

Electronic Word of Mouth: A Genre Analysis of Product Reviews on Consumer Opinion Web Sites

Irene Pollach

Vienna University of Economics and Business Administration
irene.pollach@wu-wien.ac.at

Abstract

Consumer opinion Web sites enable consumers to post reviews of products and services or view the experiences of other consumers. This form of writing can be considered a truly digital genre, as consumers were not able to share their opinions with other consumers in a structured, written format before the advent of the Internet. To identify rules and conventions established by the genre community, a sample of 358 product reviews was examined using a methodology that combines elements of case study research, corpus linguistics, and textual analysis. More precisely, the analysis focused on structure, content, audience appeals, sentence style, and word choice. The results of this analysis have implications for improving the design of consumer opinion Web sites with a view to making them more useful sources of consumer knowledge.

1. Introduction

In the past, consumers used to talk to other people when looking for opinions on a particular brand, product or company. This became known as word-of-mouth (WOM) in the marketing literature. With the advent of computer-mediated communication, these conversations moved to the WWW where consumers can share their opinions, thus engaging in electronic word-of-mouth (eWOM). Since such messages did not exist in writing before but were sent and received orally in an unstructured manner, online product reviews written by consumers for other consumers are considered a new genre. Like any digital genre, consumer opinion sites and therefore online product reviews are undergoing tremendous change and will continue to do so. Therefore, it may be useful to have a snapshot of this genre, which could serve as a starting point for tracking the changes it will undergo. The genre's evolution may be of particular interest to companies harvesting consumer opinion Web sites for marketing intelligence. Accordingly, this paper seeks to gain insights into the current nature of online product reviews and the rules established by the genre community.

2. The Nature of Electronic Word-of-Mouth

In general, word-of-mouth is defined as informal, non-commercial, oral, person-to-person communication about a brand, a product or a service between two or more consumers [1]. WOM among consumers incorporates three different activities. First, information is sought for immediate use aimed at risk reduction. Second, information is obtained and stored for future usage and, third, information is shared in order to influence other people's decisions [25]. WOM is used when buyers lack the information necessary for a purchase or when they perceive the risk associated with the purchase as high [22]. Consumers have been found to turn to personal contacts for reassurance and to loose contacts for their expertise [15].

Since people are basically willing to heed the advice of strangers, the anonymity of the WWW is by no means an obstacle to the success of eWOM. Consumer opinion Web sites have cropped up on the WWW, providing unprecedented opportunities for consumers to voice their opinions on companies, products and services in a structured, written format in the form of product reviews, complaints, discussion threads, or chats [16, 30]. This section looks at why people participate in this genre and what impact it may have on consumers and businesses.

2.1 Motivation for Participation

In commercial settings in the offline world, consumers have been found to initiate conversations with other consumers to offer advice and information without having been asked to do so [21]. Consumer opinion Web sites tap into this very desire of people to share information about topics they consider themselves to be experts on [28]. The availability of their opinion to others is particularly appealing to opinion leaders, who receive and transmit more information on topics they are interested in than other people [22]. On the Web, consumers can claim authorities they may not be and would not be able to claim in the real world, as the anonymity of the Internet even enables people to post bogus reviews on products they do not own or have never even used [18]. Apart from the social

benefits people obtain when their opinions are made available to and read by others, they may also reap economic benefits. Some consumer opinion sites offer financial incentives to reviewers [23].

2.2 The Impact of eWOM

In general, consumers are influenced by and rely on what others say about a product before they buy it [32]. If there is not enough information about the product available from other sources and the risk involved for the buyer is therefore high, the influence of WOM on consumers' purchasing decisions is also high [20]. However, product information provided by companies is less influential among consumers than information provided on consumer opinion sites or discussion boards [5]. Also, consumers consider negative WOM information more helpful than positive information in distinguishing between high-quality products and products of low quality [24]. To companies, eWOM may serve as a feedback mechanism that helps them to improve the quality of their products and to acquire new customers [12]. Companies may even offer consumer opinion forums on their own Web sites to strengthen customer loyalty and reduce service costs [7]. The feedback companies obtain should become a key component of electronic customer service [8].

3. Genre Theory

Genre analysis has been used in IS research to study communication practices occurring within IT-mediated communication systems [17]. This study applies genre theory to online product reviews posted on consumer opinion Web sites to examine the rules established by the participants in this genre. The following sections take a theoretical look at the traditional concept of genre, genres that have emerged on the Internet, and the generic nature of online product reviews.

3.1 Traditional Genre Theory

Miller suggested that only writers who are familiar with the context of a situation are able to use rhetorical strategies suitable for specific situations. Accordingly, she defined genre as "typified rhetorical actions based on recurrent situations" [27]. In their seminal work on genres of organizational communication, Yates and Orlikowski characterize genres by their shared communicative purposes and form, the latter of which includes structural text features, the communication medium, and the language system. They also hold that the communicative purpose of a genre is determined by the whole genre community rather than individual writers [39]. According to Bhatia, the notion of communicative purpose enables not only a

distinction among different genres but also between genre and subgenre, although the line between the two may be a very fine one [3]. Swales also stresses the shared communicative purpose among communicative events as the defining criterion of genre, which he defines as "communicative vehicles for the achievement of goals" [36]. Yet another view on genre is offered by Devitt et al., who draw on rhetorical genre theory to explain how and why texts are important in our lives. They see genre as a reciprocal dynamic that "reflects, constructs, and reinforces the values, epistemology, and power relationships" of a genre community [13].

