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

FRANCISCO VILLARROEL ORDENES

STEPHAN LUDWIG

KO DE RUYTER

DHRUV GREWAL

MARTIN WETZELS

Francisco Villarroel Ordenes (fvillarroelo@isenberg.umass.edu) is an Assistant Professor of Marketing, Isenberg School of Management, Department of Marketing, University of Massachusetts Amherst, 121 Presidents Drive, Amherst, 01003 MA., Tel: +1 413-545-4941. Stephan Ludwig (s.ludwig@surrey.ac.uk) is an Associate Professor of Marketing, Department of Retail Marketing and Management, University of Surrey, Guildford, Surrey GU2 7XH, UK. Ko de Ruyter (Ko.De-Ruyter.1@city.ac.uk) is a Professor of Marketing, Cass Business School, City University London, 106 Bunhill Row, EC1Y 8TZ London, United Kingdom. Dhruv Grewal (dgrewal@babson.edu) is The Toyota Chair in Commerce and Electronic Business and a Professor of Marketing, Babson College, Babson Park, MA 02457-0310, Tel: +1 781-239-3902. Martin Wetzels (m.wetzels@maastrichtuniversity.nl) is a 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. The present article is based on the first author's doctoral dissertation at Maastricht University. The authors thank the editor, associate editor for their excellent guidance and the three anonymous reviewers for their valuable feedback during the review process. Supplemental materials appear in a web appendix.

Gita Johar served as editor, and Praveen Kopalle served as associate editor for this article.

ABSTRACT

Deciphering consumer's sentiment expressions from Big Data (e.g., online reviews) has become a managerial priority to monitor product and service evaluations. However, Sentiment Analysis, the process of automatically distilling sentiment from text, provides little insight regarding the language granularities beyond the use of positive and negative words. Drawing on Speech Act Theory, this study provides a fine-grained analysis of the implicit and explicit language used by consumers to express sentiment in text. An empirical text mining study using more than 45,000 consumer reviews, demonstrates the differential impacts of activation levels (e.g., tentative language), implicit sentiment expressions (e.g., commissive language), and discourse patterns (e.g., incoherence) on overall consumer sentiment (i.e., star ratings). In two follow-up studies, we demonstrate that these speech act features also influence the readers' behavior and are generalizable to other social media contexts such as Twitter and Facebook. We contribute to research on consumer sentiment analysis by offering a more nuanced understanding of consumer sentiments and their implications.

Keywords: Consumer Sentiment, Speech Act Theory, Text Mining, Online Reviews, Sales Ranks, Social Media.

"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 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. This oversimplification hides that written language offers consumers a wider range of explicit and implicit linguistic features and patterns to express their sentiment (Gopaldas 2014). In turn, neglecting such linguistic means of sentiment expression prohibits a more accurate understanding of how verbatim consumer reviews influence the reading consumer and sales performance (Ludwig et al 2013).

We suggest that Speech Act Theory (SAT) might offer meaningful 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 features. For example, SAT introduces the linguistic properties that alter the strength of words' meanings (Holmes 1984; Sbisa 2001). In addition to the activation level inherent to emotion words (e.g., "good" vs. "awesome"; Russell and Barret 1999), phrases might exert stronger sentiment when they include certainty terms (e.g., "arrived *extremely* late"), or they might be attenuated by tentative wording (e.g., "it was *kind of* nice"). The differential effects these types of expressions have on overall sentiment strength remain uninvestigated (Packard and Berger 2016). 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). In fact, there is a limited understanding about the distinct impacts of recommendations (e.g., "You must read this book") versus statements (e.g., "We got a discount") on overall sentiment strength. Finally, consistent with research on mixed emotions (Aaker, Drolet, and Griffin 2008) and advances in text mining (Büschken and Allenby 2016), perhaps discourse patterns convey meaning, beyond that implied by the individual sentences and words. For example, sentiment incoherence (e.g., high variability of sentiment across sentences) and trends in a message might influence the overall tone of a review (Goldberg and Zhu 2006; van Dijk 1997). Drawing on SAT, we investigate the differential and asymmetric effects of explicit and implicit expressions, and the direct effects of discourse patterns on consumer sentiment strength, which enables us to offer three main research contributions.

First, we advance research on affect by empirically studying explicit sentiment expressions in online reviews, including the level of activation in emotion words (e.g., "good" vs. "awesome"), boosters (e.g., "very good") and attenuators (e.g., "kind of good"). In practice, we specify how explicit sentiment expressions relate to consumers' 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 got an upgrade in the hotel") on overall sentiment strength. Third, noting that most arguments develop across a series of sentences, we demonstrate how their relative incoherence in sentiment expressions 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 features on writers' overall sentiment strength (i.e., review star rating), and then assess them empirically using a unique data set of 45,843 online reviews. Furthermore, we demonstrate that the language features of consumer sentiment strength exhibit a stronger relation with reading consumers purchase behavior than simply valence words; we also demonstrate the generalizability of our findings in other social media contexts 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 Gopaldas (2014), 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 valence words on behaviors (Berger, Sorensen, and Rasmussen 2010; Hennig-Thurau, Wiertz and Feldhaus 2014; Ludwig et al 2013). However, we posit that these valenced words mask the effects of further language granularities, such as the strength with which consumers express their sentiment (Thelwall et al. 2010). To go beyond the simple valence, we build on SAT as an enabling framework to propose a number of novel predictions.

Speech acts are utterances that function to communicate 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 speech acts is an emotional response to a particular situation (Norrick 1978). Sentiment strength is communicated through speech acts which refer to a subject (to which the sentiment refers) and include either explicit or implicit acts and their activation level (Holmes 1984; Norrick 1978). Furthermore, sentiment strength is also conveyed through discourse patterns of explicit and implicit sentiment expression. Few consumer research studies acknowledge the importance of speech act features for deriving consumer intentions though (Thomas 1992), and existing consumer research on sentiment analysis neglects the inherent strength aspects. Past research has used 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 about here

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 explicit and implicit sentiment expressions, their activation level and higher order discourse patterns on the overall sentiment strength.

