Communications of the IIMA

The Effects of Text Complexity on Online Review Helpfulness

Michael J. Hine
Carleton University, mike.hine@carleton.ca

Follow this and additional works at: http://scholarworks.lib.csusb.edu/ciima

Part of the Management Information Systems Commons

Recommended Citation

The Effects of Text Complexity on Online Review Helpfulness

Michael J. Hine
Sprott School of Business
Carleton University
Ottawa, ON, Canada
mike.hine@carleton.ca

ABSTRACT

This research reported in this paper extends the literature on the helpfulness of online reviews. Previous research has assessed online reviews using standard unidimensional readability algorithms. This research extends previous work by investigating a multidimensional framework, and associated measures, of text complexity and its impact on the helpfulness of online reviews. Results show that as the amount of passive voice and negation in online reviews increase, the helpfulness of said review decreases. Other significant predictors of review helpfulness include word meaningfulness, lexical diversity, and the number of modifiers per noun phrase.

Keywords: online reviews, computational linguistics, Coh-Metrix

INTRODUCTION

Advancements in internet-based technologies have resulted in reduced barriers to entry for user created content. Anyone with an internet connection can easily create content through websites, message boards, online forums, blogs and a variety of different social media outlets. One area where user generated content has garnered much interest is in online product reviews. Increasingly consumers are relying on peer based product reviews to inform their purchasing decisions. A recent article in Harvard Business Reviews states that 30% of US consumers go to Amazon to access their rich repository of product information and associated reviews to start their purchasing process (Simonson & Rosen, 2014). In addition, studies commissioned by Google have reported that consumers consult an average of 10 information sources before making a final purchase decision (Simonson & Rosen, 2014). These statistics are further supported by research that has found that customer reviews can positively impact customer sales (Clemons, Gao, & Hitt, 2006). The aforementioned results along with the fact that peer-based product reviews may have more credibility than seller based information (Bickart & Schindler, 2001) has established the importance of user generated content in the form of online product reviews in a consumer's purchasing intention.

A central theme in previous online review research is the determination of what impacts the usefulness or helpfulness of a review. Research has explored impacts and relationships between many components of online review helpfulness including characteristics of the reviewers, different product types, attributes of the review, and the content of the review itself. The focus of this study is on the characteristics of the online review text that may affect its comprehension,

and thus helpfulness, by potential consumers. While some work in this area has been completed, the research is limited by the simplicity of the constructs used to assess text difficulty.

This exploratory research moves the online review literature forward by incorporating advancements in computational linguistics to explore how text coherence, structural complexity, and word complexity affect the helpfulness of online reviews. It takes a multidimensional approach to text difficulty through the use of a comprehensive suite of tools and indices called Coh-Metrix (McNamara, Graesser, McCarthy & Cai, 2014). This research is based on the premise that an online review that is more coherent, structurally simple, and uses less complex words that have fewer meanings will be more helpful to consumers.

This paper is organized as follows. First, a brief review on online review helpfulness is presented. Next, computational linguistics and coherence is introduced which forms the basis of the research model to be explored. Hypothesized relationships between the model constructs are presented followed by a methodology section and results. Concluding remarks and areas for future exploration are also included.

BACKGROUND

Online Reviews

Research on online reviews is quite varied and comprehensive. The focus of this paper research was to explore the helpfulness or usefulness of online product reviews. This body of research has explored impacts and relationships between many components of online reviews including characteristics of the reviewers, different product types, attributes of the review, and the content of the review itself.

Research in online review helpfulness that incorporates reviewer attributes primarily focus on the perceived expertise, reputation or the credibility of the reviewer. Willemsen, Neijens, Bonner, and de Ridder found a significant relationship between expertise claims of reviewers and the perceived helpfulness of online reviews (2011). Consistent with these results Baek, Ahn, and Choi found that whether the reviewer was considered one of Amazon's top 10,000 reviewers or not had a significant relationship with online review helpfulness (2013). Alternatively, the same study found no relationship between whether the reviewer revealed their name or not (an indication of authenticity and trust) and the helpfulness of the review. Wu performed a series of three experiments focused on exploring the negativity bias in an electronic word-of-mouth (eWOM) context (2013). In one of those experiments, Wu tested the moderating effect of reviewer reputation on the relationship between negative, mixed and positive reviews on review helpfulness. While the study did show that negative and mixed reviews were more helpful from a high rather than low reputation reviewer, overall there was not a significant moderating effect. Ghose and Ipeirotis reported that while past historical information about reviewers had a significant effect on review helpfulness, the direction of the relationship was inconsistent across product categories (2011).

