworks to resolve imbalanced data, to over-sample the minority (filtered) group data and, thus, preserve the diversity of the majority (unfiltered) group data while maintaining a balanced dataset (Chawla et al., 2002; Fernández et al., 2018). By over-sampling from the 8,303 genuine reviews, we produced a dataset with 58,121 genuine reviews and 58,637 fake reviews (116,758 reviews in total)—sufficient to train a model in this area (Mukherjee et al., 2013b; Ott et al., 2011; Zhang et al., 2016).

**Table 1. Descriptive Statistics: Fake Reviews**

| | | | Fake reviews | | | | |
|---|---|---|---|---|---|---|---|
| | | | **Mean** | **S.D.** | **Median.** | **Min.** | **Max.** |
| Message variables | Length | WC | 99.730 | 99.602 | 70.000 | 0.000 | 953.000 |
| | Affective processes | Affect | 9.028 | 7.409 | 7.350 | 0.000 | 100.000 |
| | Social processes | Social | 7.440 | 5.120 | 7.050 | 0.000 | 66.670 |
| | Cognitive processes | Cogproc | 9.831 | 5.225 | 9.740 | 0.000 | 50.000 |
| | Perceptual processes | Percept | 2.641 | 3.684 | 2.150 | 0.000 | 100.00 |
| Formulation variables | Readability | CLI | 7.688 | 3.316 | 7.680 | -39.610 | 59.000 |
| | Pronouns | I + we | 5.174 | 3.662 | 5.130 | 0.000 | 33.330 |
| | | Other pronouns | 7.120 | 4.456 | 6.950 | 0.000 | 50.000 |
| | Other POS tags | Adj + article + prep | 25.851 | 7.782 | 25.720 | 0.000 | 100.00 |
| | | Adverb + verb | 21.414 | 7.471 | 21.870 | 0.000 | 77.770 |
| Control variables | Rating extremity | Extremity | 1.049 | 0.876 | 1.000 | 0.000 | 3.500 |
| | Review age | Age | 900.068 | 520.763 | 836.000 | 101.00 | 2713.000 |
| | Tenure on platform | Tenure | 14.765 | 87.595 | 0.000 | 0.000 | 1429.000 |
| | | N | 8303 | | | | |

**Table 2. Descriptive Statistics: Genuine Reviews**

| | | | Genuine reviews | | | | |
|---|---|---|---|---|---|---|---|
| | | | **Mean** | **S.D.** | **Median** | **Min.** | **Max.** |
| Message variables | Length | WC | 144.744 | 121.629 | 112.000 | 0.000 | 997.000 |
| | Affective processes | Affect | 7.566 | 5.214 | 6.640 | 0.000 | 100.00 |
| | Social processes | Social | 6.772 | 3.962 | 6.470 | 0.000 | 100.00 |
| | Cognitive processes | Cogproc | 10.393 | 4.380 | 10.290 | 0.000 | 100.00 |
| | Perceptual processes | Percept | 2.905 | 2.946 | 2.590 | 0.000 | 100.00 |
| Formulation variables | Readability | CLI | 7.626 | 2.349 | 7.600 | -28.010 | 31.900 |
| | Pronouns | I + we | 5.370 | 3.091 | 5.380 | 0.000 | 50.000 |
| | | Other pronouns | 7.145 | 3.601 | 7.010 | 0.000 | 60.000 |
| | Other POS tags | Adj + article + prep | 25.666 | 5.939 | 25.640 | 0.000 | 100.000 |
| | | Adverb + verb | 21.485 | 6.026 | 21.570 | 0.000 | 100.000 |
| Control variables | Rating extremity | Extremity | 0.763 | 0.698 | 1.000 | 0.000 | 3.500 |
| | Review age | Age | 959.837 | 573.837 | 868.000 | 101.000 | 3002.000 |
| | Tenure on platform | Tenure | 184.291 | 336.593 | 0.000 | 0.000 | 2475.000 |
| | | N | 58637 | | | | |

