A Study of Manipulative and Authentic Negative Reviews

Snehasish Banerjee
Wee Kim Wee School of Communication & Information,
Nanyang Technological University.
31 Nanyang Link, WKWSCI Building,
Singapore 637718.
(65) 67905772
snehasis002@e.ntu.edu.sg

Alton Y. K. Chua
Wee Kim Wee School of Communication & Information,
Nanyang Technological University.
31 Nanyang Link, WKWSCI Building,
Singapore 637718.
(65) 67905810
altonchua@ntu.edu.sg

ABSTRACT
Given users' growing penchant to use online reviews for travel planning, the business malpractice of posting manipulative reviews to distort the reputation of hotels is on the rise. Some manipulative reviews could be positive and intended to boost own offerings, while others could be negative and meant to slander competing ones. However, most scholarly inquiry hitherto has been trained on the former. Hence, this paper investigates the extent to which linguistic cues such as readability, genre and writing style of negative reviews could help predict if they are manipulative or authentic. Analysis of a publicly available dataset of 800 negative reviews (400 manipulative + 400 authentic) indicates that manipulative reviews are generally less readable than authentic reviews. In terms of genre, although manipulative reviews should be imaginative and authentic reviews informative, spammers appear adept enough to blur the line between the two. With respect to writing style, manipulative reviews are more richly embellished with affective cues and perceptual words.

Categories and Subject Descriptors
H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – l*inguistic Processing*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *information filtering*.

General Terms
Management, Measurement, Reliability, Verification.

Keywords
Negative opinion spam, manipulative reviews, authentic reviews, linguistic analysis, logistic regression.

1. INTRODUCTION
The collaborative nature of the Internet kindled by the emergence of social media in recent years has transformed the manner in which information on products and services is being disseminated among users. Instead of being fed with information provided by advertisers and marketers, users now have access to dime a dozen opinions contributed by their online peers through what is known as user-generated content. A popular form of user-generated content that has grown exponentially in recent years includes online reviews of hotels.

Prior to making a booking, most users are inclined to browse through reviews posted by others to procure guidance on hotels [2]. Preponderance of positive reviews for a hotel creates an affirmative impression among users, which in turn may persuade them to book a room in the hotel. On the other hand, ubiquity of negative reviews for a hotel results in an inept impression, which in turn may dissuade users to stay in the hotel. Users' growing penchant to harness reviews for travel planning has however resulted as its bye-product a business malpractice known as opinion spamming [19].

Opinion spamming involves infiltrating review websites with two types of misleading reviews, namely, disruptive and manipulative [27]. The former refers to reviews that are frivolous and contain irrelevant texts that are obviously identifiable. The latter includes reviews that are maliciously written to appear authentic, and hence not easily ascribable as spam. Between the two, manipulative opinion spamming has greater repercussions because it can bamboozle users by influencing their impression towards hotels insidiously. Some hotel organizations have fallen into the temptation of manipulative opinion spamming in order to augur financial gains from the distorted reputation of hotels [1].

As contributing manipulative reviews to hoodwink users is en route towards an established business malpractice [17], scholarly interest in this theme is also on the rise. Yet two research gaps can be identified in the extant literature. First, not much scholarly inquiry has delved into identifying the predictors of review manipulation. Although manipulative and authentic reviews may not be easily distinguishable from each other, subtle linguistic nuances could set them apart. In particular, there could be telltale signs in terms of readability, genre and writing style of reviews that could help predict if they are manipulative or authentic [17], [22], [27]. Hence, shedding light on such linguistic nuances between manipulative and authentic reviews could be a significant research undertaking.

Second, most research thus far has been trained on positive manipulative opinion spamming. However, prior research shows the existence of a negativity bias such that negative reviews could have a greater impact on users' decision making vis-à-vis positive reviews [10]. Users generally tend to perceive the former as being more diagnostic than the latter [39]. To exploit such a trend, hotel organizations could conceivably be more inclined to contribute

IMCOM (ICUIMC)'14, January 9-11, 2014, Siem Reap, Cambodia.

manipulative negative reviews on their rival hotels than to post manipulative positive reviews on their own offerings. Hence, analyzing manipulative opinion spamming for negative reviews could be a timely research endeavor.

