

ORIGINAL ARTICLE OPEN ACCESS

Stylistic Signals in Professional and Amateur Language Used in C2C Markets

József Hubert  | Zsófia Kenesei  | András Bauer

Institute of Marketing and Communication Sciences, Corvinus University of Budapest, Budapest, Hungary

Correspondence: Zsófia Kenesei (zsafia.kenesei@uni-corvinus.hu)

Received: 21 May 2025 | **Revised:** 8 September 2025 | **Accepted:** 23 October 2025

Funding: Financial support was received from the National Research, Development and Innovation Office, Grant number 145937 and Corvinus University of Budapest Research Development Program.

Keywords: C2C market | emotions | expertise theory | signaling theory | text mining

ABSTRACT

In consumer-to-consumer (C2C) markets, individuals must decide whether to sell property directly or rely on professional intermediaries. While digital technologies increasingly enable direct transactions, the expertise and strategic communication skills of intermediaries continue to play an important role. Our objective in this research is to connect expertise theory with signaling theory by examining whether professionals employ stylistic elements in real estate listings that traditionally convey quality and trustworthiness, and how these differ from those used by amateurs. To address this, we introduce topic-prevalent n-grams to examine stylistic patterns across a large dataset of more than 300,000 real estate advertisements in the United States. Two consecutive studies apply natural language processing (NLP) techniques to uncover systematic differences in wording, tone, and rhetorical strategies between professional brokers and amateur homeowners. A third study, which involved a scenario-based experiment with US homeowners, illustrates how potential buyers perceive and respond to the different linguistic styles. The results show that professional brokers use language strategically to convey quality signals and enhance buyer engagement, while amateur homeowners often rely on emotional expression and negative phrasing, which can increase trustworthiness but result in less favorable responses overall. These findings highlight the central role of language style as a signal of expertise, influencing trust, perceptions of quality, and ultimately buyer decision-making. Beyond offering empirical evidence for the impact of stylistic variation, the study contributes a methodological innovation by demonstrating how topic-prevalent n-grams can be used as a powerful tool for text classification and for advancing research on language, expertise, and consumer behavior.

1 | Introduction

In markets where consumers act as sellers and want to sell their own property, particularly in the consumer-to-consumer (C2C) second-hand market, they have different selling options. They can either enter the market themselves as amateur sellers or seek assistance from intermediaries who act as professional sellers. Digital technology has enabled direct exchanges between consumers, ranging from complex products such as automobiles and real estate (Savolainen 2009) to everyday items like clothing

or electronics (Liang and Xu 2018), often diminishing the necessity for professional intermediaries (Fernando et al. 2018). While this shift has transformed many industries, professionals continue to maintain a competitive advantage over amateurs in certain contexts. Their expertise, experience, and strategic communication skills may contribute to more effective engagement with potential buyers (Larceneux et al. 2015).

The general theory of expertise (Ericsson and Smith 1991) suggests that professionals consistently outperform non-experts due

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *International Journal of Consumer Studies* published by John Wiley & Sons Ltd.

to their specialized knowledge and accumulated experience, which translates into superior performance in their respective domains (Miller et al. 2021). This expertise is often visible in multiple aspects of commerce, from pricing decisions to negotiation strategies. However, one underexplored dimension of expertise is communication style—specifically, how professionals and amateurs use language differently in promotional content (Goodwin et al. 2014). Given that initial interactions between sellers and buyers often occur through textual descriptions, how information is presented may serve as an implicit signal of quality and credibility (Spence 1973; Kirmani and Rao 2000). Signaling theory is based on the idea that in situations of asymmetric information, such as in C2C markets, sellers use observable attributes or behaviors as signals to communicate their product quality or reliability to consumers (Pfeuffer and Phua 2022; Mayayise 2024). In the context of both amateur and professional sellers, these signals can differ significantly, shaping consumer perceptions and market dynamics.

Consumer language research has experienced considerable development in the past decade (Packard and Berger 2024). Recent research has demonstrated that linguistic style can influence perceptions of trustworthiness and product quality (Moradi et al. 2024), yet the mechanisms behind these effects remain unclear. To address this gap, our study applies stylometry—a computational approach that quantifies linguistic style through features such as word usage, syntax, and structural patterns (Neal et al. 2017)—to examine whether professionals and amateurs differ in their textual patterns. By leveraging topic-prevalent n-grams, a novel text analysis technique, we systematically investigate how stylistic elements influence perceived trust and effectiveness. Our approach builds upon previous work in marketing and consumer behavior science (Hoover 2013; Stamatatos 2009), extending it to the context of digital commerce.

The objective of this research is to connect expertise theory with signaling theory and examine whether professionals utilize stylistic elements differently from amateurs in real estate listings that typically indicate quality and trustworthiness. In addition to exploring these differences, we aim to investigate the effectiveness of these signals by determining whether the reactions of potential buyers vary based on the stylistic elements present in real estate listings.

This research contributes to multiple domains. First, we enhance the understanding of expertise theory by demonstrating how linguistic patterns distinguish professionals from amateurs in a special C2C context. Second, we integrate signaling theory into consumer language research, illustrating how textual features act as implicit quality indicators for consumers. From a theoretical perspective, our experimental study demonstrates the distinct mechanisms through which professional versus amateur signals operate, specifically via the mediation of perceived trustworthiness and perceived quality. This contributes to signaling theory by showing that the outcomes of stylistic signals—quality and trust—lead to different customer reactions. Finally, we introduce a novel methodological approach—the application of topic-prevalent n-grams—to examine communication effectiveness in a way that has not been explored before in marketing science.

The paper consists of three successive studies. The context of the studies is the real estate market, one of the largest C2C markets, where traditionally, both amateurs (homeowners) and professionals (brokers) are active. First, we use stylometry to show that advertisements vary by author. Unlike the traditional lexicon-based approaches commonly found in style studies, we employed a novel method based on NLP (Natural Language Processing) to analyze topic-prevalent n-grams in Study 1, which examines style through word associations. Second, our findings advance expertise theory by highlighting the differences in language used by professionals compared to amateurs. In Study 2, we used the NRC Word-Emotion Association Lexicon and the Emotion Intensity Lexicon to provide empirical evidence of how professionals strategically incorporate emotionally charged words into their content creation. Finally, we support these results with an experiment in Study 3, demonstrating that the stylistic choices made by professionals are more effective in engaging buyers.

Our study not only contributes to consumer behavior science but also provides actionable insights for consumers who lack professional training. The findings suggest that amateurs can enhance their efficiency by adapting professional linguistic strategies, potentially automating content creation through AI-based tools to produce more persuasive and engaging content. Through a series of empirical studies, we demonstrate that strategic linguistic choices influence credibility, buyer trust, and engagement, highlighting the potential of both human expertise and AI-driven solutions in optimizing digital communication practices.

2 | Literature Review

2.1 | Professional and Amateur Sellers

Expertise theory helps explain the differences in performance between professional and amateur sellers. The theory originates from cognitive psychology and is based on research that uncovers the secret of outstanding performance (Ericsson and Smith 1991). Studies on professionals have expanded our understanding, showing that practice and experience are key sources of expertise. However, research also highlights that it is not just routine expertise but adaptive expertise that sets experts apart from non-experts (Bohle Carbonell et al. 2014).

Selling expertise refers to the depth and breadth of knowledge that salespeople have about their products, market conditions, and customer profiles (Busch and Wilson 1976; Fu and Chen 2011). This understanding significantly impacts sales effectiveness and buyer trust, highlighting the distinctions between professional sellers—who depend on their specialized knowledge and structured selling processes—and amateur sellers, who may lack formal training or thorough market insight (Schwob et al. 2023).

Professional sellers often act as “knowledge brokers,” harnessing their selling-related knowledge to access and disseminate valuable insights within networks. Verbeke et al. (2011) emphasize the importance of knowledge brokering in the sales process, as it enables sellers to create connections that improve their

effectiveness. Having sufficient experience and information about customers and competitors can enhance conversion rates in professional selling environments (Sturman 2003; Fu 2009). Moreover, such expertise fosters interpersonal trust, which has been identified as a driver of customer loyalty in commercial relationships (Guenzi and Georges 2010).

In contrast, amateur sellers typically operate with lesser expertise, potentially undermining customer trust and limiting their effectiveness. Research indicates that while expertise can elevate the perceived quality of sales interactions and positively influence purchasing behavior, amateur sellers may struggle to adeptly perceive market cues and customer needs (Fu and Chen 2011). A study by Aydin et al. (2017) highlights that critical success factors, such as intrinsic motivation and learning orientation, can help sellers enhance their performance, but they fundamentally still lag behind sellers who possess technical and experiential expertise.

The distinction between professional and amateur sellers is further pronounced in industries where local market knowledge is paramount. Fang and Hayunga (2024) found that effective real estate agents significantly impact transaction prices and marketing durations, emphasizing that local expertise surpasses mere transactional capabilities. This principle applies to amateur sellers, who often lack the nuanced understanding of market dynamics that professionals possess, leading to suboptimal selling outcomes. Curto et al. (2015) proved that real estate sales professionals in Italy can understand much more effectively the contribution of the observable qualities to the listing price and their predictive power of which house features are relevant in determining a higher price. Similarly, Miller et al. (2021) confirmed in the US market that professional salespeople can sustain a price advantage over amateurs, even after accounting for the 6% commission rate.

2.2 | The Role of Language and Style

The past decade has seen substantial progress in consumer language research concerned with “how, when, and why linguistic phenomena (e.g., rhetoric, grammar/syntax, phonetics, or semantics) reflect consumer states and traits, predict future actions, or shape consumer attitudes and behaviors” (Packard and Berger 2024, 43).