It has been found that review ratings that are further away from the average review rating of the product are interpreted as being lower in helpfulness (Baek et al., 2013). Similarly, Korfiatis, Garcia-Barriocanal, and Sanchez-Alonso reported that the higher the product review rating, the more helpful votes the review would receive (2012). Other research has explored review ratings in the context of product type delineating their results for experience goods versus search goods. An experience good is one that cannot be completely evaluated without actually interacting with it, such as running shoes. Alternatively, a search good is one whose value can be assessed before it is obtained, such as a personal computer. Mudambi and Schuff found that moderate online reviews are more helpful than either extremely positive or negative reviews for experience, but not search goods (2010).

There is considerable research on online review helpfulness that has explored the content of the review itself. The majority of said research has focused on how the content is being communicated rather than exploring what is being said. For example, Schindler and Bickart report that negative style characteristics including misspellings, bad grammar, inexpressive slang among others, are associated with less valuable reviews. The same paper reported no significant relationship between positive style characteristics and valuable reviews (2012). Results from research focused on review length (number of words) have found that longer reviews are associated with helpfulness but their impact is greater for search goods rather than experience goods (Mudambi & Schuff, 2010). Consistent with parts of the aforementioned results, Baek et al. found that review length is positively related to review helpfulness but only to a point wherein the relationship diminishes (2013). Through a content analysis of 400 online reviews, it was found that negatively valenced reviews were found more helpful but the relationship only existed for experience products and not for search products (Willemsen et al., 2011). The same research also reported that a higher number of supporting arguments and the diversity of said arguments within the product reviews is associated with more helpful reviews (Willemsen et al., 2011). Baek et al. found that helpfulness ratings increased when product reviews contained a higher percentage of negative words (2013). Concrete, as opposed to abstract reviews, have also been shown to be positively related to online product review helpfulness (Li, Huang, Tan & Wei, 2013).

Recently, researchers have begun exploring the relationship between the readability of online reviews and the helpfulness of reviews. Results have consistently reported that the readability of the review significantly predicts the helpfulness of the review (Ghose & Ipeirotis, 2011; Korfiatis et al, 2012). Simply put, a review that is easier to read and thus comprehend and interpret has more value to a potential consumer in evaluating and deciding whether to purchase a product. A variety of measures have been used in assessing the readability of online reviews including the Automated Readability Index (ARI), Coleman-Liau Index, Flesch Reading Ease (FRE), Flesch-Kincaid Grade Level, and the Gunning fog index. While all these measures are established and been used in a variety of studies across multiple disciplines, they have identified limitations that have been addressed recently by advancements in the computational linguistics domain. The goal of this manuscript is to extend the research that has taken place on readability and qualitative factors on online review helpfulness by incorporating more detailed measures of text difficulty.

Text Difficulty

The main criticism of the aforementioned readability measures is that they are unidimensional (McNamara et al., 2014) and rely primarily on word familiarity (whether reader knows the words or not) and on sentence length. While these readability measures have proved useful in decades of research, more recent advancements in computational linguistics have proposed a multilevel theoretical framework that can extend traditional measures and delve deeper in to the nuances of language and their impact on text difficulty. Specifically, Graesser and McNamara identify a six level framework for scaling text difficulty including words, syntax, textbase, situation model, the discourse genre, and the pragmatic communication level between writer and reader (2011). The framework levels are compositional components that are constructed as the process of comprehension takes place (Graesser & McNamara, 2011). It follows then, that comprehension of text can fall apart or misfire at any level of the framework. For example, a reader may not understand a particular term in the text (word level) and may not comprehend attempts at sarcasm (pragmatic communication level).

More difficult words (whether more structurally complex or less well known) can affect the overall difficulty of the text under evaluation. Similarly, the syntax, or arrangement of words and phrases can impact the level of difficulty of processing and understanding the meaning of text. Long syntactic structures and multiple embedded subordinate clauses will make it more difficult to construct meaning from the studied text (McNamara et al., 2014).

The textbase moves beyond the simpler concepts of words and syntax to include propositions and the referential links that connect the propositions together (van Dijk & Kintsch, 1983). The situation model deals with the conceptual meaning of text beyond the language and words that exist in the text and relies on context and domain knowledge (McNamara et al. 2014). Therefore, in an online review the situation model includes the information about the product in question. This could include how the product is used, the product's properties/components and the product's associated spatial representation, and any user expertise or experience that informs or improves the user experience or product performance. In both the textbase and the situation model, the concept of cohesion is of primary importance. At a broad level, cohesion refers to cues that tie or bound lexical elements of text together. Cohesion can exist in many different forms and at both the textbase and situation model level. Research in computational linguistics has established that coherence contributes significantly to the complexity of text and how easy it is to understand (McNamara et al., 2014).