To address the two research gaps, this paper conducts a linguistic analysis of manipulative and authentic negative reviews. Specifically, it seeks to predict manipulative reviews based on three aspects, namely, readability, genre and writing style. First, readability of a review refers to the effort and expertise required on the part of readers to comprehend its meaning [41]. Differences in readability of reviews tend to influence the ways they are received by users [21]. Second, with respect to genre, manipulative reviews are considered imaginative while authentic reviews are termed informative [27]. Texts of imaginative and informative genres differ considerably in their distribution of part-of-speech (POS) tags [32]. Third, writing style indicates ways users use specific types of words to construct sentences in their reviews [17]. In this context, manipulative and authentic reviews could differ in their use of affective cues, perceptual words and future tense [6], [22], [40].

This paper is significant for both theory and practice. On the theoretical front, even though much scholarly attention has delved into the classification of manipulative and authentic reviews using approaches such as Support Vector Machine or Naïve Bayes (eg. [27], [43]), little is known hitherto on the linguistic differences between them. On the practical front, moderators of review websites as well as users might lean on the findings of this paper to conjecture which reviews are likely to be manipulative, and which are likely to be authentic. This in turn could help users in more informed decision making, promoting a healthy cyber culture.

The remainder of the paper is structured as follows. The next section reviews some related works on spam and elaborates on readability, genre and writing style of reviews. The Methods section describes the dataset as well as presents the analysis procedures. This is followed by the results. The main findings gleaned from the results are discussed next. Finally, the paper concludes with notes on limitations and future research.

2. LITERATURE REVIEW

In extant literature, the study of spam activities has generally been done in the domains of web spam and email spam [3], [4]. Web spam typically can be of two types, namely, link spam and content spam. Link spam refers to spam on hyperlinks, while content spam involves adding irrelevant words in pages to trick search engine crawlers. Email spam, on the other hand, refers to unsolicited commercial advertisements in the form of emails. However, opinion spam in the form of manipulative reviews differs from all of these [24], and can arguably be one of the most difficult types to detect [19]. It is therefore no wonder that manipulative opinion spam has increasingly attracted scholarly interest.

Much of the scholarly inquiries have treated manipulative opinion spam as a classification problem. The objective has been to develop a classifier that could weed out manipulative reviews from the authentic ones. For example, in a study that attempted to detect duplicate manipulative reviews from the review website Amazon.com [19], classification was done using a list of review-centric, reviewer-centric and product-centric features. In another similar study [27], manipulative and authentic reviews were classified based on textual features such as n-grams, psycholinguistic dimensions and POS tags. Yet another study attempted to use statistical language models to weed out manipulative opinion spam [43]. Despite these ongoing scholarly efforts however, the extant literature lacks a framework of linguistic cues that could help predict if reviews are manipulative or authentic. Since readability, genre and writing style could predict review manipulation [11], [17], [18], [40], it could be a significant and timely research endeavor to study how manipulative and authentic reviews differ with respect to these dimensions.

2.1 Review Readability

Readability of a review is a measure of the amount of effort and expertise required by users to comprehend its meaning [41]. Since manipulative and authentic reviews are written with different intentions, the varying purposes could be reflected in terms of readability [17], [21].

However, there exists a conundrum in this regard. On one hand, authors of opinion spam (henceforth, spammers) might post manipulative reviews using very lucid language to make them more readable. Reviews that are more readable would certainly enhance comprehension, retention, and reading speed [9]. Manipulative reviews that are easily readable to majority of the online communities could be more influential in distorting reputation of hotels.