3.2 Genres on the WWW

In line with Yates and Orlikowski's argument that genre repertoires change when new communication media emerge [39], Crowston and Williams have observed that the WWW has modified existing genres and given rise to new ones. In their large-scale study of Web sites they identified a large number of new genres but also found instances of genres embedded in other genres [11]. Dillon and Gushrowski, who found a shared set of user expectations regarding the content of personal home pages, hold that the personal home page was the first true digital genre [14]. Shepherd and Watters have identified six different cybergenres, including home pages, brochures, resources, catalogues, search engines, and games, which they characterize according to content, form and functionality [34]. Clearly, both form and functionality of digital genres are constantly evolving along with advances in Internet technology [37]. To account for the dynamic and complex nature of digital genres, Crowston and Kwasnik have suggested a faceted classification scheme for genres, arguing that genre classification should be based on both document characteristics and the context in which the document is used [10].

3.3. Online Product Reviews as a New Genre

Online product reviews written by consumers can be considered a truly digital genre in that they are a form of writing that has only existed since the emergence of consumer opinion sites on the WWW. Previously, people shared such information only orally in the form of WOM communication with other consumers or wrote letters to companies, but did not have the opportunity to share their opinions with other consumers in a structured, written format. Also, consumers were not able to obtain product-related information from strangers.

The purpose of online product reviews is to inform potential buyers of the strengths and weaknesses of consumer products. People who share their experience help readers make purchasing decisions and may even be recognized as experts in a particular field if their product

reviews are of superior quality. Writers practicing this genre are not professional writers, let alone professional critics. Nevertheless, they seek to produce a technically accurate text that is helpful to those not owning the product they are reviewing. They may also not be native speakers of the language they are using to write the product review.

Electronic word-of-mouth occurs in a variety of formats, including not only product reviews but also discussion threads, chatrooms or complaint sites, which enable consumers to interact in the form of dialogues or polylogues. What distinguishes online product reviews from other forms of consumer interactions is that they are isolated texts unrelated to previous messages posted on the same site [30]. Further characteristics of online product reviews include persistency, asynchronicity and the potential absence of feedback.

4. Data and Methodology

Genre analysis, broadly defined as "the study of situated linguistic behaviour" [4], yields more insights when its focus is narrow, e.g. when studying a typical example of a genre, as it needs to take into account the complexities and dynamics of the world. Typically, genre studies incorporate a range of research methods, the most prominent being corpus linguistics, textual analysis, and case studies [4]. This paper makes use of these three approaches, attempting to identify the formal, structural and linguistic features of the genre of online product reviews. The results will contribute to our understanding of their nature, scope, and significance. Further, this analysis provides a starting point for analyzing future changes this genre is subject to.

A corpus of online product reviews was collected from *reviewcentre.com*, a large online product forum covering hundreds of different products. The product category of digital cameras was chosen for the analysis, as it is a highly competitive market selling information-intensive, expensive products. Therefore, consumers are likely to turn to online sources for opinions on digital cameras before making purchasing decisions and at the same time may be willing to share their own experiences once they have bought one.

Overall, a corpus of 358 product reviews was compiled, using all reviews posted in the category of digital cameras for the top 15 digital cameras from each of the top four digital-camera brands (as ranked by the forum). The 358 reviews resulted in a corpus of 64,400 words, with an average of roughly 180 words per review.

The textual data pertaining to the genre of online product reviews were analyzed both qualitatively and quantitatively. First, all messages were closely read multiple times to discover emerging themes that help to understand formal and linguistic peculiarities of this genre

[40]. Second, the texts were analyzed quantitatively using *WordSmith Tools* and a spreadsheet package. For the calculation of several quantitative features it was necessary to convert the text corpus into a word list. The large number of spelling errors in the text necessitated the manual correction of all mistakes using the spell checker of a word processing package. Ultimately, all words were lemmatized to remove apostrophes, plural endings, verb inflections, and adverb endings.

Further, to be able to interpret the textual statistics of a text corpus, it is necessary to compare it to other text corpora. For the purpose of this analysis, one issue of the *Economist* (November 30, 2002) and a corpus of 50 privacy policies originally compiled for another study [29] were used as reference corpora in order to compare the product reviews to two completely different corpora.

5. Results

The analysis looks at the genre of online product reviews from four different angles. These include (1) structure and format, (2) content, (3) appeals to audience, and (4) choice of sentence style and words, as suggested by Devitt et al. [13]. The analysis takes into account both regularities and deviations to understand the rules established by the genre community.

5.1. Structure and Format

On *reviewcentre.com*, reviewers need to register with the site in order to be able to voice their opinions using a self-selected screen name. First, they are supposed to rate the product according to pre-defined categories relevant for a particular product. For digital cameras, these categories include: "Time Digital Camera Owned", "Image Quality", "Battery Life", "Features", "Ease of Use", "Value for Money", "Overall Rating", and the question "Would you recommend it to a friend?". Users can then voice their opinions verbally in the categories "good points", "bad points", and "general comments". A typical review looks as follows:

Good Points:

Size, sexy shape, features.

Bad Points:

None I've found so far! Although I miss a view finder.

General Comments:

After having an Olympus 35mm camera for many years I was unwilling to have any other make of digital - you just cannot beat the lenses on Olympus. So when I got a Mju Mini for Christmas - WOW.

A small, light sexy beast – intuitive [sic!] to use, good display and handling. The only thing I miss is the view finder!