# 4 Results

We summarize the results for the logistic regression models in Table 3. The consistent pattern across the results indicates the model's stability. As the table shows, WC was negatively associated with the review fakeness, which supports H1. Affective and social cues were positively associated with review fakeness, whereas cognitive and perceptual cues were positively associated with review fakeness, which supports H2 to H5. Moreover, fake reviews exhibited significantly less readability than genuine reviews (the higher the readability score, the lower the readability), which supports H6. However, fake reviews used more first-person pronouns than genuine reviews, which reversely supports H7, whereas the other pronouns lacked statistical significance, which does not support H8. Moreover, fake reviews used proportionally more adjectives, articles, and prepositions compared to genuine reviews, which reversely supports H9. Finally, the proportion of adverbs and verbs was positively correlated to the filtered flag, which supports H10. We summarize the results from testing the hypotheses in Table 4.

**Table 3. Results of Logistic Regression Models**

| | | | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Coef. | S.E. | Pr(>\|z\|) | Coef. | S.E. | Pr(>\|z\|) | Coef. | S.E. | Pr(>\|z\|) |
| Control variables | Rating extremity | Extremity | 0.394 | 0.008 | *** | 0.477 | 0.009 | *** | 0.476 | 0.009 | *** |
| | Review age | Age | -0.028 | 0.001 | *** | -0.026 | 0.001 | *** | -0.027 | 0.001 | *** |
| | Tenure on platform | Tenure | -0.641 | 0.008 | *** | -0.589 | 0.008 | *** | -0.587 | 0.008 | *** |
| Message variables | Length | WC | | | | -0.367 | 0.007 | *** | -0.382 | 0.007 | *** |
| | Affective cues | Affect | | | | 2.496 | 0.139 | *** | 2.741 | 0.143 | *** |
| | Social cues | Social | | | | 4.102 | 0.172 | *** | 4.261 | 0.194 | *** |
| | Cognitive cues | Cogproc | | | | -1.669 | 0.152 | *** | -1.983 | 0.166 | *** |
| | Perceptual cues | Percept | | | | -3.352 | 0.230 | *** | -3.359 | 0.238 | *** |
| Formulation variables | Readability | CLI | | | | | | | 4.066 | 0.294 | *** |
| | Pronouns | I + we | | | | | | | 0.665 | 0.246 | * |
| | | Other pronouns | | | | | | | 0.393 | 0.222 | |
| | Other POS tags | Adj + article + prep | | | | | | | 1.689 | 0.121 | *** |
| | | Adverb + verb | | | | | | | 1.352 | 0.127 | *** |
| Adjusted McFadden | | | 0.137 | | | 0.169 | | | 0.171 | | |
| Cox & Shell | | | 0.173 | | | 0.209 | | | 0.211 | | |
| Sample size | | | No. of filtered reviews = 58121; no. of unfiltered reviews = 58637. | | | | | | | | |

Note: we rescaled all variables (except for “extremity”) by dividing by 100.
* p < 0.05, ** p < 0.01, *** p < 0.001

**Table 4. Summary of Hypothesis Testing**

| Hypothesis | Result |
|---|---|
| **H1:** Genuine reviews exceed fake reviews in length | Supported |
| **H2:** Fake reviews have more affective cues than genuine reviews. | Supported |
| **H3:** Fake reviews have more social cues than genuine reviews. | Supported |
| **H4:** Fake reviews have fewer cognitive cues than genuine reviews. | Supported |
| **H5:** Fake reviews have fewer perceptual cues than genuine reviews. | Supported |
| **H6:** Fake reviews have poorer readability than genuine reviews. | Supported |
| **H7:** Fake reviews have fewer first-person pronouns than genuine reviews. | Reversely supported |
| **H8:** Fake reviews have more other pronouns than genuine reviews. | Not supported |
| **H9:** Fake reviews have fewer adjectives, articles, and prepositions than genuine reviews. | Reversely supported |
| **H10:** Fake reviews have more verbs, and adverbs than genuine reviews. | Supported |