Texts are often put into categories (called genres) that reflect the intention and content of the text itself. For example, a common categorical scheme delineates text between narrative, expository, persuasive and descriptive with each of the said categories having additional subcategories (Brooks & Warren, 1972). Some genres of text are more difficult to comprehend than are others. For example, informational texts such as business textbooks and journal articles are more difficult to comprehend than narrative texts (reference p. 12). Finally, the communication level includes such items as the goal of the writer and reader of the text under question (Graesser & McNamara, 2011).and thus can affect the ability to accurately assess the difficulty of the text.

Multiple indices for the first five levels of the six level framework described above have been amalgamated and are available through a product called *Coh-Metrix* (see McNamara et al., 2014). Coh-Metrix is a tool developed at the Institute for Intelligent Systems at the University of Memphis, which includes many indices for cohesion, language and readability. Coh-Metrix has been used successfully in a wide range of academic research including: assessments of modified and authentic texts circulated for English as a second language students (Crossley, Louwerse, McCarthy, & McNamara, 2007); research on gender delineation in texts (Bell, McCarthy, & McNamara, 2012); and differentiation of formal/informal and spoken/written communication across genres (Dempsy, McCarthy, & McNamara, 2007; Louwerse, McCarthy, McNamara, & Graesser, 2004) among others.

Research Model and Hypotheses Development

This research focuses on particular constructs for word assessment, syntactic difficulty and cohesion (at both the textbase and situation model level) that are most relevant for the study of helpfulness of online reviews. We present the constructs in the aforementioned categories as the separation reflects the different levels of abstraction that the constructs address. The concepts of genre and pragmatic communication are not measured explicitly but rather provide the context for the derivation of the hypotheses. The research model is presented in Figure 1 followed by derived hypotheses. Fundamental to the derivation of the hypotheses is the assumption that online product reviews that are easier to interpret and understand will be more helpful to consumers. This assumption follows from the established research positively linking online review readability (using unidimensional approaches) with online review helpfulness (Ghose & Ipeirotis, 2010; Korfiatis et al, 2012).
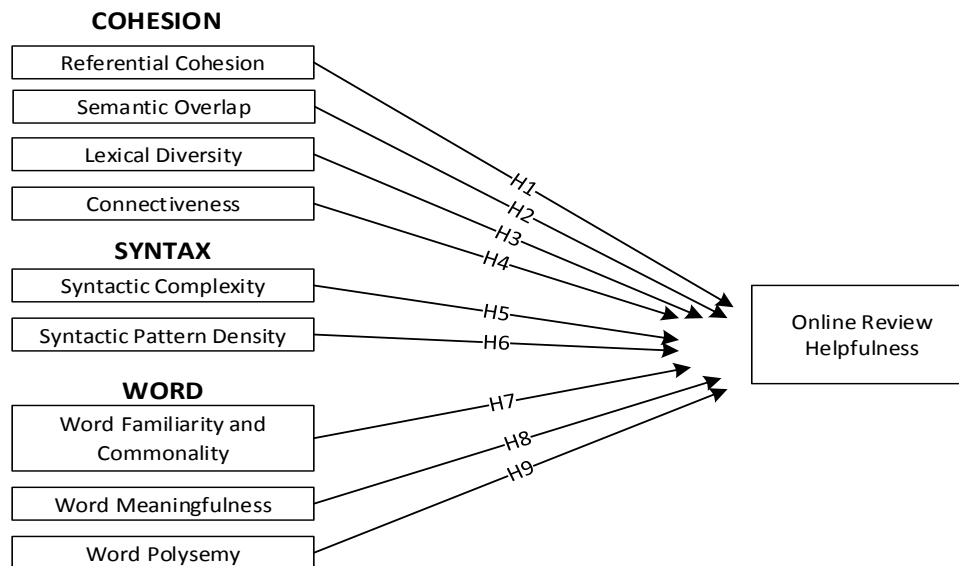
Figure 1: Online Review Text Complexity Model.

This research adopts the standard stage in consumer decision process including problem or opportunity recognition, search, evaluation of alternatives, purchase decision and post-purchased evaluation (Boone, Kurtz, MacKenzie & Snow, 2013). In the online review context, the writer of

the review is in the post-purchased evaluation stage as they have purchased the product in question and likely used and evaluated it. Similarly, the research assumes the majority of readers of the reviews and thus the persons who evaluate whether a review is helpful or not would be in the search and evaluation of alternatives stage of the consumer decision-making process.