This structure channels the writer's opinion into considering both negative and positive points, even if his/her prior opinion strongly leans towards one end of the spectrum of possible opinions. However, a few reviews do not follow this format, containing either only general comments, only good points or only bad points. Three reviews were also typed exclusively in capital letters, which is detrimental to the readability of the text. For example:

Good Points:
Bad Points:
General Comments:
 GREAT CAMERA FOR THE MONEY. COMES
 HIGHLY RECOMMENDED.

On *reviewcentre.com*, readers of product reviews can respond to a product review by indicating whether they found the review helpful or not and—if they own the same product—they can also indicate whether they agree with what the reviewer has written. These evaluations are translated into points, indicating the "respect" shown to a reviewer. Further, readers can add verbal comments to a

review. When users browse the product reviews posted for a particular product they see both the reviewer's overall rating of the product (expressed in points) and the readers' "respect" towards the reviewer, also expressed in points (see Figure 1).

5.2 Content

It would not be insightful to perform a detailed content analysis, as all reviews contain mostly positive, negative, or neutral information about digital cameras and digital photography. It seems to be more interesting to look at irregularities and deviations from the ordinary instead, as the amount of variation is inversely related to the level of consistency of discourse [31]. Apart from comments, evaluations, and personal stories (e.g. weddings, vacations, christenings) involving the product reviewed, the texts contain only few irregular features. One such feature is hyperlinks. Although people are not able to provide activated hyperlinks, five reviewers still pasted the URL into their texts. These links direct readers either to the store where they have bought their cameras or to their own online photo albums boasting pictures taken with the camera they are reviewing.

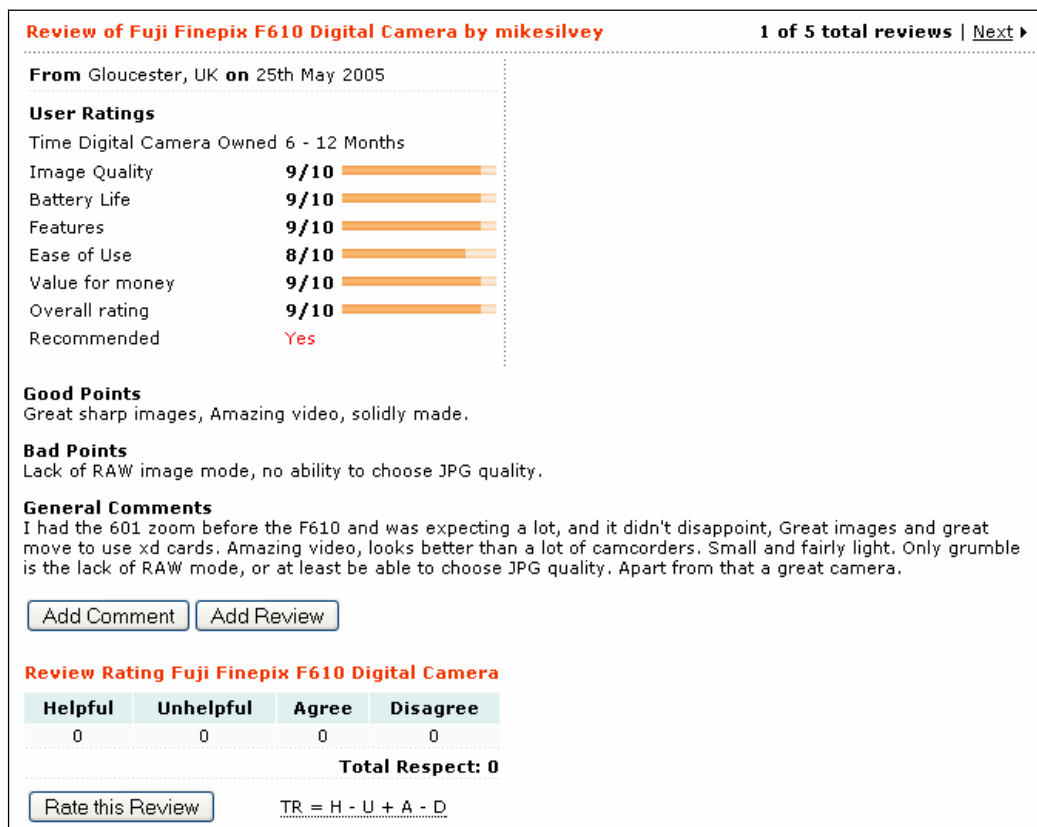


Figure 1. Exemplary Review from *reviewcentre.com*

A second deviation from the ordinary is questions. Rather than advising other people on digital cameras and photography some people ask readers for advice, e.g. *Any ideas guys?* or *Can anyone help?*. The corpus contains a total of ten such questions or cries for help. The fact that none of these were answered indicates that reviews are meant to remain in isolation and that other members of the genre community consider such questions to be violations of the genre rules and therefore do not answer them.

In a few rare cases reviewers took up what other reviewers had written. Overall, there are six instances of such intertextuality in the corpus. They occur in two different forms. First, two reviewers back their own claims by referring to other people who have voiced the same opinion: *I agree with previous comments on ...* and *other reviews all agree on*. Second, they advise people not to pay attention to negative reviews, claiming that the reviewers just do not know how to use the product properly (4 times). For example, one reviewer claims: *These people should properly read the manual before rubbishing a very good product*.