On the other hand, it is also possible for spammers to post manipulative reviews using very sophisticated language. This is because when users browse reviews, they not only go through their content but also try to gauge the competence and intelligence of users who contribute them [5], [39]. Too simplistic reviews might suggest incompetence of the contributors in writing sophisticated reviews. Given users' reluctance to follow reviews posted by incompetent contributors [30], manipulative reviews that are too simplistic could be less influential in distorting reputation of hotels. Since readability could affect the size of a review's audience [17], it is interesting to study the extent to which it could predict manipulation of reviews.

2.2 Review Genre

Writing manipulative reviews necessitates articulating events that did not take place in reality or emotions that did not exist in actuality [26], that too, in an apparently plausible and sincere manner [8]. However, texts that are written based on real experiences qualitatively differ from those based on imagined experiences [36]. Generally, there exist four genres of text, namely, conversational, task-oriented, informative and imaginative [32]. Among these, manipulative reviews are considered imaginative while authentic reviews are termed informative [27].

In particular, manipulative reviews that are supposedly imaginative tend to have different underlying POS distribution patterns compared to those of authentic reviews that are seemingly imaginative [27], [32]. Prior research has demonstrated that imaginative texts tend to include more adverbs, verbs and pronouns [25], [32]. Among pronouns, personal pronouns in the form of self references tend to be especially higher in imaginative texts than in informative texts [40]. On the other hand, informative texts tend to be richly laden with adjectives, articles, nouns and prepositions [25], [32]. It is conceivable that such nuances in POS distribution could also be observed between manipulative and authentic negative reviews.

2.3 Review Writing Style

Writing style reflects the ways users use specific types of words to construct sentences in reviews to reflect their opinions [17]. For the purpose of this paper, writing style of manipulative and authentic reviews is considered in terms of their use of affective cues, perceptual words and future tense [40], [22], [6].

First, in order to create a lasting impact, manipulative reviews could be richly embellished with exaggerated affective cues [23]. It is conceivable that manipulative reviews written by spammers to slander competing hotels could be replete with negative cues [26], [40], [42]. Second, users' physical experiences with hotel services are affected by their sensory perceptions [33]. For example, users' impression on a hotel could be a function of visual cues such as artwork and aural cues such as music [22]. These cues could be reflected through the use of perceptual words in reviews [15]. Third, given that ubiquity of negative reviews could adversely affect future sales and revenues of a given hotel [6], [12], spammers might write manipulative reviews for competing hotels not only to describe poor past experiences in the hotel, but also to express future desires of avoiding the same. As a form of exaggeration, manipulative reviews could be more fraught with negative affective cues, perceptual words and future tense vis-à-vis authentic reviews. Such nuances among reviews might help predict if they are manipulative or authentic.

3. METHODS

3.1 Dataset

One of the major bottlenecks that impede scholarly inquiry on manipulative and authentic reviews is the dearth of ground truth [19], [13], [38]. In fact, absence of dataset with ground truth has led numerous scholars to employ alternative heuristic annotation approaches to label datasets (eg. [17], [19]). Even though such approaches could be intuitive and relied upon to offer substantial approximation occasionally, they lack a compelling thrust.

Given the importance of ground truth, this paper uses the publicly available negative opinion spam dataset of 800 reviews [28]. The dataset comprises 400 manipulative reviews and 400 authentic reviews equally distributed across 20 popular hotels in Chicago.

3.2 Operationalization

Review readability is operationalized based on (1) average words per sentences (WPS), (2) Gunning-Fog Readability Index (FOG), (3) Coleman-Liau Readability Index (CLI), (4) Automated-Readability Index (ARI), and (5) Flesch-Kincaid Grade Level (FKG). Lower the number of words per sentence in a given review, higher is its readability [7], [29]. Lower values in the remaining four readability metrics suggest a more readable review. These have been widely used by the scholarly community to operationalize readability of text (eg. [9], [17], [21]).