Cohesion

As mentioned above cohesion refers to interconnections between textual elements and can take many different forms. Referential cohesion refers to overlap between nouns, pronouns or noun phrases and other content words (noun, verb, adjective or adverb) in adjacent sentences (Graesser, McNamara, & Kulikowich, 2011). For example, the following two sentences would have high referential cohesion because the noun in the second sentence (*printer*) also exists in the previous sentence (*printing*) albeit in a separate grammatical form: "*Printing* was a breeze. The *printer* also had an easy to open latch that made changing the toner very easy." High referential cohesion can assist readers in making connections within text thus avoiding the gaps that can occur at the textbase level that have been identified as contributing to increased reading time comprehension problems (McNamara & Kintsch, 1996). It is thus expected that a consumer will find reviews more helpful if there is more referential cohesion in the review text.

H1: An increase in online review referential cohesion will increase review helpfulness

Latent Semantic Analysis (LSA) (Landauer, McNamara, Dennis, & Kintsch, 2007) takes into consideration domain knowledge of the text beyond the explicit words used to provide measures of semantic overlap (Graesser et al., 2011). Fundamental to LSA is the premise that the word meaning is a product of the other words that surround it in naturalistic documents (Graesser et al., 2011). Words have higher semantic overlap when the degree of words surrounding them is similar. For example, the word "camera" would be highly associated with "lens," "photography" and "focus." An online review high in semantic overlap will be more cohesive and thus easier to understand than one that is lower in semantic overlap. It is expected that a consumer will find reviews more helpful if they have higher semantic overlap.

H2: An increase in online review semantic overlap will increase review helpfulness

Lexical diversity assesses the breadth of words usage in a text in relation to the total number of words in the text (Graesser et al., 2014). Lexical diversity is at its highest when the aforementioned ratio is 1, which represents either an extremely short text or one that is low in cohesion. New words need to be integrated into the discourse context, which requires cognitive effort and may increase reading time and comprehension difficulty. Alternatively, words that are used multiple times within the text are absorbed and integrated more easily by the reader and are indicative of a more cohesive text. It is expected that repeated use of words throughout the online review will result in a more helpful review.

H3: An increase in online review lexical diversity will decrease review helpfulness

Recall that the situation model focuses on the subject matter content. Discontinuities can occur in different situation model dimensions including causation, goals, time, space and protagonists

(Zwaan & Radvansky, 1998) resulting in a lack of cohesion (Graesser et al., 2011) and ultimately increased reading time and generation of inferences (O'Brien, Rizzella, Albrecht & Halleran, 1998; Zwaan & Radvansky, 1998). Critical to reducing or limiting the cohesion breaks is use of connectives, which are typically transitional phrases and adverbs that link the text (Graesser et al., 2011). Connectives can include "also," "moreover," "and then," "after," "during," "because," "so," "therefore," "if," and "or." For example, the following text, "The shoe had soft soles. The shoe wore out fast when used for tennis," would have lower connectiveness than "The shoe had soft soles and thus wore out fast when used for tennis." It is expected that an increase in connectiveness will result in a more helpful online review.

H4: An increase in online review connectiveness will increase review helpfulness

Syntactic Complexity

Syntactic complexity relies on the deconstruction of sentences into various lexical components such as nouns, verbs, adjectives, noun phrases, and verb phrases among others (McNamara et al., 2014) that can then be processed to determine complexity of a sentence. Not surprisingly complex syntax is more difficult to process and comprehend (Perfetti, Landi & Oakhill, 2005). Two main assessments of syntactic complexity include the number of modifiers per noun phrase and the number of words before main verb of the main clause (Graesser & McNamara, 2011), the second of which has been established to tax the working memory of the reader (Grasser, Cai, Louwerse & Daniel, 2006). For example, in the following sentences, the first is considered more syntactically complex because there are more words before the main verb of the sentence (eight words) than the second (three words): "These shoes are super squishy so I love to *run* in them. I love to *run* in these shoes because they are super squishy." It is expected that an increase in syntactic complexity to be associated with decreased online review helpfulness.

H5: An increase in online review syntactic complexity will decrease review helpfulness

The relative existence of certain word and phrase types (syntactic density) also contributes to the overall syntactic complexity of text (McNamara et al., 2014). Two types are of particular interest in this study, the existence of passive voice and negation. Both of these types of syntax have been associated with increased processing difficulty of text (Just & Carpenter, 1971, 1987). An example of negation would be, "This is not a good camera" as opposed to "This is a bad camera." Similarly, an example of passive voice would be, "The pictures produced by this camera are excellent" as opposed to active voice which would be, "This camera produces excellent pictures." It is expected that an increase in syntactic density to be associated with a decrease in online review helpfulness.

