Product-Centered Language Style in Chinese Women Online Reviews

Xian Wang1*, Congjun Mu1, Huixian Li1, Alison Noble2 and Qingyi Wang3

1 College of Foreign Languages, Shanghai Maritime University, Shanghai, China, 2 Marine Engineering Department, Antwerp Maritime Academy, Antwerp, Belgium, 3 ByteDance Ltd., Beijing, China

The relationship of language style and online review has drawn increasing academic attention recently as it can provide customers with a guide to make the purchase. Extant research attaches importance to the language style that is presented in the use of function words, instead of product-related content words. This study aimed to examine the language style generated by customers’ comments relating to the product based on content words, that is, product-centered language style (PCLS). We built a corpus of Chinese women clothes online reviews to explore the general picture and distinct features of PCLS and the distinct feature of PCLS. A content-word-centered Linguistic Inquiry and Word Count (LIWC) in terms of product performance is established. PCLS is calculated based on the language style matching (LSM) algorithm. Our results show that the PCLS in women clothes online review is featured by diverse and polarized language styles among three groups of women clothes buyers, and the prioritized arrangement of words of importance contributes to the PCLS. The findings benefit the women clothes industry in which it can help companies quickly find the distinctive and the transition of PCLS and offer an approach for companies to indiscriminately look into the significance of the product category from the linguistic perspective, which can help with product sale strategy and product design.

Keywords: language style, LSM, online reviews, women consumers, product content

INTRODUCTION

Everyday, thousands of women consumers buy clothes based on previous online reviews (Carroll, 2008), and compared with men, their intention to purchase will increase if they receive a positive response (Garbarino and Strahilevitz, 2004). The review valence, review content, extremity and length, and language style are useful information that helps post purchaser make the decision (Forman et al., 2008; Mudambi and Schuff, 2010; Pan and Zhang, 2011). The language style of online product review has been identified as a constructive tool in management prediction, such as the third party’s evaluation of participants (Romero et al., 2015), the attenuated effect of low expertise level comments (Wu et al., 2017), and the way of emotional expressions (Liu et al., 2019). However, these business applications are achieved based on the analysis of emotions inherent in the language style, but none focuses on the specific way of expression when a certain group of customers refers to the product quality. In this study, we attempted to explore the features and influences of language style relating to product quality when Chinese women clothes consumers comment online on Taobao.
The language style of online reviews has been extensively researched in recent two decades either as an information signal (Romero et al., 2015; Liu et al., 2019) or as a genre of discourse (Teso et al., 2018; Cenni and Goethals, 2021). Language style may be a tool to measure the perceived result of consumers against previous consumers (Hong et al., 2017). While Guo et al. (2020) made effort to analyze the cognitive impact based on the non-emotional content, the majority of literature relies solely on the emotional language features to describe cognitive responses to the language style of online reviews among different information receivers (Antioco and Coussement, 2018; Liu et al., 2019, Guo et al., 2020; and etc.). In addition, some scholars state that the language style of online reviews is independent of the content of reviews and put emphasis on the tokenization of emotional linguistic features, that is, the emotion-centered language style (ECLS) (Swaab et al., 2011; Ludwig et al., 2013). In the field of discourse analysis, language style is studied with a focus on the function of an acknowledged discourse instead of the internal features. For example, Ren (2018) argued that there was a mitigated language style of online reviews, and the strategy for mitigation devices is related to the positive/negative review valence. Tseng and Zhang (2020), through the example of Chinese health communication, indirectly pointed out that language style in health online reviews has its own framework.

Despite these thorough studies on the language style of online reviews (OPCs) in business research and discourse analysis, factors identified to affect the language style regarding the product are mixed in the literature. Unclassified factors such as shopping experience, price, emotions, and non-emotions (Ma and Hu, 2012) proved to contribute to language style and product-centered language style (PCLS). Guo et al. (2020) scaled down the factors by separating non-emotional from emotional content but listed a mixture of non-emotion content that contains review valence, gender, credibility, and the scientific processing system of mankind. Product attributes such as length, rating valence, and affective content directly influence the review expressions (Forman et al., 2008), which is a mixed classification as well. From these efforts, it is clear that academia intends to detail the factors inherent in the language style of online reviews that synchronize that specific way of expression, that is, the product-centered language style. However, it is not satisfactory yet as to the question: to what extent are consumers influenced by previous comments purely relating to the product? We presumed this is caused by the following: (1) it is difficult to define the product attributes, as it is related to various aspects of the product, including the performance, material, and design. Recent academic studies relating to language style in the online review are controlled by function words reflecting the emotions (Fan and Chen, 2017; Liu et al., 2018; Guo et al., 2020). (2) Compared with the description of language style from function words, which is limited in a category and in quantity, it is quite challenging to describe the product from the content words. According to Tausczik and Pennebaker (2010), content words include nouns, regular verbs, adjectives, and adverbs. (3) There are no product-centered word categories available for the comparison of language style. To make up the imbalance, this study establishes a product-centered content-word-based Linguistic Inquiry and Word Count (LIWC), a dictionary based on categories, and explores the features of the language style of women clothes, purely on product aspects (PCLS), drawing on the corpus of Chinese women clothes online reviews. In particular, we addressed the following questions:

1. Is PCLS distinguishable among women clothes buyers?
2. What linguistic features make up the PCLS variations in women online reviews?