Explicit Sentiment Expressions

In a customer review, the subject of the evaluation is a product or service and the sentiment 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., positive and negative) and activation levels (e.g., high or low) for specifying the strength of different emotions in terms of a hedonic tone and its mobilization or arousal. Explicit Sentiment expressions can also be boosted or attenuated by adding certainty words (e.g., “absolutely”) or tentative words (e.g., “apparently”; Smith and Ellsworth 1985; Sbisa 2001).

Therefore, the explicit 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 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. Consumer research lacks any quantitative assessment of these specific differential effects between boosted and attenuated sentiment expressions (Chung and Pennebaker 2007; Packard and Berger 2016) resulting on research assigning arbitrary weights (Hu, Koh, and Reddy 2014). Therefore, we phrase our hypotheses to propose that higher activation level and/or boosted explicit sentiment expressions have stronger differential effects on overall sentiment strength, compared with lower activation level and/or attenuated 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

Explicit 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). Insight into how these implicit expressions are manifest in consumers' communication is lacking though (Packard and Berger 2016). Literature on linguistics suggests that speech acts that are directive (suggestion to a third person to take an action), commissive (committing to a future action), and assertive (conveying the state of the situation) can also convey sentiment (Searle 1975). 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 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 sentiment (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 explicit sentiment expressions.

It remains unclear how the implicit expressions relate to sentiment strength. We posit that directive and commissive acts might have stronger effects on consumer sentiment strength than assertive acts. 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, 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$_2$:** Directives and commissives have stronger effects on overall sentiment strength than assertives in text-based reviews.

Discourse Patterns

Single sentences within a discourse are related and their patterns might reflect writers' sentiment towards a product or service experience (Goldberg and Zhu 2006; van Dijk 1997). Consumers encounter multiple positive and negative emotions when consuming a product or service (Aaker, Drolet, and Griffin 2008), which they verbalize across multiple sentences in a customer review. 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 disregard patterns of sentiments across sentences (Das and Chen 2007), instead they examine them at an aggregated message level (Tirunillai and Tellis 2012), or else derive it at a sentence level (Büschken and Allenby 2016; Khan, Baharudin, and Khan 2011). However, the active use of contradictory sentiment expressions (Fonic 2003) might relate to a lesser degree of conviction. Ignoring such developments across multiple sentences would fail to account for ambivalent evaluations (Otnes, Lowrey, and Shrum 1997). Then, by measuring the degree of sentiment incoherence across review sentences, we would expect that a higher degree of incoherence (i.e., ambivalence) is associated with lower overall sentiment strength. We hypothesize;

> **H_3:** An increase of sentiment incoherence across sentences has a negative effect on the overall sentiment strength in text-based reviews.

In addition to sentiment incoherence, it is possible to explore the role of other types of discourse patterns in online reviews, such as the trend of the sentiment in the review. De Saussure (2007) suggests that the message development, may further relate to the overall sentiment strength of writer. Accordingly, sentiment expressions are not randomly distributed but rather represent a set of sequentially organized propositions to explain an overall opinion. In fact, previous research acknowledges the presence of such trends in sentiment expressions but without explicating their implications (Mao and Lebanon 2009). Thus, we will conduct an exploratory analysis on the role of trend on overall sentiment strength without articulating a formal hypothesis.

STUDY 1: SENTIMENT EXPRESSIONS IN CUSTOMER REVIEWS

Setting

To examine the differential effects of explicit and implicit expressions, and the direct effect of discourse patters 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 (43,687 after removing duplicates) 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 analyzed text-based features related to sentiment 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, using the self-reported star ratings. Star ratings appear prominently in marketing and consumer research (Kronrod and Danzinger 2013; Ludwig et al. 2013), and previous text mining studies use them as proxies for sentiment strength (Pang and Lee 2005, 2008). On a five-star scale, consumers' deviations from the midpoint (i.e., three stars) indicate either relatively more negative evaluations (i.e., one or two stars) or positive evaluations (i.e., four or five stars) (Amazon 2014). A three-star (midpoint) review reflects either a truly moderate review (indifference) or a series of positive and negative sentiments that counterbalancing one another (Mudambi and Schuff 2010). In line with previous research (Chevalier and Mayzlin 2006), the star ratings were positively skewed: 54% of the reviews rated their product or service experience with 5 stars, another 27% gave 4 stars, 9% used 3 stars, and only 5% and 7% of rated 2 stars and 1 star, respectively.

To construct the text-based predictor variables, we applied the Stanford Sentence and Grammatical Dependency Parser (online available at http://nlp.stanford.edu:8080/parser/) to automatically subdivide reviews into their sentences and identify dependencies between emotion words and boosters or attenuators. The parser would first identify the presence of an emotion word and then, in cases where a booster or attenuator is present, it would automatically assess if there is a grammatical relationship (e.g., the sentence: “the hotel was very nice”, an adverb “very” is grammatically related with the adj. “nice”, therefore it boosts it).

In line with (Ghose, Ipeirotis, and Beibei 2012) we used “part of speech tagging” (automatic classification of words into part of speech, e.g., noun, adj. verbs, etc.), to retrieve the most frequent 4,071 nouns across all sentences (e.g., “staff,” “hotel”), and only kept those sentences that referred at least once to a noun indicative of the product, service or an aspect thereof to ensure the respective sentence sentiment was truly a related evaluation. To address cases of anaphora resolution (e.g., a sentence that does not contain a referee but implicitly refers to one in a next/previous sentence), we also retained adjacent sentences to any sentence with a referee (e.g., “This author keeps impressing me with the quality of his work. It is just awesome”). This was accomplished by retrieving impersonal pronouns across the sentences (e.g., “it”, “this”; Chung and Pennebaker 2007). Finally, we excluded the few cases where the service or product names contained emotion words (e.g., “The Great Gatsby”).

