"Unveiling What is Written in The Stars: Analyzing Explicit, Implicit, and Discourse Patterns of Sentiment in Social Media"

FRANCISCO VILLARROEL-ORDENES

STEPHAN LUDWIG

DHRUV GREWAL

KO DE RUYTER

MARTIN WETZELS

Paper accepted for publication in the Journal of Consumer Research

Francisco Villarroel-Ordenes, f.villarroelordenes@maastrichtuniversity.nl, PhD. Candidate, School of Business and Economics, Department of Marketing and Supply Chain Management, Maastricht University, Tongersestraat 53, 6211 LM Maastricht. Tel: +31 43 38 83861. Stephan Ludwig, s.ludwig@westminster.ac.uk, Senior Lecturer, Westminster Business School, Department of Marketing and Business Strategy, 309 Regent Street, London W1B 2HW, Tel: +44 (0)20 7911 5000. Ko de Ruyter, Ko.De-Ruyter.1@city.ac.uk, Professor of Marketing, Cass Business School, City University London, 106 Bunhill Row, EC1Y 8TZ London, United Kingdom. Dhruv Grewal, dgrewal@babson.edu, Toyota Chair in Commerce and Electronic Business and a Professor of Marketing, Babson College, Babson Park, MA 02457-0310, Tel: +01 781-239-3902. Martin Wetzels, m.wetzels@maastrichtuniversity.nl, Chair in Marketing and Supply Chain Research, School of Business and Economics, Department of Marketing and Supply Chain Management, Maastricht University. Tongersestraat 53, 6211 LM Maastricht. Tel: +31 43 38 3839.

Contribution Statement

Although researchers often use automated sentiment analysis in assessing online consumer product evaluations by counting positive and negative words, more granular sentiment expressions—such as activation levels, implicit meanings, and patterns of sentiment across sentences (e.g., in an online review)—are relatively poorly understood. These granularities aid in differentiating different degrees of sentiment strength and enable a more in-depth analysis of the relationship between sentiment expression in consumer verbatim comments and subsequent online behavior. Using Speech Act Theory as an enabling framework, this study conceptualizes the differential impacts of explicit force levels, implicit expressions, and discourse patterns on overall sentiment strength (i.e., star ratings). We demonstrate the significance of these conceptualizations in an empirical study using online consumer reviews, as well as two follow-up studies assessing their relevance for sales and generalizability across social media contexts. By zooming in on how consumers express different degrees of sentiment strength, this study offers a more in-depth understanding of online consumer behavior.

ABSTRACT

Sentiment Analysis, the process of automatically distilling sentiment from text, is often used in consumer research to assess online consumer evaluations by counting positive and negative words. However, more granular sentiment expressions—such as activation levels, implicit meanings, and patterns of sentiment across sentences (e.g., in an online review)—are relatively poorly understood. Drawing on Speech Act Theory, this study goes beyond positive and negative word counts to examine the effects of finer-grained explicit and implicit sentiment expressions, within and across sentences. We demonstrate the significance of sentiment force levels, implicit sentiment expressions, and discourse patterns on overall consumer sentiment (i.e., star ratings) in an empirical study using online consumer reviews. Two follow-up studies enhance the relevance and generalizability of the findings. As this study confirms, both implicit and explicit expressions as well as discourse patterns allude to consumers' sentiments. These expressions also drive actual purchasing behavior; and are generalizable to other social media contexts such as Twitter and Facebook. These findings contribute to research on consumer sentiment analysis by offering an in-depth understanding of how the unique speech act features constitute consumers' sentiment expressions and their implications.

Keywords: Consumer Sentiment, Sentiment Analysis, Speech Act Theory, Text Mining, Customer Reviews, Sales Ranks, Social Media, Marketing Analytics

“You do not get discoveries in the sciences by taking huge amounts of data, throwing them into a computer and doing statistical analysis of them … that’s not the way you understand things … you have to have theoretical insights.”

—Noam Chomsky, April 2014

The growing influence of online evaluations on purchasing behavior (Dimensional Research 2013; McKinsey Company 2013) has increased the interest of managers and researchers in sentiment analysis, which refers to the process of automatically distilling sentiments from text (Pang and Lee 2008). The emerging volume of research also reveals an evolution in general focus, from classifying written text by its sentiment valence (e.g., positive, negative, neutral), to measuring sentiment strength (e.g., very negative to very positive), to detailing with individual emotions (e.g., anger, fear; Pang and Lee 2005, 2008). Yet extant consumer research generally lacks such in-depth conceptualizations and instead tends to rely on single emotion word counts to measure sentiment valence (Hennig-Thurau, Wiertz and Feldhaus 2014; Tirunillai and Tellis 2012). This oversimplification hides that written language offers consumers a wider range of explicit and implicit linguistic features and patterns whereby writers can express their sentiment (Gopaldas 2014). In turn, neglecting such linguistic means of sentiment expressions prohibits a more accurate understanding of how verbatim consumer reviews influence the reading consumer and sales performance(Ludwig et al 2014).

We suggest that speech acts might offer a meaningful theoretical lens for achieving such advances (Searle 1969, 1976; Zhang, Gao, and Li 2011). Speech acts involve intentions revealed through language, and they require the recognition of a higher-order linguistic context. For example, Speech Act Theory (SAT) introduces the notion of *illocutionary force,* or linguistic properties that alter the strength of words’ meanings (Holmes 1984; Sbisa 2001). In addition to the activation level on emotion words (e.g., “good” versus “awesome”; Russell and Barret 1999), phrases might exert stronger effects when they include certainty terms (e.g., “arrived *extremely* late”), or they might be attenuated by tentative terminology (e.g., “it was *kind of* nice”). The differential effects these types of sentiments have on overall sentiment strength remain unexplored. In addition, SAT recognizes that sentiment strength can be expressed implicitly (Perrault and Allen 1980), an idea that remains underexplored in consumer literature (Kronrod and Danziger 2013). For example, we know little about the distinct impacts of recommendations (e.g., “You must read this book”) versus promises (e.g., “I will keep buying his books”) versus statements on overall sentiment strength. Finally, consistent with research on mixed emotions (Aaker, Drolet, and Griffin 2008) and advances in text modeling (Buschken and Allenby 2015), perhaps patterns in an overall discourse convey meaning, beyond that implied by the individual sentences. For example, sentiment incoherence and trends in a message (e.g., moving from negative to positive sentiment) might influence the overall tone of a review (Goldberg and Zhu 2006; van Dijk 1997). Drawing on SAT, we therefore investigate the differential, asymmetric, and direct effects of these three speech act features on consumer sentiment strength, which enables us to offer three main research contributions.

First, we advance research on affect and activation levels by empirically studying different sentiment force features in online verbatim reviews, such as the level of activation in emotion words (e.g., “good” vs. “awesome”) and the boosting and attenuating indicators (e.g., “very good”, “kind of good”). In practice, we specify how sentiment force features allow consumers to express different levels of sentiment strength. Second, our findings provide insight into how consumers can use language to convey their sentiment without using explicit, emotion-laden words (Bosco, Bucciarelli, and Bara 2004). In particular, we examine the asymmetrical effects of directive (“I recommend that you go to this hotel”) and commissive (“I will come back to this hotel") acts, relative to assertive ones ("I went to this hotel") on overall sentiment strength. Third, noting that most arguments develop across a series of sentences, we demonstrate how their relative incoherence and sentiment trends (e.g., increase in positivity) can determine the overall tone of a review (Feng and Hirst 2013; Goldberg and Zhu 2006).

In the next section, we review extant literature pertaining to consumer sentiment expressions and SAT. We formulate a set of hypotheses to assess the differential effects of the varying language choices on overall sentiment strength (i.e., review star rating), and then assess them empirically using a unique data set of more than 45,843 online reviews. Furthermore, we demonstrate that the distinct sentiment strength elements are better predictors of sales rank than sentiment valence; we also assess the generalizability of our findings in social networks where star ratings are not present. Finally, we outline theoretical and managerial implications.

CONCEPTUAL FOUNDATIONS

Consumer research recognizes the importance of sentiment in cognitive, evaluative, and behavioral settings (Baumeister et al. 2007; Richins 1997). According to Giesler (2012), sentiment fuels market dynamics, institutional changes, and economic transformations. In big data settings, consumer research that draws on psycholinguistic concepts (Pennebaker, Mehl, and Niederhoffer 2003) has assessed the impact of positive and negative sentiment words on behaviors (Berger, Sorensen, and Rasmussen 2010; Hennig-Thurau, Wiertz and Feldhaus 2014; Tirunillai and Tellis 2012). However, we posit that these valenced words might mask the effects of further language granularities, such as the strength or conviction with which consumers express their sentiment (Thelwall et al. 2010). To go beyond a simple classification of valence, we therefore seek a more in-depth understanding of consumer sentiment and discourses. We therefore build on an enabling framework from SAT to propose a substantive revision to uses of sentiment analysis in consumer research.

Speech acts are a function of the intent of the sentence in which they appear (Searle 1969; Zhang, Gao, and Li 2011). The central premise is that it is not words but their linguistic context, consisting of phrases, sentences, and discourses, that conveys the intentions of verbal messages (Searle 1969). Communicating sentiment through expressive speech acts is an emotional response to a particular situation, such as a service failure (Norrick 1978). Expressive speech acts consist of a subject (to which the sentiment refers), valence (positive or negative emotion words describing the subject), and the illocutionary force (relative strength of conveyed valence) (Holmes 1984; Norrick 1978). Few consumer research studies acknowledge the importance of speech acts for deriving consumer intentions though (Thomas 1992), and existing consumer research on sentiment analysis neglects the inherent strength aspects, resorting to either binary, positive versus negative (Homburg, Ehm, and Artz 2015; Tirunillai and Tellis 2012) or ternary, positive/negative/neutral (Das and Chen 2007; Schweidel and Moe 2014) sentiment schemes (see Table 1).

[INSERT TABLE 1]

To distill sentiments from text, it is necessary to consider the inherent strength aspect, rather than a dichotomy (Moore 2012). To improve sentiment strength assessments, such as those that might be obtained from star ratings (Tsang and Prendergast 2009), we use SAT as an enabling paradigm (Searle 1969, 1976). That is, we conceptualize and explain the distinct and collective effects of the force of the sentiment being expressed, implicit sentiment expressions, and higher order discourse patterns across sentences on the overall sentiment strength.

Sentiment force. In a customer review, the subject of the evaluation is a product or service. The review's sentiment valence might be expressed through single emotion words, with different levels of activation (e.g., "poor" vs. "horrible"; Puccinelli, Wilcox and Grewal 2010). Russell and Barret (1999) highlight the importance of both valence (i.e., unpleasant to pleasant) and activation levels (e.g., high or low) for specifying the strength of different emotions in terms of an hedonic tone (valence) and its mobilization or energy (activation). Sentiment force also can be expressed through the use of sentiment modifiers, such as certainty words (e.g., "absolutely") or tentative words (e.g., "apparently"; Smith and Ellsworth 1985). When these modifiers combine with emotion words, they can alter the force of a sentiment expression, whether by boosting it, attenuating it, or totally inverting its valence (Sbisa 2001).

Therefore, the force of sentiment expressions—as determined by the activation level in emotion words (good vs. awesome) or their combination with certainty or tentative words—should help reveal the sentiment strength expressed in a consumer's rating. To test this prediction with verbatim customer reviews, we study the differential effects of boosted versus attenuated sentiment expressions on overall sentiment strength. Text mining scholars generally assume that boosted sentiment expressions double the effect of attenuated ones (Thelwall et al. 2010; Hu, Koh, and Reddy 2014), but psycholinguistic research lacks any quantitative assessment of these specific differential effects (Chung and Pennebaker 2007). Therefore, we phrase our hypotheses to propose that higher activation level and/or boosted sentiment expressions have stronger differential effects on overall sentiment strength, compared with lower activation level and/or attenuated sentiment expressions (Sbisa 2001). Formally,

> H_1: High activation level and/or boosted sentiment expressions have stronger effects than low activation level and/or attenuated sentiment expressions on the overall sentiment strength of text-based reviews.