In what follows, we first reviewed the theory about the relationship of online review writers and post readers, the studies on PCLS. In the “Methods” section, we provided an overview of the approach called language style matching (LSM) and the new LIWC based on PCLS. In the “Results” section, we demonstrated the data and relevant explanations. In the “Discussion” section, we expound on the general features of PCLS of women reviews. Finally, we concluded that the PCLS is important to monitor and diagnose the value structure of companies’ products.

**LITERATURE REVIEW**

**Language Expectancy Theory**

The language used by review writers needs to consider the context of the relationship between the speaker and the audience (Pennebaker et al., 2003), as prior utterance affects the cognition of the recipients and may result in the synchronicity of language style, which is the linguistic scenario called the language expectancy theory (LET). LET posits that people develop appropriate communication styles in certain scenarios whereby they encode the language style of prior utterance, if affected, into their own communication styles (Burgoon and Miller, 1985).

The current academia in this regard focuses on comparing the language style matching between the speaker and the recipients in the context of different LET-related situations. Purposefully or unintentionally changing expectations will enhance or mitigate the effectiveness of the message whereby a given message will be incrementally enhanced, sharpening the recipients’ acceptance and their way of expression (Lee and Yu, 2020). Sentence structure or word choice positively or negatively influences the expectations of recipients (Averbeck and Miller, 2014). Linguistic expectations are compared to find the proper expression for crowdfunding (Parhankangas and Renko, 2017; Koh et al., 2020). In a natural disaster, the resharing of the information on Twitter plays an incremental role in self-strengthening and self-awareness (Lee and Yu, 2020). The stylistic choices made by an individual customer are intended to provoke a particular response from other customers (Semino and Culpeper, 2011). Furthermore, words are the key aspect to show the effect of language expectations, which results in the research on semantic features. For example, language density used by patients is compared to show that higher language density means the impact of certain topics on customers is more effective (Burgoon and Miller, 1985; Burgoon et al., 2002). The word diversity relating to the language complexity can demonstrate the level of information, the higher the word diversity is, the knowledge is broader (Jensen et al., 2013;
Averbeck and Miller, 2014). In communication, people often talk as appropriately as possible to achieve their expectations (Burgoon and Miller, 1985). In terms of business online review, it is important to identify and simulate that particular way of expression to ensure conformity within that group. In general, these studies are conducted on the premise that previous utterance has contextualized the thoughts and expressions of the recipients who are faced with the same scenario; however, they do not touch on the similarities and differences of adjacent recipients that may encounter different types of a product. In this research, we used LET to explain how previous expressions relating to product quality influence the expressions of the recipients.

**Product-Centered Language Style**

The content of the online review is able to manifest the review value (Willemsen et al., 2011) because in e-commerce, customers rely on limited knowledge and information they have to make purchase decisions (Steenkamp and Ter Hofstede, 2002). The attributes of the content such as length, rating valence, and affective content directly influence the review expressions (Forman et al., 2008). These review content features also influence the high and low quality of reviews in the eyes of customers (Huang et al., 2015; Liu et al., 2018). Besides, online review content is a source for customers to get various knowledge about products. Through browsing the content generated by previous writers, customers have access to social knowledge about the products they are interested in (Huang and Benyoucef, 2013).

The language style inherent in the linguistic information of online reviews offers cues for customers to assess the product (Willemsen et al., 2011). The signal and observability offered by language style are one of the components to judge the quality. Linguistic variables such as word count, argument quality, review clarity, and readability are sometimes central indicators of product quality (Maslowska et al., 2017). According to Srivastava and Kalro (2019), central linguistic factors primarily consist of “manifest” attributes such as word count, number of sentences, review rating, and number of images, which are easily observable; “latent” factors are embedded linguistic text features, in particular language style.

Recent research still focuses on ECLS from diverse business-related factors. For example, Yin et al. (2017) found that emotional arousal from a similar language style positively affects the value of the product through the price valence. The direct language contains mental energy, feelings, and motivations and presents intense emotions in post-buyers, considering the relationship of the matching level of function words (Precht, 2000, 2003). Information and emotional process can be used to manipulate by the statement that “the language style of online reviews affects how an audience perceives the reviews, independent of the content of reviews” (Ghahttarani et al., 2020). Based on these studies, some scholars have begun focusing on the relationship between products and online reviews. Product information regarding product quality is more important than emotions, for it reduces customers’ uncertainty (Pavlou and Dimoka, 2006). Useful information is obtained by customers from the non-emotional content, especially the product quality (Jiang and Benbasat, 2007; Liu et al., 2019). Customers usually add their own information to the text model to create a new comment that is in line with the previous reviews (Gernsbacher and Hargreaves, 1988; Zwaan and Radavansky, 1998). In this way, customers develop a relational tie with the cognition of the previous text in terms of the product (Wang and Karimi, 2019). The way of expression reflecting the product quality is one of the indicators that help customers make the judgment of the product quality (Guo et al., 2020). But among these studies, no one explores the features of the language style of the online review in terms of the product itself. In this study, we contributed to the literature that the PCLS is under investigation in the context of Chinese women clothes online reviews.