5.3 Appeals to Audience

The interlocutors in computer-mediated communication are only textually realized personas and therefore visually anonymous [6]. Since people may not be who they claim to be, credibility is an inherent problem in consumer-to-consumer interactions on the WWW. Product reviewers may well be manufacturers or merchants seeking to promote their products with guerilla marketing tactics or badmouthing those of others. Therefore, to be credible reviewers need to convince readers of their expertise and trustworthiness.

To examine whether authors of product reviews use appeals to credibility, the corpus was examined in light of Aristotle's classic credibility appeals of *pathos* (emotions), *logos* (reason), and *ethos* (character of the speaker). *Pathos*-based arguments attempt to persuade by eliciting emotional responses from the audience. Appeals to *logos* use sound logic and often also inductive reasoning to persuade, while *ethos*-based arguments seek to persuade by calling attention to the character of the speaker/writer, e.g. expertise, experience, authority [9].

Emotive appeals were only used sparsely in the product reviews. The only instances found were ironic remarks, typically self-mockery, e.g. *even for a novice thicko like me* or *idiot proof, so that suits me*. On *reviewcentre.com* reviewers cannot integrate or link to images, which could serve as emotional appeals, into their product reviews to enhance the credibility of their claims.

Appeals to reason were not very prevalent in the corpus either. They were presented in the form of independent, third party evaluations of the reviewer's purchasing decision. Examples of such appeals include the claim that

other people *loved* the pictures taken with this camera, awards the reviewers had won for pictures taken with their cameras, and the fact that other people bought the same camera after they had seen the reviewer's pictures.

By far the most prevalent type of credibility appeal was *ethos*. Most authors of product reviews provide information about their own history of digital photography, often in the introductory sentence of their reviews. For example, most of them state the date on which they purchased their digital camera or for how long they had used single lens reflex (SLR) cameras before they switched to digital photography. Also, they often stressed how thoroughly they had researched the market before they decided on a camera. Other information provided to demonstrate their expertise includes the number of pictures taken so far and occasions at which the camera proved invaluable.

One cannot safely say whether the authors of product reviews seek to establish their authorities consciously (e.g. to be shown more "respect") or unconsciously, but the analysis of argumentation shows that they do so mostly by appealing to their readers' trust in their experience as photographers. It is not surprising that emotive appeals are not used frequently, given that the *raison d'être* of this genre is to inform readers rather than to stir their emotions.

5.4 Sentence Style and Word Choice

5.4.1. Word and Sentence Length. Word lengths in the corpus of online product reviews and the two reference corpora were compared by looking at the distribution of word lengths in each corpus, which range from 1 to 12 characters. Figure 2 depicts the relative distribution of word lengths in the three corpora. As the diagram shows, the product reviews examined contain relatively more short words (3 to 6 letters) than the other two corpora, while the other two corpora contain relatively more words of 7 to 12 letters. These results suggest that individuals participating in computer-mediated communication (CMC) use shorter words than writers of more formal text such as articles in print media or legal documents.

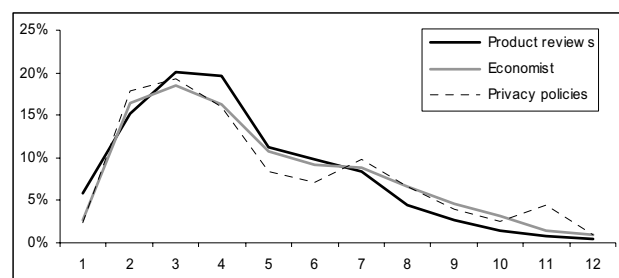


Figure 2. Relative Distribution of Word Lengths

The average length of sentences in the corpus of product reviews and the two reference corpora was calculated using *WordSmith Tools*. The average sentence lengths were 17.4 words for the product reviews, 22.5 for the *Economist* and 26.8 for the privacy policies. These results seem to mirror the level of formality inherent in these texts. Privacy policies as legal documents tend to have the longest sentences, while the *Economist* as a news and business magazine uses sentences of medium length. Online product reviews, however, use the shortest sentences. A reason for this may be that they are not written by professional writers and that they are written specifically for the Internet, which means that people often use simpler structures or do not even write in complete sentences.

5.4.2. Lexical Richness. To study the distribution of words in the corpus, a type-token analysis was performed. The type-token ratio (TTR) divides the number of distinct words in the corpus (types) by the total number of words (tokens), while the standardized type-token ratio (sTTR) is a running average based on consecutive 1,000-word chunks of text [26]. The sTTR makes text of differing lengths comparable, since shorter texts tend to have higher TTR than longer ones. In general, a high TTR suggests that the vocabulary used is rather heterogeneous, whereas a low TTR indicates that a corpus is lexically not very rich. As Table 1 shows, the corpora of product reviews and privacy policies have rather small type-token ratios, suggesting that the vocabulary used is homogenous, as are the foci of these two corpora (i.e. product reviews and privacy policies). By contrast, one issue of the *Economist* covers a broader variety of topics, thus containing more lexical variety. Thus, the results suggest that lexical richness among product reviews is low, but not as low as among privacy policies.