The way a message is formulated is supposed to convey how recipients should understand its literal meaning (Dion and Notarantonio 1992). The language style used in marketing content both impacts and reflects on consumers and companies, as discussed by Berger et al. (2022). Gao et al. (2023) demonstrated this effect in credit markets where borrowers with better-written and more readable applications received higher funding from lenders. Even small differences in language use, such as the use of the words “we” or “you and I”, can have an impact on consumer attitudes (Sela et al. 2012). Similarly, the formal or colloquial style of charitable appeals makes a difference in donor support (Pfeiffer et al. 2023). In style research, there have been studies of different styles, such as humorous vs. serious (Su et al. 2025), concrete vs. abstract

(Tong et al. 2023), or high-cognitive vs. low-cognitive (Hlee et al. 2021), but fewer studies have looked at the effect of the author of the text. The author's stylistic research is commonly found in digital and criminal forensics, cybersecurity (Brennan et al. 2012), literary studies, and digital humanities (Hoover 2013).

2.3 | Amateur and Professional Language

Even though the difference between amateurs and professionals is shrinking in many domains, research results suggest that experts' language is still different from consumers' language both in content and style (Koltay 2019). Experts tend to communicate in a more factual manner than non-experts (Kim et al. 2011) and often use more concrete language (Packard and Berger 2021). Their style typically involves more fact-based terms and fewer emotional expressions (Forgas and Tehani 2005). As noted by Solomon (1990), experts tend to provide better and richer descriptions of objects and phenomena than amateurs. In contrast, consumers may feel alienated or intimidated by the assertive language of professionally created content (Reavey et al. 2018). Further, amateurs can create messages that enhance the chance to be perceived as more trustworthy and authentic due to their use of emotionally laden language (Koltay 2019). In their study with Linguistic Inquiry and Word Count (LIWC), Kim et al. (2011) were able to differentiate between expert and novice language. Professionals created long texts and used more cognitive words; amateurs used sentences with fewer words, their descriptions were less organized and more fragmentary, and they used more emotional words.

2.4 | Emotional Language

Literature also suggests that emotionally charged expressions can have counterintuitive effects on the consumer decision-making process. For example, Rocklage and Fazio (2020) showed that in C2C communication (such as product reviews), even positive emotions might backfire. On the contrary, negative emotions, especially when expressed with high emotional intensity, can increase the perceived helpfulness of consumer reviews. Ren and Hong (2019), examining specific emotional cues, found that reviews containing fear-related language are seen as more persuasive and helpful. Advertising literature recognizes that negative emotions can serve as effective emotional appeals in persuasive messaging. Fear appeals, for example, are long recognized as potent drivers of behavior change (De Hoog et al. 2005; Witte and Allen 2000).

These findings are consistent with attribution theory (Kelley 1973), which states that individuals not only assess the content of communication but also the intentions behind others' behavior. A display of negative emotions can lead to positive social inferences (Tiedens 2001), for example, people expressing anger or sadness may be perceived as more competent (Hareli and Hess 2010). When negative emotions are expressed in text, this increases feelings of closeness (Ludwig et al. 2022), suggesting that the expression of negative emotions signals sincerity,

which, in turn, can enhance the perceived authenticity and trustworthiness of the message.

2.5 | Style as a Signal

Markets where both professionals and amateurs are present, like second-hand C2C markets, are characterized by a high degree of information asymmetry (Shen et al. 2011) since the buyer is often not aware of all the quality parameters of the property they want to buy, and in many cases, has no way of verifying them. The uncertainty originates from two sources: the uncertainty of the product quality and the uncertainty of the seller's credibility (Ghose 2009) and thus, trust plays a crucial role in the transactions (Berg et al. 2020). To overcome uncertainty originating from information asymmetry and the complex purchase situation, sellers use signals to indicate the quality of the products and the seller's credibility. Signaling theory is widely used in areas where information asymmetry exists (Kirmani and Rao 2000). Research is typically conducted in situations where it is difficult to judge the quality of the product and/or the quality of the seller before the transaction (Li et al. 2009). Signals include a money-back guarantee and branding (Lee et al. 2005), e-WOM (Tóth et al. 2022), or the broadcaster's physical characteristics (Lu and Chen 2021), while the credibility of the seller can be indicated by signals such as third-party payment or seller certification (Cardoso and Martinez 2019).

Although signaling theory originally suggests that observability is a necessary characteristic of a signal (Connelly et al. 2011) and signals are processed in isolation (Evans 2008), Steigenberger and Wilhelm (2018) proved that rhetorical signals complement substantive signals and may have a holistic effect in a high-noise environment. Moradi et al. (2024) extended their results by proving that rhetorical signals (emotional and cognitive tone, communal language, and linguistic style) play an important role in finding backers for crowdfunding projects. This result showed that less observable and recognizable elements, such as style, can also serve as signals. While the literature on the type and characteristics of signals is extensive, there is much less literature on the signalers (Bafera and Kleinert 2023). The identity

and style of the signalers are much more pronounced in style research.

3 | Overview of the Studies

Our objective with this research is to link expertise theory with signaling theory and explore if professionals use stylistic elements different from amateurs in real estate listings that traditionally signal quality and trustworthiness. Beyond exploring the differences, we also investigate whether these signals work, that is, to determine whether the reactions of potential buyers differ along the stylistic elements of real estate listings.

In this research, we conducted a detailed examination of linguistic features of real estate classified advertisements in three successive studies, which took place in two phases (Table 1). The primary aim of Phase 1 was to examine the differences in the use of language between professionals and amateurs, through the analysis of a corpus comprising over 300,000 US listings. In Phase 1, we conducted two studies. Study 1 focused on the use of stylometric markers to examine textual patterns, while Studies 2a and 2b used the NRC Word-Emotion Association Lexicon and the NRC Emotion Intensity Lexicon, respectively. Phase 2 (Study 3) built on key findings of Phase 1. While Phase 1 provides results on the differences in language use between professional and amateur sellers, Phase 2 analyzes the business relevance of this difference. In this phase, we bridged the results from Phase 1 by evaluating the impact of identified linguistic features on potential consumers. We developed a series of hypotheses that we tested with the help of an experiment with a sample of 370 US homeowners.

4 | Phase 1

Both studies of Phase 1 used data that were systematically extracted from the archives of [craigslist.org](https://www.craigslist.org), the preeminent classifieds website, using the Internet Wayback Machine (IWM, archive.org). This research involved downloading 399,208 real estate sales listings from two sections of the site: “reb” (real estate by brokers—professionals), and “reo” (real estate by owners—amateurs).

TABLE 1 | Research design.

Study	Objective of study	Method/data
Phase 1. Difference between language use of real estate brokers/homeowners		
Study 1	Measure stylistic differences in the language use of real estate brokers/homeowners	Text mining/corpus: 300,000+ US real estate classified ads
Study 2a	Emotional differences in language use of real estate brokers/homeowners with NRC Word—Emotion Association Lexicon	
Study 2b	Differences in the intensity of emotions in the use of language in real estate brokers/homeowners with NRC Emotion Intensity Lexicon	
Phase 2. Evaluating the impact of identified linguistic features on potential consumers		
Study 3	Measuring the impact of the differences in language use on potential consumers	Experiment/sample of 370 US homeowners

Following the removal of duplicates and short entries, the final size of the dataset comprised 313,721 listings, published between August 2008 and April 2022 across 411 US regions, with 229,913 listings by professionals and 136,581 by amateurs. The average number of tokenized lemmas was $M_{\text{professional}} = 102.32$ ($SD_{\text{professional}} = 71.54$) and $M_{\text{amateur}} = 97.99$ ($SD_{\text{amateur}} = 76.23$), while vocabulary richness was $M_{\text{professional}} = 70.56$ ($SD_{\text{professional}} = 43.11$) and $M_{\text{amateur}} = 68.24$ ($SD_{\text{amateur}} = 40.82$). The corpus was cleaned following established text mining protocols using Natural Language Toolkit (NLTK) libraries (Bird et al. 2009), assisted by custom Python scripts. The text preprocessing included HTML/JavaScript removal, tokenization, special character handling, stopword removal, and lemmatization—following guidelines recommended by Shankar and Parsana (2022) (see Web Appendix 1: Supporting Information).

4.1 | Method of Study 1

Stylometry is a research field in computational linguistics that focuses on the quantitative features of language (Vysotska et al. 2018). Stylometric features (or style markers) include lexical, syntactic, semantic, structural, and domain-specific levels (Neal et al. 2017) and are widely used in text classification tasks (Schuster et al. 2020), across several scientific and practical fields ranging from digital and criminal forensics to cybersecurity (Brennan et al. 2012), literary studies, and digital humanities (Hoover 2013). This study introduces stylometry to marketing science, where it has been rarely applied, and explores its use in distinguishing between professional and amateur content creators—a novel focus. Following Stamatatos (2009), we extracted lexical, syntactic, and application-specific linguistic features from the corpus (see Supporting Information).

We propose a new method—topic prevalent n-grams—to measure and compare the writing styles of the two groups. N-grams are entities that “can be defined as a set of consecutive words” that “can be useful if the researcher is interested in using the text as input for a predictive model” (Berger et al. 2020, 11). N-grams are commonly used for author identification in different scientific and practice-oriented fields, but we did not find any studies in the literature that connected n-gram analysis with groups differentiated by expertise or domain knowledge.

We propose the following method for our research. Building on two consecutive words (or bigrams) we developed the topic prevalent n-grams method that allowed us to explore recurring n-gram patterns, providing deeper insights into the stylistic signatures of the two subcorpora under study. Appendix 1 details a formal definition of the topic prevalent n-grams method. Here we present an example to illustrate the concept for clearer initial understanding. Take “satellite image” as an exemplary bigram. We found that this bigram occurred in 0.67% of all broker postings and only 0.02% of all owner postings. Hence, we call a satellite image a “professional prevalent bigram” if the difference between 0.67% and 0.02% exceeds a certain θ threshold (to be found empirically). In our corpus, we identified 2,857,358 unique bigrams. Based on the algorithm in Appendix 1, the empirical threshold was set at 0.00003, yielding 128,249 owner-prevalent and 186,334 broker-prevalent bigrams.