# 5 Discussion

Under the psychological cues metrics, fake reviews exhibit characteristics similar to deceivers themselves as we hypothesized. Moreover, observations on the review length and readability further support this argument. However, contrary to what previous literature suggests as deceptive writing characteristics, we observed that fake reviewers used more first-person pronouns, though we did not find a significant difference in the extent to which they used other pronouns. This finding diverges from what we would expect from deceivers (DePaulo et al., 2003; Ickes et al., 1986; Pennebaker, 2011). Additionally, contrary to our hypothesis, fake reviews used proportionally more adjectives, articles, and prepositions, which increased their apparent credibility.

From a macro viewpoint, all the selected variables at the message level indicate that fake reviews align closely with a deceptive writing style. We found that fake reviews were shorter in length and used more affective and social cues but fewer cognitive and perceptual cues. All these message-level variables relate to stage at which a writer forms an intention and selects information, and we found these variables to closely align with deceptive reviews. On the contrary, for the formulation-level variables, although fake reviews lacked readability compared to genuine reviews and used proportionally more verbs and adverbs as we expected, we found no significant difference in the extent to which they used other pronouns. Also, our finding that fake reviews used first-person pronouns and adjectives, articles, and prepositions to a greater extent than genuine reviews contradicts existing knowledge on deceptive texts. We found that the features of fake reviews were more consistent in message level variables than the formulation level variables. Thus, we can conclude that, although fake reviews might vary in their expression and formulation depending on the circumstances or writers might carefully articulate them to influence readers to perceive them as more credible, writers cannot as easily manipulate a review's actual content and the intention it exhibits through the message-level variables. Thus, we can deem message-level variables as better factors to predict fake online reviews.

While we found different results to what expected regarding first-person pronouns (e.g., "I" and "we") and other pronouns, we can justify them based on two main arguments. First, we expected that fake reviews would use proportionally fewer first-person pronouns but proportionally more other pronouns to dissociate themselves from falsehoods and put the target on others (Newman et al., 2003). However, genuine review writers may adjust their writing style to be more analytic and evaluative and less personal and subjective via using as many other pronouns as fake review writers but with fewer first-person references. This arguments rests on the logic that the pronouns a writer uses largely depends on the context (e.g., writers widely use third-person pronouns in scientific texts without compromising their credibility). Thus, considering the context we examine here, genuine review writers might use fewer first-person pronouns when paying attention to objective perspectives to evaluate a restaurant's food or service rather than focusing on personal experience or mere subjective feelings (Li et al., 2019; Ott et al., 2011). Second, fake review writers may have made deliberate attempts to use first-person pronouns to manipulate readers. Since they constitute a foundational linguistic element in English grammar, writers would find it relatively simple and beneficial to manipulate how they used first-person pronouns. Liu (2020) supports this proposition in identifying fake review writers as more inclined to deliberately use more first-person pronouns to sound more credible through convincing readers that they wrote their review based on actual experience. Thus, fake reviewers may choose to "attach" themselves to the specific "storytelling" aspect of online reviews rather than detach themselves from their lies to create a more convincing story (Liu, 2020).

Moreover, contrary to H1, we found an association between fake reviews and proportionally more adjectives, articles, and prepositions. We may have found such a result due to fake review writers' conscious attempts to align their writing style with genuine review writers, which involves using descriptive information to balance out their misleading false statements (Tausczik & Pennebaker, 2010). As a result, they may use proportionally more adjectives, articles, and prepositions than genuine review writers to provide excessive details about their imaginative experience. Although Tausczik and Pennebaker (2010) suggested that too much detail might lead fake review writers to include inaccuracies in their reviews, our results suggest that fake review writers do not consider such risks. We also note that Banerjee and Chua (2014b) found that more authentic reviews contain fewer articles and prepositions and found insignificant differences in how many adjectives they use, which largely coincides with our results. In Tables 5 and 6, we show reviews that exemplify the above discussion.