H6: An increase in online review syntactic pattern density will decrease review helpfulness

Word Information

How hard a text is to read and comprehend is affected by what words are contained in the text. Words that are more common in the English language and more familiar to a reader of an online review will be processed more easily and thus the review will be easier to comprehend. It is

expected that reviews with more familiar and common words will increase the helpfulness of the review.

H7: An increase in online review word commonality and frequency will increase review helpfulness

Also related to word difficulty is word meaningfulness, which reflects the degree to which the word is associated with other words (Toglia & Battig, 1978). Words high in meaningfulness invoke larger numbers of word associations than words that are low in meaningfulness. An example of a word with high meaningfulness rating would be "sex" while "aural" is a word with low meaningfulness. Words high in meaningfulness are typically less difficult to process and thus increase the readability of text. It is expected that reviews with more meaningful words will increase the helpfulness of the review.

H8: An increase in online review word meaningfulness will increase review helpfulness

Polysemy denotes the number of meanings a word may have (Graesser et al., 2011). For example, the word "ditch" has two common meanings, one referring to a narrow strip of land that is lower than the land on either side of it and the other referring to intentionally leaving someone somewhere without telling the affected person. High word polysemy increases the potential for different word interpretations and thus is often an indication of text ambiguity (Graesser et al., 2011). It is expected that online reviews with higher polysemy to have lower review helpfulness.

H9: An increase in online review word polysemy will decrease review helpfulness

METHODOLOGY

Sample Description

Our sample consisted of product reviews that were crawled from Amazon.com. Two *search good* product categories and two *experience good* product categories were selected. Product categories were selected based on their use or recommendation in previous studies on online reviews (Willemsen et al., 2011). The search product categories selected were digital cameras and laser printers and the experience product categories selected were sunscreen and running shoes. Crawlers were created using import.io, a free web-based data extraction tool, and subsequently mined each of the product categories resulting in 3864 running shoes, 9840 sunscreens, 338 digital cameras, and 7608 laser printers. The results for each product category were then sorted based on the helpfulness total votes each individual product received. Then the ten most voted on products for each of the product categories were selected. Additional crawlers were created and reviews for each of the selected products were mined. The online reviews were imported into Excel, and several Visual Basic for Applications algorithms were applied to the data. These algorithms performed the following operations: elimination of any duplicate reviews; elimination of any reviews that did not have any helpfulness total votes; and elimination of any reviews that were written by a person who was not an Amazon Verified Purchaser.

Recently Amazon has added the Verified Purchaser feature to their online review system. This offers assurance that the reviewer poster has purchased the product being reviewed through Amazon or one of their channel partners. To our knowledge, no other research on online reviews using Amazon as a source has incorporated this important feature. Given the preponderance of manipulated reviews (research by Hu, Bose, Koh & Liu has estimated it at near 10% (2012)), whatever avenues exist to validate the legitimacy of online reviews should be explored.

Executing the steps above resulted in 619 sunscreen reviews, 1497 running shoe reviews, 2201 laser printer reviews, and 4255 digital camera reviews. All data was collected during January and February of 2014. A further culling of the data based on only including reviews that had at least four helpfulness votes. This is consistent with previous research in online reviews. The final sample size and average review lengths for the data used for this study is presented in table 1 below.

Product	Number of Reviews	Average Review Length
Experience: Running Shoes	268	93.52
Experience: Sunscreen	135	121.04
Search: Printer	1063	205.99
Search: Camera	474	208.49
Total	1940	185.15

Table 1: Online Review Descriptive Statistics.

Variables

As mentioned previously the dependent variable in this study is the helpfulness of online reviews. Consistent with previous research, this variable is operationalized in the form of a helpfulness ratio:

$$HR = \frac{Helpfulness\ Votes}{Total\ Votes}$$

All independent variables in this study are operationalized through *Coh-Metrix. Referential Cohesion* was assessed using the coreference localized stem overlap measure (CRFSOl) which is one the most liberal of the cohesion measures available in Coh-Metrix. CRFSOl finds all nouns in a sentence and then searches the previous sentence for an overlap (match) of a content word (noun, verb, adjective or adverb). Overlaps must share common lemma (for example, party/parties, goose/geese, mark/marked). If any match occurs, the result is 1; else, it is 0. Then scores are averaged over all sentences in the review. For example, if there are four sentences in a review (and thus three comparisons) and two pairs of sentences overlap CRFSOl would be 2/3 or .67.

Latent Semantic Analysis was assessed by comparing semantic overlap (similarity) between adjacent sentences (LSASSl). LSASSl determines whether words in adjacent sentences are implicitly similar or related in meaning. For example, "dog" in one sentence would have high semantic overlap with "fur" and "barking" in an adjacent sentence. Overlap is averaged across all sentences in the review resulting in an assessment between 0 and 1.