4.2 | Results of Study 1

Study 1 evaluated the effectiveness of known stylometric features (see Appendix 2) and topic-prevalent n-grams in classifying texts by group membership. Using a decision tree, Multinomial Naive Bayes (MNB), and logistic regression classifiers, our study provides model-free evidence on the effectiveness of various stylometry features and prevalent n-grams. We compared the performance of the decision tree, MNB, and logistic regression (Table 2).

Examining the role of univariate analysis in distinguishing classified ads written by professionals from those written by amateurs, our findings reveal that among the stylometric features assessed (Stamatatos 2009)—word count, vocabulary richness, part-of-speech (POS) tags, sentiment, bag of words—none demonstrated significant explanatory power in identifying group membership, with the notable exception of topic prevalent n-grams. This method achieved a predictive accuracy of 84.01%–88.17%, highlighting its effectiveness. These results are in line with the literature; multiple studies showed the explanatory power of n-grams to be stronger than that of other stylometric variables (Eder and Górski 2023; Stamatatos 2009; Wright 2017).

4.3 | Method of Study 2

Emotions strongly influence buying decisions (Jørgensen 2016; Levy et al. 2013) and pathos (the emotional tone of the text) is a common tool used in advertisements (Arndt et al. 2013, 2017; Pryce and Oates 2008). While Study 1 found no major sentiment differences, we propose that the key difference lies in the expression of emotions and their intensity. Thus, it is not the valence of emotions (positive, negative, or neutral tone) that may serve as distinctive characteristics, but rather the specific emotions and the intensity of these emotions.

We reanalyzed the Study 1 database using NRC lexicons. In selecting the NRC Lexicon for Study 2 over other lexicons such as the LIWC, we prioritized the NRC Lexicon's ability to identify and categorize emotions in text, offering two valuable tools: the Word—Emotion Association Lexicon (Mohammad and Turney 2013), tagging words with 8 core emotions, and the Emotion Intensity Lexicon (Mohammad 2017), scoring emotional strength. The decision to use the NRC Emotion Lexicons was further influenced by their construction method; the scores within these lexicons were derived from crowdsourced evaluations, reflecting the perspectives of the general population (Mohammad and Turney 2013).

4.4 | Results of Study 2

4.4.1 | Study 2a

By using the NRC Emotion Lexicon (Mohammad and Turney 2013), we were able to quantify the presence of words related to 8 core emotions in the advertisements. This approach provided a new perspective on understanding the emotional appeals used, compared to the stylometry-based method used in Study 1. The data shows that emotional language is significantly

TABLE 2 | Performance metrics of classifiers on different stylometric markers.

	Decision tree				Multinomial naïve bayes				Logistic regression			
	P	R	F1	A	P	R	F1	A	P	R	F1	A
Word count	0.613	0.991	0.757	0.612	0.612	1.0	0.612	0.759	0.612	1.0	0.612	0.760
Unique word count (vocabulary richness)	0.619	0.961	0.753	0.614	0.612	1.0	0.612	0.759	0.612	1.0	0.612	0.760
Part of speech tags	0.682	0.834	0.750	0.660	0.672	0.830	0.648	0.743	0.658	0.9006	0.653	0.761
Sentiment	0.623	0.953	0.753	0.618	—	—	—	—	0.610	0.9801	0.604	0.752
Bag of words (no tiers)	0.616	0.981	0.756	0.613	0.612	1.0	0.612	0.759	0.611	1.0	0.612	0.759
Bag of words: 10 tiers	0.650	0.855	0.739	0.630	0.639	0.811	0.605	0.715	0.626	0.9377	0.620	0.751
Prevalent bigrams	0.8766	0.900	0.888	0.861	0.931	0.765	0.822	0.840	0.868	0.8960	0.853	0.882

Note: Decision tree max_depth = 10.

Abbreviations: A: accuracy, P: precision, R: recall.

more frequent in amateur listings than in professional listings—particularly for negative emotions. Words associated with anger appear more frequently in amateur listings (0.9287 vs. 0.5726), with a noticeable effect size ($d = -0.27$). Similarly, expressions of disgust are more commonly found in listings created by amateurs (0.6308 vs. 0.4079), with a substantial effect size ($d = -0.23$). (This pattern likely reflects how amateurs describe undesirable features more directly, using terms like “sewer,” “toilet,” or “open sewage”, while professionals tend to avoid such language. The elevated presence of disgust-related terms may indicate unfiltered descriptions).

On the other hand, the slightly higher frequency of words related to trust in professional listings (4.06 vs. 3.94), despite a small effect size ($d = 0.03$), may indicate a professional strategy to promote reliability and assurance. We draw similar conclusions from the noticeable cautiousness of professionals in using negative emotional words across all four negative categories: anger (amateurs: 0.9287, professionals: 0.5726, $d = -0.27$), disgust (amateurs: 0.6308, professionals: 0.4079, $d = -0.23$), fear (amateurs: 0.7222, professionals: 0.4859, $d = -0.19$), and sadness (amateurs: 0.8853, professionals: 0.5737, $d = -0.24$) (see Web Appendix 2a: [Supporting Information](#) for full results).

4.4.2 | Study 2b

In addition to the use of emotionally charged words, the emotional intensity of the word choices of both groups was assessed using the NRC Emotion Intensity Lexicon (Mohammad 2017).

Incorporating the findings of Study 2a, an examination of the emotional intensity scores associated with basic emotions in real estate advertisements revealed a less frequent use of emotionally charged language by professionals. However, where agents do choose to use emotional lexicon, there is a clear preference for words that evoke stronger responses for positive emotions. This pattern is exemplified by the emotional intensities of anticipation and joy, where professionals exhibit higher mean scores (anticipation intensity: $M_{\text{professional}} = 0.4917$, $M_{\text{amateur}} = 0.4448$, Cohen's

$d = 0.42$; joy intensity: $M_{\text{professional}} = 0.3879$, $M_{\text{amateur}} = 0.3457$, Cohen's $d = 0.38$). Conversely, in the context of negative emotions, larger Cohen's d values are observed for disgust and fear, indicating that amateurs tend to use these emotions in a more intensive way in their listings (disgust intensity: $M_{\text{professional}} = 0.4407$, $M_{\text{amateur}} = 0.5369$, Cohen's $d = -0.37$; fear intensity: $M_{\text{professional}} = 0.3270$, $M_{\text{amateur}} = 0.3518$, Cohen's $d = -0.16$) (see Web Appendix 2b: [Supporting Information](#) for full results).

5 | Phase 2

5.1 | Hypotheses

Our research examines whether stylistic differences between professional and amateur-generated advertisements influence consumer perceptions and decision-making. Study 1 identified how these two seller groups differ in their writing style, while Study 2 explored the emotional characteristics of their messaging. In Study 3, building on prior research into language's role in persuasion and signaling, we examine how stylistic differences between professional and amateur real estate advertisements influence consumer perceptions and behavior. Drawing on expertise theory and signaling theory, we explore whether these styles affect behavioral intention directly, as well as indirectly through perceived trustworthiness and product quality. The following hypotheses, illustrated in Figure 1, outline our expectations regarding these relationships.

Prior research suggests that language plays a crucial role in shaping consumer interest and willingness to take action (Pryce and Oates 2008). Professionals equipped with domain expertise and market experience are expected to create advertisements that align with best practices in persuasive communication (Hebdzyński 2023). In contrast, amateurs, lacking formal training, may produce less strategically framed messages (Curto et al. 2015; Miller et al. 2021). One key factor in advertising effectiveness is behavioral intention—whether the message compels consumers to seek further information or proceed with a purchase (Pryce and Oates 2008). Building on expertise theory,

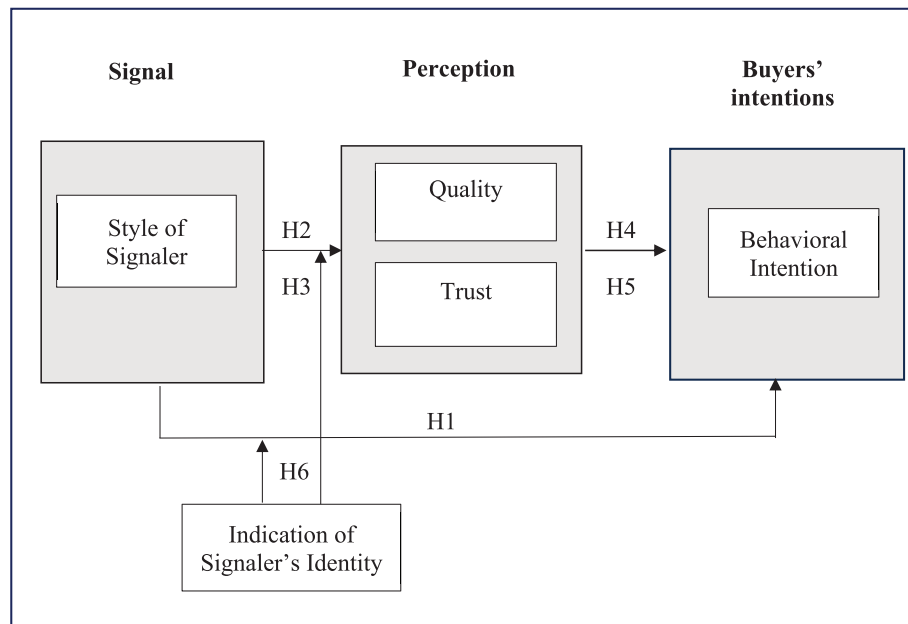


FIGURE 1 | Overview of the hypotheses.

we propose that professional-style advertisements are more effective in driving consumer engagement.

H1. *Advertisements written in a professional style have a stronger impact on potential buyers' intention to view or purchase the house compared to those written in an amateur style.*

In addition to the intention to act, it is also worth looking at the factors that help to sell the property. Based on signaling theory, two fundamental factors were identified that signals must communicate: quality of the product and reliability/trustworthiness of the signaler (Connelly et al. 2011). Trust between parties in commercial transactions can lower transaction costs and make the process faster and easier. This phenomenon has been widely studied in personal selling (Swan and Nolan 1985); direct selling (Young and Albaum 2003), and business-to-business sales (Valtakoski 2015). Trust in complex product markets can become essential when the information asymmetry between the buyer and seller is high. The real estate market is a specific market where buyers are usually private individuals, but sellers are either private individuals or professional agents. We have evidence that information asymmetry in the real estate market is a problem and can lead to misbehavior by agents (Rutherford et al. 2005). On the other hand, we only have qualitative results (DeMarco 2016) or speculative evidence based on the sellers' views (Stamsø 2015) to support the idea that buyers trust amateurs more than professional salespeople.