**Table 5. Review Examples**

| Sample flagged (fake) review | Sample unflagged (genuine) review |
|---|---|
| *I took my boyfriend here for his birthday, from the moment we walked in we were treated like family. The atmosphere was warm and welcoming. I had the vegetarian goulash (sp?) which was awesome, my boyfriend had the skirt steak which was equally as good. The service and entertainment a guitarist and violinist were great, very friendly and fun. Although the food and staff were all excellent there was one thing that stood above the rest, the owner Bronco, he made our night something we will never forget. You're the best Bronco and we WILL be back. P.S. watch out for the holy water.* | *Pretty God-danged good. I cannot say that in any other day of my life have I eaten 4 and a half hot dogs and/or sausages. Selma Hayek - grilled, a little too charred for my taste, but yummy - 3 stars Mandarin chicken sausage with siracha mustard - sweet and tangy, mixed with heat - 4 stars Hot Doug - Chicago style greatness - 5 stars Foie gras dog - next time, i'm having it steamed and with no sea salt, but 5 stars Lamb and cognac dog with peppercorn cheese - really really cheesy! had to brush some off, but great sauce! 5 stars. ....I cant wait to go back and try the rest.* |

**Table 6. Examples of Fake and Genuine Review**

| Message level | | | Formulation level | | |
|---|---|---|---|---|---|
| | **Fake example** | **Genuine example** | | **Fake example** | **Genuine example** |
| **Length** | 105.00 | 109.00 | **Readability** | 9.12 | 7.03 |
| **Affective cues** | 9.52 | 5.50 | **First-person pronouns** | 8.57 | 5.50 |
| **Social cues** | 13.33 | 0.92 | **Other pronouns** | 7.62 | 2.76 |
| **Cognitive cues** | 4.76 | 11.93 | **Adjectives, articles, and prepositions** | 25.72 | 18.34 |
| **Perceptual cues** | 1.90 | 8.26 | **Adverbs and verbs** | 21.90 | 12.84 |

## 5.1 Implications for Research

In this study, we systematically categorize and examine the effects of various crucial psycholinguistic cues that researchers have suggested to have predictive value to identify fake reviews. We examined the relationship between these cues and fake reviews using logistic regression and found that fake reviews aligned more with deceptive writing styles for the message-level variables, were generally shorter in length, and possessed more affective and social cues but fewer cognitive and perceptual cues. Judging from the results, fake reviews resemble genuine writing styles at the formulation level as they lack some characteristics (fewer first-person pronouns, more other pronouns, and fewer adjectives, articles, and propositions) that researcher have established deceptive writings to exhibit.

Accordingly, with this study, we make two important theoretical implications. First, we contribute to the general body of knowledge on fake review identification in the online restaurant review context. We tested and ratified some important psychological fake review characteristics that previous studies have proposed against a real-world review dataset from Yelp.

Second, while much research has studied fake review characteristics, it has mostly taken a machine learning approach, and we lack empirical evidence for the sufficiency of psycholinguistic characteristics in fake reviews identification in conjunction with a theoretical foundation. In this study, we developed a research model based on Levelt's (1993) language-production model that splits language formulation into two stages. The model provides a framework for understanding from which stage the psycholinguistic characteristics that fake reviews exhibit originate and, thus, the differences present between fake and genuine reviews and the psychology behind it. To the best of our knowledge, this study constitutes an initial effort to use a fundamental language-production model to examine online reviews against the psycholinguistic characteristics. Thus, future research will benefit from a proven framework in addition to our results. Moreover, in identifying that the message-level variables offer more insights on review fakeness than formulation-level variables, our findings suggest that future research should examine fake reviews' main characteristics. More specifically, future research needs to explore the message-level variables to determine their predictive capability.