Table 1. Corpora Types, Tokens and TTR

	Product Reviews	Economist	Privacy Policies
Documents	358	1 issue	50
Tokens	64,400	53,089	60,255
TTR	7.12%	16.32%	5.00%
sTTR	39.90%	49.74%	33.41%

5.4.3. Word Frequencies. A frequency analysis was performed, looking at the 100 most frequent content words in the corpus. For this purpose all grammatical words (e.g. articles, conjunctions, prepositions, pronouns etc.) and other words without semantic content (e.g. numbers, brand names) were removed. Also, photography terms were removed, as a high frequency of such terms is to be expected and the focus of the analysis is on the genre of product reviews in general rather than reviews of digital cameras. The remaining 100 most frequent words, which occurred from 39 to 676 times, were grouped into 6

categories according to their meaning plus 2 categories containing the remaining verbs/nouns and adjectives. Table 2 shows for each category the number of different words (types), the three most frequent words, the total number of words (tokens), and the type-token ratio.

Table 2. Word-List Types, Tokens, and TTR

Category	Types	Top Three	Tokens	TTR
Verbs/nouns	36	use, quality, take	3,885	0.93%
Positive	19	good, easy, great	2,357	0.81%
Negative	8	not, no, problem	1,241	0.64%
Emphasis	8	very, only, real(ly)	1,046	0.76%
Consumption	7	buy, price, need	761	0.92%
Adjectives	10	small, little, full	599	1.67%
Time	7	time, last, day	501	1.40%
Expression	5	recommend, say, think	344	1.45%
TOTAL	100	good, use, not	10,695	0.94%

Clearly, verbs and nouns with general meanings make up the largest proportion in terms of both types and tokens. As is also evident from Table 2, positive words, negative words, and words of emphasis play an important role in product reviews. The fact that the TTR of these three categories is lower than those obtained for general verbs/nouns (0.93%) and the total sample (0.94%), indicates that each of these words occurs relatively more often, and words of consumption do so slightly as well. The noteworthy frequency of time-related words can be put down to the fact that people tend to state in their reviews when they bought the product or for how long they have had it.

5.4.4. Computer-Assisted Semantic Analysis. To assess the semantic content of all product reviews, several so-called "dictionaries" (in the sense of word lists) originally compiled for the *General Inquirer* project [19, 35] were compared against the words used in the product reviews. These dictionaries capture all words that relate to a certain semantic category but are not mutually exclusive. To account for lexical ambiguities such as polysemes (words with several related meanings) and homographs (words with unrelated meanings sharing the same orthographic form), the dictionary entries give the probability with which words carry each of their possible meanings and therefore belong to certain dictionaries. For this study, only words that belong to a category with a probability of 99% were considered [33].

Of the 182 dictionaries included in the *General Inquirer*, 13 were considered potentially relevant to the study of this genre (see Appendix for details on these dictionaries). They were compared against a lemmatized list of all words and their frequencies in the product-review corpus to determine which semantic categories are the most prevalent in the corpus. This kind of analysis differs from the analysis of word frequencies above (5.4.3) in that the former looked at which words (and

concepts) were used most often, thus pursuing an inductive strategy, whereas this analysis adopts a deductive approach by comparing the entire range of words used against predefined lists of words, thereby corroborating the findings from 5.4.3.

Table 3 summarizes the findings from this analysis. It gives the number of words each *General Inquirer* dictionary contains (GI Words), the percentage of these words represented in the corpus (Types), the frequency with which these types occur in the corpus in absolute terms (Tokens), the average number of instances of the GI words (Token/Word), and the type-token ratios.

The type-token ratios for the GI dictionaries give insights into the distribution of GI words by calculating the average frequency with which a type occurs in the corpus. A high TTR indicates that a greater variety of GI words is used, while a low ratio suggest that only a few GI words (or maybe just one) are used very often. In fact, the five lowest TTR (approx. 5% or lower) can all be put down to just on GI word used more than twice as often as the second most frequent GI word in that category. These words include: *very* (Overstatement), *small* (Understatement), *excellent* (PosAff, Evaluation), and *seem* (If).

Table 3. *General Inquirer* Analysis

	GI Words	Types	Tokens	Token/Word	TTR
Negative	1,947	13.2%	2,184	1.12	11.8%
Positive	1,472	19.2%	1,954	1.33	14.5%
Econ	367	22.6%	992	2.70	8.4%
Quality	246	20.7%	427	1.74	11.9%
Know	235	31.5%	1,000	4.26	7.4%
NegAff	133	16.5%	142	1.07	15.5%
Fail	113	8.8%	31	0.27	32.3%
Overstatement	72	68.1%	1,384	19.22	3.5%
PosAff	67	40.3%	558	8.33	4.8%
Evaluation	39	74.4%	582	14.92	5.0%
Try	36	19.4%	20	0.56	35.0%
Understatement	24	79.2%	438	18.25	4.3%
If	10	70.0%	137	13.70	5.1%

For a category to be highly represented in the corpus, the results obtained for the relative number of types and the token-word ratios would have to be high. In the present analysis those categories most prevalent in the corpus include *Understatement*, *Evaluation*, *If*, and *Overstatement*. Not only was a high proportion of these GI words found in the corpus (approx. 70%), but the GI words also resulted in a very high number of tokens relative to the number of GI words (13.70 and above on average). Those categories least represented in the corpus include *Fail*, *Try*, *NegAff*, and *Negative*.

5.4.5. Negation. Since product reviewers are encouraged to deal with negative aspects of the product they are reviewing, it is worth examining how reviewers report negative information about products they rate positively overall. In 52 instances, for example, reviewers coupled

words like *problem*, *gripe*, *complaint*, *niggles*, *drawback*, *quirks*, *downer*, *down point*, *bug bear*, *criticism*, and *irritation* with mitigating adjectives like *only*, *occasional*, *small*, *slight* and *minor* to downplay negative points. Also, semantically neutral words such as *thing*, *point* or *comment* carry negative meaning when coupled with *bad* or *negative*.