Following these cleaning steps we started operationalizing explicit sentiment expressions and their activation level by drawing on the circumplex model by Russel and Barret (1999). First, we created four text-mining dictionaries for emotional valence (positive vs. negative) and relative activation levels (high vs. low). First we used the emotion dictionaries developed for the LIWC program (Pennebaker et al. 2007) which offers reliable convergence between the positive and negative dimensions it extracts from text-based contents (Pennebaker et al. 2007). In line with Netzer et al. (2012), we then enriched these dictionaries with more, context specific words that had positive and negative meaning, gleaned from online emotion dictionaries such as emoticons from Pcnet (Zhang, Gao, and Li 2011). For the activation level of negative valenced words we used the LIWC sub dictionaries, specifically sadness (i.e., negative and low on activation (ENL); Russel and Barret 1999), anger and anxiety (i.e., negative and high on activation (ENH); Russel and Barret, 1999). Finally, the LIWC dictionaries do not distinguish between levels of activation (high vs. low) in the positive emotion dictionary so that these were assigned through manual coding. Two independent coders, unfamiliar with the study purpose were instructed to classify the 516 positive words as either “high on activation” (EPH) or “low on activation” (EPL), with no neutral option (Krippendorff’s alpha = 83%; discrepancies resolved through discussion). To assess the robustness of our self-constructed positive emotion dictionaries, we compared them with the dictionary of affection in language by Whissell (2009), which automatically assigned an activation score per word or text on a continuous scale (from “low” (1) to “high” (3)). We conducted a one-way analysis of variance to determine the statistical differences in the level of activation score between our positive low and positive high categories (as classified by the coders). We found significant differences with $M_{\text{high\_activation}}$ of 2.11 and 1.95 $M_{\text{low\_activation}}$($F = 9.701$, $p < .01$).

Following this categorization into negative and positive valence as well as into high and low levels of activation innate to emotion words themselves, we accounted for the activation level infused through boosters and attenuators words (certainty, tentative and negation words retrieved from LIWC dictionaries) appearing in the same sentence. We used the Stanford dependency parser (Stanford Parser 2014) to detect grammatical dependencies between emotion words, boosters and attenuators. Following the approach by Taboada et al. (2011), certainty words appearing alongside low activated emotion words were considered “boosters” turning, for example, a low activated “sad” into highly activated “very sad”. Similarly, an emotion word high on activation (e.g., “great”) would be reclassified as low on activation if it was accompanied by an “attenuation” (e.g., “hardly great”). Finally, the Stanford Dependency Parser allowed us to account for negations which are considered specific type of attenuation (Sbisa 2001). Although we do not propose formal hypotheses about negations, we need to control for their sentiment inverting implication whereby a positive expression (e.g., "good") would become negative (e.g., "not good"). Therefore, we measured four variables signifying explicit sentiment expression (with their activation level) and the same variables on their negated form as follows:

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

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

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

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

where PH_i and PL_i represent the positive, high and low activated proportions for review *i*, respectively, and Neg_PH_i and Neg_PL_i represent the negation of positive high and low activated proportions, respectively.

These equations feature three binary variables for each sentence *j*: (B_{ij}), which refers to the presence (1) or not (0) of a booster (e.g., "!!"); (A_{ij}), which reflects whether there is an attenuator (1) or not (0) (e.g., "potentially"); and (N_{ij}), which indicates whether any grammatical dependency with a negation exists (1) or not (0). For example, PH_i indicates the positive and high activated proportion in review *i*, operationalized as the sum of positive emotion words high on activation (EPH_{ij}), including those that occur in combination with a booster (EPH_{ij} * B_{ij}) and the sum of positive emotion words which are low on activation yet are combined with a booster (EPL_{ij} * B_{ij}) divided by the total word count ($WCount_{ij}$) in review *i* and sentence *j* (*m* denotes the review number; n denotes the sentence number in review *i*), and subsequently divided by the total amount of sentences in review *i*, ($SCount_i$). In equation 2, we constructed the positive and low activation proportion in review *i* in a similar fashion, just this time accounting for the occurrence of attenuation words with positive words either high or low on activation. Equations 3 and 4 describe our operationalization of negated positive expressions. We derive NH_i (the high activated and negative proportion), NL_i (the low activated and negative proportion) for review *i* using the same approach.

To assess the internal validity of these explicit sentiment expressions, we used SentiStrength (Thelwall et al. 2010), a state-of-the-art tool to predict sentiment from short texts (for a recent application to marketing research, see Tang, Fang, and Wang (2014)). Note that SentiStrength's purpose is classifying short text into positive and negative from 1 to 5 and -1 to -5 respectively, while in our case we want assess the differential effects of positive (negative) high and low proportions (Thelwall et al. 2010). We used SentiStrength at the sentence level and then computed the average of the positive and negative sentiment strength variables at a review level. The results indicate correlations of .534 and .562 for aggregated measures of explicit sentiment and sentiment strength.

Having derived our measurements of explicit sentiment expressions, we extracted the *implicit* sentiment expressions namely commissives (C), directives (D), and assertives (A), which convey sentiment without emotion words (e.g., "I recommend this book"). Following a linguistics approach (Villarroel Ordenes et al. 2014), we developed "regular expression codes" (i.e., REGEX; Feldman and Sanger 2007) for word combinations that convey implicit sentiment. First we retrieved all the sentences that were not identified as explicit sentiment expressions, also void of any emotion word, (170,694 sentences) and extracted a random 1% sample of them from the book and hotel review data set. Two independent coders coded the main speech act in each sentence into assertive, directive, or commissive, as well as the respective valence of that speech act (i.e., positive, neutral, or negative). They also copied out the specific word (or word combination) that determined the valence for them (please see web appendix A for the coding instructions). The coders achieved a Krippendorff's alpha of 74% for the type of the speech act and 92% for the valence (disagreements were resolved in a post discussion). Based on these identified word combinations we developed a list of regular expression codes (REGEX) for implicit sentiment expressions (please see web appendix B for illustrative examples). We used this new list of REGEX to automatically retrieve 8,578 sentences (16% of all reviews) without emotion words in our data set. The final variables for the proportion of implicit sentiment expressions and their valence in the review texts were computed as follows:

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

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

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

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

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

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

where PC_i represents the proportion of positive commissive in review *i*. Same as above, it is computed by dividing the sum of positive commissive words in a review sentence ($PosCommissive_{ij}$), by the total number of words in that same review sentence ($WCount_{ij}$). We then aggregate the proportion of positive commissives at a review level by dividing the sum of commissives proportions by the total number of sentences *j* in review *i* (*m* = review number; n = sentence number at review *i*). We derived the proportion of the other implicit speech acts in the same manner, using the respective REGEXs for each.