**Studies Related to Language Style Matching**

Language style matching is a technique of text analysis that measures the matching style degree of two or several people in their utterances (Ireland et al., 2011). People who have the same interest or background are likely to make utterances that mirror the sentence structure they have previously heard or read, and the verbal and non-verbal behaviors after a social interaction becomes synchronized (Ireland and Pennebaker, 2010). Such phenomenon is called LSM, which shows the matching level of identical behaviors between one individual and another, either the verbal or the written style (Gonzales et al., 2010).

Language style matching has been extensively used as an approach in various business discourse texts to study the impact and characteristics of consumer comments (Ireland et al., 2011). Using this linguistic method, it is found that figurative language of consumer reviews leads to favorable attitudes to an enjoyable purchase experience (Kronrod and Danziger, 2013); post-reviewers will note down figurative language instead of literal language to show their persuasive influence (Wu et al., 2017); the psychological processes can be identified to examine qualitative features of review text (Liu et al., 2018); mitigated perceived expectation/expectancy on certain products is congruent with the negative text acquired by post-reviewers (Burgoon and Poire, 2010; Liu et al., 2018). In recent research, LSM gives rise to the function words used in an online review, by which emotions’ roles are thoroughly studied (Fan and Chen, 2017; Liu et al., 2018; Guo et al., 2020). For example, the sentiment in online consumer reviews negatively or positively affects the degree of synchrony of expressed texts made by recipients (Salehan and Kim, 2016; Yin et al., 2017). The engagement in email references of the librarian’s answers is measured to indicate the holistic assessment of virtual reference services (Agee, 2019). In addition to function words, the impact of topic and focus of interest on post-customers has drawn wide attention with the approach of LSM. Lin and He (2009) proposed a joint sentiment/topic model detecting how simultaneous sentiment is likely to reflect the change of topic theme, the socioeconomic characteristics. The attention focus plays a key role in families, communities, and at work when people happen to talk about a common theme (Chung and Pennebaker, 2007; Ireland et al., 2011). In this study, applying...
the LSM method in women clothes online reviews in the context of PCLS, we attempted to explore the linguistic characteristics of product-centered texts, that is, what semantic expression makes one group’s reviews different from those of another group.

Language style matching is an algorithm developed mainly to calculate verbal cohesiveness based on an automated textual analysis of words, which can predict the cohesiveness of the language used by various groups (Tausczik and Pennebaker, 2010). Language features including word count, sentence patterns, and verb tense were also related to the groups' cohesiveness and performance (Gonzales et al., 2010). According to Gonzales et al. (2010), the LSM algorithm is calculated in the following procedures. First, count the absolute value of the proportions of words in a given word category out of the total number of words in each group. Second, divide this value by the sum of each group, where the difference in word usage adjusted for combined word usage in that category between group 1 and group 2 is provided. Then, the result is subtracted by 1 to obtain the similarity.

In this study, we measured PCLS based on the LIWC for product content in the online reviews. Therefore, we employed a refined LSM calculation based on the formula created by Gonzales et al. (2010).

\[
\text{LSM affordable} = 1 - \frac{|(G1 - G2)|}{(G1 + G2 + 0.000001)}.
\]

In the above formula, “G1” and “G2” stand for the written texts in the word category of “affordable” by each group of customers. The denominator of 0.000001 is to avoid empty sets in cases of a zero LIWC category’s score. G1 is obtained by calculating the percentages of total content-and-sentiment words in the word category among the total number of words. The result is an 8-dimensional language-style score for each review.

METHODS

Overview of the Research Method
In this study, we proposed an LIWC for content words with the LSM approach in an attempt to demonstrate how the PCLS of a given group of Chinese women clothes online reviews differs from that of the other two groups. The overview of the method is seen in Figure 1. It is composed of two parts. The first part is to build a novel LIWC dictionary for content words, ready for the calculation of LSM. The second part is to calculate the language style of three groups of Chinese women clothes customers based on the LSM algorithm, and the results are used to compare general features of PCLS, that is, whether convergence and divergence of language style exist in the online review in terms of age group. To be specific, there are four procedures: (1) new content words representing the linguistic content words (product-centered words) are established to adapt to the online women clothes reviews; (2) the product-centered words are based on words that reflect the product categories most talked about in women online reviews; (3) the LSMs of three groups of women customer reviews are calculated; (4) the results of LSMs are compared with each other to show the level of cohesiveness and likelihood. We gave special consideration to the PCLS in terms of LSMs because customers believe that product quality can bring them a more useful assessment of the product than the sentiment does (Pavlou and Dimoka, 2006).

Reviews of Chinese Women Clothes
To make the samples as wide as possible with regards to Chinese women online reviews, we established a corpus of 32,667 online reviews written by women clothes buyers, from Taobao, an online e-commerce platform of Alibaba in China, spanning from 2018 to 2020. Information relating to product quality includes brand, manufacturer, size, color, texture, etc. More importantly, the online reviews vary with a host of factors such as age, community, and personal hobbies. Among these factors, age is one of the most influencing factors in women clothes purchases (Mirowsky and Ross, 2001). Our data are extracted based on the age group as set out by Taobao. The Taobao division of the age group is set by its own system of division, which is not in line with the division standards of the prevailing standards, for example, the standards of the World Health Organization (2010). Therefore, we followed the typical Taobao division of demography by age and classify them into three age groups: 18–25, 25–45, and 45–65, respectively, representing young, middle, and seniors. In each age group, the online reviews are expressed under considerable and enormous brands of women clothes. To address the impact of brands on online reviews, we selected those brands’ reviews by sales volume not lower than the top 10 brands, which are listed on Taobao’s web page. Then, the reviews are randomly taken according to brands as shown in Table 1.