Based on signaling theory and in accordance with Moradi et al. (2024), we believe that style can function as a signal. From the results of Studies 2a and 2b, we know that amateurs are willing to display negative emotions, signaling the feeling that they do not want to hide the negative aspects of the property. This may imply that the creator of these advertisements can be trusted more (Ludwig et al. 2022). All this leads us to hypothesize that stylistic elements of an amateur-type advertisement signal the trustworthiness of the seller.

H2. *Creators of amateur-style listings are viewed as more trustworthy by potential buyers than those who create professional-style listings.*

On the other hand, however, advertisements are only effective if the property displayed is sufficiently attractive. Given that professionals have the background and competence to produce attractive listings (Gay and Zhang 2014; Palmon and Sopranzetti 2017), it is important to examine whether the perceived quality of the property advertised is, in fact, affected by the style of the creator. From the results of Studies 2a and 2b, we know that professionals use words with a positive charge that are more emotionally intense and use fewer words with a negative charge. Thus, they are more focused on signaling the qualities of the property and emphasizing them as intensively as possible. Based on these results, we propose that the stylistic elements of listings created by professionals signal higher quality to the potential buyer.

H3. *The quality of properties in advertisements by professionals is seen as superior compared to those created by amateurs.*

An important stage in the customer journey of buying a property is whether potential buyers are interested in the property and respond to the advertisement in some way. Our hypotheses so far have aimed to test this relationship. However, the question of what internal process leads to this point remains to be answered. We propose that both trustworthiness and perceived quality may influence the buyer's behavioral intention. We know that trust has both a direct effect on the buyer-seller relationship driven by the seller's expertise (Kennedy et al. 2001; Wood et al. 2008) and has a mediating role for future buying intentions by softening the risk of continuation of the relationship (Garbarino and Johnson 1999).

Based on signaling theory, we propose that style as a signal has an effect on the action of the potential buyers through the mediation of the credibility of the signaler.

H4. *Perceived trustworthiness of the seller in amateur-style advertisements positively influences the consumer's behavioral intention.*

According to signaling theory, not only the trustworthiness of the signaler but also the perceived quality of the product can increase the willingness to purchase (Connelly et al. 2011). The role of perceived product quality as a mediator between signals and willingness to buy is a phenomenon that has been explored in previous research. Treiblmaier and Garaus (2023) showed that blockchain labels as signals enhanced perceived quality, which subsequently led to a higher willingness to purchase the product. In a similar vein, Tsui (2012) found that advertising that enhances perceived quality increases consumers' willingness to pay. Wells et al. (2011) demonstrated, based on signaling theory, that the quality of a website and the credibility of its signals contribute to a greater intention to purchase by enhancing the perceived quality of the product. Based on these results, we propose that the higher the perceived quality of the professional-style listings, the more likely the buyer is to take an action.

H5. *Perceived product quality in professional-style advertisements positively influences the consumer's behavioral intention.*

The final aspect we wanted to explore is whether factors beyond style influence buyers' perceptions and intentions. The effectiveness of real estate listings can be impacted not only by the content of the advertisement but also by the identity of the source. Identity signals refer to communicative actions, displays, or behavioral patterns that indicate an individual's affiliation with a recognizable and categorizable social group (Smaldino 2022). Revealing or concealing one's identity can lead to distorted perceptions and may alter the listener's opinion (Le Forestier and Lewis Jr 2024).

Research on source credibility primarily focuses on factors such as the attractiveness, gender, or similarity between the source and the listener (Pornpitakpan 2004). However, aspects like the expertise of the source are seldom examined. In the video game market, Cox and Kaimann (2015) demonstrated that reviews from professional critics positively influence sales significantly more than consumer reviews do. This leads us to propose a moderation effect. We assume that revealing the advertiser's identity may enhance the primary influence of stylistic presentation.

H6. *The effect of stylistic signals on (a) behavioral intention, (b) trustworthiness, and (c) perceived quality is moderated by the visibility of the advertiser's identity (professional vs. amateur).*

5.2 | Method of Study 3

5.2.1 | Design and Procedure of Study 3

To test our hypotheses, we designed a computer-based experiment with a 2 (listing style: professional vs. amateur) $\times$ 2 (indication of advertiser: yes vs. no) between-subjects design. Behavioral intention, trustworthiness, and perceived quality of the real estate served as dependent variables.

We collected data from Prolific; the final sample consisted of 370 participants (mean age = 46.36, SD = 12.91; 49.7% female). To be able to match our experimental results with the results from Study 1, we screened our respondents based on two criteria. First, only US residents were eligible. As the corpus of our data in Phase 1 was a US-based dataset, we wanted to limit the data collection to the USA. Second, only those respondents who were owners of real estate were reached. We considered property ownership to be important because we expected respondents who owned property to have some experience in buying property. In the final sample, based on the respondents' answers, 40.3% had bought real estate in the past 5 years, 42.4% more than 5 years ago, and 17.3% had never bought real estate. We used this variable as a control in the analysis to see if experience in buying real estate affected the results. Respondents were randomly assigned to one of the treatments. The listings were created for the research, and the effects of the manipulations designed in the listings were measured.

5.2.2 | Stimulus Materials and Measures

We used text as the manipulation material (Berger et al. 2022). In the questionnaire, real estate listings were created and tested as manipulations. Our objective was to create listings that described an average piece of real estate but in different styles. It should be noted that in this case, it was not sufficient to follow the usual scenario-based design where the same scenario text is used and only words or phrases are changed (e.g., instead of 'nice' home, it was not sufficient to use 'wonderful' home). In this case, based on the topic-prevalent n-gram method, the whole text had to reflect the style of the listing. Thus, it is not only the individual words that are important but also the word combinations and phrases (bigrams and n-grams). Accordingly, the elaboration of the stimuli was a longer process, and the algorithm developed in Phase 1 was called upon to help us create a typical broker/owner-type listing (professional/amateur style).

The initial property for which we created listings was a Mediterranean-style 880 sq. ft. house with two bedrooms, worth around \$500,000. We chose the Mediterranean style, as it was mentioned in both professionally created and amateur listings, in different contexts. Basic parameters were set to reflect the "average" home on the US market, ensuring that our stimuli resonated with typical buyer expectations. Although location and price are the two main pieces of information required by home buyers (Nowak and Smith 2017), our interest is in the effect of listings; thus, we controlled for these two characteristics of the property.

The text of the listings for the house was manipulated in the experiment following an iterative process based on the findings of Phase 1. With the help of the algorithm, we collected the most typical listings and created a list of the most frequently occurring items in a listing (e.g., bedroom, furnished, bathroom). These key items were run in the algorithm to find out in which word combinations the word occurred in professional and amateur listings. Then we ran the word combinations again, and so on until we were able to create a listing that reflected the "typical" professional and amateur style for the

same property. The listings were then scored by the algorithm to see if they matched the respective style. The iteration was continued until the listings were placed by the algorithm in the two extreme quartiles, thus ensuring the strongest possible difference. Whichever version met this goal, we also looked at emotions and tried to swap words to reflect the emotional characteristics of the owner/broker classified ads. We aimed to cover as many emotional categories as possible so that we ended up with a good overlap with the original sample for six of the eight emotional categories.

After the iterative process, we tested the listings for language and content with the help of homeowners and real estate experts. We also tested the listings on a sample of students.

The visual design of the stimuli largely followed the design of [craigslist.org](https://www.craigslist.org) with some minor modifications (small picture, table highlighting main features). To give respondents the same mental picture of the text, an average photo was placed next to the text, and the main basic data was highlighted in a table. However, we made sure that the size of the photo was much smaller than the text itself so that the focus was placed on the text. To manipulate whether the advertiser's identity appeared in the listing, we simply displayed it in the title bar or omitted whether the seller was a broker or owner. The final manipulations can be found in Web Appendix 3: [Supporting Information](#).

To measure the dependent variables, we used 7-point Likert scales. Behavioral intention was measured with two items: "I would love to see this house for real" and "I would love to buy this house" ($r=0.78$, $p<0.001$). Perceived quality of the real estate was measured with three items: "This house is unattractive/attractive, poor quality/great quality, ordinary/glamorous" ($\text{Alpha}=0.82$). Trustworthiness of the advertiser (Kennedy et al. 2001) was measured with two items: "I believe what this person (who created this ad) says and that he/she would not try to take advantage of the customer", and "This person is straightforward and honest even though his/her interests are involved" ($r=0.83$, $p<0.001$).

5.3 | Results of Study 3

5.3.1 | Manipulation Check

To test if the broker/owner identity manipulations were effective, we asked respondents if a broker or owner created the ad (if it was indicated): 95% of those for whom it was indicated that the listing was created by a broker guessed well, while 76% of those for whom it was an owner created one guessed well (Pearson $\text{Chi}^2<0.001$).

The manipulation check for the style manipulation was based on the algorithm in Study 1. We ran the algorithm to decide whether the listing created was in broker or owner style. As described above, the text of the ads was adjusted until the algorithm found it to be in a suitable broker or owner style. The numerical result of this is given in Web Appendices 4a and 4b: [Supporting Information](#).

5.3.2 | Effect of Advertisement Style and Identity of the Seller

First, we performed a 2×2 multivariate analysis of covariance (ANCOVA) where advertisement style and identity of the advertiser served as independent variables and behavioral intention, perceived trustworthiness, and perceived quality as dependent variables. We used age, gender, educational level, and previous experience in buying a home as controls. Figure 2 provides a visualization of the results.