Implicit Sentiment Expressions. Expressive speech acts are not a prerequisite to convey sentiment (Pinker, Nowak, and Lee 2008). It also can be conveyed implicitly, through expressions in which the speaker alludes to an act or notion without explicitly stating it (Searle 1975), such as when a brand implicitly conveys its intentions (Kronrod and Danziger 2013). Insight into how these implicit expressions are manifest in consumers' communication is lacking though (McGraw, Warren, and Kan 2015). Literature on linguistics suggests that speech acts that are directive (suggestion to take an action), commissive (committing to a future action), and assertive (conveying the state of the situation) also communicate expressive acts of sentiment (Searle 1975). In addition, Schellekens, Verlegh, and Smidts (2010) find that such implicit sentiment expressions are common in online customer reviews, often as suggestions, commands, or requests for action by peers. Directive speech acts, such as "You should stay here" or "I wouldn't recommend you to read it," can be associated with positive and negative feelings (e.g., D'Andrade and Wish 1985). Commissive speech acts instead involve the speaker promising, intending, or vowing to do something in the future (Searle 1976), though they also can denote negative emotions (e.g., "I will never read another book from this author") or positive ones ("We'll come back for sure"). Finally, assertive speech acts represent a state of affairs (Searle 1975)—such as "We got a discount" or "We waited for over an hour"—and thus implicitly convey positive or negative sentiment, without the use of any emotion words.

Although all of these speech acts can convey sentiment without using emotionally valenced words, it remains unclear which expressions represent the strongest conduits to sentiment extremity. We posit that directive and commissive acts might have stronger effects on consumer sentiment strength than assertive acts, because directive acts encompass a form of active exercise of power towards readers, and commissive acts imply the reviewer assumes the ability to commit to an action (Austin 1962; Searle 1976), rather than just providing a simple description of circumstances or characteristics (Searle 1976; Austin 1962). Assertive acts therefore are the least powerful and generally presented as a true-or-false statement (Searle 1976). Thus, we hypothesize;

> H_2: Directive and commissive speech acts have stronger effects on overall sentiment strength than assertive acts in text-based reviews.

Discourse Patterns. Single sentences within a discourse are related. Their patterns might reflect how writers communicate about a product or service experience to express their sentiments (Goldberg and Zhu 2006; van Dijk 1997). Consumers experience multiple positive and negative emotions when consuming a product or service (Aaker, Drolet, and Griffin 2008), which they might verbalize across multiple sentences in a customer review. These patterns in turn define sentiment expressions, because language is not limited to single sentences but instead reflects the combined dynamics of the various sentences (Schweidel and Moe 2014). Accordingly, Auramäki, Lehtinen, and Lyytinen (1988) suggest that different patterns within a discourse, such as incoherence and trend, may indicate more positive or negative sentiments.

Current sentiment analysis methods neglect sentiment expressions that contain mixed positive and negative emotions (Das and Chen 2007), examine them at an aggregated message level (Tirunillai and Tellis 2012), or else derive sentiment at a sentence level (Buschken and Allenby 2015; Khan, Baharudin, and Khan 2011). However, the active use of contradictory expressions (Fonic 2003) might produce arguments with a lesser degree of conviction. Ignoring such developments across multiple sentences would fail to account for ambivalent evaluations, yet consumer research shows that such evaluations can have negative impacts on sentiment toward a product or service (Otnes, Lowrey, and Shrum 1997). We hypothesize;

H_3: Sentiment incoherence across sentences in text-based reviews has a negative effect on the overall sentiment.

For exploratory purposes (Dubois, Rucker and Galinsky 2016) we also consider other types of discourse patterns in online reviews. De Saussure (2007) suggests that a positive (negative) trends in the message development , may further relate to the overall sentiment of writer. According to this prediction, sentiment expressions are not randomly distributed but rather represent a set of propositions, sequentially organized to explain an overall opinion. In fact, previous research acknowledges the presence of such positive or negative trends in sentiment expressions but without explicating their implications (Mao and Lebanon 2007). In turn we assess the implications of these two types of trend on overall sentiment strength.

STUDY 1: ROLE OF ONLINE CUSTOMER REVIEWS

Setting

To examine the differential effects of the features of various speech acts on sentiment strength, we collected review data from three online customer review sites (Amazon.com, Bn.com, tripadvisor.com) through Monzenda, a web scraping software service. The data included text-based comments and associated star ratings from 45,843 customer reviews posted about 1,618 products and services (Bn.com, 527 books and 3,746 reviews; Amazon.com, 1,091 books and 18,060 reviews; Tripadvisor.com, 81 hotels and 24,037 reviews). With this data set, we could analyze text-based sentiment expressions across two different contexts, books and hotels, and thus consider how consumers express their sentiments about both products and services.

Measure Development

We measure the dependent variable, consumers' overall sentiment strength toward a product or service, with self-reported star ratings. Star ratings appear prominently in marketing and consumer research (Kronrood and Danzinger 2013; Ludwig et al. 2013), and previous text mining studies use them as proxies for sentiment strength (Pang and Lee 2005, 2008). In online retailing studies, star ratings also serve to represent consumers' deviations from three stars, such that on a five-star scale, one or two stars indicate negative evaluations, and four or five stars signal positive evaluations (Amazon 2014). A three-star (midpoint) review might reflect either a truly moderate review (indifference) or a series of positive and negative sentiments that cancel one another out (Mudambi and Schuff 2010). Any star ratings other than the midpoint are relative deviations from it, toward extremely negative (one star) or extremely positive (five stars) ratings. In line with previous research (Chevalier and Mayzlin 2006), the star ratings in our data were skewed toward positive reviews: 54% of the reviews offered 5 stars, another 27% rated 4 stars, 9% used 3 stars, and only 5% and 7% of the reviews were rated 2 and 1 stars, respectively.

To assess the text-based predictor variables, or speech act features, we first preprocessed the reviews related to each industry by removing duplicate reviews (2,156). Then we applied the Stanford Sentence Parser (online available at http://nlp.stanford.edu:8080/parser/) to divide entire reviews into sentences automatically (we removed sentences with 1 or 0 words) and identify grammatical dependencies between emotion words and sentiment force features (e.g., certainty words). Furthermore, by using part of speech tagging, which can retrieve the grammatical classes of words, we identified the most frequent noun and noun phrases in the data set (Ghose, Ipeirotis, and Beibei 2012). We retrieved the most frequent 4,071 nouns (e.g., "staff," “hotel”), then focused our analysis on sentences that contained at least one noun associated with a product or service feature as a sentiment subject. To deal with anaphora resolution (i.e., when a sentence does not have an explicit referee but implicitly refers to one), we also considered sentiment in adjacent sentences to a referee such as, “This author keeps impressing me with the quality of his work. It is just awesome.” To do this we used the linguistic inquiry and word count (LIWC) dictionary to retrieve indirect pronouns (e.g., “it”, “this”; Chung and Pennebaker 2007). Finally, in a few cases, when the names of books or hotels used an emotion word, we did not take it into consideration as a sentiment of the consumer.

We followed a supervised text-mining technique, using dictionaries provided by the LIWC program (Pennebaker et al. 2007), which can automatically retrieve expressive emotion words. Originally developed to analyze emotional writing, the LIWC program offers strong, reliable convergence between the dimensions it extracts from text-based contents and ratings performed by human coders (Pennebaker et al. 2007). Similar to Netzer et al. (2012), we enriched the dictionaries with more words that had positive and negative meaning, gleaned from online emotion dictionaries such as SentiwordNet and emoticons from Pcnet (Zhang, Gao, and Li 2011).

Following Russell and Barret (1999), we then categorized the LIWC emotion word dictionaries on the basis of relative activation levels. Specifically, expressive negative emotion words could be assigned to an expressive, negative, low activation (ENL) category, which included only negative emotion words pertaining to the sadness dictionary, or to an expressive, negative high-activation (ENH) category that included the anger and anxiety word dictionaries. As LIWC does not provide the positive emotion dictionary categorized into different levels of activation, we had to manually categorize it into low and high activation expressive groups (EPL and EPH, respectively). Two independent coders, unfamiliar with the study purpose but under the supervision of the second author, performed the classifications of 516 positive words (Krippendorff's alpha = 83%; discrepancies resolved through discussion). We validated this classification with the dictionary of affection in language (Whissell 2009), which computes an activation degree per word or text on a continuous scale (1 to 3). With the activation scores for the entire list of positive emotion words, we ran a one-way analysis of variance to determine the statistical differences in the level of activation for the positive low and positive high categories (as classified by the coders). We found significant differences ($F = 9.701$, $p < .01$), with mean scores of 2.11 and 1.95 for the high and low categories, respectively.

We used the Stanford dependency parser (Stanford Parser 2014) to detect grammatical dependencies between expressive words and sentiment force elements (boosting, attenuation, negation). Boosting and attenuation words were selected from the certainty and tentative categories of the LIWC. We also identified negations (e.g., "not good") through the Stanford Parser, without using a dictionary. Using a two-step approach (Taboada et al. 2011), we first determined if the high and low activation words were accompanied by a booster word (B), in which case they stayed in or moved to the high activation category, respectively (e.g., "very good"). We also considered if they were accompanied by an attenuation word (A), in which case high activation words moved into the low activation category, and low activation terms stayed in that category. Next, we identified cases in which the positive or negative boosted and attenuated expressions were negated (N). Negations represent a specific type of attenuation (Sbisa 2001). Although we do not propose formal hypotheses about negations, we assess their role. Therefore, we derived four main sentiment strength variables and their negated forms (e.g., positive high and its negation positive high). Each variable represents a proportion score for each sentence j (proportion of words) and review i (proportion of sentences). We calculated them as follows:

$$PH_i = [\sum_{\substack{i=0 \\ j=0}}^{\substack{i=m \\ j=n}} \frac{EPH_{ij}+EPH_{ij}* B_{ij}+EPL_{ij}*B_{ij}}{WCount_{ij}}] \Big/ SCount_i, \tag{1}$$

$$PL_i = [\sum_{\substack{i=0 \\ j=0}}^{\substack{i=m \\ j=n}} \frac{EPL_{ij}+EPH_{ij}* A_{ij}+EPL_{ij}*A_{ij}}{WCount_{ij}}] \Big/ SCount_i, \tag{2}$$

$$Neg\_PH_i = N_{ij}[\sum_{\substack{i=0 \\ j=0}}^{\substack{i=m \\ j=n}} \frac{EPH_{ij}+EPH_{ij}* B_{ij}+EPL_{ij}*B_{ij}}{WCount_{ij}}] \Big/ SCount_i, \text{ and} \tag{3}$$

$$Neg\_PL_i = N_{ij}[\sum_{\substack{i=0 \\ j=0}}^{\substack{i=m \\ j=n}} \frac{EPL_{ij}+EPH_{ij}* A_{ij}+EPL_{ij}*A_{ij}}{WCount_{ij}}] \Big/ SCount_i, \tag{4}$$

where PH_i and PL_i represent the high and low positive proportion for review i, respectively, and Neg_PH_i and Neg_PL_i represent the negation of positive high and low expressions, respectively.