To validate our measurement of implicit sentiment expressions, we assessed the precision (quality of the extraction) of the REGEX (Feldman and Sanger 2006). We used a random subsample from all the sentences extracted by the REGEX (8,578). Then, we asked two independent coders to manually classify them into the three speech acts and their valence. The coders achieved 92% of agreement on the speech act variable and 88.6% (measured by Krippendorff's alpha). The REGEX classification compared with the coders resulted on an average precision of 80.03%. Results were good indicators for our operationalization. Web appendix B provides more details.

Next, we moved to discourse patterns of sentiment across all the sentences in a review. We derived these discourse patterns for all reviews with more than two sentences (reviews with only one or two sentences did not have any discourse pattern as one needs at least three sentences to identify trends and (in)coherences). First, we computed the overall difference between positive and negative proportions of each sentence in each review *($DifPosNeg_{ij}$)* by deducting all explicit and implicit negative sentiment expression proportions from the positive ones (outlined in equations 1–10). Consistent with our previous rationale, we assigned weights to each of proportion variable to account for their differential, asymmetric effects. Rather than assigning arbitrary weights, we obtained the log-odds coefficients of each proportion variable by regressing them on the consumer sentiment strength first (i.e., the star-rating of the review), then we multiplied each proportion by its respective coefficient before deducting negative from positive. So for example, instead of using just the sum of positive directives (or multiply positive directives with an arbitrary weighting), we multiplied it with the exponential of the log-odds coefficient 0.07 equal to1.072 and representing the probability of staying on a higher star rating category) obtained from Model 2 (please see table 3). The difference between positive and negative sentiment proportion in each sentence of each review was therefore computed as follows:

$$DifPosNeg_{ij} = \sum_{\substack{i=0 \\ J=0}}^{\substack{i=m \\ j=n}} [PHL_{ij} * \beta_{PH} + PL_{ij} * \beta_{PL} + PC_{ij} * \beta_{PC} + PD_{ij} * \beta_{PD} + PA_{ij} * \beta_{PA}] - [NH_{ij} * \beta_{INH} + NL_{ij} * \beta_{INL} + NC_{ij} * \beta_{NC} + ND_{ij} * \beta_{ND} + NA_{ij} * \beta_{NA}], \textbf{(11)}$$

With these weighted sentence level sentiment proportions, we then operationalized the respective incoherence of positivity within each review as the standard deviation (SD) in positivity across review's sentences;

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

Finally, we operationalized the trend of positivity by regressing the sentence number (SentNumb$_{ij}$) on positivity (*DifPosNeg*$_{ij}$) for each review *i* separately (i.e., $DifPosNeg_{ij} = \alpha + \beta * SentNum_{ij}$). The resulting beta-coefficient of each of the ordinary least squares regressions signifies the overall trend in positivity within the respective review. A β coefficient closer to 0 signifies a more stable positivity trajectory, a negative β indicates a decreasing trend in positivity and a positive one an increasing trend in positivity. Since we are interested in the sentiment implications of positive and negative trends separately, we split this trend variable into negative *(NT$_i$)* and positive values *(PT$_i$)*. This approach preserves the continuous nature of our trend variable while avoiding reducing it to a categorical dichotomization (Rucker, McShane and Preacher 2015).

For robustness purposes we also conducted a sensitivity analysis on the effect of our weighting approach by comparing it against a computation of $DifPosNeg_{ij}$ without weighting the speech acts by their coefficients. The results remained the same as in our final model 3 in table 3 (please see web appendix C for more information). We summarize all our variables as well as the way they were operationalized in table 2.

Insert table 2 about here

Control Measures

Following related research on sentiment analysis, we controlled for a number of additional linguistic aspects. Firstly, we accounted for sentiment subjectivity (Pang and Lee 2008) by measuring the proportion of first-person pronouns (FP_i) (e.g., “I” or “we”) in each review. Barasch and Berger (2014) suggest that such first-person pronouns are reflective of writers’ self-focus and personal involvement which may indicate a greater overall strength in their described sentiment. Second, following Chevalier and Mayzlin (2006), we included dummy variables to control for the popularity differences between review sites on which each review *i* features (D_RS_i) with barnesandnoble.com denoted as 0 and Amazon.com denoted as 1. Finally, we controlled for the total number of sentences in a review as a separate control variable ($TSent_i$).

Analysis

Given that sentiment strength represents an ordinal variable (with star-ratings running from 1 to 5), we used an ordinal logit model (Farley, Hayes, and Kopalle 2004) to assess our hypotheses. For robustness purposes we compared our model choice to a partial proportional odds model, which allows the coefficient sizes to vary across star categories (i.e., multinomial logit) (Williams 2006). We found that despite an increase in model fit assessed with the Akaike Information Criteria (AIC), the coefficient interpretation and significance remains the same, so we opted to use the more parsimonious ordinal logit model.

We next specified a series of ordinal logistic regression models to estimate the effect of explicit and implicit sentiment expressions, together with discourse patterns on the star rating. For interpretability, we standardized all predictor variables before conducting the analysis. We relied on Knime 3.2 to estimate the models, beginning with the four proportions of explicit sentiment expressions positive and negative, and four separate negation variables that represent the negated versions in Model 1 (a). Noting the positive (negative) coefficients of the negated variables in Model 1 (a), and in line with Sbisa (2001), we then aggregated the negated positive and negated negative proportions with explicit expressions with low activation (positive and negative) in Model 1 (b).

We then introduced the implicit sentiment expressions (six variables) in Model 2 and accounted for the discourse pattern variables, incoherence, positive trend and negative trend in Model 3 (please note that since such patterns need at least three sentences our final sample for Models 3 excluded 2,283 and 947 reviews for hotels and books respectively). To ensure comparability, the covariates remained the same for all consecutive models; only number of sentences was added to Model 3 as a separate covariate due to our specific interest to control for more (less) extensive discourses (see table 3).

In line with the requirement of ordinal logit models, we found that all our intercept estimates are unique and significantly different from their adjacent cut points at $p < .01$ (Godes and Silva 2012). Using AIC, we confirmed that the implicit expression and discourse pattern explanatory variables added explanatory power to the final model (see table 3, Models 1–3).