The analysis indicated that advertisement style had a significant effect on behavioral intention, with professionally created content driving higher behavioral intention ($M_{\text{professional_style}}=4.70$ vs. $M_{\text{amateur_style}}=4.26$; $F(1,369)=7.9$, $p<0.01$) supporting H1. Similarly, advertisement style had a significant main effect on trustworthiness, but in the opposite direction, suggesting a stronger effect of the amateur style advertisements ($M_{\text{professional_style}}=3.89$ vs. $M_{\text{amateur_style}}=4.20$, $F(1,369)=4.66$, $p<0.05$), supporting H2. The results also supported that classified ads created in professional style had a higher perception of the real estate quality ($M_{\text{professional_style}}=4.64$ vs. $M_{\text{amateur_style}}=4.24$, $F(1,369)=6.54$, $p<0.05$); thus, H3 was confirmed.

In addition, we found that identity of the seller did not affect behavioral intention ($M_{\text{indicated}}=4.38$ vs. $M_{\text{not_indicated}}=4.56$, $F(1,369)=1.31$, ns.), trustworthiness ($M_{\text{indicated}}=4.05$ vs. $M_{\text{not_indicated}}=4.05$, $F(1,369)=0.06$, ns.), and perceived quality ($M_{\text{indicated}}=4.36$ vs. $M_{\text{not_indicated}}=4.52$, $F(1,369)=1.14$, ns.). As the effects of the identity were highly insignificant, we accept H6a–H6c. The interaction effect of style and identity of advertisers on the three dependent variables was insignificant. Thus, we reject H6a–H6c.

5.3.3 | Test of Mediation on Behavioral Intention

We used bootstrapping with repeated extraction of 5000 samples (Hayes 2017, PROCESS Model 4, two mediators) to test H4 and H5. We used the style of ads as the independent variable, perceived quality and trustworthiness as mediators, and behavioral intention as the dependent variable. We controlled for the potential effects of age, gender, education, and previous buying experience.

Based on the analysis, we found evidence for full mediation (direct effect: $b=-0.11$; $t=-1.89$; $p=0.06$; 95% CI = $[-0.22$ to $0.004]$), none of the control variables were significant. The results indicate a total effect of style ($b=-0.34$; $\text{BootSE}=0.15$; 95% CI = $[-0.64$ to $-0.05]$), while an indirect effect of trust ($b=0.028$; $\text{BootSE}=0.02$; 95% CI = $[0.003$ to $0.06]$) and an indirect effect of quality ($b=-0.37$; $\text{BootSE}=0.14$; 95% CI = $[-0.65$ to $-0.09]$) resulting in a competing mediation. Based on these results, we accept H4 and H5; both perceived quality and trust mediated the effect of style on behavioral intention. On the other hand, even though respondents viewed classified ads created by amateurs as more trustworthy—leading to a higher intention to engage—the impact of the perceived quality of ads created by professionals was even stronger. Therefore, we confirm Hypothesis (H1),



FIGURE 2 | Results of Study 3.

not as a direct effect, but as an indirect effect through the mediation of quality and trust.

6 | General Discussion

Our research explores a market where both professional and amateur sellers operate, a phenomenon that has gained relevance with the rise of digital platforms and the expansion of online marketplaces. This dynamic is particularly evident in second-hand goods, digital services, and

consumer-to-consumer transactions, where sellers range from experienced professionals to individuals with little to no formal sales expertise.

Such markets are often characterized by information asymmetry, as sellers typically know more about a product’s qualities and potential defects than buyers do. Additionally, many purchases in these markets involve a degree of financial or functional risk, making trust a key factor from the initial stages of interaction. Our study aimed to determine whether expert and amateur linguistic styles in advertisements are distinct and whether these

differences impact buyer perceptions of trust and quality, as well as their behavioral intention.

The research consists of findings from three interconnected studies focused on the real estate market, which serves as a significant C2C environment involving both amateur sellers (homeowners) and professional brokers. In Study 1, we employed a stylometric approach to analyze topic-relevant n-grams. This analysis revealed systematic stylistic differences between authors, allowing us to identify nuanced word-association patterns that distinguish individual writing styles. Building on these insights, Study 2 further explored expertise theory by comparing the language used by professionals and amateurs. It demonstrated that professionals intentionally incorporate emotionally charged language into their advertisements. Finally, Study 3 provided experimental evidence suggesting that these professional stylistic strategies result in higher buyer engagement, thus reinforcing the connection between linguistic style and market effectiveness.

In the first phase of the research, we used two studies to explore the linguistic and stylistic differences between professional broker and amateur homeowner classified ads. In this phase, we explored the linguistic differences between listings created by professionals and amateurs. We analyzed a dataset of 229,000 broker listings and 136,000 homeowner listings from [craigslist.org](https://www.craigslist.org). By extracting stylometric features and applying algorithms from literary studies and forensics, we developed a method called topic prevalent n-grams. Using logistic regression, this method successfully differentiated the two groups with an 88.17% accuracy.

In Studies 2a and 2b, we investigated the use of emotional language through the NRC Lexicon. The analysis showed that professionals generally use fewer emotionally charged words across seven of eight emotional categories, except 'trust.' However, they tend to select words with greater intensity in all categories except for the negative emotions of 'fear' and 'disgust.'

In Phase 2, we assessed how real estate listings impact potential buyers. Employing a 2×2 between-subjects experimental design, we manipulated the listing styles and the presence of the advertiser's identity. Drawing from the results of Studies 1 and 2, we created listings in both professional and amateur styles that were presented with and without the advertiser's identity. In addition, we measured behavioral intention, perceived quality, and trustworthiness of the advertiser. The results indicated that the listing style significantly affected all three dependent variables, whereas the advertiser's identity did not.

Furthermore, mediation analysis revealed that listing style influences behavioral intention through perceived quality and trust, demonstrating full mediation. This suggests that listings in professional style are seen as higher quality, leading to a higher behavioral intention. At the same time, amateur-style listings cause higher trust, leading to higher intention. Among the two competing effects, the quality route seemed to be stronger, indicating that higher quality perceptions outweighed

weaker trust perceptions, leading to greater efficiency in professional advertisements.

7 | Theoretical Contributions

Our study contributes to consumer language research, expertise theory, and signaling theory by demonstrating how stylistic differences influence advertising effectiveness. Research on rhetoric (Moradi et al. 2023; Steigenberger and Wilhelm 2018) and style (Pfeiffer et al. 2023) has shown that linguistic choices impact content effectiveness. We extend this research by identifying significant differences in writing styles between professional and amateur sellers. While prior studies have primarily focused on word selection and grammatical structures (Pfeiffer et al. 2023) or function words (Boghrati et al. 2023; Ireland and Pennebaker 2010; Moradi et al. 2024), our research introduces a different approach.

We developed a topic-prevalent n-grams algorithm adapted from authorship attribution models used in criminology (Brennan et al. 2012) and literary studies (Hoover 2013). This is the first study to apply this method to distinguish writing styles based on expertise. Our findings highlight that style is more than just word choices—it involves the relationship between words and how they structure persuasive messages. This innovation enhances marketing research by demonstrating the effectiveness of simple NLP tools, such as n-grams, in identifying distinct writing patterns. Our findings also highlight that style is more than rules or lexicons; style may also mean how the words relate to each other.

Beyond methodology, our research expands style research by examining the role of author expertise in shaping stylistic choices. Studies have explored contrasts such as formal vs. colloquial (Pfeiffer et al. 2023), humorous vs. serious (Su et al. 2025), and concrete vs. abstract (Tong et al. 2023). However, few studies have examined how expertise influences writing style in a business context. Our findings suggest that professionals adopt stylistic strategies that enhance persuasion and consumer engagement.

In addition to style research, our study contributes to expertise theory, originally developed in cognitive psychology (Ericsson and Smith 1991). This theory explains how expertise develops through experience and deliberate practice. In markets where both professionals and amateurs operate, such as C2C second-hand markets, the performance gap between these groups remains debated. While some studies suggest that professionals secure higher prices (Curto et al. 2015; Miller et al. 2021), others find no significant price premium (Hendel et al. 2009).

Our findings indicate that amateur-style advertisements are perceived as more trustworthy, while professional-style listings signal higher product quality and are more likely to drive consumer action in the initial phase of the buying process.

Prior research suggests that experts communicate differently from amateurs. Experts use factual, concrete language (Kim et al. 2011; Packard and Berger 2021) and rely more on objective

statements rather than emotions (Forgas and Tehani 2005). By linking emotional language research with expertise theory, we provide evidence that experts not only write differently but do so for strategic reasons. In analyzing the presence and strength of emotions, we found that emotional language is significantly more frequent in amateur listings than in professional ones, particularly for negative emotions. In terms of emotional intensity, professionals are generally less likely to use emotionally charged language, but when they do, there is a clear preference for words that evoke stronger responses to positive emotions. This approach appears to be strategic and aimed at maintaining a professional tone so as not to detract from the appeal of the properties; negative emotional expressions could have a negative impact on potential buyers' perceptions (Haag et al. 2000).

Our study also contributes to signaling theory by demonstrating that linguistic style functions as a market signal. While traditional signals include pricing, branding, and third-party endorsements, we show that stylistic elements also shape buyer perceptions. Our findings indicate that professional-style advertisements signal higher product quality, while amateur-style advertisements signal higher seller trustworthiness. This aligns with research emphasizing that effective signals must be both observable and interpretable (Connelly et al. 2011; Kirmani and Rao 2000). However, writing style operates as a more subtle, often unconscious, signal, influencing buyers even when they do not consciously recognize its effects. Most prior research has focused on the types of signals rather than who sends them (Bafera and Kleinert 2023). We extend this by showing that the signal sender—expert vs. non-expert—affects both product quality perceptions and seller credibility.