Review genre is operationalized on the basis of the POS distributions of manipulative and authentic reviews. Specifically, the following are considered, (1) pronouns, (2) personal pronouns, (3) verbs, (4) adverbs, (5) nouns, (6) adjectives, (7) articles, and (8) prepositions. The first four could be higher in manipulative reviews while the next four could be higher in authentic reviews [25], [32], [40]. However, POS tags such as conjunctions and auxiliary verbs were not admitted for analysis due to lack of literature support. For measuring the POS tags, Stanford Parser's POS tagger was utilized [20].

Writing style of a given review is operationalized as the proportion of (1) negative cues, (2) perceptual words, and (3) future tense in reviews. These were computed using the Linguistic Inquiry and Word Count (LIWC) algorithm [31], a popular automated text analysis tool. The applicability of LIWC has been demonstrated in numerous studies on manipulation and deception (eg. [16], [37]).

3.3 Data Analysis

This paper includes 16 independent variables (IVs) for analysis, the five readability indicators, the eight POS tags and the three writing style metrics. The values for all the IVs were standardized by converting percentages into z-scores [26]. On the other hand, the dependent variable comprises review manipulation. Since the objective of this paper is to predict manipulative reviews from the authentic ones, the dependent variable was dummy-coded such that 1 (0) indicates manipulative (authentic) reviews.

Given the dichotomous nature of the dependent variable, binomial logistic regression was used for data analysis [14]. It converts the dependent variable into its logit equivalent and applies maximum likelihood estimation. The coefficients of logistic regression estimate the odds ratio for the IVs in the model to examine the extent to which those could predict manipulative reviews.

4. RESULTS

The descriptive statistics (Table 1) based on the 16 IVs suggest that manipulative reviews were generally less readable than authentic reviews. The former was more verbose with higher values of readability indices on an average. Based on POS tags, manipulative reviews were richer in pronouns, personal pronouns, verbs, adverbs and prepositions, while authentic reviews had more nouns, adjectives and articles. Moreover, manipulative reviews were more richly embellished with negative cues, perceptual words and future tense compared to authentic reviews.

For the logistic regression, result of the Omnibus test indicates acceptable performance of the model ($\chi^2 = 202.55$; df = 16; - 2 log likelihood = 906.48; $p < 0.001$). To further ascertain performance of the model, the more stringent Hosmer-Lemeshow goodness-of-fit test was also performed. A non-significant result ($\chi^2 = 3.89$; df = 8; $p = 0.87$) suggests that the model prediction is not significantly different from the observed values, thereby confirming that the model fits well with the data. Although no specific counterpart of R^2 in multiple regression is available in logistic regression, two commonly reported pseudo-R^2 statistics include the Cox and Snell R^2 and the Nagelkerke R^2. For the model, Cox and Snell R^2 was 0.22 and Nagelkerke R^2 was 0.30. This suggests that around 22 % to 30 % of the variability in review manipulation could be explained by the model.

Table 2 summarizes the extent to which the 16 IVs could predict if reviews were manipulative or authentic. In terms of review readability, WPS, CLI and ARI emerged as significant predictors. In fact, the descriptive statistics of all readability indicators suggest that manipulative reviews were invariably less readable than authentic ones. With respect to review genre, the POS tags that turned out to be significant predictors of review manipulation include personal pronouns, verbs, nouns, articles and prepositions. In terms of writing style, results indicate that the use of negative affective cues and perceptual words could significantly predict if reviews were manipulative or authentic.

be adept enough to write manipulative reviews in a manner so as to blur the line between imaginative and informative texts.