Hypotheses Testing

Before testing H_1, we assessed the effect of the negations for each main variable (PH_i, PL_i, NH_i, NL_i). Please 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.

The parameter estimates provided support for most of the hypotheses. Using Wald z tests, Model 1, confirmed that in line with H_1, consumers' use of explicit positive expressions that are high on activation and/or boosted (PH_i) has significantly stronger positive effects on their overall sentiment strength than positive expressions low on activation and/or attenuated (PL_i). This effect is consistent across the product and service contexts (for books β_{PH} .93 vs. β_{PL} .11, Wald z = 29.45, $p < .01$; and for hotels β_{PH} .89 vs. β_{PL} .05, Wald z = 36.27, $p < .01$). Similarly, for hotels, the use of explicit negative expressions which are high on activation and/or boosted (NH_i)

has significantly stronger negative effects on sentiment strength than the use of explicit negative expressions which are low on activation and/or attenuated (NL_i) (β_{NH} -.44 vs. β_{NL} -.40, Wald z = -1.98, $p < .05$). However, for books, contrary to our expectations, explicit negative expressions which are low on activation had a significantly stronger effect than the high activation counterparts (β_{NH} -.31 vs. β_{NL} -.37, Wald z = 3.45, $p = .01$). We elaborate on these results in the discussion section.

Insert table 3 about here

The results of Model 2 supported our overall prediction that directives and commissive in consumer reviews, relate to stronger overall consumer sentiment compared to assertives. Considering implicit "positive" expressions only, directives had a relatively stronger effect than assertives across product and service reviews (for books $\beta_{DP} = .28$ vs. $\beta_{AP} = .05$ respectively, Wald z = 6.28, $p < .01$; for hotels $\beta_{DP} = .07$ vs. $\beta_{AP} = -.001$ respectively, Wald z = 3.30, $p < .01$), as do commissives for hotels only (hotels $\beta_{CP} = .09$, Wald z = 5.57, $p < .01$). Similarly, we found that directives implicitly conveying negative sentiment have stronger effects on consumer sentiment strength than assertives (for books $\beta_{DN} = -.28$ vs. $\beta_{AN} = -.06$, Wald z = -7.33, $p < .01$; for hotels $\beta_{DN} = -.15$ vs. $\beta_{AN} = -.07$, Wald z = -4.06, $p < .01$). We also found a statistically stronger effect by negative commissives as opposed to negative assertive acts in support of H_2.

Examining the effects of discourse patterns on consumers' overall sentiment strength we found that, consistent with H_3, an increase on sentiment incoherence (e.g., frequent changes in positivity across the sentences of the review) relates to an overall more negative consumer sentiment strength (for books: $\beta_{SD} = -.17$, $p < .01$; for hotels $\beta_{SD} = -.16$, $p < .01$). Furthermore, our exploratory analysis into the effects of sentiment trends across the sequence of sentences in a review revealed that an increase of positive trend (i.e., more positive sentiment expressions at the end of the review) significantly relates to an overall more negative consumer sentiment strength (for books β_{PT} = -.14, p < .01; for hotels β_{PT} = -.14, p < .01). Conversely an increase of negative trends (more negative sentiment expressions towards the end of the review) significantly related to a more positive sentiment overall (for books β_{NT} = .12, p < .01; for hotels β_{NT} = .08, p < .05).

Regarding our control variables, in line with Chevalier and Mayzlin (2006) we found that the review site had a significant influence on the sentiment strength of the consumer reviews (β_{D_RS} = .06, p < .01) with more positive reviews on Amazon.com. The use of personal pronouns had a significant positive effect on overall consumer sentiment strength for hotels (β_{FP} = .19, p < .01) and a significant negative effect in the case of book reviews (β_{FP} = -.10, p < .01). Finally, the total number of sentences per review had a significant negative effect on overall consumer sentiment strength for hotel reviews (β_{TSent} = -.07, p < .01) and a non-significant effect for book reviews (β_{TSent} = -.01, p = .10).

Robustness Check

Crucially, the inherently endogenous relationship between written expressions inside the reviews and the self-reported consumer star rating prevents us from making causal implications. 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 coder scored no more than 25 reviews and each review was scored by 10 coders. The correlation between our sentiment strength variable (star rating) and the AMT average was .84 ($p < .01$). We then replicated our final Model 3 using ordinary least squares regression to explain the externally coded sentiment strength with our speech act features. The results (please see table 4) corroborated our hypotheses.

Insert table 4 about here

STUDY 2: RELEVANCE OF SENTIMENT STRENTH

Setting

To assess the relevance of our speech act features beyond predicting the overall sentiment strength of consumer reviews, we considered their implications for consumers purchase decision making and in turn sales next. Specifically, we examined the implications of weekly changes in the overall sentiment and the strength of verbatim consumer reviews (derived using our independent variables from Study 1 for products' sales rank fluctuations on online retail sites. We expect that our approach to decode the consumer sentiment strength 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 the influence of consumer reviews (1) just using their *valence*, as the difference between positive and negative emotion words, in line with most research so far (Model A) or (2) *sentiment*, derived using our more fine-grained approach in Model 3 from Study 1 (Model B). We tested and compared the direct influences of these 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. In line with Chevalier and Mayzlin (2006), we collected a longitudinal data set with 352 books available on both sites, with an average of 9.2 weeks of observations. We gathered, from both sites, the weekly sales rank of each book, prices 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 the Appendix).