This study uses the recent and now accepted vocd algorithm (Malvern & Richards, 1997) for measuring *Lexical Diversity* (LDVOCD). Basic approaches to lexical diversity such as type token ratio (TTR) simply compute the number of unique words in a sample divided the total number of words. Because of the high correlation between TTR and length of the sample, the vocd approach addresses the potential confound by using estimation algorithms (McCarthy & Jarvis, 2010).

For this study, *connectives* are operationalized using the CNCAll measure. This measure gives an occurrence score per 1000 words for all connectives. These include connectives that are additive (also, moreover), temporal (and then, after, during), causal (because, so) and logical operators (therefore, if, and, or) (Graesser & McNamara, 2011).

Two measures were used to assess *syntactic complexity*; one assessing the mean number of words before the main verb or left embeddedness (SYNLE) and the other assess the average number of modifiers per noun phrase (SYNNP). For syntactic pattern density DRPVAL, which assess the density of passive voice in text, and DRNEG, which assessing the incidence of negation in text, were selected.

Word frequency measures assess how often a content word occurs in the English language. Coh-Metrix relies on the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995) which contains frequency information for close to 18 million words. For the purposes of this study, a measure (WRDFRQmc) that takes a log transformation of the least frequent content word in each sentence and then averages that result across all sentences of a review was selected. The reason for this choice is that it has been established that the least frequent word in a text passage can often limit how easy a text is to comprehend (McNamara et al., 2014).

Coh-Metrix relies on two databases for additional information on psychological and semantic dimensions of words. For the purposes of this study Coh-Metrix relies on the MRC Psycholinguistic database (Coltheart, 1981) for the measures for word familiarity and word meaningfulness. Coh-Metrix relies on the WordNet database (Fellbaum, 2005) to calculate average polysemy for all content words in an online review.

Word familiarity is assessed using the WRDFamC measure. Word familiarity is a rating of how familiar an adult finds a content word. Similarly, word meaningfulness is a rating of how meaningful a content word is to an adult. Word meaningfulness is assessed using the WRDMEAc measure.

Polysemy is assessed using the WRDPOLc measure. Recall that polysemy refers to the number of meanings for a word. For example, the word "duck" has at least two meanings, one referring to an animal, and the other referring to moving your body towards the ground. Coh-Metrix relies on the WordNet (Fellbaum, 2005) to calculate average polysemy for all content words in an online review.

Multiple linear regression was used to test the identified hypotheses. All independent variables were forced into the model simultaneously given there is theoretical justification for the inclusion of the predictors (Field, 2013). All fundamental assumptions of linear regression were

met. Because of the exploratory nature of this research, we only included control variables that appear in the majority of studies on online review helpfulness. Because of their identified importance in previous studies, word count (DESWC), and product type (Experience/Search) have been added to the regression model as control variables.

RESULTS AND DISCUSSION

Results for the first regression model are presented in Table 2 below.

	Standardized B	t	Significance
Constant		-1.315	.189
DESWC	.169	6.909	.000**
Product Type	-.109	-5.072	.000**
CRFSO1	.013	.431	.666
LSASS1	.045	1.523	.128
LDVOCD	.245	9.895	.000**
CNCAll	.031	1.475	.140
SYNLE	-.038	-1.810	.070
SYNNP	.078	3.493	.000**
DRPVAL	-.051	-2.477	.013*
DRNEG	-.162	-7.551	.000**
WRDFAMc	.047	1.795	.073
WRDFRQmc	.026	1.227	.220
WRDMEAc	.076	2.893	.004**
WRDPOLc	.018	.831	.406
Dependent Variable: Helpfulness Ratio R^2=.204; *p<=.05; **p<=.01			

Table 2: Model 1: Regression Results for Text Complexity on Online Review Helpfulness.

As identified in Table 2, of the cohesion measures, only lexical diversity (LDVOCD) significantly predicts online review helpfulness but it is in the opposite direction of what was hypothesized. Hypotheses H1, H2, H3 and H4 are thus rejected. The average number of modifiers per noun phrase (SYNNP) was a significant predictor of review helpfulness but in the opposite direction of what was hypothesized. The mean number of words before the main verb (SYNLE) was insignificant. Hypothesis H5 was thus rejected. Hypothesis H6 regarding syntactic pattern density was fully supported with both the density of passive voice (DRPVAL) and the incidence of negation (DRNEG) being significant in the hypothesized direction. Finally, of the word measures only, word meaningfulness (WRDMEAc) was significant with word familiarity (WRDFAMc), word frequency (WRDFRQmc), and word polysemy (WRDPOLc) being insignificant. Hypothesis H8 is accepted and hypotheses H7 and H9 are rejected. The independent variables explained 20.4% variance in the model. A reduced model was created that

only includes significant predictors from the original model; these are presented in Table 3 below. The model reduction resulted in a slight decrease in R^2 to 19.6%.