Syntactically, negative information was often presented together with positive information and linked with the contrastive conjunction *but*, e.g. *It is true that the camera has limitations but ...* or *This camera is not 110% perfect but ...* Also the argument that every product has downsides or that certain problems occur with every digital camera were used.

5.4.6. Personal Pronouns. Looking at pronouns can give insights into the extent to which people talk about themselves, about their audience, or about third parties. To ensure that only pronouns that unmistakably refer to human beings are included in the analysis, third-person plural pronouns (they, their/s, them/selves) were excluded. The results in Table 4 show that personal pronouns occur quite frequently, averaging 8 personal pronouns per product review (excluding references to the third person plural).

In particular, the frequent use of "I" (on average 4 per review) indicates that writers talk frequently about themselves. There are only a total of 42 references to the first person plural, suggesting that writers talk about their personal experience rather than their families'. Writers also address their audiences directly or generically, but to a far lesser extent than they talk about themselves. In fact, a total of 2,035 first-person pronouns are used, but only 755 second-person pronouns. Third parties are referred to with personal pronouns only to a miniscule extent.

Table 4. Breakdown of Pronouns

I	1,550
you	629
my	359
your, yours, yourself	126
me, mine, myself	126
we, us, our(s), ourselves	42
(s)he, him, his, her(s), him/herself	20
Total	2,852
Pronouns per review	8

5.4.7. Formality. The language of computer-mediated communication tends to be colloquial in nature and often reads as if it was spoken. Electronic discourse thus has characteristics of both oral and written language [38]. Evidence for the informality of CMC language found in the corpus includes abbreviated word forms (e.g. *addl*, *cam*, *yrs*), non-standard spellings (e.g. *pix*, *coz*, *w/o*), and colloquial contractions (e.g. *kinda*, *dunno*, *gonna*). To save writing time, apostrophes are frequently omitted in

contracted forms (e.g. *dont*, *thats*) and messages are written in lowercase throughout. Further, the corpus contains a small number of conventionalized acronyms, used in CMC to shorten commonly used expression. Only three such acronyms were found in the online product reviews, including *BTW* (once), *IMO* (twice), and *IMHO* (once). These observations suggest that people seek to make their writing more efficient by using short forms, but are not avid users of Internet lingo.

It seems that people do not take the time to proofread their messages and subject them to a spell check before they post them. Apostrophes are frequently misplaced (e.g. *is'nt*, *it's successor*) and both spelling errors and typographical mistakes (e.g. *unfotunately*, *amature*, or *enthousiatic*) abound. This suggests that the error tolerance in this genre community is high, valuing content over form. However, while writers probably feel that the anonymity of the Web protects them from embarrassment, such errors will impact the literate reader's impression, since language is the only means of self-presentation in online product reviews.

In addition to shortened forms and errors, the language of online product reviews is characterized by interjections characteristic of spoken language rather than written. These interjections include *ah/oh well*, *bugger*, *erm*, *hey*, *man*, *no*, *oh yeah*, *wow*, and *yes*. For example:

- *Yes it was more expensive, but:*
- *No, no, I dont [sic!] work for Olympus!*
- *So when I got a Mju Mini for Christmas – WOW*

To some extent, these patterns mirror monologic or dialogic speech, reflecting a conversation rather than a piece of writing. This seems to confirm the notion that the language of CMC is a hybrid of spoken and written discourse.

5.4.8. Paralinguistic Features. Participants in computer-mediated communication have developed orthographic strategies designed to compensate the impersonality of written discourse. When using these non-verbal cues "[t]he writer tries to enforce a univocal interpretation on prose that is otherwise open to many interpretations" [6]. In particular, capitalization, spelling, and punctuation are used to express what disembodied words on a computer screen cannot convey [38], e.g. emotions or emphasis.

One such means is iconic sequences of ASCII characters ("emoticons") intended to add positive or negative tones to utterances or to indicate irony [2]. Only twelve emoticons were found in the corpus (9 positive, 2 ironic, 1 negative). To signal emphasis of certain words visually, people use capital letters for individual words or put them in between asterisks (e.g. *GREAT*, **really**). Similarly, people overuse punctuation marks, in particular exclamation marks, to express enthusiasm for the product they bought (e.g. *Recomended!!!!* [sic!]), but also question marks (e.g. *So what???*), or combinations of the

two (e.g. *What's up with this??!!!*) to express anger or disappointment. To convey emotions in computer-mediated conversations people also use conventionalized acronyms [2]. The only such acronym found in the corpus was *LOL*, expressing laughter. In general, non-verbal cues expressing emotions were not very prevalent in the corpus, suggesting that people take their task as reviewers seriously, using neutral, non-emotive language.

6. Discussion and Conclusion

Writing is both a cognitive act and a social practice. Writing adequately, therefore, means adapting one's words to the expectations of the interlocutors [40]. These expectations can also be viewed as the rules implicitly established by the genre community. The online product reviews studied also seem to adhere to such implicit genre rules regarding content, format, and language.

Reviewers tend to take their task as critics seriously, stating not just that they love or hate the product but reporting problems in great detail or describing in reasonable detail how useful the product was to them on a certain occasion. Typically, product reviews contain the product's good points, its bad points and general comments, all of which the site encourages users to provide. It is common for reviewers to provide evidence for the expertise they claim, e.g. by specifying since when they have used the product and for what purposes they have used it. Notably, these reviews remain in isolation. They are generally not linked textually or hypertextually to other relevant information. Also, authors do generally not seek to encourage readers to respond to what they have written.