Content-Word Linguistic Inquiry and Word Count
According to Pennebaker (2015), the LIWC is one of the core parts of LSM. LIWC is used to count and classify words. With the LIWC program, each word can be encoded and compared to see if it belongs to one of the words classified under a certain category in the dictionary. For example, “linen” is coded as a content word of material category. Then, “linen” is compared to see if “linen” is in the dictionary. LIWC produces the percentage of each LIWC category and usage frequency.
TABLE 1 | Data collection source.

| Group No. | No. of reviews | Age span | Top brand(s) by volume | Age group based on Taobao division |
|-----------|----------------|----------|------------------------|------------------------------------|
| G1        | 10,553         | 18–25    | Peacebird, Tyakasha, Handu Group, La Chapelle, etc. | Young                              |
| G2        | 11,096         | 25–45    | Ochirly, Prich, Hopeshow, Uniqlo, etc. | Middle                             |
| G3        | 11,018         | 45–65    | Cadence, Ms. Tang, Cabenzi, Sommeliy, etc. | Seniors                            |

The updated LIWC program (Pennebaker, 2015) is a 90-plus category dictionary composed of almost 6,400 words, word stems, and emotions which additionally include one or more word categories or sub-dictionaries. Most prominent in the dictionary is “21 standard linguistic dimensions (e.g., percentage of words in the text that are pronouns, articles, and auxiliary verbs) and 41-word categories tapping psychological constructs (e.g., affect, cognition, biological processes, and drives)” (Pennebaker, 2015, p. 4).

The amazing part of LIWC is that it expands into aspects of words and gets updated. Psychology and children’s language LIWC (Hartley et al., 2003) have been included. Recently, the causal words and insight words under the category of cognition are included as well to show the level of appraisal (Kross and Ayduk, 2008). As for the research on an online review, current LSM is based on the emotional aspect, in particular the function words of the emotional aspect in the LIWC (Leonczyc, and Aleksandra, 2016; Fan and Chen, 2017). Although some scholars have argued the social influence on consumers from the perspective of economic status, gender, etc., the reliability of language style is dependent on the data from emotional functional words (Willemsen et al., 2011; Cheng and Ho, 2015). In brief, there is no LIWC categorized by words in regard to product quality at present.

According to Tausczik and Pennebaker (2010), content words and function words play an equal role in the evaluation of language style. Style or function words are those words such as pronouns, prepositions, articles, conjunctions, and adverbs. Content words include nouns, regular verbs, adjectives, and adverbs, which convey the content of communication (Haspelmath, 2001). In this study, based on the features of our data and business practice, we developed a word dictionary of content words (Table 2) by a combination of relevant studies (Yu, 2012; Zhang, 2019). Our dictionary is an 8-category product content dictionary, which is affordability, brand, color, clothes function, materials, size, garment style, and workmanship, respectively. For example, words such as “slim,” “warm,” and “young” are grouped into the category of clothes function while words such as “linen,” “silk,” and “texture” are classified into the material category.

**RESULTS**

**General Results**

Based on our corpus of 32,667 online reviews in terms of three age groups, we acquired Table 3, which shows the LIWC score, LSM scores, and LSM mean.

To test the validity of these data, we conducted a standard deviation (SD) for 8-category words of three groups. The SD ranges from 0.11 to 0.5, which means that the data collected are acceptable (refer to Figure 2).

**The Diverse and Polarized Product-Centered Language Style**

The disparity of each group in terms of the specific categories of the product indicates that the PCLS of women clothes buyers is diverse in general. From Table 3, the LIWC within the young women (G1) shows three high figures, namely, function, style, and color, at 5.51, 4.41, and 2.57, respectively, while other aspects stay at a stunning figure lower than 1, with price at 0. This overwhelming way of expression with regards to a product indicates that young women clothes buyers, though cognitively influenced by previous customers, are likely to expressively comment on the function and style of a product than on price. Likewise, we can get the linguistic features of middle-aged women (G2) and senior women (G3). The PCLS of middle-aged women is evenly distributed with three high figures, namely, function (10.03), style (6.99), and workmanship (3.34), which manifests that middle-aged women have a more equilibrium view on clothes and tend to pose a higher demand for quality (workmanship) than young women (G1). As for G3, the PCLS is more evenly distributed with four high figures, that is, function (9.41), material (6.97), size (5.57), and style (3.14), which shows that senior women buyers hold a comprehensive look at clothes with a special concern for the size and style.