These equations feature three binary variables for each sentence *j*: (B_{ij}), which refers to the presence (1) or not (0) of a boosting element (e.g., "!!"); (A_{ij}), which reflects whether there is an attenuating speech element (1) or not (0); and (N_{ij}), which indicates whether any grammatical dependency with a negation exists (1) or not (0). Thus for example, PH_i indicates the positive high proportion in review *i*, operationalized as the sum of positive high activation words (EPH_{ij}), boosted positive high activation words $EPH_{ij} * B_{ij}$, and boosted positive low activation words $EPL_{ij} * B_{ij}$ divided by the word count ($WCount_{ij}$) in review *i* and sentence *j* (*m* = review number; n = sentence number at review *i*), and subsequently divided by the sentence count in review i, $SCount_i$. In Equation 2, we also include a final component to represent attenuated positive low activation words. When N = 1 in Equations 3 and 4, the positive high and low cases are being negated. Finally, we derive NH_i (high negative proportion) and NL_i (low negative proportion) for review *i* and similarly use the sum of negative emotion words divided by total words, subsequently divided by the sentence number.

To assess the validity of our first sentiment force measurement, we used SentiStrength (Thelwall et al. 2010), a state-of-the-art application to predict sentiment from short texts (for a recent application to marketing research, see Tang, Fang, and Wang 2014). The results indicate correlations of .58 and .45 for aggregated measures of positive and negative sentiment, respectively, which validate these measures.

With regular expression codes (i.e., regex; Feldman and Sanger 2007), we extracted the implicit sentiment expressions conveyed for commissive (C), directive (D), and assertive (A) speech acts and their respective valence. Following a linguistics approach (Villarroel et al. 2014), we developed the regular expression codes for implicit sentiment by first retrieving a random sample of 1% of sentences, void of any emotion, from the book and hotel industries (558 and 1,018, respectively). Two independent coders, under the supervision of the first author, identified word patterns that conveyed sentiment implicitly in these sentences. They coded the main speech act in the sentence (assertive, directive, commissive), its valence (positive, negative, neutral), and the specific word patterns that determined the valence of these speech acts (see Online Appendix A for the coding instructions). The coders achieved a Krippendorff's alpha of 77% for the valence and 94% for n the type of the speech act (disagreements were resolved in a meeting between the first author and the coders). Based on the word patterns identified by the coders we developed regular expression codes for implicit sentiment

Online Appendix B contains illustrative examples of regular expression codes for identifying implicit sentiment expressions. With these regex codes, we retrieved 8,578 sentences without emotion words (16% of the reviews). The aim of the regular expressions is to capture not all implicit sentiment expressions but rather a representative sample for the hypotheses tests.

Then we can compute:

$$CP_i = \left[\sum_{\substack{i=0 \\ j=0}}^{\substack{i=m \\ j=n}} \frac{Pos\_Commissive_{ij}}{WCount_{ij}}\right] \Big/ SCount_i, \qquad (5)$$

$$CN_i = \left[\sum_{\substack{i=0 \\ j=0}}^{\substack{i=m \\ j=n}} \frac{Neg\_Commissive_{ij}}{WCount_{ij}}\right] \Big/ SCount_i, \qquad (6)$$

$$DP_i = \left[\sum_{\substack{i=0 \\ j=0}}^{\substack{i=m \\ j=n}} \frac{Pos\_Directive_{ij}}{WCount_{i,j}}\right] \Big/ SCount_i, \qquad (7)$$

$$DN_i = \left[\sum_{\substack{i=0 \\ j=0}}^{\substack{i=m \\ j=n}} \frac{Neg\_Directive_{ij}}{WCount_{ij}}\right] \Big/ SCount_i, \qquad (8)$$

$$AP_i = \left[\sum_{\substack{i=0 \\ j=0}}^{\substack{i=m \\ j=n}} \frac{Pos\_Assertive_{ij}}{WCount_{ij}}\right] \Big/ SCount_i, \text{ and} \qquad (9)$$

$$AN_i = \left[\sum_{\substack{i=0 \\ j=0}}^{\substack{i=m \\ j=n}} \frac{Neg\_Assertive_{ij}}{WCount_{ij}}\right] \Big/ SCount_i, \qquad (10)$$

where CP_i represents the commissive positive proportion in review *i*, operationalized as $Pos\_Commissive_{ij}$ divided by the word count $WCount_{ij}$, and subsequently divided by the number of sentences in review i (*m* = review number; n = sentence number at review *i*). We derived CN_i (commissive negative proportion) and the other forms of implicit speech acts using the sum of speech acts that implicitly convey negative sentiment, divided by the total words.

The validation of these measures relied on Schweidel and Moe's (2014) approach, such that the first author checked a subsample of 200 retrieved implicit sentiment expressions. The Krippendorff's alpha between the classification performed with the regular expression and the manual classification by the author reached 88%, indicating that the operationalization of implicit sentiment valence worked well.

Next, we noted two main discourse patterns of sentiment across customer reviews: incoherence (standard deviation of the sentiment intensity) and trend (slope across sentences). We derived these macro speech acts in all reviews with more than one sentence (reviews with one sentence had an incoherence value of 0; reviews with one or two sentences had a trend value of 0). We computed the overall sentiment proportion of each sentence, or $DifPosNeg_{ij}$, by accounting for all explicit and implicit sentiment expressions in Equations 1–10. Consistent with our previous rationale, we assigned weights to each of the previously described measures of sentiment from the Equations. Assigning weights to the individual sentiment strength measures is consistent with our hypothesized arguments about the differential, asymmetric effects of speech act features. Thus, we first computed the coefficients of the previously defined measures (we used the exponential of the log-odds coefficient provided in the results, which indicates the probability of reaching a higher star rating category) to explain sentiment strength, then multiplied them by the proportion at a sentence level. This procedure to aggregate varying sentiment strengths in a single measure is also more precise than current sentiment analysis research (Hu, Koh, and Reddy 2014), which assigns weights of 2 and -2 for strong positive and negative terms, accordingly,

$$DifPosNeg_{ij} = \sum_{\substack{i=0 \\ j=0}}^{\substack{i=m \\ j=n}} [PHL_{ij} * \beta_{IPH} + PL_{ij} * \beta_{IPL} + CP_{ij} * \beta_{CP} + DP_{ij} * \beta_{DP} + AP_{ij} * \beta_{AP}] - [NH_{ij} * \beta_{INH} + NL_{ij} * \beta_{INL} + CN_{ij} * \beta_{CN} + DN_{ij} * \beta_{DN} + AN_{ij} * \beta_{AN}]$$

(11)

We next operationalized the incoherence of positivity as variation in sentiment across a sequence of sentences (SDi)), using the standard deviation in $DifPosNeg_{ij}$ across all sentences in a review:

$$SD_i = SD(DifPosNeg_{ij}) \tag{12}$$

Finally, to keep consistency with our exploratory analysis of trend, we calculated the slope of positivity as the beta coefficient of an ordinary least squares regression, $DifPosNeg_{ij} = \alpha + \beta * SentNum_{ij}$, where $SentNum_{ij}$ is sentence number j in review i. A rate of change close to 0 signifies a stable trajectory. Then, we split this variable in two groups with positive and negative trend values. This approach preserves the continuous nature of our variable trend and avoid the costs of a categorical dichotomization (Rucker, McShane and Preacher 2015). Thus, our model includes the following slope measure of sentiment:

$$Trend_i = cov\left(Sentnum_j, DifPosNeg_j\right)/var\left(SentNum_j\right). \quad (13)$$

To validate the effect of these sentiment pattern measures across sentences, we ran a sensitivity analysis of the computation of $DifPosNeg_{ij}$, in which we computed an alternative $DifPosNeg_{ij}$ that does not assign any weights to the individual measures of sentiment force and implicit sentiment. We then recalculated the values of the standard deviation across sentences and trends. We summarize the speech act features in Table 2.

[INSERT TABLE 2]

Control Measures

We controlled for the use of first-person pronouns (Pennebaker et al. 2007) to account for the degree of subjectivity in the review. According to Barasch and Berger (2014), first-person pronouns ("I" or "we") allow people to focus on themselves, therefore providing a more experiential or subjective viewpoint. Furthermore, previous research in sentiment analysis has shown that subjective sentences are generally containing the opinion and sentiment of the writer (Pang and Lee 2008). We measured the proportion of first-person pronouns, relative to the number of words in each sentence, then aggregated this calculation to the review level. We also controlled for the review site (D_RS_i) in the case of book reviews (Barnes & Noble = 0, Amazon = 1). Finally, regarding our discourse measurements (incoherence and trends), we controlled for the total amount of sentences in a review ($TSent_i$).

Analysis

For more fine-grained analyses of the proposed method capabilities across two linguistic domains, and to avoid cofounding effects in our measures, we proceeded with two separate models for books and hotels. Because we interpret sentiment strength as an ordinal variable, we continued our analysis using an ordinal logit model (Farley, Hayes, and Kopalle 2004) and assessed the hypotheses with a comparison of the differential and asymmetric effects of the different speech act features (i.e., Wald test between coefficients). We also tested the robustness of our models according to the Akaike information criterion (AIC) value. In accordance with the proportional odds assumption (Harrell 2001; Williams 2006), we sought to corroborate our results by specifying partial proportional odds to allow some coefficients to vary across star categories (i.e., multinomial logit). We find that despite an increase in model fit, the interpretation of the coefficient does not change, so we opted for ordinal logit, which offers a more parsimonious model and accounts for the order in our dependent variable.

Consistent with SAT, we operationalized the text-based sentiment strength of a customer review as a function of sentences and their speech act features: sentiment force, implicit sentiment expressions, and discourse patterns of sentiment across sentences. With a sequential approach, we tested the hypotheses. Model 1 thus included sentiment force variables, together with the control variables:

$$l(\beta) = \log P\,(SR_i | PH_i, PL_i, NH_i, NL_i, Neg_i, FP_i, D\_RS_i, \mu_i; \alpha, \beta), \qquad (14)$$

where SR_i is the star rating, PH_i and PL_i are the proportions of positive high and low sentiment across all sentences in review i, and INH_i and INL_i are the means of the negative high and low sentence proportions in review i, respectively. Moreover, Neg_i refers to the four independent variables that represent the negated versions of the four proportions (e.g., Neg_PH_i is the negation of positive high expressions). We added two control variables to account for the proportion of non-emotion sentences ($NSent_i$) and the proportion of first-person pronouns (FP_i), along with a dummy variable for the review site in the book setting (D_RS_i)

In Model 2, to avoid multicollinearity, and after identifying the effect of negations in overall sentiment, we aggregated the negation variables into positive and negative low proportions (PL and NL, respectively). We then included implicit sentiment expressions conveying positive and negative sentiment and the control variables as predictor variables in Model 2:

$$l(\beta) = \log P\,(SR_i | PH_i, PL_i, NH_i, NL_i, CP_i, DP_i, AP_i, CN_i, DN_i, AN_i, FP_i, D\_RS_i, \mu_i; \alpha, \beta), \qquad (15)$$

where CP_i, DP_i, AP_i and CN_i, DN_i, AN_i represent the sentence-level means of the sentiment proportions for commissive, directive, and assertive language, respectively, proceeded by a P if the implicit sentiment is positive or N if negative.

Finally, in Model 3, we added the discourse patterns as predictive features of overall sentiment strength, together with the total number of sentences (control variable):

$$l(\beta) = \log P\,(SR_i | PH_i, PL_i, NH_i, NL_i, CP_i, DP_i, AP_i, CN_i, DN_i, AN_i, SD_i, PT_i, NT_i, TSent_i, FP_i, D\_RS_i, \mu_i; \alpha, \beta),$$

(16)

such that the incoherence SD_i, positive trend PT_i and negative trend NT_idiscourse patterns offer potential sentiment predictors. To provide a better reflection of the discourse patters, this last model focuses only on reviews with more than 3 sentences (2283 and 947 reviews were excluded for hotels and books respectively). We conducted three ordinal logistic regressions to assess the individual hypotheses.