Results

Changes in the sentiment strength of the review texts in the previous week (t – 1) exerted a significant influence on the corresponding 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 ($\beta_{Sentiment\ Amazon}$ = -.027, $p < .01$). 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 ($\beta_{Sentiment\ BN}$ = .024, $p < .05$). Using just the changes in valence is not as good for predicting changes in sales (see Model A). For example, while changes in valence in the reviews featured on Amazon exhibit a significant influence on subsequent sales

($\beta_{Valence\ Amazon}$ = -.020, $p < .05$), the predicted effect is less stark. Changes in valence in the consumer reviews on BN.com are only marginally significantly related to the changes in sales performance of products on Amazon.com ($\beta_{Valence\ BN}$ = .017, $p = .063$). Accounting for changes in the more fine-grained model of sentiment strength results in a significantly better model fit (Model 2, Wald χ^2 = 80.69, Model 1 Wald χ^2 = 53.65, $p < .001$). We did not find any significant influence of any of the implicit sentiment expressions on sales with the exception of negative directives (e.g., "do not buy this book") (please see details on the web appendix D). Negative directives increase the sales rank of the product on Amazon.com, hence effectively discouraging consumers from purchase and reducing sales ($\beta_{Negative\ Directive\ Amazon}$ = .029, p < .05). These results are in line with previous research by Ludwig et al. (2013) who suggest that, trying to avoid informational overload in low-involvement purchase decisions, consumers resort to heuristic processing and hence screen for the most easily accessible indicators. Therefore, accounting for explicit sentiment expressions with a more nuanced categorization of explicit emotion words into low and high activation, including boosters and attenuators, as well as implicit sentiment expressions and discourse patterns, rather than a basic differentiation between positive and negative emotion words, ultimately improves the estimate of consumers purchase behavior on online retail sites featuring these reviews.

STUDY 3: GENERALIZABILITY OF SENTIMENT STRENGTH

Setting

To add generalizability to our results, we scraped 1,716 verbatim consumers' online service evaluations from Twitter and Facebook across six product and service categories (financial services, travel, retail, news media, health services and electronics). Two independent coders scored their perceived sentiment strength of each message on a scale from 1 to 5 (Krippendorff's alpha = 77.9%; discrepancies resolved through discussion). We derived our speech act features in the same manner as for study 1 and 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 platforms. We also included five dummy variables to control for the industry types.

In this new context, we had to modify the regular expressions (REGEX) from Study 1 by altering the contextual verbs. For example, the regular expression "you should + buy" indicated a directive act in Study 1, whereas in this social media context including for example the evaluation of news media, we used "you should + watch" instead. Using these REGEX adaptations, we retrieved 8% of product evaluations that included at least one of the six implicit sentiment expressions (whereas we had 16% in our study on customer reviews). With this smaller sample, we decided to aggregate all implicit speech acts (commissive, directive, and assertive) into just two categories, implicit positive and implicit negative. Although comments were shorter (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 discourse patterns for those that had three sentences or more. Replicating Model 3 then we assessed the generalizability of our results from Study 1.

Results

We report the effects in table 5. In line with H_1, we found a significantly stronger effect of explicit sentiment expressions which are high on activation and/or boosted compared with the ones low on activation and/or attenuated (for explicit positive β_{PH} = .60 vs. β_{PL} = .31, Wald z = 2.66, p = .01; and for explicit negative β_{NH} = -.42 vs. β_{NL} = -.03, Wald z = -3.34, p < .01).

Insert table 5 about here

Although we found consistency in the coefficients directions for implicit positive and negative expressions, we did not find any significant effects. For the discourse pattern measures, we obtained evidence that incoherence in positivity across the sentences within comments have a negative, weakly significant impact on sentiment strength (β_{SD} = -.18, p = .06). However, neither positive nor negative trend has a significant impact. These results can largely be explained by the limited amount of comments that exceed 2 sentences (26%). As to our control variables, there are significant differences in sentiment across the industries and Facebook evaluations are more negative than those on Twitter. Finally, we again benchmarked our sentiment model, derived using the more nuanced sentiment strength model with the valence one (derived using the proportion of positive emotion words and negative emotion words per comment). According to table 5, our sentiment model, including the nuanced speech act features is significantly better at predicting the comments' overall sentiment rated by the coders than the valence model (AIC = 3301.7 vs. AIC = 3435.22).

DISCUSSION

Extending Extant Research

By theorizing about speech acts, this article informs sentiment analysis and provides a deeper understanding of how consumers express sentiment in online reviews, together with assessing the implications for subsequent consumer behavior. By empirically examining the hypothesized relationships, their relevance and generalizability, we extend extant research in three important ways.

First, where prior consumer research has relied on simple positive and negative emotion word frequencies, we offer a more nuanced and theoretically robust approach to decode consumer sentiments in verbatim reviews and posts. By accounting for activation level differences, innate to emotion words (Russell and Barret 1999), and the strengthening and weakening effects of boosters and attenuators (Pennebaker et al. 2007), we augment previous approaches and to improve the prediction of consumer sentiment strength and its implications for purchase decisions across multiple online contexts. In particular, compared with explicit positive sentiment expressions which are low on activation and/or attenuated, the use of explicit positive expressions which are high on activation and/or boosted doubles the probability that a consumer rates her experience one point higher on a 1 to 5, star rating scale. Contrary to our expectations, we did not find the same differential effects for explicit negative sentiment expressions across contexts. While we found a similar significant difference for hotel reviews, for book reviews explicit negative expressions actually have a stronger negative effect on overall consumer sentiment if they are low on activation and/or attenuated. A possible explanation may be that in book reviews explicit sentiment expressions at least partially blend with the type of book that is described, so “sad” might be actually a sought after feature for a tragedy book, and “disgusting” might describe a desirable antagonist character in a horror book. These contextual limitations are also in line with research on affect suggesting that the use taxonomic applications of emotions, might not work across contexts in the same way (Russell and Barret 1999). This dependency on word taxonomies associated with sentiment is an important finding to be considered in future research using sentiment analysis. Overall, in line with Russell and Barret (1999) and Sbisa (2001), we empirically demonstrate the importance of considering the relationship between activation levels innate to emotion words in combination with boosters and attenuators.

Second, SAT suggests that assertive commissive and directive expressions can implicitly convey the speaker’s sentiment, without any explicit emotion words (Searle 1975). Such implicit sentiment expressions are quite frequent and appeared in 16% of consumer reviews in our data set (only the ones recalled by our REGEX). We predicted and found that such *emotionless*, implicit acts relate asymmetrically to consumers’ overall sentiment. Specifically, we found that positive (negative) directives and commissives exerted stronger effects on overall sentiment strength than did assertives. The linguistic context suggests that generic assertions in 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 contribute to conceptualizations of implicit sentiment expressions (Feldman 2013; Montoyo, Martínez-Barco, and Balahur 2012) and add to current research on implicit language (Packard and Berger 2016), in that we introduce and validate a theoretical framework of emotionless expressions.