The language used in these product reviews has relatively few instances of the language typical of CMC. The sparse use of paralinguistic features suggests that reviewers are careful not to make their reviews too informal and thus appear unprofessional. At the same time, the texts are less formal than news features and legal texts, as the narrower lexical range and shorter sentences and words suggest. Words of emphasis, de-emphasis, and vagueness as well as words judging the quality of the product are very prevalent in online product reviews, suggesting that reviewers are enthusiastic about the product but are careful not to present their opinions as the universal truth. They also tend to include themselves using expressive verbs in their texts (e.g. *I think*) rather than the categorical present tense to describe how things are.

Another point to consider in genre analysis is how or why the texts have been textualized the way they are. In the present study, the writers' state of strong emotional arousal may be the reason why they participate in this genre in the first place. The frequent use of overstatements, understatements and words of emphasis as well as

orthographic phenomena such as overpunctuation, capitalization, and emoticons suggest that the genre participants are strongly emotionally involved with the subject matter. They need to express strong emotions verbally or sometimes orthographically to voice their satisfaction or dissatisfaction with a product and to recommend it strongly or to advise people not to buy it.

Noteworthy deviations from the genre conventions found in the online product reviews include: Three reviews typed in capital letters, ten questions directed at the audience, five hyperlinks, six intertextual comparisons, five acronyms (*IMO*, *IMHO*, *BTW*, *LOL*), and twelve emoticons. These features were found in 37 product reviews, i.e. 10.33% of all reviews. One review contained three irregularities, two reviews included two such features, and 34 reviews had one irregular feature each. In the genre community studied, there are no direct penalties for violating genre conventions. Readers can post comments and show their "respect" towards a reviewer, but the usefulness of a review is not necessarily reduced when linguistic genre conventions are violated and therefore it seems unlikely that readers use these facilities to penalize reviewers for violating genre rules.

Although the findings may not be generalizable to reviews of products other than digital cameras, the results of the genre analysis still have implications for the design of consumer opinion Web sites and in particular product-review sites. In particular, some of the deviations of the genre rules can presumably be put down to the design of the Web site. For example, questions directed at the audience would not appear in product reviews if the site had a discussion forum as well. Further, if the site supported activated hyperlinks more people would make use of this facility to direct other users to their own picture galleries. Also, giving users the possibility to make personal profiles available to others would help reviewers to provide information about themselves in a more structured manner. Alternatively, the site could add a field on top of the review box where people enter for how long they have used the product. It is worth noting that orthographic errors were so prevalent in the corpus that they can hardly be considered a deviation from genre rules. Although texts made available on a Web site are potentially planned and prepared beforehand, this does not seem to be the case with online product reviews. Rather, they appear to be the spontaneous product of high spirits. Offering a spell checker to writers would not only raise their credibility but would also make the texts better suited for corporate data mining activities, which could help companies to improve the quality of their products and services.

References

- [1] Arndt, J. (1967). "The Role of Product-Related Conversations in the Diffusion of a New Product", *Journal of Marketing Research*, 4, 291-295.
- [2] Baym, N.K. (1998). "The Emergence of On-Line Community", *Cybersociety 2.0. Revisiting Computer-Mediated Communication and Community*. Ed. S.G. Jones. Thousand Oaks, Sage. 35-68.
- [3] Bhatia, V.K. (1993). *Analysing Genre: Language Use in Professional Settings*, Harlow: Longman.
- [4] Bhatia, V.K. (2002). "Applied Genre Analysis: A Multi-Perspective Model", *Ibérica*, 4, 3-19.
- [5] Bickart, B. and Schindler, R.M. (2001). "Internet Forums as Influential Sources of Consumer Information", *Journal of Interactive Marketing*, 15(3), 31-40.
- [6] Bolter, J.D. (1996). "Virtual Reality and the Redefinition of Self", *Communication and Cyberspace. Social Interaction in an Electronic Environment*. Eds. L. Strate, R. Jacobson and S.B. Gibson. Cresskill, NJ: Hampton Press. 105-119.
- [7] Chiou, J.-S. and Cheng, C. (2003). "Should a Company Have Message Boards on its Web Sites?", *Journal of Interactive Marketing*, 17(3), 50-61.
- [8] Cho, Y. et al. (2002). "An Analysis of Online Customer Complaints: Implications for Web Complaint Management", *Proceedings of the 35th Hawaii International Conference on System Sciences*, Los Alamitos: IEEE Press.
- [9] Cragan, J.F. and Shields, D.C. (1998). *Understanding Communication Theory*. Boston: Allyn and Bacon.
- [10] Crowston, K. and Kwasnik, B.H. (2004). "A Framework for Creating a Facetted Classification for Genres: Addressing Issues of Multidimensionality", *Proceedings of the 37th Hawaii International Conference on System Sciences*, Los Alamitos: IEEE Press.
- [11] Crowston, K. and Williams, M. (1997). "Reproduced and Emergent Genres of Communication on the World Wide Web", *The Information Society*, 16(3), 201-215.
- [12] Dellarocas, C. (2003). "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms", *Management Science*, 49(10), 1407-1424.
- [13] Devitt, A., Reiff, M.J., and Bawarshi, A. (2004). *Scenes of Writing: Strategies for Composing with Genres*. New York: Longman.
- [14] Dillon, A. and Gushrowski, B. (2000). "Genres and the Web: Is the Personal Home Page the First Uniquely Digital Genre?", *Journal of the American Society for Information Science*, 51(2), 202-205.
- [15] Duhan, D.F. et al. (1997). "Influences on Consumer Use of Word-of-Mouth Recommendation Sources", *Journal of the Academy of Marketing Science*, 25(4), 283-295.
- [16] Evans, M. et al. (2001). "Consumer Interaction in the Virtual Era: Some Qualitative Insights", *Qualitative Market Research*, 4(3), 150-159.
- [17] Firth, D. and Lawrence, C. (2003). "Genre Analysis in Information Systems Research", *Journal of Information Technology Theory and Application*, 5(3), 63-77.