## 5.2 Implications for Practice

Our findings have immense practical value in helping one identify fake posts in the online review context. In particular, our findings may prove beneficial for review filtering system designers and, subsequently, review readers. For filter-algorithm designers, the categorization of psycholinguistic characteristics with guidance from the language-production model may help them to design algorithms that identify fake reviews based on anomalies in any grouped stage rather than relying on individual characteristic variables. Specifically, they should pay attention to message-level variables and adjust algorithms to consider psychological cues (affective, social, cognitive, and perceptual cues).

For review readers, we identify that one cannot fabricate message-level variables as easily as formulation-level variables. Our findings suggest that review readers need to pay special attention to a review's content rather than the language appeal and review structure. Finally, our findings can assist review writers whose reviews platforms have flagged as fake to alter their writing style when sharing their genuine experiences.

## 5.3 Limitations and Future Research

As with any study, this one has several limitations. First, we acknowledge that one cannot compare the analytical approach that we followed to existing fake review-classification models. Given this research's explanatory nature, future research needs to confirm and the predictive value of the linguistic characteristics that we highlight.

Second, since we used labeled data from Yelp in our logistic regression, we had no possible way to ascertain the labels' accuracy for this specific dataset even though studies in the field commonly treat this set of labeled data as ground truth (Luca & Zervas, 2016). Hence, the capabilities of Yelp's filtering algorithm limit our study. Moreover, empirical studies that use Yelp's dataset have featured both false positives (genuine reviews that it flagged as) and false negatives (fake reviews that it did not filter) (Luca & Zervas, 2016). However, in general, since researchers have considered Yelp's proprietary algorithm highly accurate and widely recognized its reliability (Mukherjee et al., 2013b; Weise, 2011), this limitation may not have posed a significant issue in our study. With that said, future research could look into it further. For instance, future research could use data from various filtering algorithms to verify our findings or take a different approach to conduct experimental studies to obtain first-hand online review data.

Third, the dataset we used covered reviews from October, 2004, to September, 2012. Although completely analyzing potential changes in fake reviews' styles falls beyond our scope in this paper, both fake and genuine review writers may have changed after 2012. However, as we found from reviewing the literature, we still found results that support the vast majority of earlier studies. The dataset we used represents the best that we could obtain for our purposes since Yelp has ceased publicizing filtered reviews. We acknowledge this limitation and encourage future studies to validate our findings if Yelp or any other online review platforms decide to publish their dataset again in the future.

Fourth, the dataset we used contained only Yelp reviews for restaurants in Chicago, USA. Thus, the observations on fake reviews' psycholinguistic characteristics may subject to variations in different cultures, locations, and business types. Hence, the patterns observed might differ for restaurants in areas other than Chicago and for other business types (e.g., hotels) in Chicago and other areas (Tausczik & Pennebaker, 2010). Further studies should verify and generalize our findings via analyzing online reviews for other regions, cultures, and business types.

Fifth, we focused on linguistic summary variables rather than the composite variables themselves. For example, we considered cognitive cues rather than its components (e.g., insight, causation, etc.). While such an approach more than suited our exploratory focus in this study, future studies may further emphasize the composite variables to provide more insights.

Finally, a related limitation concerns the fact that we used LIWC 2015. Although LIWC 2015 constitutes an effective and valid tool to measure reviews' psycholinguistic characteristics (Ott et al., 2011; Pennebaker et al., 2015), LIWC (and, indeed, other similar measurement programs) cannot perfectly cover all the measurements to an exhaustive extent. Furthermore, due to LIWC's closed-sourced nature, we could not further investigate the extent of the measurement that LIWC provides for our purposes in this study. We could have manually evaluated the reviews in lieu of using LIWC to reduce this limitation's impact on our study; however, we could not have practically done so due to our dataset's large size. Future studies could take alternative approach for cross-validation purposes on a smaller and more workable dataset.