Third, we underscore the necessity of considering the message development itself (van Dijk 1997) and contribute to conceptualizations of sentiment dynamics (Schweidel and Moe 2014) by exploring how discourse patterns within reviews reflect consumers' sentiments. A consumer's overall sentiment is likely negative if the positivity of the sentiment expressions (explicit and implicit) are incoherent across the sequence of sentences in a review. 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 is associated with more negative reviews. Our exploratory analysis of positive and negative trends similarly yields interesting results. On the one hand, we found that positive trends (e.g., increasing positivity towards the end of a review) 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 confronts upsetting topics. On the other hand, negative trends are associated with more positive reviews. This finding is consistent with empirical studies 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 staff"; De Ascaniis 2013).

Corroborating Extant Research

The 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 found that explicit 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 expressions. 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 influence future sales ranks also emphasize the importance of improving sentiment analysis. First, we corroborate research by Chevalier and Mayzlin (2006) 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 explicit sentiment expressions (hence the effects of activation and valence).

The result that particularly negative directives impact sales is also in line with the findings of this article, 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" (Büschken and Allenby 2016) and in turn giving researchers and managers a more comprehensive understanding of the individual intentions 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. According to Chung and Pennebaker (2007), this finding might reflect the difference in the use of singular versus plural. First person plural (e.g., "we" or "us") 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). Therefore, our finding corroborates that the use of first person pronouns in singular is more associated with negative reviews compared to plural.

LIMITATIONS AND FURTHER RESEARCH

We note the massive potential for further studies on how different patterns of sentiment can drive subsequent consumer behavior. 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 emotion word (e.g., from positive to negative or vice versa). 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 other 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 expressions 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 article; 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 2009; 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. SAT reflects considerations of the referee or subject and the proposition (Searle 1969), so a book review likely features a combination of the reader's experience with the character and story plot, whereas sentiment toward a hotel more commonly is conveyed in terms of the customer experience. The study of sentiment expression in language could benefit from a deeper understanding of the context in which sentiment is presented. Widening its application in contexts such as communication within organizations (e.g., emails) (Schrage 2016) and conversations (e.g., sales negotiations and online chats) would contribute to the development of the field. Additional research using nested logit models could seek to uncover the relation between sentiment and its linguistic context (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 commissives, directives, and assertives, but not an exhaustive compilation of 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 (i.e., recall) 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.

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 explicit sentiment expressions with high and low activation including language granularities such as negations, certainty and tentative words. However, we believe that these 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. Please find more information regarding this type of analysis in web appendix E.

Eight, we encourage researchers to further explore discourse patterns. Our study provides an exploratory analysis concerning 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. Furthermore, the impact of lack of sentiment as opposed to positive versus negative trend should be investigated. 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. Further research could also focus on discourse dynamics within or between discourses such as consumer reviews or online conversations between customer and employees.

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. 2013; He and Bond 2015), such that at low levels of activation, reviews drive sales, but at very high levels of activation, they do not. It would be insightful 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 expressions simultaneously to describe product and service experiences.

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). This article illustrates the importance of speech act features for deriving writer’s sentiment strength as well as the sales implications. 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. Furthermore, as we show in Study 3, our findings can be extrapolated to other social media contexts in which consumers share product and service experiences. As such, this can contribute to latest engagement strategies suggesting the alignment of brand communications with online consumer sentiment expressions (Magids, Zorfas, and Leemon 2015).

Finally, this study provides better understanding of the linguistic markers of sentiment, spanning both sentence 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; speech act theory provides guidelines for including additional, explicit, implicit and context related cues. At the intersection of linguistic, consumer research and text mining, these theory-driven improvements are particularly relevant, considering the growing amount of insights that will stem from unstructured content.

DATA COLLECTION INFORMATION

The data for Study 1 were acquired through Monzenda, a web scraping software service. The second author supervised the collection of this data by the firm InSites in summer 2012. The analysis of this data was performed by the first author. The data for the robustness check was collected and analyzed by the first author. The data for Study 2 was collected in 2010 by the second author and analyzed together with first author during this review process. Finally, for study three the data was specifically collected by the social media firm InSites, who kindly collaborated with our project. The first author did the analysis for this last dataset.

Appendix: 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.

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:

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

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 web 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$_{it-1}$	-0.027	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) $_{it-1}$	-0.062	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.

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), 286-99.

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.

Büschken, Joachim and Greg M. Allenby (2016), "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, 508-14.

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), 229-59.

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), 156-65.

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].

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), 82-9.

———, 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), 191-98.

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.

Gergen, Kenneth J., and Mary M. Gergen (1988), “Narrative and the Self as Relationship,” *Advances in Experimental Social Psychology*, 21, 17-56.

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.

Godes, David, and José C. Silva (2012), “Sequential and Temporal Dynamics of Online Opinion,” *Marketing Science*, 31(3), 448-73.

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.

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), 1509-27.

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*, 52(5), 629-41.

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), 1141–57.

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*, 142-50.

Maks, Isa, and Piek Vossen (2012), "A Lexicon Model for Deep Sentiment Analysis and Opinion Mining Applications," *Decision Support Systems*, 53(4), 680-88.

McGraw, A. Peter, Caleb Warren, and Christina Kan (2015), "Humorous Complaining," *Journal of Consumer Research*, 41(5), 1153–71.

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.

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.

Packard, Grant, and Jonah Berger (2016), "How Language Shapes Word of Mouth's Impact" *Journal of Marketing Research*, In-Press.

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*, 115-24.

——— 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), 1243-54.

———, 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, 50-9.

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.

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), 666-78.

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, Ed. 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), 174-84.

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

Villarroel Ordenes, 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), 509-21.