To further show the general trend of PCLS, we conducted the linear regression as shown in Figure 3. Figure 3 shows the linear regression of the correlation of three LSMS. The X-axis stands for the three age groups, while Y-axis means the correlation coefficient. It is seen that the PCLS presents a downward trend in terms of three groups of women online buyers, which means that the PCLS diverges or drops in matching in terms of age groups.
TABLE 3 | Sample LSM for online review of female garments.

| LIWC category | LIWC score | LSM scores | Mean LSM |
|---------------|------------|------------|----------|
|               | G1         | G2         | G3       | G1/G2   | G1/G3   | G2/G3   |         |
| Affordable    | 0.00       | 1.82       | 1.39     | 0.00    | 0.00    | 0.87    | 0.29    |
| Brand         | 0.74       | 1.22       | 2.79     | 0.75    | 0.42    | 0.61    | 0.59    |
| Color         | 2.57       | 1.22       | 1.05     | 0.64    | 0.58    | 0.92    | 0.71    |
| Functional    | 5.51       | 10.03      | 9.41     | 0.71    | 0.74    | 0.97    | 0.81    |
| Material      | 0.37       | 2.43       | 6.97     | 0.26    | 0.10    | 0.52    | 0.29    |
| Size          | 0.37       | 1.22       | 5.57     | 0.46    | 0.12    | 0.36    | 0.32    |
| Style         | 4.41       | 6.99       | 3.14     | 0.77    | 0.83    | 0.62    | 0.74    |
| Workmanship   | 0.74       | 3.34       | 1.74     | 0.36    | 0.59    | 0.89    | 0.55    |

The Priority of Product-Centered Language Style

In this study, we posited that the arrangement of product categories is an indicator of the language style a customer employs to demonstrate his or her identity, and in other words, the priority of texts/words relating to product categories affects the way of expression.

To acquire the PCLS of women clothes buyers in terms of priority and importance, we applied the results of LSM to the statistical analysis function of percent rank and get the importance list of the eight categories of women clothes reviews (Table 4).

Table 4 shows that the overall importance of the eight categories is rated based on the results of LSMs, with a breakdown of percentage from 100 to 0%. One way to interpret Table 4 is that by the same percentage of importance, three groups show a different priority/sequence of product categories. For example, the priority of G1/G2 is clothes style, brand, clothes function, color, size, workmanship, material, and affordability, which are different from that of G1/G3, that is, clothes style, clothes function, workmanship, color, brand, size, material, and affordability. If the significance is set as 50% manually to show the high (>$50%) and low attention focus (<50%), we can see that color and function are the high “categories” shared by the three groups. The second way to interpret Table 4 is to look up the percentage of a certain product category, for example, “material.” Fortunately, the category of “material” is the only one that marks the same ratio at 14.2%, signifying the “only” shared interest of focus among the three groups. As the percentage of product categories is indicative of the focus of interest from customers, we can see that the different percentage by significance shows a clear-cut priority with regards to the way of expression of women clothes reviewers.

DISCUSSION

The first question of this study is to describe the key language features of women clothes online reviews when they comment on issues relating to a product. As shown in the previous section, diverse and polarized are the key features. We attributed the diversity of PCLS to the phenomenon called the insider language style of a speech community. Fairclough, 2003 defined insider languages as the marked language that serves to construe and reinforce the identity of a speech community. In online reviews, customers usually add their own information to the text model, which actually creates a new comment in line with the previous reviews (Gernsbacher and Hargreaves, 1988; Zwaan and Radvansky, 1998). Insider customers are sensitive to the language style of their community and follow the text rules that demonstrate such factors as the density of words and the text of social cognition (Cheng and Ho, 2015) or social variables such as gender, religion, and social position (Deckert and Vickers, 2011). To be specific, the young women clothes buyers (G1) identify themselves as the group of people who cares least about money,
TABLE 4 | Product-centered language style (PCLS) priority rankings between groups.

| Rank | Significance (%) | Category | G1/G2 LSM | Category | G1/G3 LSM | Category | G2/G3 LSM |
|------|------------------|----------|-----------|----------|-----------|----------|-----------|
| 1    | 100.00           | Clothes style | 0.77     | Clothes style | 0.83 | Clothes function | 0.97 |
| 2    | 85.70            | Brand     | 0.75     | Clothes function | 0.74 | Color     | 0.92 |
| 3    | 71.40            | Clothes function | 0.71 | Workmanship | 0.59 | Affordable | 0.87 |
| 4    | 57.10            | Color     | 0.64     | Color     | 0.58 | Workmanship | 0.69 |
| 5    | 42.80            | Size      | 0.46     | Brand     | 0.42 | Clothes style | 0.62 |
| 6    | 28.50            | Workmanship | 0.36 | Size      | 0.12 | Brand     | 0.61 |
| 7    | 14.20            | Material  | 0.26     | Material  | 0.1  | Material  | 0.52 |
| 8    | 0.00             | Affordable | 0        | Affordable | 0    | Size      | 0.36 |

which is reinforced within the community of G1 and eventually forms a special language style.

In addition to diversity, we found that women reviews show a polarized trend by division of the group. We calculated the average mean of three LSM (0.49, 0.42, and 0.70) and then apply the correlation function. The results are 0.78, 0.04, and 0.37. By the definition of the correlation function, a higher correlation value means higher relatedness of the two variables. The G1/G3 correlation (0.04) shows that G1 and G3 have the least matching in language style, which can well explain the group disparity between young and seniors, whereas the G1/G2 correlation (0.78) shows relatively high synchrony in language style, which represents the smooth transition from young- to middle-aged woman. This result is of significance to clothes manufacturers in that products designed for young women can be branched out to middle-aged women in business promotion strategy, which may greatly increase the demography coverage of target customers.