Table 1. Descriptive statistics of the dataset

	IVs	Manipulative (Mean ± SD)	Authentic (Mean ± SD)
Readability	WPS	16.97 ± 4.32	14.81 ± 7.12
	FOG	10.94 ± 2.20	9.90 ± 3.18
	CLI	7.34 ± 1.49	6.95 ± 1.82
	ARI	7.07 ± 2.45	6.08 ± 3.75
	FKG	7.15 ± 2.05	6.30 ± 3.01
Genre	Pronoun	12.96 ± 3.47	11.58 ± 3.64
	Pers. pronoun	8.82 ± 3.06	7.50 ± 3.17
	Verb	14.23 ± 2.69	13.35 ± 2.87
	Adverb	5.00 ± 2.02	4.84 ± 2.18
	Noun	23.00 ± 3.31	25.13 ± 4.22
	Adjective	7.89 ± 2.57	8.34 ± 2.75
	Article	9.37 ± 2.15	9.56 ± 2.50
	Preposition	13.17 ± 2.63	12.74 ± 2.70
Writing Style	Neg. cues	1.88 ± 1.37	1.63 ± 1.18
	Perceptual	2.25 ± 1.47	1.83 ± 1.44
	Future	0.96 ± 0.75	0.85 ± 0.79

Table 2. Predictors of review manipulation

	IVs	β	SE	Wald	Exp(β)
Readability	WPS	0.80	0.18	19.55	2.23***
	FOG	0.14	0.12	1.54	1.16
	CLI	1.64	0.34	22.89	5.16***
	ARI	-2.14	0.46	21.84	0.12***
	FKG	0.28	0.23	1.55	1.33
Genre	Pronoun	-0.02	0.05	0.13	0.98
	Pers. pronoun	0.17	0.05	9.64	1.18**
	Verb	0.14	0.04	13.44	1.15***
	Adverb	0.04	0.04	0.75	1.74
	Noun	-0.07	0.03	4.98	0.94*
	Adjective	0.02	0.04	0.36	1.02
	Article	0.08	0.04	3.60	1.09*
	Preposition	0.09	0.04	5.88	1.10*
Writing Style	Neg. cues	0.18	0.07	7.60	1.20**
	Perceptual	0.28	0.06	21.55	1.32***
	Future	0.03	0.11	0.05	1.03

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.01$

5. DISCUSSION

Based on the results shown in Table 2, three categories of IVs in predicting review manipulation can be identified. Specifically, they include (1) positively associated IVs, (2) negatively associated IVs, and (3) unassociated IVs.

The positively associated IVs in decreasing order of strength are CLI, WPS, perceptual words, negative cues, personal pronouns, verbs, preposition and articles. The readability indicators of CLI and WPS generally tend to be higher for manipulative reviews than for authentic reviews. This suggests that readability of reviews could be an important predictor of review manipulation [17]. Furthermore, manipulative reviews appeared to be more richly laden with perceptual words and negative cues. This could be vestige of the hyperbolic tone that spammers use in manipulative reviews to bamboozle users' impression about hotels [22], [23]. The dominance of personal pronouns in manipulative reviews reveals the lack of guilt among spammers. Apparently unperturbed by any ethical constraints, they perhaps do not bother to dissociate themselves from their manipulative comments about hotels [34], [35], [40]. Consistent with prior research such as [25] and [32], manipulative reviews contained more verbs compared to authentic reviews. However, contrary to the literature, the former's richness with respect to prepositions and articles surpassed those of the latter. This suggests that spammers tend to

The negatively associated IVs in decreasing order of strength are ARI and nouns. It is interesting to note that the readability indicator of ARI appears to be contradictory to that of CLI. This could be attributed to the particularity of CLI as readability metric. It is no wonder that prior studies have often advocated the concurrent use of multiple metrics to avoid unique idiosyncratic errors [9]. Further scholarly investigation could be useful to confirm if such a difference between the readability metrics is generalizable. The richness of reviews in nouns as an indication of authentic reviews is however expected from prior research [25], [32].

The unassociated IVS include FOG, FKG, pronouns, adverbs, adjectives and future tense. The non-significance of FOG and FKG with respect to the dependent variable could be attributed to the uniqueness of these readability metrics. Moreover, no significant difference in the use of pronouns, adverbs and adjectives between manipulative and authentic reviews further affirms conscious effort on the part of spammers to blur the line between imaginative and informative texts. Interestingly, the non-significance of future tense indicates that spammers generally do not use substantial future tense in their manipulative reviews to deceive users. Perhaps they are still unaware of this potentially subtle way to influence users' perceptions insidiously.