Hypotheses Testing

The estimates of the four cut points (i.e., intercepts in ordinal logit) across the three models, which indicate the latent variable values for establishing the five sentiment strength groups, provided an increasing trend, from negative to positive. Similar to previous marketing studies (e.g., Godes and Silva 2012), for conciseness, we do not report the cut points, but all of them are significantly different from their adjacent cut points at $p < .01$.

Before testing H_1, as a robustness check, we assessed the effect of the negations for each main variable (PH, NH, PL, and NL) (see Table 3, Model 1a). Noting the positive (negative) coefficients of the negated variables, and in line with Sbisa (2001), we aggregated them for books and hotels as attenuated sentiment expressions. For hotels, the negation of positive high and low became negative low, and the negation of negative high and low became positive low. For books, the negation of positive high and low and of negative high all became negative low, whereas the negation of negative low became positive low.

In line with H_1, Model 1b in Table 3 shows that positive high (PH) expressions had a significantly stronger positive effect on sentiment strength than did positive low (PL) expressions for books and hotels (books PH .93 vs. PL .11, Wald z = 29.45, $p < .01$; hotels PH .89 vs. PL .06, Wald z = 36.27, $p < .01$). Similarly, for hotels, negative high (NH) expressions had a significantly stronger effect on sentiment strength than did negative low (NL) ones (NH -.44 vs. NL -.40, Wald z = -1.98, $p < .05$). However, for books, the negative low expressions had significantly stronger effects (NH -.31 vs. NL -.37, Wald z = 3.45, $p = .01$). These results support H_1 but also leave room for discussion about the reversed, stronger effect of negative low over negative high expressions for book reviews.

[INSERT TABLE 3]

In line with H_2, the results in Table 3 show an overall incremental effect of positive (negative) assertive versus commissive or directive statements on sentiment strength. For positive sentiment for example, directive acts have a stronger effect than assertive ones (books .28 vs. .05, Wald z = 6.28, $p < .01$; hotels .07 vs. -.00, Wald z = 3.30, $p < .01$), as do commissive acts for hotels only (hotels .09, Wald z = 5.57, $p < .01$). For negative sentiment, we find that directive acts have stronger effects than assertive acts (books -.28 vs. -.06, Wald z = -7.33, $p < .01$; hotels -.15 vs. -.07, Wald z = -4.06, $p < .01$), commissive acts have stronger effects than assertive ones (books -.17, Wald z = -8.25, $p < .01$; hotels -.15, Wald z = -5.92, $p < .01$).The results thus support H_2.

As noted previously, before testing H_3, we used the exponential of the log-odds coefficient of Table 3 (Model 2) as weights to create an aggregated sentiment strength variable for each sentence, $DifPosNeg_{ij}$. For example, the coefficient of PH (.88) resulted in a value of 2.41, so we used this weight to compute the aggregated sentiment score per sentence. This procedure, applied to all the variables in the aggregation, also received validation from an alternative sensitivity analysis, in which we assigned $DifPosNeg_{ij}$ no weight; the results were consistent (see Online Appendix C). In support of our prediction, sentiment expressed through inconsistent language had a significant negative effect on overall sentiment strength (hotels β_{SD} = -.16, p < .01; books: β_{SD} = -.17, p < .01), in support of H_3. Furthermore, while our exploratory analysis of positive trend across the sequence of sentences in the review indicated a more negative sentiment overall (hotels β_{PT} = -.14, p < .01, books β_β_{PT} = -.14, p < .01), negative trend indicated a more positive sentiment overall (hotels β_{NT} = .08, p < .01, books β_{NT} = .12, p < .01).

We controlled for the book review site to ensure generalizability. The site was significant (β_{D_RS} > .06, p < .01 across four models): Amazon reviews tended to be more positive than Barnes & Noble reviews. The use of personal pronouns had a significant positive effect for hotels (β_{FPP} = .19, p < .01) and a negative one for books (β_{FPP} = -.10, p < .01). Finally, the covariate total number of sentences per review had a significant negative effect for hotels (β_{TSent} = -.07, p < .01) and a negative non-significant effect on overall sentiment strength for books (β_{TSent} = -.01, p < .01).

Robustness Check

The endogenous relationship between written expressions and the self-reported consumer star rating prevents us from making causal conclusions. Therefore, we tested Model 3 with a random subsample of the books data set (1,925 reviews, or approximately 10% of the data). In line with previous research (Ghose, Ipeirotis, and Beibei 2012), we paid participants on Amazon's Mechanical Turk (AMT) to code this data set into sentiment strength categories, ranging from 1 to 5. Each review was coded by 10 persons, and no worker could code more than 25 reviews. With the reported sentiment strength scores, we computed the average sentiment strength per review and used an ordinary least squares regression to explain the average sentiment strength, using our speech act predictors. The results (Table 4) corroborated our hypotheses. The correlation between the star ratings and the AMT average was .84 ($p < .01$).

[INSERT TABLE 4]

STUDY 2: RELEVANCE OF SENTIMENT STRENGTH

Setting

Changes in the sentiment expressed in verbatim consumer reviews might lead to differential sales. Specifically, we expect that a more fine-grained approach to decode the overall sentiment of reviews can reveal the influence on sales ranks, such that overall positive (negative) sentiments should increase (decrease) sales performance, even after controlling for changes in the number of reviews, price changes, or time-invariant effects (e.g., product type, popularity).

Following an approach outlined by Chevalier and Mayzlin (2006), we tested and compared the influence of sentiments in consumer reviews that we derived using (1) just *valence*, in the form of positive and negative emotion words (Model A) or (2) *sentiment*, derived using our more fine-grained approach (Model B) from Study 1 (Model 3). We tested the influences on sales performance across a sample of consumer reviews written for books released between April 15 and May 5, 2010 on both Amazon.com and Barnes&Noble.com. We collected a longitudinal data set with 352 matched books, with an average of 9.2 weekly observations. We gathered, from both sites, the weekly sales rank of each book, price charged, total number of reviews featured on the product site in a given week, and the review texts of all reviews posted.

We followed Chevalier and Mayzlin's (2006) approach for cleaning and establishing the data set for analysis (for more details, see appendix 1).

Results

Changes in the sentiment strength of the review texts in the previous week (t – 1) exerted a significant influence on the log of sales rank difference, across Amazon.com and BN.com, in the following week (t) (see Model B). When more reviews appear on Amazon.com's product page from one week to the next and invoke more positive sentiment overall, sales of the reviewed product improve on Amazon.com compared with BN.com ($\beta_{sentimentAmazon}$ = -.028, $p < .01$). The coefficient is negative in this case, because decreases in sales ranks actually imply more sales. Conversely, a positive change in sentiment strength in the reviews on BN reduces sales at Amazon ($\beta_{sentimentBN}$= .024, $p < .05$). Using just the changes in valence is not viable for predicting changes in sales (cf. Model A). For example, changes in valence in the reviews featured on Amazon exhibit a significant influence on subsequent sales ($\beta_{valenceAmazon}$= -.020, $p < .05$), but the predicted effect is less stark. If we used just valence, there would be no significant effect of changes in valence in the consumer reviews on BN.com on sales performance on Amazon ($\beta_{valenceBN}$ = .017, $p = .063$). Only by using the changes in sentiment strength do we uncover the significantly better Model prediction (Model 2, Wald χ^2 = 80.69, Model 1 Wald χ^2 = 53.65, $p < .001$). Therefore, adding variables from SAT to decode the sentiment of verbatim consumer reviews improves predictions of subsequent sales performance. We do not find any significant influence of any of the implicit sentiment expressions on sales (please see details on the Online Appendix D), with the exception of negative directives (e.g., "do not buy this book").

Such negative directives on the product site increase the sales rank of the product, hence the sales are decreased ($\beta_{\text{negative directive Amazon}}$= .029, $p < .05$). These results are in line with previous research by Ludwig et al. (2013) who suggest that, trying to avoid informational overload, they resort to heuristic processing and hence screen for the most easily accessible indicators, which are affect word cues (hence the effects of sentiment and valence).

STUDY 3: GENERALIZABILITY OF SENTIMENT STRENGTH TO SOCIAL MEDIA

Setting

To add external validity to our results, we scrapped online consumer and service evaluations from Twitter and Facebook across six product and service categories (e.g., financial services, travel, retail). We retrieved 1,716 evaluations and asked two independent coders, supervised by the first author, to rate the sentiment strength from 1 to 5. In addition to measuring the differential effects of the sentiment variables (sentiment force, implicit sentiment, and discourse patterns), we controlled for the social media platform by adding a dummy variable (1 = Facebook; 0= Twitter), noting that Schweidel and Moe (2014) indicate that sentiment can vary across venue format. The sample of product and services evaluations is smaller than the sample of customer reviews from websites in Study 1, so we did not separate the models by industries but rather added five dummy variables to control for differences across industries. The only measurement modification pertained to the regular expression codes used to detect implicit sentiment expressions; with this smaller sample, we decided to aggregate the individual speech acts (commissive, directive, and assertive) into implicit positive and implicit negative categories.

Despite the shorter comments (i.e., 1.69 and 2.71 sentences on average per product and service evaluation in twitter and Facebook respectively, versus 8.12 sentences in customer reviews), we still considered the shape parameters incoherence and trend. Replicating Model 3 from Study 1 then enabled us to corroborate the differential effects of our sentiment predictors.

Results

In line with H_1 (Table 5), we found a stronger, significant effect on overall sentiment strength for positive high compared with positive low expressions (.60 vs. .31, Wald z = 2.66, p = .01). Negative high expressions also had a stronger negative effect than negative low ones on overall sentiment strength (-.42 vs. -.03, Wald z = -3.34, $p < .01$).

[INSERT TABLE 5]

In this new context, we had to modify the general expressions from Study 1 by altering the contextual verbs associated with the regular expressions. For example, the regular expression for assertive acts in Study 1 indicated “should + buy,” whereas in the social media context, we referred to news media and measured “should + watch.” Despite our adaptations, we could retrieve only 8% of product evaluations that reported at least one of the six implicit sentiment expressions (cf. 16% of customer reviews). Therefore, we aggregated the three positive variables into one “implicit positive” measure and the three negative ones into an “implicit negative” variable. Although we found consistency in the coefficients (e.g., negative impact of indirect negative), we did not find any significant effects. For the sentiment pattern measures, we obtained evidence that incoherence had a negative, marginally significant impact on sentiment (β_{SD}= -.18, p = .06). Neither positive nor negative trend had a significant impact, but this was largely influenced by the low amount of evaluations with more than 3 sentences (26%). The industry dummies were significant, and Facebook evaluations were more negative than those on Twitter.

Finally, we benchmarked the sentiment approach with a basic sentiment proportion measure, derived using the numbers of positive versus negative words per review (obtained from the LIWC dictionaries of positive and negative emotions). According to Table 5, our theory-driven model (AIC = 3300.016) offered stronger predictive power than did the valence-based model (AIC = 3435.22) for determining consumer sentiment strength.

DISCUSSION

Extending Extant Research

By zooming in on how consumers' written reviews reflect differential, asymmetric sentiment strength, and how sentence patterns might exert direct negative effects on overall sentiment expressions, we contribute to the literature on consumer sentiments expression and improve predictions of subsequent consumer behavior. By empirically validating the hypothesized relationships and addressing their relevance and generalizability, we extend extant research in three ways.