This polarized feature can also be explained by the high and low language style matching in women clothes reviews with regards to product quality. According to Tausczik and Pennebaker (2010), the LSM score is on a scale of 0–1.0. A score of zero indicates no matching in the language style while 1.0 means high cohesiveness. Scores of 0.60 or below are usually considered low matching, and scores of 0.85 or above are deemed as signs of high matching. High and low matching levels have been proved valid in different fields. For example, high-level matching exists among customers of different nations who speak different languages when they comment on the product (Schuckert et al., 2015). Librarians can linguistically generate high-level matching feedback to the comments of the audience (Agee, 2019). The desire to enhance or mitigate the impact of language style may influence recipients who eventually produce a high or low synchronized way of expression (Giles et al., 2007). Hilti et al. (2020) proved that a certain degree of high LSM exists in two genders in teenagers’ social media writing. Table 3 shows that the only category of high matching level (/>0.85) appears in G2/G3, which are clothes function, color, and price (0.97, 0.92, 0.87) while there is none in G1/G2 LSM and G1/G3 LSM. This points out that G2/G3 online reviews influence each other higher and stronger in the categories (garment function, color, and price) than other product categories, which, in the eyes of a company, may locate the “shared” aspects of a product among different customer groups.

We assumed that the declining trend of language style matching degree in Figure 3 is largely attributed to the weakened expectancy of information or the stronger role of insider language. As language expectancy theory puts it, if people sense the style of community speech and get affected, they may translate the language style of previous comments into their own language style (Burgoon and Miller, 1985). But this LET is conditioned by the ad hoc fact that customers are able to identify the language style. In case of failure to the identification of prior comments, the chances of encoding are slim. Since G1 is unlikely to identify the encoded language style of G3, which may cause the low initiation of language expectancy by a host of factors such as leadership, honesty, status, age, and culture (Chung and Pennebaker, 2007), G1 may not mimic the text model of G3 and eventually comment very differently.

For the second question on the makeup of PCLS, we referred to the concept of word of significance. Groves et al. (2000) stated that the way a potential participant responds to a product depends on the importance placed on an element of the product. It is inferred that if the importance is placed on a host of linguistic elements with different degrees of priority, and such importance is identified by the post-customer and rearranged into his online review with extra information, the importance placed on a host of elements is, in fact, the linguistic salience of prioritized product features by importance. Since people who think more about product quality will repeat the elements of product quality and be controlled in their text expression (Tausczik and Pennebaker, 2010), customers can identify and track what information they are interested in which is linked to product quality.

The fact that customers use priority PCLS to gain useful information is related to psychological linguistics. The depth of thinking or attention is reflected in the language style, which links thoughts (Tausczik and Pennebaker, 2010). In other words, the language style is reflected in thinking that correlates to how frequently a word or a group of words are mentioned. In essence, language style is closely related to attention focus, which is one of the key factors that contribute to the construction of language style (Pennebaker, 2011). The self-reference of pronouns “I” in the political advertisement is a representation of the idea of being independent (Gunsch et al., 2000). The use of “we” is linked with “better” marriage (Simmons et al., 2005). The style differences of the verb tense are the result of the speaker’s attitude toward reality (Pasupathi, 2007). These studies attach importance to using tokenized words to differ the language style, but none
focuses on the arrangement of the tokenized words, let alone the PCLS priority.

It is interesting to see from Table 4 that the ranking/sequence of G2/G3 against that of G1/G2 is 2 against 4, which represents the ratio of the language matching degree of middle-aged women/senior women vs. that of young women/middle-aged women. The higher priority of “color” in terms of middle-aged women/senior women implies that people of older ages like to dress in brighter and glossier clothes. We reasoned this to be the combination of individual focus and situational focus. The language style of a certain group is featured with the two characteristics as follows: the uniqueness of the language style of a person and the shadowed influence from other groups (Krapp, 2002; Hidi and Boscolo, 2006). Individual focus results from an individual’s enthusiasm and tendency toward specific information and objects. Situational focus is activated by the surrounding environment such as the emotions of others. In the case of women of older ages, they are influenced by the contextual situation of the middle aged group (G2) in the choice of “color.” Also, the fact that senior women mention more frequently the word “color” is attributed to the individual personality because the self-focus inherent in language use that “differs with age, sex, personality, and mental health” (Tausczik and Pennebaker, 2010. p. 36) can be identified from their linguistic characteristics.

CONCLUSION

Recently, scholars studied the language style used in product online reviews (Huang et al., 2015; Liu et al., 2018), solely based on the function words, which demonstrates that the emotional language style embedded in function words is related to the product. However, a large part of online reviews is concerned about product quality instead of emotions. In this study, we built a corpus of Chinese women clothes online reviews to explore the PCLS. A novel LIWC dictionary covering eight categories of product quality is sorted and established, which, in the practical sense, adds a new dimension to the existing LIWC dictionary.

With the product-category LIWC and the approach of LSM, we are able to find that the PCLS of women clothes buyers lies in (1) the diverse but polarized way of expression among different groups of buyers. The correlation function of all LSMs shows that young and seniors have the least matching in language style while middle age and seniors show a high similarity in PCLS; (2) despite the polarized language style, there exists a priority of language style relating to a certain product category, and such priority of PCLS is influenced by how customers weigh product categories in their heart. In this regard, PCLS provides a new dimension to study the language style of online reviews in the context of the product.