6. CONCLUSION

Given that manipulative opinion spamming is fast becoming a well-established industrial malpractice, this paper conducted a linguistic analysis of manipulative and authentic reviews. In particular, it investigated the extent to which linguistic differences in terms of readability, genre and writing style between manipulative and authentic reviews could help predict review manipulation. Drawing from a publicly available dataset of 800 negative reviews (400 manipulative + 400 authentic), results indicate that manipulative reviews are generally less readable than authentic reviews. In terms of genre, although manipulative reviews should be imaginative and authentic reviews informative, spammers appear adept enough to blur the line between the two. With respect to writing style, manipulative reviews are more richly embellished with affective cues and perceptual words.

The findings from this paper offer implications for both theory and practice. On the theoretical front, it represents one of the earliest attempts to conduct a linguistic analysis of manipulative and authentic negative reviews. Even though much scholarly attention has delved into the classification of such reviews using approaches such as support vector machine or Naïve Bayes (eg. [27], [43]), little is known hitherto on the specific linguistic differences between them. Hence, this paper represents a modest attempt to develop a framework that could help distinguish between manipulative and authentic negative reviews.

On the practical front, users may lean on the findings of this paper to conjecture which reviews are likely to be manipulative and which are likely to be authentic. This in turn can prevent users from being victims of manipulative opinion spamming, thereby aiding more informed decision making. Moderators of review websites may also tap into the findings of this paper to automatically recommend reviews that are potentially authentic and weed out those that are likely to be manipulative, thereby promoting a healthy cyber culture.

However, the findings of the paper are constrained by the secondary dataset used for analysis [28]. The dataset comprises reviews only for popular hotels in Chicago. This may reduce the generalizability of the findings. Nevertheless, this paper offers a few potential directions for future research. A possible direction of investigation could include analyzing the extent to which linguistic differences could predict manipulation in reviews posted for lesser known budget hotels. Yet another research direction might involve comparing negative manipulative reviews with positive manipulative reviews. Furthermore, given that this study was limited to reviews for hotel services, future research should consider investigating if such linguistic patterns could also be detected in manipulative and authentic reviews contributed for products. Future research might also investigate the extent to which manipulative reviews are able to influence users' purchase intentions. Such studies are necessary to extend the line of research from linguistic analysis to users' decision making.

7. REFERENCES

[1] Allen, E. 2012. Dear staff. We could do with some positive comments: Hotel boss is caught telling his workers to post fake reviews on TripAdvisor. *MailOnline* (2012, October 9). Retrieved February 27, 2013 from http://www.dailymail.co.uk/news/article-2214974/Hotel-boss-caught-telling-workers-post-fake-reviews-TripAdvisor.html

[2] Bart Y, Shankar V, Sultan F, and Urban, G. L. 2005. Are the drivers and role of online trust the same for all web sites and consumers? A large-scale exploratory empirical study. *Journal of Marketing* 69, 4 (Oct. 2005), 133–152. DOI= http://dx.doi.org/10.1509/jmkg.2005.69.4.133

[3] Castillo, C., Donato, D., Becchetti, L., Boldi, P., Leonardi, S., Santini, M., and Vigna, S. 2006. A reference collection for web spam. *SIGIR Forum* 40, 2 (Dec. 2006), 11–24. DOI= http://dx.doi.org/10.1145/1189702.1189703

[4] Chirita, P.A., Diederich, J., and Nejdl, W. 2005. MailRank: Using ranking for spam detection. In *Proceedings of the ACM International Conference on Information and Knowledge Management* (Oct. 2005). ACM, 373-380. DOI= http://dx.doi.org/10.1145/1099554.1099671

[5] Connors, L., Mudambi, S. M., and Schuff, D. 2011. Is it the review or the reviewer? A multi-method approach to determine the antecedents of online review helpfulness. In *Proceedings of the Hawaii International Conference on System Sciences* (Jan. 2011). IEEE, 1-10. DOI= http://dx.doi.org/10.1109/HICSS.2011.260

[6] Duan, W., Gu, B., and Whinston, A.B. 2008. Do online reviews matter? An empirical investigation of panel data. *Decision Support Systems* 45, 4 (Nov. 2008), 1007-1016. DOI= http://dx.doi.org/10.1016/j.dss.2008.04.001

[7] Foltz, P. W., Laham, D., and Landauer, T. K. 1999. Automated essay scoring: Applications to educational technology. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 939-944.