# 6 Conclusion

In this paper, we use a labeled Yelp review dataset to examine differences in psycholinguistic styles between fake and genuine reviews as they manifest in the various stages of the language-production process. Our results suggest that fake reviews exhibit deceptive writing styles in their content but less so in their expressive style. Our findings suggest that message-related features could have potential effectiveness in detecting fake online reviews while filtering algorithms based on language expression might not be as effective. Our findings provide both theoretical and practical implications for effectively filtering fake online reviews and support the overarching goal to increase online consumer reviews' trustworthiness.

## References


Anderson, E., & Simester, D. (2013). *Deceptive reviews:* The influential tail. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.393.802&rep=rep1&type=pdf

Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of online consumer reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce, 17*(2), 99-126.

Banerjee, S., & Chua, A. Y. (2014a). A study of manipulative and authentic negative reviews. In *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication.*

Banerjee, S., & Chua, A. Y. (2014b). A theoretical framework to identify authentic online reviews. *Online Information Review*.

Banerjee, S., & Chua, A. Y. (2014c). Understanding the process of writing fake online reviews. In *Proceedings of the 9th International Conference on Digital Information Management.*

Banerjee, S., Chua, A. Y., & Kim, J.-J. (2015*).* Using supervised learning to classify authentic and fake online reviews. In *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication.*

Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. Academic Press.

Bott, L., & Williams, E. (2018). Psycholinguistic approaches to lying and deception. In J. Meibauer (Ed.), *The Oxford handbook of lying*. Oxford University Press.

Burgoon, J. K., Blair, J. P., Qin, T., & Nunamaker, J. F. (2003). Detecting deception through linguistic analysis. In *Proceedings of the International Conference on Intelligence and Security Informatics.*

Burke, P. A., & Dollinger, S. J. (2005). "A picture's worth a thousand words": Language use in the autophotographic essay. *Personality and Social Psychology Bulletin, 31*(4), 536-548.

Chau, P. Y. K., Ho, S. Y., & Yao, Y. (2011). The effects of malfunctioning personalized services on users' trust and behaviors. In *Proceedings of the Pacific Asia Conference on Information Systems.*

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321-357.

Chen, P.-Y., Dhanasobhon, S., & Smith, M. D. (2008). All reviews are not created equal: The disaggregate impact of reviews and reviewers at Amazon.com. *SSRN*. Retrieved from https://ssrn.com/abstract=918083

Cheung, M. Y., Luo, C., Sia, C. L., & Chen, H. (2009). Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. *International Journal of Electronic Commerce, 13*(4), 9-38.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*(3), 345-354.

Dellarocas, C., & Narayan, R. (2007). Tall heads vs. long tails: Do consumer reviews increase the informational inequality between hit and niche products? *SSRN*. Retrieved from https://ssrn.com/abstract=1105956

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74-118.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*(1), 88-106.

Fan, H., & Lederman, R. (2018). Online health communities: How do community members build the trust required to adopt information and form close relationships? *European Journal of Information Systems, 27*(1), 62-89.

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research, 61*, 863-905.

Fontanarava, J., Pasi, G., & Viviani, M. (2017). Feature analysis for fake review detection through supervised classification. In *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics.*

Gal, S., & Woolard, K. A. (1995). Constructing languages and publics: Authority and representation. *Pragmatics, 5*(2), 129-138.

Garrett, M. F. (1975). The analysis of sentence production. In In G. H. Bower, (Ed.), *Psychology of learning and motivation* (vol. 9, pp. 133-177). Elsevier.

Geng, S., Niu, B., Feng, Y., & Huang, M. (2020). Understanding the focal points and sentiment of learners in MOOC reviews: A machine learning and SC-LIWC-based approach. *British Journal of Educational Technology, 51*(5), 1785-1803.

Granhag, P. A., & Strömwall, L. A. (2004). *The detection of deception in forensic contexts*. Cambridge University Press.

Gretzel, U., & Yoo, K. H. (2008). Use and impact of online travel reviews. *Information and communication technologies in tourism 2008*, 35-46.

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2007). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes, 45*(1), 1-23.

Hartwig, M., & Bond, C. F., Jr. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology, 28*(5), 661-676.