As pointed out, the language style of online review is to boost more effective communication. In this regard, the language style of online review may be more effective to convey customer expressions if it is used with the body signal (Poggi et al., 2013) and facial expressions (Heisler et al., 2003). This stance on effective signals may be taken into account for the future study regarding the PCLS.

This research contains practical benefits not only for women clothes industry but also for consumer management dealing with online reviews. Consumers, regardless of men or women, use language style in online consumer reviews to classify reviews and show their identity (Schindler and Bickart, 2012). The PCLS of women clothes online reviews, featuring diversity and polarity among different groups, can help companies quickly find the polarized PCLS and the transition of PCLS, which can be an aid to the product promotion strategy and product line design. Moreover, the priority PCLS offers a platform for companies to indiscriminately look into the importance of the product category from the linguistic perspective, where feedback (online reviews) of each product category can be quantitatively weighed by the same significance. In general, this quantitative PCLS can help companies identify those crucial product categories and improve product design and promotion strategy.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

XW and CM are the principal authors of this article, contributed equally to the conception of the article and to the review and critical analysis of relevant scientific literature, wrote a first version of the manuscript, arranged new sections, and produced a final version of the manuscript incorporating changes in response to the helpful suggestions of the reviewers. HL contributed to the conception, to the review and critical analysis of scientific literature on the subject, collecting data and building the corpus and she edited the figures, tables, and references in each version of manuscript. AN contributed to review the different versions of the manuscript and to polish the language. QW applied the mathematical method of percent rank function and interpreted the conceptualize of the manuscript. All authors contributed equally to the conception of the article and to the review and critical analysis of relevant scientific literature, wrote a first version of the manuscript, arranged new sections, and produced a final version of the manuscript incorporating changes in response to the helpful suggestions of the reviewers. HL contributed to the conception, to the review and critical analysis of scientific literature on the subject, collecting data and building the corpus and she edited the figures, tables, and references in each version of manuscript. AN contributed to review the different versions of the manuscript and to polish the language. QW applied the mathematical method of percent rank function and interpreted the conceptualize of the manuscript. All authors contributed equally to the conception, to the review and critical analysis of scientific literature, wrote a first version of the manuscript, arranged new sections, and produced a final version of the manuscript incorporating changes in response to the helpful suggestions of the reviewers.

FUNDING

This study was supported by the National Social Science Fund of China (Grant Nos. 18CYY0231 and 17BYY103).

REFERENCES

Agee, A. (2019). Language style matching as a measure of librarian/patron engagement in email reference transactions. J. Acad. Librarr. 45:102069. doi: 10.1016/j.acalib.2019.102069

Averbeck, J. M., and Miller, C. (2014). Expanding language expectancy theory: the suasory effects of lexical complexity and syntactic complexity on effective

Antico, M., and Coussement, K. (2018). Misreading of consumer dissatisfaction in online product reviews: writing style as a cause for bias. Int. J. Inform. Manage. 38, 301–310. doi: 10.1016/j.ijinfomgt.2017.10.009