[8] Friedman, H. S., and Tucker, J. S. 1990. Language and deception. In H. Giles, and W. P. Robinson, Eds., *Handbook of Language and Social Psychology*. John Wiley.

[9] Ghose, A., and Ipeirotis, P. G. 2011. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering* 23, 10 (Oct. 2011), 1498-1512. DOI= http://dx.doi.org/10.1109/TKDE.2010.188

[10] Gibson, B., and Oberlander, E. 2008. Wanting to appear smart: Hypercriticism as an indirect impression management strategy. *Self and Identity* 7, 4 (Sep. 2008), 380-392. DOI= http://dx.doi.org/10.1080/15298860701668986

[11] Giles, H. 2009. The process of communication accommodation. In N. Coupland, and A. Jaworksi, Eds., *The New Reader in Sociolinguistics*, 276-286. Macmillan.

[12] Godes, D., and Mayzlin, D. 2004. Using online conversations to study word-of-mouth communication. *Marketing Science* 23, 4 (Fall 2004), 545-560. DOI= http://dx.doi.org/10.1287/mksc.1040.0071

[13] Gokhman, S., Hancock, J., Prabhu, P., Ott, M., and Cardie, C. 2012. In search of a gold standard in studies of deception. In *Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection*, 23–30.

[14] Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C. 1998. *Multivariate Data Analysis*, 5th Edition. Prentice Hall.

[15] Hall, W. S., Scholnick, E. K., and Hughes, A. T. 1987. Contextual constraints on usage of cognitive words. *Journal of Psycholinguistic Research* 16, 4 (Jul. 1987), 289-310. DOI= http://dx.doi.org/10.1007/BF01069284

[16] Hancock, J. T., Curry, L. E., Goorha, S., and Woodworth, M. 2008. On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes* 45, 1 (Mar. 2008), 1-23. DOI= http://dx.doi.org/10.1080/01638530701739181

[17] Hu, N., Bose, I., Koh, N. S., and Liu, L. 2012. Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems* 52, 3 (Feb. 2012), 674-684. DOI= http://dx.doi.org/10.1016/j.dss.2011.11.002

[18] Ireland, M. E., and Pennebaker, J. W. 2010. Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology* 99, 3 (Sep. 2010), 549-571. DOI= http://dx.doi.org/10.1037/a0020386

[19] Jindal, N., and Liu, B. 2008. Opinion spam and analysis. In *Proceedings of the ACM International Conference on Web Search and Web Data Mining* (Feb. 2008). ACM, 219-230. DOI= http://dx.doi.org/10.1145/1341531.1341560

[20] Klein, D., and Manning, C. D. 2003. Accurate unlexicalized parsing. In *Proceedings of the Association for Computational Linguistics*, 423-430.

[21] Korfiatis, N., García-Bariocanal, E., and Sánchez-Alonso, S. 2012. Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications* 11, 3 (May-Jun. 2012), 205-217. DOI= http://dx.doi.org/10.1016/j.elerap.2011.10.003

[22] Lin, I. Y. 2004. Evaluating a servicescape: The effect of cognition and emotion. *International Journal of Hospitality Management* 23, 2 (Jun. 2004), 163-178. DOI= http://dx.doi.org/10.1016/j.ijhm.2003.01.001

[23] Maurer, C., and Schaich, S. 2011. Online customer reviews used as complaint management tool. In *Information and Communication Technologies in Tourism*, 499-511. Springer.