First, to decode consumer sentiments and their influence, prior consumer research has relied on simple word frequencies, such as the number of positive or negative emotion words in verbatim customer reviews and posts (Hennig-Thurau, Wiertz and Feldhaus 2014). By accounting for activation level differences, innate to sentiment expressions (Russell and Barret

1999), and the influences of certainty (Pennebaker et al. 2007), we augment such approaches. In particular, disentangling positive or negatively valenced emotions and the degree of certainty with which they are expressed significantly improves estimates of consumer sentiments. On the one hand, compared with positive low activation and/or attenuated sentiment expressions, the use of positive high activation and or boosted expressions doubles the probability of a higher star rating designation by a consumer. On the other hand, compared with negative low activation and/or attenuated sentiment expressions, the use of negative high activation and/or boosted expressions did not double the probability of a lower star rating designation by a consumer. In particular, we failed to find a significant difference between highly activated and/or boosted and low activated and/or attenuated negative expressions in book reviews. Sentiment expression in book reviews thus depends at least partially on the context, so “sad” might be an appreciated feature for a tragedy genre, and “disgusting” might describe the antagonist character. Our findings which reflecting differences in hotels compared to books is also in line with research on affect suggesting that the use taxonomic structures regarding to emotion, might not work across contexts in the same way (Russell and Barret 1999). This is an important finding for research in sentiment analysis, which is highly dependent on word taxonomies associated with sentiment and activation. Overall, in line with Russell and Barret (1999) and Sbisa (2001), we empirically demonstrate the importance of considering the nuanced relationship among sentiment force features (i.e., activation level, certainty, tentative and negations), and overall sentiment in online consumer reviews.

Second, SAT predicts that assertive, commissive, and directive speech acts implicitly convey the speaker’s sentiment, without using explicit emotion words (Searle 1975). We predict and find that such “emotionless,” implicit acts relate asymmetrically to consumers’ overall sentiment. Implicit sentiment features appeared in about 16% of consumer reviews, and in line with our hypotheses, we found that positive (negative) directive and commissive acts exerted stronger effects on sentiment strength than did assertive acts. The linguistic context suggests that generic assertions hotel reviews (e.g., “We stayed in a superior double room,” “Rooms were clean”) may not really have an effect on the overall sentiment as they are only aligned with general expectations. Furthermore, commissive language tended to be used more in hotel but less in book reviews, likely because it is generally less common to commit to read a book again (once in a life product experience), whereas returning to a certain hotel is a likely option. Our findings also contribute to conceptualizations of implicit sentiment expressions (Feldman 2013; Montoyo, Martínez-Barco, and Balahur 2012), in that we introduce and empirically validate a theoretical framework of emotionless speech acts.

Third, we underscore the necessity of considering the message development (van Dijk 1997) and contribute to conceptualizations of sentiment dynamics (Schweidel and Moe 2014) by exploring how sentence-level developments reflect consumers’ sentiments. A consumer’s overall sentiment is likely negative if the development of the sentiment expressions in a review (explicit and implicit) are incoherent. In line with SAT and discourse literature (van Dijk 1997), as well as the concept of consumer ambivalence (Otnes, Lowrey, and Shrum 1997), we verify that relative incoherence across all review sentences decreases the overall positivity of the sentiment. Our exploratory analysis of positive and negative trends across sentences drove consistent and interesting results. On the one hand, we found that positive trends reflect a more negative consumer sentiment overall. Smyth (1998) justifies the association between more negative reviews and positive trends (e.g., decreasing negativity) on the inherent curative process of writing, which provides assimilation and understanding of the negative event. This is also in line with Pennebaker and Seagal (1999), who conceptualizes writing as a process by which people can confront upsetting topics. On the other hand, negativity trends are associated with more positive reviews. This finding consistent with empirical research suggesting that positive reviews start with their most activated emotions (e.g., “the hotel was a disaster”) and then dilutes through a constellation of supporting statements (e.g., “I had an issue with the personnel”; De Ascaniis 2013).

Corroborating Extant Research

Consumer review phenomena stimulate extensive, insightful research to uncover relations between text-based sentiments and retail performance, yet we still lack a good synthesis of the divergent sentiment analysis approaches (Schweidel and Moe 2014). In this empirical, theory-driven approach, we achieve some corroboration of extant research findings though. For example, in line with Barasch and Berger (2014) and Schweidel and Moe (2014), we confirm that the general presence of positive emotion words relates to more positive consumer sentiment overall. However, we find that specific sentiment expressions can also be context dependent in terms of the product/service and the social media platform (Schweidel and Moe 2014). For example, while implicit sentiment expressions through commissive language are very frequent in hotel reviews (e.g., “I will come back”), they are rather an exception the book evaluations (i.e., it is rather uncommon to say “I will read this book again”). In addition the heterogeneity across platforms plays an important role in how consumers express their sentiment. Product evaluations in online reviews are in average 8 sentences long, while in twitter and Facebook 1.6 and 2.7 sentences in average. As such, social media platforms force consumers to be more explicit and brief regarding their overall sentiment strength. This is in line with the significant effects of explicit and highly activated and/or boosted sentiment force indicators and the non-significant effect of implicit sentiment expressions on overall sentiment strength. The latest changes in Twitter and Facebook providing consumer more character spaces and new emoticons might be a response to the need of a more complete sentiment expression (Bloomberg 2016; Wired 2016).

Our findings that weekly sentiment changes in the verbatim consumer reviews (derived using our algorithm) influence future sales ranks also emphasize the importance of improving the accuracy or precision of sentiment analysis. First, we corroborate research by Chevalier and Mayzlin (2006; modeling details provided in appendix 1) by finding that sales on online retail sites are significantly influenced by price fluctuations. Furthermore, in line Ludwig et al. (2013), who suggest that book reviews are processed heuristically, we corroborate that consumers avoid informational overload and resort to heuristic processing, screening for the most easily accessible indicators, which are affect word cues (hence the effects of sentiment and valence). The result that particularly negative directives (being the strongest class of speech acts) impact sales is also in line with the findings of this paper, which suggest that more negative will always hurt sales more, meanwhile positivity (especially if it gets too much) gets scrutinized at some point.

We corroborate and support the latest marketing research on text mining by suggesting that the focus should extend beyond single words, to include the discourse patterns of sentences and entire paragraphs. This suggestion goes in line with moving sentiment analysis research from a “bag of words” to a “bag of sentences” (Buschken and Allenby 2015) and in turn giving researchers and managers a more comprehensive understanding of the individual and aggregated intentions (speech acts) included in product and service evaluations.

Finally, our findings link to research in psycholinguistics (Pennebaker et al. 2007). In hotel reviews, consumers use first-person pronouns with more positive sentiment, whereas in book reviews, their usage shows the opposite connection. According to Chung and Pennebaker (2007), this finding might reflect the difference in singular versus plural first-person pronouns. First person plural relates more to shared positive experiences whereas singular (e.g., “I” or “myself”) pronouns connect more to negative experiences and depression (Chung and Pennebaker 2007). In fact, we found that hotel reviews showed an almost equal use of first person pronouns in singular and plural (a ratio of 1:1), while book reviews were characterized by the use of mainly first person pronouns singular compared with plural (a ratio of 2:1). This different use of plural versus singular in the two review contexts explains why we find a positive impact in hotels and a negative one on books.

LIMITATIONS AND FURTHER RESEARCH

We note the massive potential for further studies on how different patterns of sentiment can drive subsequent consumer behavior (Ludwig et al. 2014). By theorizing about speech acts, this article informs sentiment analysis, resulting in a greater understanding of how consumers express sentiment in product and service reviews. Several limitations of our study also provide worthwhile avenues for continued research.

First, consumer research often uses direct inverses of the sentiment of a negated valence word (e.g., from positive to negative or vice versa; Ghose, Ipeirotis, and Beibei 2012). Our more granular revision of negations instead showed that for book reviews, negations of negative high expressions (e.g., “not horrible” or “not too bad”) have attenuation effects but do not reverse the meaning completely. Unlike a logical negation, a phrase such as "the service wasn't horrible" does not translate to its equivalent in positive terms, such as "it was amazing." Building on this finding, research should zoom in on the differential impacts of negations in customer reviews and social media, which could enhance understanding of the language in user-generated content.

Second, we propose a new, metric-based approach to improve understanding of sentiment expression and its components, but we do not establish a new class of probability models for sentiment analysis. This important task is beyond the scope of our paper; it also is being addressed by recent developments in computer linguistics and machine learning. In this sense, we view our work as complementary: It provides a theoretical basis for a better elaboration of sentiment analysis and other models derived from language. Regarding our dictionary approach, further research could assess the diverse implications of word taxonomies as the ones suggested by Tausczik and Pennebaker (2010) and Whissell (2009). Further research could also incorporate our findings and assess their implications in other context such as sentiment in voice or videos (Poria et al 2016) and also through other learning algorithms, such as support vector machines and hidden Markov models (Mao and Lebanon 2007; Thelwall et al. 2010).

Third, despite finding relative differences in how sentiment is expressed in book versus hotel reviews, we did not test specifically whether the different contexts prompted different sentiment expressions. According to SAT, linguistic propositions reflect considerations of the referee or subject (Searle 1969), so a book review likely features a combination of the reader's experience with the character, story, and plot, whereas sentiment toward a hotel more commonly is conveyed in terms of the customer experience. Additional research could seek to uncover the relation between sentiment and its linguistic context, possibly with nested logit models (Farley, Hayes, and Kopalle 2004).

Fourth, Luna and Peraccio (2005) note the importance of considering multiple consumer languages in marketing decisions. Although our approach only focuses on English reviews, it would be interesting to study how sentiment is expressed in different languages or different English-speaking countries, to identify implications for decoding consumer sentiments. Further research could apply SAT to assess how different types of speech acts, translated into various languages, exert distinct effects on the overall sentiment expression.

Fifth, sentiment connotations in customer reviews are not always literal. Ironic or sarcastic connotations use subtleties to communicate meanings opposite those of the actual words (Gopaldas 2014; McGraw, Warren and Kan 2015). Further research might investigate linguistic properties that characterize ironic statements, to help identify the sentiment orientation of user-generated content and enable companies to avoid erroneous sentiment predictions.

Sixth, we used regular expressions to retrieve commissive, directive, and assertive speech acts, not an exhaustive compilation of non-expressive speech acts that implicitly convey sentiment. This current approach indicated that 16% of the reviews contained at least one of these speech acts. Further text mining studies might improve the retrieval mechanisms for detecting implicit sentiment expressions. Although the automated classification of speech acts is a relatively new area (Zhang, Gao, and Li 2011), developments in the detection of varying speech acts might reveal additional implications of consumers' reviews. A recent meta-analysis (Purnawirawan et al. 2015) indicates that review valence is key for influencing further consumer recommendations, though a focus on explicit valenced language might mask the effect of commissive, directive, and assertive language.

Seventh, further research could look into the individual effects of certainty and tentative words (boosters and attenuators) when combined with valenced words (i.e., control condition) and their differential impact on sentiment. Our analysis provided an aggregated overview of positive/negative high vs. low including features of language such as negations, certainty and tentative words. However, we believe that these more granular components and other function words can be studied individually in further research. It would contribute to understand how the interaction of content words together with booster and attenuators has an impact on consumers' emotional states and behaviors.

Eight, we encourage researchers to further explore discourse patterns such as trend. Our study provides an exploratory analysis, regarding broad types trend (positive and negative), however there might be more specific types of trends such as from positive to negative, from negative to more negative or from positive to more positive, that are worth studying. Literature in argumentation patterns (e.g., consequential argumentation; Walton 1999), narrative (e.g., genre; Gergen and Gergen 1988) and also psychology literature (e.g., writing as a curative process; Pennebaker and Seagal 1999) could be helpful for researchers interested in this topic.