	Standardized B	**t**	**Significance**
Constant		.684	.494
DESWC	.174	7.173	.000**
Product Type	-.104	-4.915	.000**
LDVOCD	.245	10.004	.000**
SYNNP	.055	2.649	.008**
DRPVAL	-.051	-2.484	.013*
DRNEG	-.159	-7.525	.000**
WRDMEAc	.108	5.142	.000**
Dependent Variable: Helpfulness Ratio R^2=.196; *p<=.05; **p<=.01			

**Table 3: Model 2: Regression Results for Text Complexity
on Online Review Helpfulness.**

The results for the cohesion measures are interesting. Specifically the research found that an increase in lexical diversity in online reviews results in an increase in online review helpfulness, which is in the opposite direction to what was initially hypothesized. Recall that lexical diversity assesses the relative amount of unique words to the total number of words in the text and is sensitive to text length (i.e., Shorter text, in general with have higher lexical diversity than longer text). Given the relatively small average length of online reviews, it could be that the impact of additional information provided by unique words in the review is outweighing the impact of the cohesion provided by using fewer unique words relative to text length. The relative short text length of online reviews could also explain why the other cohesion measures were not significant in predicting online review helpfulness. The nonsignificant cohesion results may also be the result of the reverse cohesion effect (O'Reilly & McNamara, 2007) that sometimes occurs. This effect purports that if a reader is already very knowledgeable about the discourse topic they will not benefit from high cohesion text while a low knowledge reader will benefit from cohesive text. So, in this study if the readers of online reviews are already very knowledge about the product they are reading the review of, high cohesion should not have a positive effect on them judging whether the review is helpful or not. Alternatively, a consumer who is just starting out doing research on a product that they know nothing about should value highly cohesive text and be more likely to judge it as being helpful. It also has been reported that reading skill can overcome the benefits of cohesive text (O'Reilly & McNamara, 2007).

One of the two measures of syntactic complexity was significant in predicting online review helpfulness. Specifically, the number of modifiers per noun phrase (SYNNP) was significant, while the words before the verb of the main clause (SYNLE) were not significant. However, the direction of the SYNNP variable was opposite of what was hypothesized. Recall the premise was that text that was more complex would be associated with less helpful reviews. This result indicates that a more complex syntactic structure is positively associated with online review helpfulness. This is not entirely surprising given the goal of author of the online review is to provide information about a product. In doing so, an author may use many modifiers to richly

describe the product in question (the noun) or a characteristic of a product that would result in a high SYNNP assessment. Both measures of syntactic pattern density were significant in the hypothesized direction. The use of passive voice (DRPVAL) in online review text significantly predicts a decrease in online review helpfulness. Similarly, the use of negation (DRPNEG) in online review text significantly predicts a decrease in online review helpfulness. The use of active voice and the reduction of negations should result in online reviews that are easier to understand and more engaging to the potential consumer.

All of the word level measures were insignificant except for word meaningfulness (WRDMEAc). It is surprising that word meaningfulness is significant while word familiarity (WRDFAMc) and word frequency (WRDFRQmc) are not as they all, at a high level, deal with individual word difficulty. Finally, word polysemy (WRDPOLc) was insignificant in predicting online review helpfulness. It could be that because online reviews are so precisely focused on describing a consumer's experience with a single product and are thus densely populated with words reflecting the product domain that words that are assessed as having multiple meanings are easily interpreted by the reader of the online review. In addition, if readers of online reviews are already familiar with the product category in question, their product knowledge network will help scope the interpretation of the polysemic word to its proper meaning.

It is recommended that future work to gain greater insight into the role of text difficulty in online reviews proceed in an experimental setting where the length of reviews and the knowledge of the readers can be controlled more explicitly. This would allow a matching of subjects unfamiliar / familiar with a product to evaluate online reviews of different lengths (for example, short/med/long) and thus test some of the explanations provided for the unexpected results from this study. Additional work could explore the mediating effect of product type (experience/search) on some of the constructs presented in this paper. Future research will investigate the role of positive and negative affect in online review text on review helpfulness using Pennebaker, Francis, and Booth's Linguistic Inquiry Word Count (LIWC) (2001).