[24] Mukherjee, A., Liu, B., and Glance, N. 2012. Spotting fake reviewer groups in consumer reviews. In *Proceedings of the ACM International Conference on World Wide Web* (Apr. 2012). ACM, 191-200. DOI= http://dx.doi.org/10.1145/2187836.2187863

[25] Nakamura, J. 1991. The relationship among genres in the LOB corpus based upon the distribution of grammatical tags. *JACET Bulletin* 22, 44-74.

[26] Newman, M. L., Pennebaker, J. W., Berry, D. S., and Richards, J. M. 2003. Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin* 29, 5 (May 2003), 665-675. DOI= http://dx.doi.org/10.1177/0146167203029005010

[27] Ott, M., Choi, Y., Cardie, C., and Hancock, J.T. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the Association for Computational Linguistics: Human Language Technologies*, 309-319.

[28] Ott, M., Cardie, C., and Hancock, J. T. 2013. Negative Deceptive Opinion Spam. In *Proceedings of North American Association for Computational Linguistics*, 497-501.

[29] Otterbacher, J. 2009. 'Helpfulness' in online communities: A measure of message quality. In *Proceedings of the ACM International Conference on Human factors in Computing Systems* (Apr. 2009). ACM, 955-964. DOI=http://dx.doi.org/10.1145/1518701.1518848

[30] Pavlou, P.A., and Dimoka, A. 2006. The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research* 17, 4 (Dec. 2006), 392–414. DOI= http://dx.doi.org/10.1287/isre.1060.0106

[31] Pennebaker, J. W., Chung, C.K., Ireland, M., Gonzales, A., and Booth, R. J. 2007. The development and psychometric properties of LIWC 2007. Austin, TX: LIWC.net.

[32] Rayson, P., Wilson, A., and Leech, G. 2001. Grammatical word class variation within the British National Corpus sampler. *Language and Computers* 36, 1, 295-306.

[33] Schiffman, H. R. 2001. *Sensation and Perception*, 5th Edition. Wiley.

[34] Vartapetiance, A., and Gillam, L. 2012. I don't know where he is not: Does deception research yet offer a basis for deception detectives? In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, 5-14.

[35] Vrij, A. 2000. *Detecting lies and deceit: The psychology of lying and its implications for professional practice*. John Wiley.

[36] Vrij, A., Edward, K., Roberts, K. P., & Bull, R. 2000. Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior* 24, 239-264.

[37] Vrij, A., Mann, S., Kristen, S., and Fisher, R.P. 2007. Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior* 31, 5 (Oct. 2007), 499-518. DOI= http://dx.doi.org/10.1007/s10979-006-9066-4

[38] Wu, G., Greene, D., and Cunningham, P. 2010. Merging multiple criteria to identify suspicious reviews. In *Proceedings of the ACM Conference on Recommender Systems*. ACM, 241-244. DOI= http://dx.doi.org/10.1145/1864708.1864757

[39] Wu, P.F., van der Heijden, H., and Korfiatis, N. 2011. The influences of negativity and review quality on the helpfulness of online reviews. In *Proceedings of the International Conference on Information Systems*.

[40] Yoo, K.H., and Gretzel, U. 2009. Comparison of deceptive and truthful travel reviews. In *Information and Communication Technologies in Tourism*, 37-47. Springer.

[41] Zakaluk, B. L., and Samuels, S. J. 1988. *Readability: Its past, present and future*. International Reading Association.

[42] Zhou, L., Burgoon, J. K., Nunamaker, J. F., and Twitchell, D. 2004. Automated linguistics based cues for detecting deception in text-based asynchronous computer-mediated communication. *Group Decision and Negotiation* 13, 1 (Jan. 2004), 81-106. DOI= http://dx.doi.org/10.1023/B:GRUP.0000011944.62889.6f

[43] Zhou, L., Shi, Y., and Zhang, D. 2008. A statistical language modeling approach to online deception detection. *IEEE Transactions on Knowledge and Data Engineering* 20, 8 (Aug. 2008), 1077-1081.