- [18] Gelb, B.D. and Sundaram, S. (2002). "Adapting to Word of Mouse", *Business Horizons*, 45(4), 21-25.
- [19] *General Inquirer*, <http://www.wjh.harvard.edu/~inquirer/>
- [20] Gilly, M.C. et al. (1998). "A Dyadic Study of Interpersonal Information Search", *Academy of Marketing Science*, 26(2), 83-100.
- [21] Harris, K. and Baron, S. (2004). "Consumer-to-Consumer Conversations in Service Settings", *Journal of Service Research*, 6(3), 287-303.
- [22] Haywood, K.M. (1989). "Managing Word of Mouth Communications", *The Journal of Services Marketing*, 3(2), 55-67.
- [23] Hennig-Thurau, T. et al. (2004). "Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?", *Journal of Interactive Marketing*, 18(1), 39-52.
- [24] Herr, P.M., Kardes, F.R., and Kim, J. (1991). "Effects of Word-of-Mouth and Product-Attribute Information on Persuasion", *Journal of Consumer Research*, 17(4), 454-462.
- [25] Lampert, S.I. and Rosenberg, L.J. (1975). "Word of Mouth Activity as Information Search: A Reappraisal", *Academy of Marketing Science*, 3(4), 337-354.
- [26] Lebart, L., Salem, A. and Berry, L. (1998). *Exploring Textual Data*. Dordrecht: Kluwer Academic Publishers.
- [27] Miller, C.R. (1984). "Genre as Social Action", *Quarterly Journal of Speech*, 70, 151-167.
- [28] Nah, F. et al. (2002). "Knowledge Management Mechanisms in E-Commerce: A Study of Online Retailing and Auction Sites", *The Journal of Computer Information Systems*, 42(5), 119-128.
- [29] Pollach, I. (2004). "Online Privacy Statements — Are They Worth Reading?", *Innovations Through Information Technology*. Ed. M. Khosrow-Pour. Hershey: Idea Publishing. 217-220.
- [30] Pollach, I. (2005). "Trust, Quality, and Motivation in Consumer-to-Consumer Interactions on the WWW", *Managing Modern Organizations with Information Technology*. Ed. M. Khosrow-Pour. Hershey: Idea Publishing. 425-428.
- [31] Potter, J. and Wetherell, M. (1994). "Analyzing Discourse", *Analyzing Qualitative Data*. Eds. A. Bryman and R.G. Burgess. London: Routledge. 47-66.
- [32] Price, L.L. and Feick, L.F. (1984). "The Role of Interpersonal Sources in External Search: An Informational Perspective", *Advances in Consumer Research* Vol. 11. Ed. T.C. Kinnear. 250-253.
- [33] Scharl, A., Pollach, I. and Bauer, C. (2003). "Determining the Semantic Orientation of Web-Based Corpora", *Intelligent Data Engineering and Automated Learning*. Lecture Notes in Computer Science Vol. 2690. Eds. J. Liu et al. London: Springer. 840-849.
- [34] Shepherd, M. and Watters, C. (1999). "The Functionality Attribute of Cybergenres", *Proceedings of the 32nd Hawaii International Conference on System Sciences*, Los Alamitos: IEEE Press.
- [35] Stone, P.J. et al. (1966). *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge: MIT Press.
- [36] Swales, J.M. (1990). *Genre Analysis: English in Academic and Research Settings*, Cambridge: Cambridge University Press.
- [37] Toms, E.G. (2000). "Recognizing Digital Genre", *Bulletin of the American Society for Information Science and Technology*, 27(2), 20-22.
- [38] Werry, C.C. (1996). "Linguistic and Interactional Features of Internet Relay Chat", *Computer-Mediated Communication. Linguistic, Social and Cross-Cultural Perspectives*. Ed. S.C. Herring. Amsterdam: John Benjamins. 47-63.
- [39] Yates, J. and Orlikowski, W.J. (1992). "Genres of Organizational Communication: A Structural Approach to Studying Communication and Media", *Academy of Management Review*, 17(2), 299-326.
- [40] Zucchermaglio, C. and Talamo, A. (2003). "The Development of a Virtual Community of Practices Using Electronic Mail and Communicative Genres", *Journal of Business and Technical Communication*, 17(3), 259-284.

Appendix – General Inquirer Dictionaries

Dictionary	Contains words indicating ...
Positive	positive connotations
Negative	negative connotations
PosAff	positive feelings
NegAff	negative feelings
Overstatement	emphasis
Understatement	de-emphasis
Try	actions taken to reach goals
Fail	that goals have not been achieved
If	doubt, uncertainty and vagueness
Know	(un)awareness, (un)importance, (un)certainly
Econ	concepts of business and economics
Evaluation	judgment and evaluation
Quality	qualities or degrees of qualities