CONCLUSION

Overall some indications of the importance of applying a more robust and comprehensive model of text complexity to understand what makes an online review helpful in the eyes of a potential consumer was supported. The results indicate that increased use of passive voice and negation in online reviews results in lower helpfulness ratings. It is recommended that writers of online review to write in an active voice and avoid negation where possible. While there were other significant results, they were in the opposite direction of what was hypothesized and require further investigation before concrete recommendations can be made to provide prescriptive advice to online review writers.

REFERENCES

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database [CD-ROM]. Philadelphia: Linguistic Data Consortium.

Baek, H., Ahn, J. & Choi, Y. (2013). Helpfulness of online consumer reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce*, *17*(2), 99-126.

Bell, C. M., McCarthy, P. M., & McNamara, D. S. (2012). Using LIWC and Coh-Metrix to investigate gender differences in linguistic styles. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation, and resolution* (pp. 545-556). Hershey, PA: IGI Global. doi: 10.4018/978-1-60960-741-8.ch032

Bickart, B., & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, *15*(3), 31-40. doi: 10.1002/dir.1014

Boone, L. E., Kurtz, D. L., MacKenzie, H., & Snow, K. (2013). *Contemporary marketing (3rd ed.)*. Toronto, ON: Nelson Education

Brooks, C., & Warren, R. P. (1972). *Modern rhetoric*. New York: Harcourt Brace Jovanovich.

Clemons, E., Gao, G., & Hitt, L. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, *23*(2), 149-171. Doi: 10.2753/MIS0742-1222230207

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33*, 497-505. Doi: 10.1080/14640748108400805

Crossley, S. A., Louwerse, M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal, 91,* 15-30. doi: 10.1111/j.1540-4781.2007.00507.x

Dempsey, K. B., McCarthy, P. M., & McNamara, D. S. (2007). Using phrasal verbs as an index to distinguish text genres. In D. Wilson & G. Sutcliffe (Eds.), *Proceedings of the twentieth International Florida Artificial Intelligence Research Society Conference* (pp. 217-222). Menlo Park, CA: AAAI Press.

Fellbaum, C. (2005). WordNet(s). In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., pp. 665-670). Oxford, UK: Elsevier, 665-670

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London, UK: Sage.

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

Graesser, A. C., Cai, Z., Louwerse, M. M., & Daniel, F. (2006). Question understanding aid (QUAID): A Web facility that tests question comprehensibility. *Public Opinion Quarterly, 70*(1), 3-22.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, *3*, 371-398. doi: 10.1111/j.1756-8765.2010.01081.x

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223-234.

Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, *52*, 674-684. doi: 10.1016/j.dss.2011.11.002

Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 244-253. doi: 10.1016/S0022-5371(71)80051-8

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston, MA: Allyn & Bacon.

Korfiatis, N., Garcia-Barriocanal, E. & Sanchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, *11*, 205-217. doi: 10.1016/j.elerap.2011.10.003

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum.

Li, M., Huang, L., Tan, C-. H., & Wei, K-. K. (2013). Helpfulness of online product reviews as seen by consumers: Source and content features. *International Journal of Electronic Commerce*, *17*(4), 101-136.

Louwerse, M. M., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2004). Variation in language and cohesion across written and spoken registers. In K. D. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 843-848). Mahwah, NJ: Erlbaum.

Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 58-71). Clevedon, UK: Multilingual Matters.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*, 381-392. doi: 10.3758/BRM.42.2.381

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. New York, NY: Cambridge University Press.

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247-287. doi: 10.1080/01638539609544975

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, *34*, 185-200.

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1200-1210.

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, *43*, 121-152. doi: 10.1080/01638530709336895

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001* [Computer software]. Mahwah, NJ: Erlbaum Publishers.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227-247). Oxford: Blackwell. doi: 10.1002/9780470757642.ch13

Schindler, R. M., & Bickart, B. A. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behavior*, *11*, 234-243. doi: 10.1002/cb.1372

Simonson, I., & Rosen, E. (2014, January-February). What marketers misunderstand about online reviews. *Harvard Business Review*, 23-25.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic.

Willemsen, L. M., Neijens, P. C., Bronner, F., & de Ridder, J. A. (2011). "Highly recommended!" The content characteristics and perceived usefulness of online consumer reviews. *Journal of Computer-Mediated Communication*, *17*(1), 19-38. doi: 10.1111/j.1083-6101.2011.01551.x

Wu, P. F. (2013). In search of negativity bias: An empirical study of perceived helpfulness of online reviews. *Psychology & Marketing*, *30*, 971-984. doi: 10.1002/mar.20660

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162-185.

This Page Was Left Blank Intentionally.