While research has documented the effectiveness of different signals (Bafera and Kleinert 2023; Connelly et al. 2011; Kirmani and Rao 2000), the underlying process remains less explored. By integrating signaling theory with expertise theory, we provide a clearer understanding of how professional and amateur writing styles impact consumer decision-making. The finding that professional-style messages enhance perceived quality while amateur-style messages increase trust underscores the importance of dual signaling pathways (Connelly et al. 2011). This nuanced finding advances signaling theory by demonstrating that various stylistic cues do not function in isolation; instead, they can create competing mediating effects—specifically, trust versus perceived quality—on consumer behavior. Importantly, our study does not analyze these mediating mechanisms separately; rather, it explores their simultaneous effects within a single empirical framework. By simultaneously testing these effects, we demonstrate how stylistic signals can generate competing pathways, and how their relative strength ultimately shapes consumer action. This contributes to signaling theory by providing novel evidence that signals may interact, rather than function independently, and that their effectiveness must be understood in terms of their combined influence. In contrast to prior research that highlighted quality as a mediator of willingness to pay (Tsui 2012; Wells et al. 2011) or examined trust in isolation (Garbarino and Johnson 1999), our findings enrich signaling theory by showing that the relative dominance of quality signals over trust signals explains why professional-style communication is more effective in C2C markets.

Finally, by confirming that quality signals dominate trust signals in shaping consumer action, our study refines the understanding of how buyers prioritize cues in high-stakes, information-rich contexts such as real estate. This complements prior findings that different signals gain salience at different stages of the decision-making process (Moradi et al. 2024). In early filtering stages, quality-based stylistic signals appear decisive, while trust-based cues may play a larger role later in relational interactions. Thus, our research deepens theoretical insights into when and why particular signals are more effective.

8 | Practical Implications

Our study demonstrates the significant role of style in shaping consumer decision-making. Expert sellers instinctively adopt more effective linguistic patterns, leading to stronger engagement than amateur sellers. This suggests that stylistic choices are not minor details but key factors in influencing consumer behavior, reducing selling time, and enhancing market efficiency.

Our methodology also holds important implications for marketing and content optimization. Stylometry, a widely used tool in forensic and literary analysis, has been underutilized in consumer behavior studies. Our research highlights how stylometry can be used to measure and refine stylistic effectiveness, allowing businesses to create data-driven recommendations for improving textual content. Additionally, author-style recognition systems could help intermediary platforms categorize content, validate quality, and provide tailored suggestions for enhancing effectiveness.

Finally, advancing current research could assist intermediaries in developing tools that bridge the stylistic gap between professionals and amateurs. Our research shows that stylistic signals and emotional language are vital for lead generation. The findings reveal that these elements significantly influence buyer engagement and shape perceptions of trust and quality. This underscores the potential to enhance content strategies for more effective lead generation.

9 | Limitations and Future Research Directions

While our research provides valuable insights, certain limitations should be acknowledged to clarify the scope of the results. The validity of Phase 1 could be strengthened by applying our methodology to other markets where both professional and amateur sellers coexist, such as used cars or high-tech gadgets. We analyzed real estate listings as a unified dataset without considering factors such as property price, neighborhood, or broker experience levels. These elements may influence stylistic differences and warrant further exploration. Similarly, we did not differentiate between independent agents and those working for larger firms, which could impact advertising strategies. Future research should examine these distinctions to gain a more nuanced understanding of stylistic variation.

Our study relied on publicly available data, limiting the ability to assess actual sales outcomes. Future studies could incorporate

sales cycle data, user profiles, or transaction success rates to enhance the managerial implications of our findings. A further limitation of the dataset is that it is a US-based, English-language corpus, which represents a contextual boundary, but also opens a promising avenue for future research to examine whether language-emotion associations hold across cultural and multilingual contexts. Additionally, style in advertising may influence multiple stages of the selling process, but our study primarily focuses on the initial phase. We identified a seemingly contradictory relationship: amateur-style advertisements increased trust, while professional-style text signaled higher quality and generated greater interest. Further research should explore whether trust and quality signals can be balanced, enabling sellers to optimize both rather than choosing between them.

Since professional language style is systematically distinct, it is reasonable to assume that these differences result from a learning process. Investigating how professionals develop and refine their stylistic expertise could provide valuable insights into how amateurs can improve their communication strategies over time. Additionally, our study classified homeowner/broker writing differences as stylistic without assessing content-specific priorities—for example, which property features each group highlights most often. The n-gram approach used in our study captures style but not content, which may limit generalizability across industries.

Another limitation involves geographical and linguistic factors. Our study analyzed US real estate listings in English, meaning that results may not directly apply to markets with different business models, economic conditions, or languages. Conducting similar studies in other cultural and linguistic contexts could expand the relevance of our findings.

Finally, our data collection in Phase 2 (Study 3) was conducted online using Prolific, limiting participation to US homeowners. Online surveys are subject to external distractions, response biases, and platform-specific limitations, which could affect data quality. Additionally, our research focused solely on textual content, while images, videos, and virtual reality elements are increasingly important in modern digital marketing. Future studies should explore the interaction between text and visuals to provide a more comprehensive understanding of advertising effectiveness.

Ethics Statement

This study was conducted with the approval of the Research Ethics Committee of Corvinus University of Budapest. The ethics approval was granted under file number KRH/23/2024.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data for studies 1, 2a, and 2b come from archive.org and are publicly available. All databases for the studies are available upon reasonable request from the authors.

References

- Arndt, A., D. M. Harrison, M. A. Lane, M. J. Seiler, and V. L. Seiler. 2013. "Can Agents Influence Property Perceptions Through Their Appearance and Use of Pathos?" *Housing Studies* 28, no. 8: 1105–1116.
- Arndt, A., D. M. Harrison, M. A. Lane, M. J. Seiler, and V. L. Seiler. 2017. "Real Estate Agent Target Marketing: Are Buyers Drawn Towards Particular Real Estate Agents?" *Journal of Housing Research* 26, no. 1: 39–52.
- Aydin, S., E. Koc, and N. Kaya. 2017. "The Influence of Critical Success Factors on B2B Sales Performance and the Mediating Role of Social Intelligence." *Business Management and Strategy* 8, no. 2: 1–25.
- Bafera, J., and S. Kleinert. 2023. "Signaling Theory in Entrepreneurship Research: A Systematic Review and Research Agenda." *Entrepreneurship Theory and Practice* 47, no. 6: 2419–2464.
- Berg, L., D. Slettebø, I. Kjørstad, and T. G. Rosenberg. 2020. "Trust and the Don't-Want-To-Complain Bias in Peer-to-Peer Platform Markets." *International Journal of Consumer Studies* 44, no. 3: 220–231.
- Berger, J., A. Humphreys, S. Ludwig, W. W. Moe, O. Netzer, and D. A. Schweidel. 2020. "Uniting the Tribes: Using Text for Marketing Insight." *Journal of Marketing* 84, no. 1: 1–25.
- Berger, J., G. Packard, R. Boghrati, et al. 2022. "Marketing Insights From Text Analysis." *Marketing Letters* 33: 365–377.
- Bird, S., E. Klein, and E. Loper. 2009. *Natural Language Processing With Python: Analyzing Text With the Natural Language Toolkit*. O'Reilly Media, Inc.
- Boghrati, R., J. Berger, and G. Packard. 2023. "Style, Content, and the Success of Ideas." *Journal of Consumer Psychology* 33, no. 4: 688–700.
- Bohle Carbonell, K., R. E. Stalmeijer, K. D. Könings, M. Segers, and J. J. G. van Merriënboer. 2014. "How Experts Deal With Novel Situations: A Review of Adaptive Expertise." *Educational Research Review* 12: 14–29.
- Brennan, M., S. Afroz, and R. Greenstadt. 2012. "Adversarial Stylometry: Circumventing Authorship Recognition to Preserve Privacy and Anonymity." *ACM Transactions on Information and System Security* 15, no. 3: 1–22.
- Busch, P., and D. T. Wilson. 1976. "An Experimental Analysis of a Salesman's Expert and Referent Bases of Social Power in the Buyer-Seller Dyad." *Journal of Marketing Research* 13, no. 1: 3–11.
- Cardoso, S., and L. F. Martinez. 2019. "Online Payments Strategy: How Third-Party Internet Seals of Approval and Payment Provider Reputation Influence the Millennials' Online Transactions." *Electronic Commerce Research* 19: 189–209.
- Connelly, B. L., S. T. Certo, R. D. Ireland, and C. R. Reutzel. 2011. "Signaling Theory: A Review and Assessment." *Journal of Management* 37, no. 1: 39–67.
- Cox, J., and D. Kaimann. 2015. "How Do Reviews From Professional Critics Interact With Other Signals of Product Quality? Evidence From the Video Game Industry." *Journal of Consumer Behaviour* 14, no. 6: 366–377.
- Curto, R., E. Fregonara, and P. Semeraro. 2015. "Listing Behaviour in the Italian Real Estate Market." *International Journal of Housing Markets and Analysis* 8, no. 1: 97–117. <https://doi.org/10.1108/IJHMA-01-2014-0003>.
- De Hoog, N., W. Stroebe, and J. B. De Wit. 2005. "The Impact of Fear Appeals on Processing and Acceptance of Action Recommendations." *Personality and Social Psychology Bulletin* 31, no. 1: 24–33.
- DeMarco, J. V. 2016. "The Decision-Making Processes of Nontraditional or for Sale by Owner (FSBO) Home Sellers Engaged in the Listing and Sale of Their Homes." State University of New York at Albany.