Hauch, V., Blandón-Gitlin, I., Masip, J., & Sporer, S. L. (2015). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review, 19*(4), 307-342.

Ho, S. Y., & Chau, P. Y. K. (2013). The effects of location personalization on integrity trust and integrity distrust in mobile merchants. *International Journal of Electronic Commerce, 17*(4), 39-72.

Ho, S. Y., & Lim, K. H. (2018). Nudging moods to induce unplanned purchases in imperfect mobile personalization contexts. *MIS Quarterly, 42*(3), 757-778.

Ho, S. Y., & Richardson, A. (2013). Trust and distrust in open source software development. *Journal of Computer Information Systems, 54*(1), 84-93.

Hu, F., Xu, X., Wang, J., Yang, Z., & Li, L. (2017). Memory-enhanced latent semantic model: Short text understanding for sentiment analysis. In *Proceedings of the International Conference on Database Systems for Advanced Applications.*

Hunt, K. M. (2015). Gaming the system: Fake online reviews v. consumer law. *Computer Law & Security Review, 31*(1), 3-25.

Ickes, W., Reidhead, S., & Patterson, M. (1986). Machiavellianism and self-monitoring: As different as “me” and “you”. *Social Cognition, 4*(1), 58-74.

Jindal, N., & Liu, B. (2007). Review spam detection. In *Proceedings of the 16th International Conference on World Wide Web.*

Karami, A., & Zhou, B. (2015). Online review spam detection by new linguistic features. In *iConference 2015 Proceedings*.

Lane, L. W., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological Science, 17*(4), 273-277.

Levelt, W. J. (1993). *Speaking: From intention to articulation* (vol. 1). MIT press.

Levy, N. (2017). The bad news about fake news. *Social Epistemology Review and Reply Collective, 6*(8), 20-36.

Li, Y., Kuan, K. K. Y., & Liu, N. (2019). Exploring the linguistic characteristics of online consumer reviews by top reviewers and ordinary reviewers. In *Proceedings of the Pacific Asia Conference on Information Systems.*

Lim, E.-P., Nguyen, V.-A., Jindal, N., Liu, B., & Lauw, H. W. (2010). Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on Information and knowledge management.*

Lin, C. H., Hsu, P. Y., Cheng, M. S., Lei, H. T., & Hsu, M. C. (2017). Identifying deceptive review comments with rumor and lie theories. In *Proceedings of the International Conference on Swarm Intelligence.*

Lin, I. Y. (2004). Evaluating a servicescape: The effect of cognition and emotion. *International Journal of Hospitality Management, 23*(2), 163-178.

Little, A., & Skillicorn, D. B. (2008). Detecting deception in testimony. In *Proceedings of the IEEE International Conference on Intelligence and Security Informatics.*

Liu, B. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.

Luca, M. (2016). Reviews, reputation, and revenue: The case of Yelp.com. *SSRN.* Retrieved from https://ssrn.com/abstract=1928601

Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science, 62*(12), 3412-3427.

Malbon, J. (2013). Taking fake online consumer reviews seriously. *Journal of Consumer Policy, 36*, 139-157.

Maurer, C., & Schaich, S. (2011). Online customer reviews used as complaint management tool. In R. Law, M., Fuchs, & F. Ricci (Eds.), *Information and communication technologies in tourism* 2011. Springer.

Moon, S., Kim, M.-Y., & Iacobucci, D. (2021). Content analysis of fake consumer reviews by survey-based text categorization. *International Journal of Research in Marketing, 38*(2), 343-364.

Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013a). *Fake review detection: Classification and analysis of real and pseudo reviews.* Retrieved from http://www2.cs.uh.edu/~arjun/papers/UIC-CS-TR-yelp-spam.pdf

Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013b). What Yelp fake review filter might be doing? In *Proceedings of the 7th international AAAI Conference on Weblogs and Social Media.*

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

Ott, M., Cardie, C., & Hancock, J. (2012). Estimating the prevalence of deception in online review communities. In *Proceedings of the 21st International Conference on World Wide Web.*

Ott, M., Cardie, C., & Hancock, J. T. (2013). Negative deceptive opinion spam. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics.*

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. *arXiv.* Retrieved from https://arxiv.org/abs/1107.4557

Park, E. (2019). Motivations for customer revisit behavior in online review comments: Analyzing the role of user experience using big data approaches. *Journal of Retailing and Consumer Services, 51*, 14-18.