A final avenue for further research is to explore curvilinear effects related to extreme positive (negative) reviews or extreme variations or trends. Previous research shows curvilinear valence effects (Ludwig et al. 2014; He and Bond 2015), such that at low levels of activation, reviews drive sales, but at very high levels of activation, they do not (because review readers assumed the review writers were being irrational). It would be interesting to connect the potential curvilinear effects of incoherence with research on ambivalence, though little is known about extreme ambivalence or when consumers use high positive and negative language simultaneously to describe product and service experiences. Further analysis of the non-linear effects of incoherence (ambivalence) in customer reviews would be insightful.

IMPLICATIONS

The sheer volume of unstructured, text-based sentiments has led to intensified efforts to gauge their impact and integrate their insights into marketing (Gopaldas 2014). The latest managerial evidence (Magids, Zorfas, and Leemon 2015) suggests that online consumer sentiments represent an enormous opportunity to create new value, so companies should pursue emotional connections as a key strategy. This article illustrates the importance of speech act features for analyzing sentiment, not just to derive the writer's sentiment but also to predict its value for subsequent sales. Our Study 2 findings—that weekly sentiment changes in verbatim consumer reviews influence readers' reactions (i.e., changes in sales ranks)—emphasize the importance of moving from sentiment valence to sentiment strength.

To improve implications, researchers need to discern sentiment appropriately, rather than relying on simple valence. Sentiment is continuous (rather than either positive or negative) and requires consideration of its granular, explicit and implicit conveyance in writing. Researchers then can achieve better results in terms of decoding writers' willingness to act and readers' reactions. As we show in Study 3, the findings can be extrapolated to other contexts in which consumers share product and service experiences, without assiging stars. Our Study 2 highlights that improvements in sentiment classification have important applications for sales forcasting.

Finally, this study provides better understanding of the linguistic markers of sentiment, spanning both word use and message development. Our research offers a theory-based approach to improve understanding of consumer sentiment. This study delineates and validates general cues at each level; the speech act framework provides further guidelines for including additional, context-specific, and independent cues. At the intersection of linguistic and consumer research, these theory-driven improvements are particularly relevant, considering the growing amount of potential research insights that will stem from online, unstructured content.

REFERENCES

Aaker, Jennifer, Aimee Drolet, and Dale Griffin (2008), "Recalling Mixed Emotions," *Journal of Consumer Research*, 35(2), 268–78.

Amazon, 2014. (Accessed July 24, 2014), [available at https://www.amazon.co.uk/gp/help/customer/display.html?nodeId=13841791].

Auramäki, Esa, Erkki Lehtinen, and Kalle Lyytinen (1988), "A Speech-Act-Based Office Modeling Approach," *ACM Transactions on Information Systems (TOIS)*, 6(2), 126–52.

Barasch, Alexandra and Jonah Berger (2014), "Broadcasting and Narrowcasting: How Audience Size Affects What People Share," *Journal of Marketing Research*, 51(3), 286-299.

Baumeister, Roy F., Kathleen D. Vohs, C. Nathan DeWall, and Liqing Zhang (2007), "How Emotion Shapes Behavior: Feedback, Anticipation, and Reflection, Rather than Direct Causation," *Personality and Social Psychology Review*, 11(2), 167–203.

Berger, Jonah, and Katherine L. Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research*, 49 (2), 192-205.

———, Alan T. Sorensen, and Scott J. Rasmussen (2010), "Positive Effects of Negative Publicity: When Negative Reviews Increase Sales," *Marketing Science*, 29(5), 815–27.

Bloomberg, 2016. (Accessed June 1, 2014), [available at http://www.bloomberg.com/news/articles/2016-05-16/twitter-to-stop-counting-photos-and-links-in-140-character-limit].

Bosco, Francesca, Monica Bucciarelli, and Bruno G. Bara (2004), "The Fundamental Context Categories in Understanding Communicative Intention," *Journal of Pragmatics*, 36(3), 467–88.

Buschken, Joachim and Greg M. Allenby (2015), "Sentence-Based Text Analysis for Customer Reviews." *Marketing Science*, Forthcoming.

Cambria, Erik, Jie Fu, Federica Bisio, and Soujanya Poria (2015), "AffectiveSpace 2: Enabling Affective Intuition for Concept-Level Sentiment Analysis," In AAAI, 508-514.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43(3), 345–54.

Chomsky, Noam (2012). (Talks at Google, accessed Dec 1 2015) [available at https://www.youtube.com/watch?v=Y3PwG4UoJ0Y].

Chung, Cindy and James Pennebaker (2007), "The Psychological Functions of Function Words," in *Social Communication*, K. Fiedler, ed. New York: Psychology Press, pp 343-59.

D'Andrade, R. G., and M. Wish (1985), "Speech Act Theory in Quantitative Research on Interpersonal Behavior," *Discourse Processes*, 8(2), 229-259.

Das, Sanjiv and Mike Chen (2007), "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," *Management Science*, 53(9), 1375–88.

De Ascaniis, Silvia, and Ulrike Gretzel (2013), "Communicative functions of Online Travel Review titles: a pragmatic and linguistic investigation of destination and attraction OTR titles," *Studies in Communication Sciences*, 13(2), 156-165.

De Saussure, Louis (2007), "Pragmatic Issues in Discourse Analysis," *Critical Approaches to Discourse Analysis Across Disciplines*, 1(1), 179–95.

Dimensional Research (2013), (Accessed December 10, 2014) [http://marketingland.com/survey-customers-more-frustrated-by-how-long-it-takes-to-resolve-a-customer-service-issue-than-the-resolution-38756].

Dubois, David, Derek D. Rucker, and Adam D. Galinsky (2016), "Dynamics of Communicator and Audience Power: The Persuasiveness of Competence versus Warmth," *Journal of Consumer Research*, 43(1), 68-85.

Farley, John U., Andrew F. Hayes, and Praveen K. Kopalle (2004). "Choosing and Upgrading Financial Services Dealers in the US and UK," *International Journal of Research in Marketing*, 21(4), 359-75.

Feldman, Ronen (2013) "Techniques and Applications for Sentiment Analysis," *Communications of the ACM*, 56(4), 82-89.

———, Ronen and James Sanger (2007), *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press.

Feng, Vannessa and Graeme Hirst (2013), "Patterns of Local Discourse Coherence as a Feature for Authorship Attribution," *Literary and Linguistic Computing*, 29(2), 191-198.

Fonic, Nick (2003), "From Speech Acts to Speech Actions," in *Ideas About Social Reality: Extensions, Criticisms, and Reconstructions*, John Searle and B. Smith, ed. Cambridge: Cambridge University Press, 34–51.

Ghose, Anindya, Panagiotis Ipeirotis, and Li Beibei (2012), "Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content," *Marketing Science*, 31(3), 493–520.

Giesler, Markus (2012), "How Doppelganger Brand Images Influence the Market Creation Process: Longitudinal Insights from the Rise of Botox Cosmetic," *Journal of Marketing*, 76 (6), 55–68.

Godes, David, and José C. Silva (2012), "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science*, 31(3), 448-473.

Goldberg, Andrew B. and Xiaojin Zhu (2006), "Seeing Stars When There aren't Many Stars: Graph-Based Semi-Supervised Learning for Sentiment Categorization," in Proceedings of HLT-NAACL 2006 Workshop on Textgraphs: Graph-Based Algorithms for Natural Language Processing, 45–52.

Gopaldas, Ahir (2014), "Marketplace Sentiments," *Journal of Consumer Research*, 41(4), 995–1014.

Harrell, F.E. (2001), *Regression Modeling Strategies*. New York: Springer-Verlag.

He, Stephen X., and Samuel D. Bond, (2015), "Why is the crowd divided? Attribution for dispersion in online word of mouth," *Journal of Consumer Research*, 41 (6), 1509-1527.

Hennig-Thurau, Thorsten, Caroline Wiertz and Fabian Feldhaus (2014), "Does Twitter Matter? The Impact of Microblogging Word of Mouth on Consumers' Adoption of New Movies," *Journal of the Academy of Marketing Science*, 43(3), 375-94.

Holmes, Janet (1984), "Modifying Illocutionary Force," *Journal of Pragmatics*, 8, 345–65.

Homburg, Christian, Laura Ehm, and Martin Artz (2015), "Measuring and Managing Consumer Sentiment in an Online Community Environment," *Journal of Marketing Research*, Forthcoming.

Hu, Nan, Noi Sian Koh, and Srinivas K. Reddy (2014), "Ratings Lead You to the Product, Reviews Help You Clinch It? The Mediating Role of Online Review Sentiments on Product Sales," *Decision Support Systems*, 57, 42–53.

Khan, Aurangzeb, Baharun Baharudin, and Khairullah Khan (2011), "Sentiment Classification Using Sentence-Level Lexical Based Semantic Orientation of Online Reviews," *Trends in Applied Sciences Research*, 6(10), 1819–1157.

Kronrod, Ann, and Shai Danziger (2013), "'Wii Will Rock You!' The Use and Effect of Figurative Language in Consumer Reviews of Hedonic and Utilitarian Consumption," *Journal of Consumer Research*, 40(4), 726–39.

Long, J. S. (1997), *Regression Models for Categorical and Limited 5 Dependent Variables*, Thousand Oaks, CA: Sage Publications.

Ludwig, Stephan, Ko De Ruyter, Mike Friedman, Elisabeth C. Brüggen, Martin Wetzels, and Gerard Pfann (2013), "More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates," *Journal of Marketing*, 77(1), 87–103.

Luna, David, and Laura A. Peracchio (2005), “Advertising to Bilingual Consumers: The Impact of Code-Switching on Persuasion,” *Journal of Consumer Research*, 31(4), 760-65.

Magids, Scott, Alan Zorfas and Daniel Leemon (2015), “The New Science of Customer Emotions,” *Harvard Business Review*, 76(November), 66-74.

Mao, Yi and Guy Lebanon (2009), “Generalized Isotonic Conditional Random Fields,”*Machine Learning*, 77, 961-68.

Maas, Andrew L., Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts, (2011), “Learning word vectors for sentiment analysis,” *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 142-150.

Maks, Isa, and Piek Vossen (2012), “A lexicon model for deep sentiment analysis and opinion mining applications,” *Decision Support Systems*, 53(4), 680-688.

McGraw, A. Peter, Caleb Warren, and Christina Kan (2015), “Humorous Complaining,” *Journal of Consumer Research*, 41(5), 1153–1171.

McKinsey Company (2013), “Ratings Revisited: Textual Analysis for Better Risk Management,” (Accessed July 15, 2014), [avaliable at http://www.mckinsey.com/insights/risk_management/ratings_revisited_textual_analysis_for_better_risk_management].

Montoyo, Andres, Patricio Martínez-Barco, and Alexandra Balahur (2012), “Subjectivity and Sentiment Analysis: An Overview of the Current State of the Area and Envisaged Developments,” *Decision Support Systems*, 53(4), 675–79.

Moore, Sarah G. (2012), “Some Things Are Better Left Unsaid: How Word of Mouth Influences the Storyteller,” *Journal of Consumer Research*, 38(6), 225–48.

Mudambi, Susan M., and David Schuff (2010), “What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com,” *MIS Quarterly*, 34(1), 185–200.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), “Mine Your Own Business: Market-Structure Surveillance Through Text Mining,” *Marketing Science*, 31(3), 521–43.

Norrick, Neal R. (1978), “Expressive Illocutionary Acts,” *Journal of Pragmatics*, 2, 277–91.

Otnes, Cele, Tina M. Lowrey, and L. J. Shrum (1997), “Toward an Understanding of Consumer Ambivalence,” *Journal of Consumer Research*, 24(1), 80-93.

Pang, Bo, and Lillian Lee (2005) "Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales," *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 115-124.

——— and ——— (2008), "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.

Pennebaker, James W., and Janel D. Seagal (1999), "Forming a story: The health benefits of narrative," *Journal of clinical psychology*, 55(10), 1243-1254.