- Dion, P., and E. Notarantonio. 1992. "Salesperson Communication Style: The Neglected Dimension in Sales Performance." *Journal of Business Communication* 29, no. 1: 63–77.
- Eder, M., and R. L. Górski. 2023. "Stylistic Fingerprints, POS – Tags, and Inflected Languages: A Case Study in Polish." *Journal of Quantitative Linguistics* 30, no. 1: 86–103.
- Ericsson, K. A., and J. Smith. 1991. "Definition of Performance and Expertise: A Comparison in: Toward a General Theory of Expertise (1991)." Cambridge University Press.
- Evans, J. S. B. 2008. "Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition." *Annual Review of Psychology* 59, no. 1: 255–278.
- Fang, L., and D. K. Hayunga. 2024. "The Impact of Real Estate Agents' Expertise on House Prices and TOM." *Real Estate Economics* 52, no. 1: 45–72.
- Fernando, A. G., B. Sivakumaran, and L. Suganthi. 2018. "Comparison of Perceived Acquisition Value Sought by Online Second-Hand and New Goods Shoppers." *European Journal of Marketing* 52, no. 7/8: 1412–1438.
- Forgas, J. P., and G. Tehani. 2005. "Affective Influences on Language Use: Mood Effects on Performance Feedback by Experts and Novices." *Journal of Language and Social Psychology* 24, no. 3: 269–284.
- Fu, F. Q. 2009. "Effects of Salesperson Experience, Age, and Goal Setting on New Product Performance Trajectory: A Growth Curve Modeling Approach." *Journal of Marketing Theory and Practice* 17, no. 1: 7–20.
- Fu, J. R., and J. H. Chen. 2011. "The Impact of Seller Expertise and a Refund Guarantee on Auction Outcome: Evidence From an Online Field Experiment of Camera Lens Market." *International Journal of Electronic Business Management* 9, no. 3: 187–195.
- Gao, Q., M. Lin, and R. Sias. 2023. "Words Matter: The Role of Readability, Tone, and Deception Cues in Online Credit Markets." *Journal of Financial and Quantitative Analysis* 58, no. 1: 1–28.
- Garbarino, E., and M. S. Johnson. 1999. "The Different Roles of Satisfaction, Trust, and Commitment in Customer Relationships." *Journal of Marketing* 63, no. 2: 70–87.
- Gay, S., and A. Zhang. 2014. "Expertise Value Added in the Real Estate Market." SSRN Scholarly Paper 2515157. <https://doi.org/10.2139/ssrn.2515157>.
- Ghose, A. 2009. "Internet Exchanges for Used Goods: An Empirical Analysis of Trade Patterns and Adverse Selection." *MIS Quarterly* 33, no. 2: 263–291.
- Goodwin, K., B. Waller, and H. S. Weeks. 2014. "The Impact of Broker Vernacular in Residential Real Estate." *Journal of Housing Research* 23, no. 2: 143–161.
- Guenzi, P., and L. Georges. 2010. "Interpersonal Trust in Commercial Relationships: Antecedents and Consequences of Customer Trust in the Salesperson." *European Journal of Marketing* 44, no. 1/2: 114–138.
- Haag, J., R. Rutherford, and T. Thomson. 2000. "Real Estate Agent Remarks: Help or Hype?" *Journal of Real Estate Research* 20, no. 1–2: 205–216.
- Hareli, S., and U. Hess. 2010. "What Emotional Reactions Can Tell Us About the Nature of Others: An Appraisal Perspective on Person Perception." *Cognition and Emotion* 24, no. 1: 128–140.
- Hayes, A. F. 2017. "Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach." Guilford publications.
- Hebzdzyński, M. 2023. "Quality Information Gaps in Housing Listings: Do Words Mean the Same as Pictures?" *Journal of Housing and the Built Environment* 38, no. 4: 2399–2425.
- Hendel, I., A. Nevo, and F. Ortalo-Magné. 2009. "The Relative Performance of Real Estate Marketing Platforms: MLS Versus FSBOMadison. Com." *American Economic Review* 99, no. 5: 1878–1898. <https://doi.org/10.1257/aer.99.5.1878>.
- Hlee, S., H. Lee, C. Koo, and N. Chung. 2021. "Will the Relevance of Review Language and Destination Attractions Be Helpful? A Data-Driven Approach." *Journal of Vacation Marketing* 27, no. 1: 61–81.
- Hoover, D. L. 2013. "Quantitative Analysis and Literary Studies." *A Companion to Digital Literary Studies*, edited by R. Siemens and S. Schreibman, 517–533. John Wiley & Sons.
- Hutto, C., and E. Gilbert. 2014. "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." in Proceedings of the International AAAI Conference on Web and Social Media 8, no. 1, 216–225.
- Ireland, M. E., and J. W. Pennebaker. 2010. "Language Style Matching in Writing: Synchrony in Essays, Correspondence, and Poetry." *Journal of Personality and Social Psychology* 99, no. 3: 549–571.
- Jørgensen, C. J. 2016. "The Space of the Family: Emotions, Economy and Materiality in Homeownership." *Housing, Theory and Society* 33, no. 1: 98–113.
- Kelley, H. H. 1973. "The Processes of Causal Attribution." *American Psychologist* 28, no. 2: 107–128.
- Kennedy, M. S., L. K. Ferrell, and D. T. LeClair. 2001. "Consumers' Trust of Salesperson and Manufacturer: An Empirical Study." *Journal of Business Research* 51: 73–86.
- Kim, K., J. Bae, M. W. Nho, and C. H. Lee. 2011. "How Do Experts and Novices Differ? Relation Versus Attribute and Thinking Versus Feeling in Language Use." *Psychology of Aesthetics, Creativity, and the Arts* 5, no. 4: 379–388.
- Kirmani, A., and A. R. Rao. 2000. "No Pain, no Gain: A Critical Review of the Literature on Signaling Unobservable Product Quality." *Journal of Marketing* 64, no. 2: 66–79.
- Koltay, T. 2019. *Amateur-Professional*, 1–6. International Encyclopedia of Media Literacy.
- Larceneux, F., T. Lefebvre, and A. Simon. 2015. "What Added Value Do Estate Agents Offer Compared to FSBO Transactions? Explanation From a Perceived Advantages Model." *Journal of Housing Economics* 29: 72–82.
- Le Forestier, J. M., and N. A. Lewis Jr. 2024. "When and Why People Conceal Their Identities." *Nature Reviews Psychology* 3, no. 7: 489–498.
- Lee, B.-C., L. Ang, and C. Dubelaar. 2005. "Lemons on the Web: A Signalling Approach to the Problem of Trust in Internet Commerce." *Journal of Economic Psychology* 26, no. 5: 607–623. <https://doi.org/10.1016/j.joep.2005.01.001>.
- Levy, D., L. Murphy, and C. K. Lee. 2013. "Influences and Emotions: Exploring Family Decision-Making Processes When Buying a House." In *The Microstructures of Housing Markets*, 110–128. Routledge.
- Li, S., K. Srinivasan, and B. Sun. 2009. "Internet Auction Features as Quality Signals." *Journal of Marketing* 73, no. 1: 75–92.
- Liang, J., and Y. Xu. 2018. "Second-Hand Clothing Consumption: A Generational Cohort Analysis of the Chinese Market." *International Journal of Consumer Studies* 42, no. 1: 120–130.
- Lu, B., and Z. Chen. 2021. "Live Streaming Commerce and Consumers' Purchase Intention: An Uncertainty Reduction Perspective." *Information & Management* 58, no. 7: 103509.
- Ludwig, V. U., B. Berry, J. Y. Cai, N. M. Chen, D. L. Crone, and M. L. Platt. 2022. "The Impact of Disclosing Emotions on Ratings of Interpersonal Closeness, Warmth, Competence, and Leadership Ability." *Frontiers in Psychology* 13: 989826.
- Mayayise, T. O. 2024. "Investigating Factors Influencing Trust in C2C e – Commerce Environments: A Systematic Literature Review." *Data and Information Management* 8, no. 1: 100056.