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

Authors	Context	Outcome Variable of Sentiment	Explicit Sentiment Expressions	Implicit Sentiment Expressions	Discourse Patterns
Pang and Lee (2005)	Improve accuracy in sentiment analysis	Four-star classification	Sentence polarity (positive and negative)	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 and negations words 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 and neutral words dictionary. Polarity modifiers (e.g., not) and shifters (e.g., "very, lack of")	Considered 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. 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 assesses 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 and 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 and negative words dictionary	Not considered for analysis	Not considered for analysis

Tirunillai and Tellis (2012)	Uses reviews valence to predict stock price	Positive v/s negative	Positive and negative words dictionary	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)	Measures sentiment in phrases with a scale from -3 to 3. Negation is considered	Not considered for analysis	Not considered for Analysis
Maks and Vossen (2012)	Political texts	Positive, negative, both and neutral	Positive, negative and neutral words	Indirect expressive verbs (e.g., boast) 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 and negative words dictionary	Not considered for analysis	Not considered for analysis
Xiong and Bharadwaj (2013)	Uses news' sentiment to predict stock prices	Positive v/s negative	Positive and 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
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
Cambria et al. (2015)	Improve accuracy in sentiment and emotion analysis	Positive v/s negative; also single emotions (e.g., grief or joy)	Positive and negative. Also 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 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 Articles
Explicit Sentiment Expressions	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 Expressions	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): Explicit Sentiment Expressions		Model 1 (b): Explicit Sentiment Expressions		Model 2: Implicit Sentiment Expressions		Model 3: Discourse Patterns	
Variables	**Hotel**	**Books**	**Hotel**	**Books**	**Hotel**	**Books**	**Hotel**	**Books**
Positive High (*PH*$_i$)	0.90**	0.92**	0.89**	0.93**	0.88**	0.93**	1.03**	1.12**
Negative High (*NH*$_i$)	-0.44**	-0.31**	-0.44**	-0.31**	-0.43**	-0.31**	-0.37**	-0.26**
Positive Low (*PL*$_i$)	0.03**	0.10**	0.05**	0.11**	0.04**	0.11**	0.06**	0.18**
Negative Low (*NL*$_i$)	-0.37**	-0.30**	-0.40**	-0.37**	-0.39**	-0.36**	-0.37**	-0.32**
Neg_Positive High (*Neg_PH*$_i$)	-0.07**	-0.10**						
Neg_Negative High (*Neg_NH*$_i$)	0.01	-0.07**						
Neg_Positive Low (*Neg_PL*$_i$)	-0.15**	-0.14**						
Neg_Negative Low (*Neg_NL*$_i$)	0.10**	0.05**						
Positive Commissive (*PC*$_i$)					0.09**	0.05**	0.13**	0.07**
Positive Directive (*PD*$_i$)					0.07**	0.28**	0.09**	0.38**
Positive Assertive (*PA*$_i$)					-0.001	0.05**	-0.001	0.04**
Negative Commissive (*NC*$_i$)					-0.15**	-0.17**	-0.12**	-0.11**
Negative Directive (*ND*$_i$)					-0.15**	-0.28**	-0.13**	-0.22**
Negative Assertive (*NA*$_i$)					-0.07**	-0.06**	-0.05**	-0.05*
Incoherence (*SD*$_i$)							-0.16**	-0.17**
Positive Trend (*PT*$_i$)							-0.14**	-0.14**
Negative Trend (*NT*$_i$)							0.08*	0.12**
Total Sentences (*TSent*$_i$)							-0.07**	-0.01
First Person Pronouns (*FP*$_i$)	0.24**	-0.06**	0.24**	-0.06**	0.24**	-0.06**	0.19**	-0.10**
Dummy Review Site (*D_RS*$_i$)		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 1: Coefficients in models 1, 2, and 3 are log-odd probabilities; the dependent variable was the ordinal star rating. All variables were standardized. Note 2: For model fit comparisons we also ran model 3 using the entire samples (without excluding reviews with less than 3 sentences) resulting in a better model fit.

TABLE 4. Robustness Check, Amazon Mechanical Turk

Variables	Books Model 3
Positive High (PH_i)	0.26**
Negative High (NH_i)	-0.15**
Positive Low (PL_i)	0.06**
Negative Low (NL_i)	-0.17**
Positive Commissive (PC_i)	0.01
Positive Directive (PD_i)	0.09**
Positive Assertive (PA_i)	0.02
Negative Commissive (NC_i)	-0.10**
Negative Directive (ND_i)	-0.12**
Negative Assertive (NA_i)	-0.03†
Incoherence (SD_i)	-0.05†
Positive Trend (PT_i)	-0.05*
Negative Trend (NT_i)	0.03
Total Sentences ($TSent_i$)	-0.01
First Person Pronouns (FP_i)	-0.01
Dummy Review Site (D_RS_i)	0.05**
Intercept	0.74**
R-Squared	0.27

†$p < .1$ *$p < .05$. **$p < .01$.

Note: 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 OLS regression.

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

Variables (Standardized)	Valenced Model	Sentiment Model
Positive Valence Proportion	0.42**	
Negative Valence Proportion	0.05	
Positive High (*PH_i*)		.60**
Negative High (*NH_i*)		-.42**
Positive Low (*PL_i*)		.31**
Negative Low (*NL_i*)		-.03
Implicit Positive		.03
Implicit Negative		-.07
Incoherence (*SD_i*)		-.18†
Positive Trend (*PT_i*)		.04
Negative Trend (*NT_i*)		.03
Total Sentences (*$TSent_i$*)		-.08**
First Person Pronouns (*FP_i*)		.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's 77.9%; disagreement was resolved through discussion). All variables were standardized before the ordinal regression.

HEADINGS LIST

1. **CONCEPTUAL FOUNDATIONS**
2. Explicit Sentiment Expressions
2. Implicit Sentiment Expressions
2. Discourse Patterns
1. **STUDY 1: SENTIMENT EXPRESSIONS IN CUSTOMER REVIEWS**
2. Setting
2. Measure Development
2. Control Measures
2. Analysis
2. Hypotheses Testing
2. Robustness Check
1. **STUDY 2: RELEVANCE OF SENTIMENT STRENGTH**
2. Setting
2. Results
1. **STUDY 3: GENERALIZABILITY OF SENTIMENT STRENGTH**
2. Setting
2. Results
1. **DISCUSSION**
2. Extending Extant Research
2. Corroborating Extant Research
1. **LIMITATIONS AND FURTHER RESEARCH**
1. **IMPLICATIONS**