———, Cindy K. Chung, Molly Ireland, Amy Gonzales, and Roger Booth (2007), *The Development and Psychometric Properties of LIWC2007*. Austin, TX: LIWC.net.

———, Matthias R. Mehl, and Kate G. Niederhoffer (2003), "Psychological Aspects of Natural Language Use: Our Words, Our Selves," *Annual Review of Psychology*, 54(1), 547–77.

Perrault, C. Raymond and James Allen (1980), "A Plan-Based Analysis of Indirect Speech Acts," *Computational Linguistics*, 6(3), 167–82.

Pinker, Steven, Martin Nowak, and James J. Lee (2008), "The Logic of Indirect Speech," *Proceedings of the National Academy of Sciences of the United States of America*, 105(3), 833–38.

Poria, Soujanya, Erik Cambria, Newton Howard, Guang-Bin Huang, and Amir Hussain (2016), "Fusing audio, visual and textual clues for sentiment analysis from multimodal content," *Neurocomputing*, 174, 50-59.

Puccinelli, Nancy M., Keith Wilcox, and Dhruv Grewal (2015), "Consumers' Response to Commercials: When the Energy Level in the Commercial Conflicts with the Media Context," *Journal of Marketing*, 79(2), 1-18.

Purnawirawan, Nathalia, Martin Eisend, Patrick De Pelsmacker, and Nathalie Dens (2015), "A Meta-Analytic Investigation of the Role of Valence in Online Reviews," *Journal of Interactive Marketing*, 31 (August), 17-27.

Richins, Marsha L. (1997), "Measuring Emotions in the Consumption Experience," *Journal of Consumer Research*, 24(2), 127-46.

Rucker, Derek D., Blakeley B. McShane, and Kristopher J. Preacher (2015), "A researcher's guide to regression, discretization, and median splits of continuous variables," *Journal of Consumer Psychology*, 25(4), 666-678.

Russell, James A. and L.F. Barret (1999), "Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant," *Journal of Personality and Social Psychology*, 76(5), 805–19.

Sbisa, Marina (2001), "Illocutionary Force and Degrees of Strength in Language Use," *Journal of Pragmatics*, 33(12), 1791–814.

Schellekens, Gaby A. C., Peeter W. J. Verlegh, and Ale Smidts (2010), "Language Abstraction in Word of Mouth," *Journal of Consumer Research*, 37(2), 207–23.

Schumaker, Robert P., Yulei Zhang, Chun-Neng Huang, and Hsinchun Chen (2012), "Evaluating Sentiment in Financial News Articles," *Decision Support Systems*, 53(3), 458-64.

Schweidel, David, and Wendy Moe (2014), "Listening In on Social Media: A Joint Model of Sentiment and Venue Format Choice," *Journal of Marketing Research*, 51(4), 387-402.

Searle, John R. (1969), *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

——— (1975), "Indirect Speech Acts," in ACM SIGART Bulletin, *Syntax and Semantics*, P. Cole and J. L. Morgan, eds. Academic Press, New York, 59–82.

——— (1976), "A Classification of Illocutionary Acts," *Language in Society*, 5(1), 1–23.

Smith, Craig A., and Phoebe C. Ellsworth (1985), "Patterns of Cognitive Appraisal in Emotion," *Journal of Personality and Social Psychology*, 48(4), 813-38.

Smyth, Joshua M (1998), "Written emotional expression: effect sizes, outcome types, and moderating variables," *Journal of consulting and clinical psychology*, 66(1), 174-184.

Stanford Parser (2014), (Accessed July 15, 2014), [avaliable at http://nlp.stanford.edu:8080/parser/].

Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede (2011), "Lexicon-Based Methods for Sentiment Analysis," *Computational Linguistics*, 37(2), 267–307.

Tang, Tanya, Erik Fang, and Feng Wang (2014), "Is Neutral Really Neutral? The Effects of Neutral User-Generated Content (UGC) on Product Sales," *Journal of Marketing*, 78(4), 1–51.

Tausczik, Yla R. and James W. Pennebaker (2010), "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods," Journal of Language and Social Psychology, 29 (1), 24-54.

Thelwall, Mike, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas (2010). "Sentiment Strength Detection in Short Informal Text," *Journal of the American Society for Information Science and Technology*, 61(12), 2544–58.

Thomas, Gloria (1992), “The Influence of Processing Conversational Information on Inference, Argument Elaboration, and Memory,” *Journal of Consumer Research,* 19(1), 83-92.

Tirunillai, Seshadri and Gerard Tellis (2012), “Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance,” *Marketing Science,* 31(2), 198–215.

Tsang, Alex and Gerard Prendergast (2009), "Is a ‘Star’ Worth a Thousand Words?: The Interplay Between Product-Review Texts and Rating Valences," *European Journal of Marketing,* 43(11/12), 1269-80.

Van Djik, Teun (1997). *Discourse as Structure and Process*. London: Sage Publications Ltd.

Villarroel, O., Francisco, Babis Theodulidis, Jamie Burton, Thorsten Gruber, and Mohamed Zaki, (2014), “Analyzing Customer Experience Feedback Ussing Text Mining: A Linguistic Based Approach,” *Journal of Service Research*, 17(3), 278-95.

Whissell, Cynthia (2009), “Using the Revised Dictionary of Affect in Language to Quantify the Emotional Undertones of Samples of Natural Language,” *Psychological Reports*, 105(2), 509-521.

Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann (2009), "Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis," *Computational Linguistics,* 35(3), 399-433.

Williams, Richard (2006), “Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables,” *Stata Journal*, 6(1), 58-82.

Wired, 2016. (Accessed June 1, 2014), [available at http://www.wired.com/2016/02/facebook-reactions-totally-redesigned-like-button/].

Xiong, Guiyang, and Sundar Bharadwaj (2013), "Asymmetric Roles of Advertising and Marketing Capability in Financial Returns to News: Turning Bad into Good and Good into Great," *Journal of Marketing Research,* 50(6), 706-24.

Zhang, Renxian, Dehong Gao, and Wenjie Li (2011), “What Are Tweeters Doing: Recognizing Speech Acts in Twitter,” in *Analyzing Microtext*, San Francisco: AAAI Press, 86–91.

TABLE 1. Empirical Studies Using Sentiment Analysis and Considerations of SAT

TABLE 1
Comparison of Empirical Studies Using Sentiment Analysis and Their Consideration of SAT Features

Authors	Context	Outcome Variable of Sentiment	Valence Condition	Illocutionary Force Features	Implicit Sentiment	Speech Act Patterns
Pang and Lee (2005)	Improve accuracy in sentiment analysis	Four star classification	Sentence polarity (positive/negative)	Not considered for analysis	Not considered for analysis	Not considered for analysis
Das and Chen (2007)	Use sentiment to predict stock prices	Positive, negative and neutral	Positive / negative / neutral words dictionary	Negations handled through a dictionary	Not considered for analysis	Not considered for analysis
Wilson, Wiebe, and Hoffmann (2009)	Improve accuracy in sentiment analysis	Positive, negative, both and neutral	Positive / negative / neutral / both words dictionary	Polarity modifiers (e.g., not) and shifters (e.g., very, lack of)	Considered to some extent by the analysis of context words	Not considered for analysis
Khan, Baharudin and Khan (2011)	Improve accuracy in sentiment analysis	Positive, negative and neutral	Positive, negative and neutral sentences (Bag of sentences)	Subjective or opinionated words, negations, shifters, boosters and attenuators	Not considered for analysis	Contextual features of sentence structure
Maas et al. (2011)	Improve accuracy in sentiment analysis	Positive v/s negative	Based on word similarities and polarity probability	It assess the strength of word similarities	Not considered for analysis	Not considered for analysis
Taboada et al. (2011)	Improve accuracy in sentiment analysis	Positive v/s negative	Positive / negative words	Word strength considering, part of speech, negations, boosters and attenuators	Not considered for analysis	Not considered for analysis
Berger and Milkman (2012)	Use sentiment to predict e-WOM	Positive v/s negative	Positive / negative words dictionary	Not considered for analysis	Not considered for analysis	Not considered for analysis
Tirunillai and Tellis (2012)	Uses reviews valence to predict stock price	Positive v/s negative	Positive / negative words dictionary	Not considered for analysis	Not considered for analysis	Not considered for analysis

Ghose, Ipeirotis, and Beibei (2012)	Uses hotel reviews to design hotel rankings	From -3 (very negative) to +3 (very positive)	Not considered for analysis	Considered by measuring sentiment in phrases with a scale from -3 to 3 (AMT). Negation considered	Not considered for analysis	Not considered for Analysis
Maks and Vossen (2012)	Political texts	Positive, negative, both and neutral	Positive / negative / neutral words	Not Considered for Analysis	Indirect expressive verbs to detect subjectivity	Not considered for analysis
Schumaker, Zhang, Huang, and Cheng (2012)	Uses news' sentiment to predict stock prices	Positive, negative and neutral	Positive / negative words dictionary	Not considered for analysis	Not considered for analysis	Not considered for analysis
Xiong and Bharadwaj (2013)	Uses news' sentiment to predict stock prices	Positive v/s negative	Positive / negative words dictionary	Negations and modifiers handled through a dictionary	Not considered for analysis	Not considered for analysis
Schweidel and Moe (2014)	Validation of an aggregated online sentiment measure	Positive, negative and neutral	Manually coded posts ; validated through positive / negative words dictionary	Not considered for analysis	Not considered for analysis	Not considered for analysis
Homburg, Ehm, and Artz (2015)	Social Media Virtual Communities	Positive v/s negative	Manually coded words into positive and negative	Not considered for analysis	Not considered for analysis	Not considered for analysis
Cambria et al. (2015)	Improve accuracy in sentiment and emotion analysis	Positive v/s negative; also single emotions (e.g., grief or joy)	Positive / Negative and twenty four emotion words for clustering	Punctuation, negation, boosters and attenuators, emoticons, single emotions (e.g., joy)	Not considered for analysis	Not considered for analysis
Poria et al. (2016)	Improve accuracy in sentiment analysis in text and videos	Positive, negative and neutral	Word, audio and video polarity	Word polarity ranging from -1 to 1, single emotions (e.g., joy), negations, modifiers	Facial expressions and voice strength	Not considered for analysis

TABLE 2. Review of Construct Definitions, Examples, and Representative Studies

Speech Act Features	Construct	Definitions	Word and Sentence Patterns	Examples	Representative Papers
Sentiment force	High	Consumer is strongly expressing positive (negative) sentiment.	High activation words; High activation + certainty words; Low activation +certainty words	I was amazing; It was really amazing; It was really good.	Searle (1976); Holmes (1982); Sbisa (2001)
	Low	Consumer is weakly expressing positive (negative) sentiment.	Low activation words; Low activation + tentative words; High activation +tentative words; Negations + high and low activation	It was nice; It was kind of nice; I was kind of awesome; It wasn't bad; It wasn't horrible.	
Implicit Sentiment	Directive	Consumer is (not) recommending to other consumers.	First Person Pronoun + Conditional + Directive Verb	I will recommend it; I suggest you to go; I advise you to buy.	Pinker, Nowak, and Lee (2008); Searle (1975, 1976)
	Commissive	Consumer is (not) committing to (re)patronage in the future.	First Person Pronoun + Future tense + Contextual verb	I will come back; I would read it again; I will continue buying.	
	Assertive	Consumers are making an affirmative (negative) statement about their experience.	First Person Pronoun + Assertive Verb + Contextual Noun(phrase)	We had a view; We didn't have to wait; I read it in a day.	
Discourse Patterns of Sentiment	Incoherence	Consumer level of sentiment ambivalence in a review.	Degree of variation of positivity in reviews of 2 or more sentences	The service was amazing. However the breakfast was kind poor. Not sure if we will come back.	van Dijk 1997; Auramäki, Lehtinen, and Lyytinen (1988); (Fonic 2003)
	Positive Trend	Consumer incremental positivity as the review unfolds.	Sentiment slope in reviews of 3 or more sentences	The service was horrible. We were not expecting it. But for that price is okay.	van Dijk 1997; de Saussure (2007)
	Negative Trend	Consumer detrimental positivity as the review unfolds.	Sentiment slope in reviews of 3 or more sentences	The service was great. We were expecting it. The price was too high though.	van Dijk 1997; de Saussure (2007)