- Miller, N., A. Sanchez, M. Sklarz, and A. Vamosiu. 2021. "Saving Real Estate Commissions at Any Price: Does Having a Real Estate Agent Influence the Sales Price of a Home?" *Journal of Housing Research* 30, no. 2: 175–206.
- Mohammad, S. M. 2017. "Word Affect Intensities." arXiv preprint arXiv: 1704.08798.
- Mohammad, S. M., and P. D. Turney. 2013. "Crowdsourcing a Word-Emotion Association Lexicon." *Computational Intelligence* 29, no. 3: 436–465.
- Moradi, M., M. Dass, D. Arnett, and V. Badrinarayanan. 2024. "The Time-Varying Effects of Rhetorical Signals in Crowdfunding Campaigns." *Journal of the Academy of Marketing Science* 52, no. 2: 370–398.
- Moradi, M., M. Dass, and P. Kumar. 2023. "Differential Effects of Analytical Versus Emotional Rhetorical Style on Review Helpfulness." *Journal of Business Research* 154: 113361. <https://doi.org/10.1016/j.jbusres.2022.113361>.
- Neal, T., K. Sundararajan, A. Fatima, Y. Yan, Y. Xiang, and D. Woodard. 2017. "Surveying Stylometry Techniques and Applications." *ACM Computing Surveys* 50, no. 6: 1–36.
- Nowak, A., and P. Smith. 2017. "Textual Analysis in Real Estate." *Journal of Applied Econometrics* 32, no. 4: 896–918.
- Packard, G., and J. Berger. 2021. "How Concrete Language Shapes Customer Satisfaction." *Journal of Consumer Research* 47: 787–806.
- Packard, G., and J. Berger. 2024. "The Emergence and Evolution of Consumer Language Research." *Journal of Consumer Research* 51, no. 1: 42–51.
- Palmon, O., and B. Sopranzetti. 2017. "On the Relationship Between the Number of a Broker's Real Estate Listings and Transaction Outcomes." *Review of Quantitative Finance and Accounting* 49, no. 1: 65–89.
- Pfeiffer, B. E., A. Sundar, and E. Cao. 2023. "The Influence of Language Style (Formal vs. Colloquial) on the Effectiveness of Charitable Appeals." *Psychology & Marketing* 40, no. 3: 542–553.
- Pfeuffer, A., and J. Phua. 2022. "Stranger Danger? Cue-Based Trust in Online Consumer Product Review Videos." *International Journal of Consumer Studies* 46, no. 3: 964–983.
- Pornpitakpan, C. 2004. "The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence." *Journal of Applied Social Psychology* 34, no. 2: 243–281.
- Pryce, G., and E. Oates. 2008. "Rhetoric in the Language of Real Estate Marketing, Housing Studies." *Housing Studies* 23, no. 2: 319–348.
- Reavey, B., M. Puzakova, A. Larsen, and K. Trina. 2018. "The Multidimensionality of Anthropomorphism in Advertising: The Moderating Roles of Cognitive Business and Assertive Language." *International Journal of Advertising* 37, no. 3: 440–462.
- Ren, G., and T. Hong. 2019. "Examining the Relationship Between Specific Negative Emotions and the Perceived Helpfulness of Online Reviews." *Information Processing & Management* 56, no. 4: 1425–1438.
- Rocklage, M. D., and R. H. Fazio. 2020. "The Enhancing Versus Backfiring Effects of Positive Emotion in Consumer Reviews." *Journal of Marketing Research* 57, no. 2: 332–352.
- Rutherford, R. C., T. M. Springer, and A. Yavas. 2005. "Conflicts Between Principals and Agents: Evidence From Residential Brokerage." *Journal of Financial Economics* 76, no. 3: 627–665.
- Savolainen, R. 2009. "The Information Needs of Prospective Homebuyers: An Exploratory Study of Apartment Purchases in Finland." *International Journal of Consumer Studies* 33, no. 5: 566–571.
- Schuster, T., R. Schuster, D. J. Shah, and R. Barzilay. 2020. "The Limitations of Stylometry for Detecting Machine-Generated Fake News." *Computational Linguistics* 46, no. 2: 499–510.
- Schwob, A., R. de Kervenoael, V. Kirova, and T. Vo-Thanh. 2023. "Casual Selling Practice: A Qualitative Study of Non-Professional Sellers' Involvement on C2C Social Commerce Platforms." *Information Technology & People* 36, no. 2: 940–965.
- Sela, A., S. C. Wheeler, and G. Sarial-Abi. 2012. "We Are Not the Same as You and I: Causal Effects of Minor Language Variations on Consumers' Attitudes Toward Brands." *Journal of Consumer Research* 39, no. 3: 644–661.
- Shankar, V., and S. Parsana. 2022. "An Overview and Empirical Comparison of Natural Language Processing (NLP) Models and an Introduction to and Empirical Application of Autoencoder Models in Marketing." *Journal of the Academy of Marketing Science* 50, no. 6: 1324–1350.
- Shen, C., J. Chiou, and B. Kuo. 2011. "Remedies for Information Asymmetry in Online Transaction: An Investigation Into the Impact of Web Page Signals on Auction Outcome." *Internet Research* 21, no. 2: 154–170.
- Smaldino, P. E. 2022. "Models of Identity Signaling." *Current Directions in Psychological Science* 31, no. 3: 231–237.
- Solomon, G. E. A. 1990. "Psychology of Novice and Expert Wine Talk." *American Journal of Psychology* 103: 495–517.
- Spence, M. 1973. "Job Market Signaling." *Quarterly Journal of Economics* 87, no. 3: 355–374.
- Stamatatos, E. 2009. "A Survey of Modern Authorship Attribution Methods." *Journal of the American Society for Information Science and Technology* 60, no. 3: 538–556.
- Stamsø, M. A. 2015. "Selling a House and the Decision to Use a Real Estate Broker in Norway." *Property Management* 33, no. 2: 173–186.
- Steigenberger, N., and H. Wilhelm. 2018. "Extending Signaling Theory to Rhetorical Signals: Evidence From Crowdfunding." *Organization Science* 29, no. 3: 529–546.
- Sturman, M. C. 2003. "Searching for the Inverted U – Shaped Relationship Between Time and Performance: Meta-Analyses of the Experience/Performance, Tenure/Performance, and Age/Performance Relationships." *Journal of Management* 29, no. 5: 609–640.
- Su, L., L. Gong, and Y. Huang. 2025. "Humorous or Serious? The Interaction Effect of Language Style and Tourism Activity Type on Tourist Well-Being." *Journal of Sustainable Tourism* 33, no. 3: 1–28.
- Swan, J. E., and J. J. Nolan. 1985. "Gaining Customer Trust: A Conceptual Guide for the Salesperson." *Journal of Personal Selling & Sales Management* 5, no. 2: 39–48.
- Tiedens, L. Z. 2001. "Anger and Advancement Versus Sadness and Subjugation: The Effect of Negative Emotion Expressions on Social Status Conferal." *Journal of Personality and Social Psychology* 80, no. 1: 86–94.
- Tong, Z., R. Yu, and H. Xiao. 2023. "How Should Cities Communicate? The Interaction Effect of City Stereotypes and Advertising Language on Travel Intention." *Journal of Destination Marketing & Management* 27: 100755.
- Tóth, Z., M. Mrad, O. S. Itani, J. Luo, and M. J. Liu. 2022. "B2B eWOM on Alibaba: Signaling Through Online Reviews in Platform-Based Social Exchange." *Industrial Marketing Management* 104: 226–240.
- Treiblmaier, H., and M. Garaus. 2023. "Using Blockchain to Signal Quality in the Food Supply Chain: The Impact on Consumer Purchase Intentions and the Moderating Effect of Brand Familiarity." *International Journal of Information Management* 68: 102514.

- Tsui, H. C. 2012. "Advertising, Quality, and Willingness-To-Pay: Experimental Examination of Signaling Theory." *Journal of Economic Psychology* 33, no. 6: 1193–1203.
- Valtakoski, A. 2015. "Initiation of Buyer–Seller Relationships: The Impact of Intangibility, Trust and Mitigation Strategies." *Industrial Marketing Management* 44: 107–118.
- Verbeke, W., B. Dietz, and E. Verwaal. 2011. "Drivers of Sales Performance: A Contemporary Meta-Analysis. Have Salespeople Become Knowledge Brokers?" *Journal of the Academy of Marketing Science* 39: 407–428.
- Vysotska, V., V. Lytvyn, M. Hrendus, S. Kubinska, and O. Brodyak. 2018. "Method of Textual Information Authorship Analysis Based on Stylometry." In *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, 2, 9–16, IEEE.
- Wells, J. D., J. S. Valacich, and T. J. Hess. 2011. "What Signal Are You Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions." *MIS Quarterly* 35, no. 2: 373–396.
- Witte, K., and M. Allen. 2000. "A Meta-Analysis of Fear Appeals: Implications for Effective Public Health Campaigns." *Health Education & Behavior* 27, no. 5: 591–615.
- Wood, J. A., J. S. Boles, W. Johnston, and D. Bellenger. 2008. "Buyers' Trust of the Salesperson: An Item-Level Meta-Analysis." *Journal of Personal Selling & Sales Management* 28, no. 3: 263–283.
- Wright, D. 2017. "Using Word n-Grams to Identify Authors and Idiolects: A Corpus Approach to a Forensic Linguistic Problem." *International Journal of Corpus Linguistics* 22, no. 2: 212–241.
- Young, L., and G. Albaum. 2003. "Measurement of Trust in Salesperson–Customer Relationships in Direct Selling." *Journal of Personal Selling & Sales Management* 23, no. 3: 253–269.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Appendix S1:** Supporting Information.

Appendix 1

Let a corpus C consist of k documents D_j , such that $\sum_{j=1}^k D_j = C$. Each document D_j is a sequence of words reduced to their dictionary form (lemmatization). We define an n -gram n_i as a contiguous sequence of n words within C . The corpus is partitioned into l topics T_l , where each T_l is a subset of documents from C and $\bigcup_{l=1}^l T_l = C$.

The probability of an n -gram occurring in the corpus is given by $P(n_i) = \frac{\sum_{j=1}^k 1(n_i \in D_j)}{k}$, where 1 is the indicator function, equal to 1 if n_i is present in D_j and 0 otherwise.

The probability of n_i within a specific topic T_l is $P_{T_l}(n_i) = \frac{\sum_{j=1}^{k(l)} 1(n_i \in T_l)}{k(l)}$, where $k(l)$ is the number of documents in topic T_l .

An n -gram n_i is considered to be a prevalent n -gram of topic T_l if for every other topic T_m , $m \neq l$ the difference in probabilities $P_{T_l}(n_i) - P_{T_m}(n_i) > \theta$ for every $m \neq l$ where θ is a threshold to be determined empirically. See suggested algorithm for determining optimal value for θ in Appendix 2.

In our corpus, we identified 2,857,358 unique bigrams and following the steps of the algorithm detailed in Appendix 2 empirical value for the threshold was set to be 0.00003. Using this value, we found 128,249 amateur prevalent bigrams and 186,334 professional prevalent bigrams.

```

1. Procedure FindOptimalThreshold(N, T, ε)
2. θ_opt ← 0
3. max_accuracy ← 0
4. For θ in 0 to 1 step ε:
5. For each topic Tl in T:
6. N_prevalent ← {ni ∈ N | P_Tl(ni) > θ}
7. accuracy_Tl ← Classify(Tl, N_prevalent)
8. If accuracy_Tl > max_accuracy
9. max_accuracy ← accuracy_Tl
10. θ_opt ← θ
11. Return θ_opt

```

Formal definition of topic prevalent n -grams and proposed algorithm for finding optimal threshold for classification accuracy. *Source:* Own elaboration.

Appendix 2

Type	Description	Generation method
Part-of-speech tags	Serves as a syntactic style marker, identifying 29 different types such as verbs, nouns, and adjectives	Generated using the pos_tag function of the NLTK tag package (Bird et al. 2009)
Sentiment	Quantifies sentiment scores to measure the polarity (positive, negative, or neutral) and intensity (ranging from −1 to 1) of emotions in the text	Uses NLTK's sentiment vader (Hutto and Gilbert 2014) to summarize text and analyze variance in scores.
ngrams	Acts as a lexical style marker, with bigrams representing a contiguous sequence of two words from the dataset and trigrams representing a sequence of three words	(Custom Python code)
Bag of words	Another lexical style indicator, focusing on the occurrence of the 1000 most frequent words within the listings	(Custom Python code)
Vocabulary richness	Measures the number of unique words per listing, indicating lexical diversity	(Custom Python code)

Note: Linguistic features and generation method.