Pennebaker, J. W. (2011). *The secret life of pronouns*. Bloomsbury.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Retrieved from https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf

Plotkina, D., Munzel, A., & Pallud, J. (2020). Illusions of truth—experimental insights into human and algorithmic detections of fake online reviews. *Journal of Business Research, 109*, 511-523.

Proyer, R. T., & Brauer, K. (2018). Exploring adult playfulness: examining the accuracy of personality judgments at zero-acquaintance and an LIWC analysis of textual information. *Journal of Research in Personality, 73*, 12-20.

Rayson, P., Wilson, A., & Leech, G. (2002). Grammatical word class variation within the British National Corpus sampler. In *Proceedings of the 21st International Conference on English Language Research on Computerised Corpora.*

Ren, Y., & Ji, D. (2017). Neural networks for deceptive opinion spam detection: An empirical study. *Information Sciences, 385*, 213-224.

Schmid, H. B. (2014). Expressing group attitudes: On first person plural authority. *Erkenntnis, 79*(9), 1685-1701.

Segal, D. (2011). A rave, a pan or just a fake? *The New York Times*. Retrieved from https://www.nytimes.com/2011/05/22/your-money/22haggler.html

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24-54.

Tian, G., Lu, L., & McIntosh, C. (2021). What factors affect consumers' dining sentiments and their ratings: Evidence from restaurant online review data. *Food Quality and Preference, 88*, 104060.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons.

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior, 24*(4), 239-263.

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science, 20*(1), 28-32.

Wang, G., Xie, S., Liu, B., & Philip, S. Y. (2011). Review graph based online store review spammer detection. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining.*

Weise, K. (2011). A lie detector test for online reviewers. *Bloomberg Business Week*. Retrieved from https://www.bloomberg.com/news/articles/2011-09-29/a-lie-detector-test-for-online-reviewers

Wu, G., Greene, D., Smyth, B., & Cunningham, P. (2010). Distortion as a validation criterion in the identification of suspicious reviews. In *Proceedings of the First Workshop on Social Media Analytics.*

Yang, Y., Yan, Y., Qiu, M., & Bao, F. (2015). Semantic analysis and helpfulness prediction of text for online product reviews. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics.*

Yoo, K. H., & Gretzel, U. (2009). *Comparison of deceptive and truthful travel reviews.* In W. Höpken, U. Gretzel, & R. Law (Eds.), *Information and communication technologies in tourism 2009*. Springer.

Zhang, D., Zhou, L., Kehoe, J. L., & Kilic, I. Y. (2016). What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *Journal of Management Information Systems, 33*(2), 456-481.

Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group Decision and Negotiation, 13*(1), 81-106.

Zhou, L., & Sung, Y.-W. (2008). Cues to deception in online Chinese groups. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences.*

## About the Authors

**Boran Wang** is an honours student at the School of Electrical and Information Engineering of the University of Sydney. She has been studying online consumer reviews for the past two years and has a special interest in the interactions between humans and information systems. As such, her thesis focuses on information assimilation in the medical industry.

**Kevin K.Y. Kuan** is a Senior Lecturer in the School of Computer Science at the University of Sydney. His research focuses on the interplay between human behavior and information systems at both individual and organizational levels, including human-computer interaction, online consumer reviews, and information systems adoption and assimilation. His research has been published in leading information systems journals such as *Journal of Management Information Systems*, *Journal of the Association for Information Systems*, *European Journal of Information Systems*, and *Information and Management*.

Transactions on Human – Computer Interaction