TABLE 3. Study 1 Results Ordinal Logit Model

MODELS	Model 1 (a): Negations Separated		Model 1 (b): Negation Aggregated		Model 2: Indirect Speech Acts		Model 3: Discourse Patterns	
Variables	Hotel	Books	Hotel	Books	Hotel	Books	Hotel	Books
Positive High (PH)	0.90**	0.92**	0.89**	0.93**	0.88**	0.93**	1.03**	1.12**
Negative High (NH)	-0.44**	-0.31**	-0.44**	-0.31**	-0.43**	-0.31**	-0.37**	-0.26**
Positive Low (PL)	0.03**	0.10**	0.05**	0.11**	0.04**	0.11**	0.06**	0.18**
Negative Low (NL)	-0.37**	-0.30**	-0.40**	-0.37**	-0.39**	-0.36**	-0.37**	-0.32**
Neg_Positive High (Neg_PH)	-0.07**	-0.10**						
Neg_Negative High (Neg_INH)	0.01	-0.07**						
Neg_Positive Low (Neg_IPL)	-0.15**	-0.14**						
Neg_Negative Low (INL)	0.10**	0.05**						
Commissive Positive (CP)					0.09**	0.05**	0.13**	0.07**
Directive Positive (DP)					0.07**	0.28**	0.09**	0.38**
Assertive Positive (AP)					-0.00	0.05**	-0.001	0.04**
Commissive Negative (CN)					-0.15**	-0.17**	-0.12**	-0.11**
Directive Negative (DN)					-0.15**	-0.28**	-0.13**	-0.22**
Assertive Negative (AN)					-0.07**	-0.06**	-0.05**	-0.05**
Incoherence (SD)							-0.16**	-0.17**
Positive Trend (PT)							-0.14**	-0.14**
Negative Trend (NT)							0.08**	0.12**
Total Sentences (TSent)							-0.07**	-0.01
First Person Pronouns (FP)	0.24**	-0.06**	0.24**	-0.06**	0.24**	-0.06**	0.19**	-0.10**
Dummy Review Site (D_RS)		0.07**		0.07**		0.07**		0.06**
AIC Ordinal-Logit	46908.2	45918.1	47009.4	45906.7	46713.4	45474.3	44604.8	40508.3
Sample size	24033	19654	24033	19654	24033	19654	23086	17371

†*p* < .1 **p* < .05. ***p* < .01.

Note: Coefficients in Model 1, 2, and 3 are log-odd probabilities; the dependent variable was the ordinal star rating. Validation results are beta coefficients from ordinary least squares regressions, and the dependent variable was an average response from 1 to 5, according to 10 Amazon Mechanical Turk participants per review. All variables were standardized before the ordinal regression.

TABLE 4. Robustness Check, Amazon Mechanical Turk

Variables	Books Model 3
Positive High (PH)	0.25**
Negative High (NH)	-0.14**
Positive Low (PL)	0.06**
Negative Low (NL)	-0.17**
Commissive Positive (CP)	0.01
Directive Positive (DP)	0.09**
Assertive Positive (AP)	0.02 †
Commissive Negative (CN)	-0.10**
Directive Negative (DN)	-0.12**
Assertive Negative (AN)	-0.03
Incoherence (SD)	-0.05**
Positive Trend (T)	-0.03
Negative Trend(NT)	0.03 †
Total Sentences (TSent)	-0.01
First Person Pronouns (FP)	-0.01
Dummy Review Site (D_RS)	0.05**
Intercept	0.74**
R-Squared	0.27

†$p < .1$ *$p < .05$. **$p < .01$.

TABLE 5. Generalizing to Other Social Media (Study 3: Facebook and Twitter)

MODELS Variables (Standardized)	Valenced Word Counts	Model 3
Positive Proportion	0.42**	
Negative Proportion	0.05	
Positive High (PH)		.60**
Negative High (NH)		-.42**
Positive Low (PL)		.31**
Negative Low (NL)		-.03
Indirect Positive Proportion		.03
Indirect Negative Proportion		-.07
Incoherence (SD)		-.18 †
Positive Trend(PT)		.04
Negative Trend(NT)		.03
Total Sentences (TSent)		-.08**
First Person Pronouns		.16*
Dummy Retail	0.02	.18**
Dummy Health	0.21	.26**
Dummy Media	-0.19	.24**
Dummy Telecommunication	-0.17	.19**
Dummy Travel	-0.30	-.00
Dummy Social Media Type	-0.07	.33**
AIC Ordinal-Logit	3435.22	3301.7

†*p* < .1 **p* < .05. ***p* < .01.

Note: The coefficients in Models 1, 2, and 3 are log-odd probabilities; the dependent variable was the coded star rating (two independent coders, Krippendorff 55.4%; disagreement was resolved through discussion). All variables were standardized before the ordinal regression.

Appendix 1: Methodological Details for Study 2

We aimed to follow the approach suggested by Chevalier and Mayzlin (2006) as closely as possible. Accordingly, we first cleaned the sample first. Amazon updates sales ranks daily for products that achieve rankings of 100,000 or below; for all others, it updates them monthly. Therefore, we removed all books below a sales rank of 100,000 during the observation period. Barnes & Noble records sales ranks up to 650,000 and updates all of these products daily. We removed any books for which there was no sales rank recorded on BN during the observational period. We also removed books that did not launch on both sites in the same week. This data screening reduced our sample to 352 books with an average of 9.2 weekly observations. Neither site supplies actual book sales, so we approximated weekly sales with the natural log of the weekly sales ranks (Chevalier and Mayzlin 2006). We also took the natural log of the weekly book price and the total number of reviews on the respective book site. Using the log odds coefficients to predict the review sentiment derived using positive and negative valence (i.e., proportion of positive and negative words per review obtained from the LIWC dictionaries of positive and negative emotions) and the sentiment strength from our algorithm in model 3, we established two overall scores per review:

$$Valence_i = 0.43 * Proportion\ Positive\ Valence_i - 0.26 * Proportion\ Negative\ Valence_i \quad (2.1)$$

$$Sentiment\ Strength_i = 1.07 * PH_i - 0.28 * NH_i + 1.3 * PL_i - 0.32 * NL_i + 0.05 * CP_i + 0.33 * DP_i + 0.05 * AP_i - 0.16 * CN_i - 0.26 * DN_i - 0.05 * AN_i - 0.15 * SD_i - 0.06 * TP_i + 0.05 * TN_i - 0.02 * TotakSent_i - 0.09 * FPP_i + 0.07 * D\_RS_i \quad (2.2)$$

We then aggregated the overall sentiment scores across all consumer reviews for the same book (z) in a given week (t) to derive a mean level of valence and sentiment for each book in each week separately, one for Amazon.com and one for BN.com. In addition to the influences of the time-varying drivers of sales performance (e.g., price), we expect unobservable, fixed (time-invariant) effects to correlate with the independent variables (e.g., author's fame). Omitting these effects would bias the coefficients of our model. Moreover, potential subtle differences between the two retail sites, in terms of their users' preferences and structure, may exist. To overcome such biases, we difference the records of sales ranks across sites and across time, then deduct the previous (lagged) level of each explanatory variable from the current one (Chevalier and Mayzlin 2006). To capture the influence of the explanatory variables, at the week and book difference levels, on weekly changes in sales differentials, we specified a hierarchical linear model (HLM), which accounts for weekly interdependencies between observations for the same book and simultaneously allows for investigations of cross-level effects (Long 1997). With multiple weeks observed for each book, the HLM approach also controls appropriately for the possibility that changes in the sentiment of the reviews, the number of reviews posted, and the price changes on the same book site may be more similar than they are for changes on other book sites. Therefore, for sentiment the model is estimated as follows:

$$\Delta[\ln(rank\ Amazon.com_{zt}) - [\ln(rank\ BN.com_{zt})] = \beta_0 + \beta_1\Delta(Price\ Amazon.com_{zt-1}) + \beta_2\Delta\ln(Price\ BN.com_{zt-1}) + \beta_3\Delta\ln(Review\ Amount\ Amazon.com_{zt-1}) + \beta_4\Delta\ln(Review\ Amount\ BN.com_{zt-1}) + \beta_5\Delta\ln(sentiment\ Amazon.com_{zt-1}) + \beta_5\Delta\ln(sentiment\ BN.com_{zt-1}) + \mu_{0t} + \mu_{1t}week + \epsilon_{zt-1} \quad (2.3)$$

In this model, z is the book, and t indicates the week. Our dependent variable is the change from the previous week in the difference between Amazon and BN for the ln sales rank. Following Chevalier and Mayzlin (2006), for the fixed portion of our model, we control for the respective changes in price and the amount of reviews on each site in the previous week (t – 1) to maintain causality implications. This approach also eliminates book site–specific fixed effects. We allow for a random slope (u_{1t}) for each week, to account for the typical decline in sales along the product life cycle, and we assume an independent covariance structure for the random effects (u_{0t}; u_{1t}). Note that we have also conducted tests for the implicit speech acts influence on sales (i.e., assertives, commissives, directives, positive and negative trends, and incoherence) yet failed to find any significant influence with the exception of negative directives (e.g., "do not buy this book") in the consumer reviews on Amazon which increase the sales rank of the respective book site (e.g., decrease the sales), in the book review setting (please see online Appendix for results).

Model A: Valence N (reviews) = 3249, groups (books) = 352, min obs per group 1, max 16, average 9.2, Wald χ^2 =53.65, LL= -2502.92

Variables	Coefficient	Std.Err	z	P>|z|
Δ Valence Amazon $_{it-1}$	-0.020	0.009	-2.180	0.029
Δ Valence BN$_{it-1}$	0.017	0.009	1.860	0.063
Δ Amazon.com (Price) $_{it-1}$	0.145	0.051	2.850	0.004
Δ BN.com (Price) $_{it-1}$	-0.063	0.035	-1.820	0.069
Δ Amazon.com (# of reviews) $_{it-1}$	-0.006	0.016	-0.370	0.714
Δ BN.com (# of reviews) $_{it-1}$	-0.014	0.010	-1.360	0.175
Week	0.013	0.002	5.450	0.000

Model B: Sentiment N (reviews) = 3249, groups (books) = 352, min obs per group 1, max 16, average 9.2, Wald χ^2 =80.69, LL= -2489.67

Variables	Coefficient	Std.Err	z	P>|z|
Δ Sentiment Amazon$_{it-1}$	-0.028	0.011	-2.54	0.011
Δ Sentiment BN$_{it-1}$	0.024	0.011	2.16	0.031
Δ Amazon.com (Price) $_{it-1}$	0.153	0.051	3.01	0.003
Δ BN.com (Price) $_{it-1}$	-0.063	0.035	-1.80	0.071
Δ Amazon.com (#of reviews) $_{it-1}$	-0.016	0.010	-1.56	0.119
Δ BN.com (# of reviews) $_{it-1}$	-0.005	0.016	-0.35	0.729
Week	0.013	0.002	5.49	0.001

Notes: The final sample is the set of 352 books launched on both sites in April–May 2010. The dependent variable is Δ[ln(rankAmazon.com$_{it}$) – ln(rank BN.com$_{it}$)]. All variables are standardized.