Heisler, M., Heisler, M., and Heisler, M. (2003). Expanding language expectancy theory: the suasory effects of lexical complexity and syntactic complexity on effective
message design. Commun. Stud. 65, 72–95. doi: 10.1080/10510974.2013.775955
Burgoon, J. K., and Poire, B. L. (2010). Effects of communication expectancies, actual communication, and expectancy disconfirmation on evaluations of communicators and their communication behavior. Hum. Commun. Rev. 20, 67–96. doi: 10.1111/j.1460-9669.2009.00316.x
Burgoon, M., and Miller, G. R. (1985). An Expectancy Interpretation of Language and Persuasion. London: Routledge.
Burgoon, M., Denning, V. P., and Roberts, L. (2002). Language expectancy theory. Persuasion Handbook Dev. Theory Pract. 2002, 117–136. doi: 10.4135/9781412976046.n7
Carroll, S. (2008). The practical politics of step-stealing and textual poaching: youtube, audio-visual media and contemporary swing dancers online. Convergence 14, 183–204. doi: 10.1177/1354856507087943
Cenni, I., and Goethals, P. (2021). Business responses to positive reviews online: face-work on tripadvisor. J. Prognost. 180, 38–50. doi: 10.1016/j.progma.2021.04.008
Cheng, Y. H., and Ho, H. Y. (2015). Social influence’s impact on reader perceptions of online reviews. J. Bus. Res. 68, 883–887. doi: 10.1016/j.jbusres.2014.11.046
Chung, C., and Pennebaker, J. W. (2007). The psychological functions of function words. Soc. Commun. 2007, 343–359.
Deckert, S. K., and Vickers, C. H. (2011). An Introduction to Sociolinguistics: Society and Identity. London: A&C Black.
Fan, K. C., and Chen, Y. L. (2017). ”). Exploratory study on corporate brand image and customer satisfaction on consumer purchase behavior: a case study of UNIQLO”. In Proceedings of the 2017 International Conference on Applied System Innovation (ICASI), (IEEE).
Fairclough, N. (2003). Analysing Discourse: Textual Analysis for Social Research. Hove: Psychology Press.
Forman, C., Ghose, A., and Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets. Inform. Soc. 19, 291–313. doi: 10.1177/09625443080926893
Garbarino, E., and Strahilevitz, M. (2004). Gender differences in the perceived risk communicators and their communication behavior. Hum. Commun. Res.
Gusch, M. A., Brownlow, S., Haynes, S. E., and Mabe, Z. (2000). Differential forms of accommodation in teenagers’ social media writing: convergence patterns in mixed-gender conversations. J. Quant. Lang. 2020:1807853. doi: 10.1080/09296174.2020.1807853
Hong, H., Xu, D., Wang, G. A., and Fan, W. (2017). Understanding the determinants of online review helpfulness: a meta-analytic investigation. Dec. Support Syst. 102, 1–11. doi: 10.1016/j.dss.2017.06.007
Huang, A. H., Chen, K., Yen, D. C., and Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. Comput. Hum. Behav. 48, 17–27. doi: 10.1016/j.chb.2015.01.010
Huang, Z., and Benyousef, M. (2013). From E-commerce to social commerce: a close look at design features. Electr. Commerce Res. Appl. 12, 246–259. doi: 10.1016/elera.2012.12.003
Ireland, M. E., and Pennebaker, J. W. (2010). Language style matching in writing: synchrony in essays, correspondence, and poetry. J. Personali. Soc. Psychol. 99, 549–571. doi: 10.1037/a0020386
Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., and Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. Psychol. Sci. 22, 39–44. doi: 10.1177/0956797610392928
Jensen, M. L., Averbeck, J. M., Zhang, Z., and Wright, K. B. (2013). Credibility of anonymous online product reviews: a language expectancy perspective. J. Manage. Inform. Syst. 30, 293–324. doi: 10.2753/MISO742-122300109
Jiang, Z., and Benbasat, I. (2007). The effects of presentation formats and task complexity on online consumers’ product understanding. Mis Q. 2007, 473–500. doi: 10.2307/25148804
Koh, Y., Lee, M., Kim, J., and Yang, Y. (2020). Successful restaurant crowdfunding: the role of linguistic style. Int. J. Contemp. Hospit. Manag. 32, 3051–3066. doi: 10.1108/IJCHM-02-2020-0159
Krapp, A. (2002). Structural and dynamic aspects of interest development: theoretical considerations from an ontogenetic perspective. Learn. Instr. 12, 383–409. doi: 10.1080/0959475720100011-1
Kronrod, A., and Danziger, S. (2013). Wiwi will rock you!” the use and effort of figurative language in consumer reviews of hedonic and utilitarian consumption. J. Consumer Res. 40, 726–739. doi: 10.1086/671998
Kross, E., and Ayduk, O. (2008). Facilitating adaptive emotional analysis: distinguishing distanced-analysis of depressive experiences from immersed-analysis and distraction. Personali. Soc. Psychol. Bull. 34, 924–938. doi: 10.1177/0146167208315938
Lee, C. H., and Yu, H. (2020). The impact of language on retweeting during acute natural disasters: uncertainty reduction and language expectancy perspectives. Industr. Manage. Data Syst. 120, 1501–1519. doi: 10.1108/IMDS-12-2019-0711
Leonczyk, and Aleksandra. (2016). The role of language proﬁciency in the perception of I2 taboo words by late bilingual speakers. Eur. Yearbook 16, 25–58. doi: 10.1057/eurosla.16.022eo
Lin, C., and He, Y. (2009). “Joint sentiment/topic model for sentiment analysis,” in Proceeding of The 18th ACM Conference on Information and Knowledge Management 09, 355–360.
Liu, A. X., Xie, Y., and Zhang, J. (2019). It’s not just what you say, but how you say it: the effect of language style matching on perceived quality of consumer reviews. J. Mem. Lang. 72, 699–717. doi: 10.1016/j.jmemlang.2017.09.003
Liu, X., and Hu, Z. M. (2012). Online review of research in 2006-2011: thematic analysis and distraction. Industr. Manage. Data Syst. 122, 924–938. doi: 10.1108/IMDS-12-2019-0711
Liu, Z., and Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. Commun. Res. 37, 191–199. doi: 10.1177/0010413509343021
Gonzales, A. L., Hancock, J. T., and Pennebaker, J. W. (2010). Language style matching in predicting a social dynamics in small groups. Commun. Res. 37, 3–19. doi: 10.1177/0010413509343418
Groves, R. M., Singer, E., and Corning, A. (2000). Leverage-salience theory of survey participation: description and an illustration. Public Opin. Quart. 64, 299–308. doi: 10.12688/gatesopenres.12753.1
Gunsch, M. A., Brownlow, S., Haynes, S. E., and Mabe, Z. (2000). Differential forms of linguistic content of various of political advertising. J. Broad. Electr. Med. 28, 56–75. doi: 10.1177/1073137X99021002
Guo, J., Wang, X., and Wu, Y. (2020). Positive emotion bias: role of emotional content from online customer reviews in purchase decisions. J. Retaili. Consumer Serv. 52:101891. doi: 10.1016/j.jretconser.2019.10.1891
Hartley, J., Pennebaker, J. W., and Fox, C. (2003). Using new technology to assess the academic writing styles of male and female pairs and individuals. J. Techn. Writ. Commun. 33, 243–261. doi: 10.2190/9VPN–RRX9–G0UF–CJSX
Haspelmath, M. (2001). “Word classes and parts of speech,” in International Encyclopedia of the Social and Behavioral Sciences, eds P. B. Baltes and N. J. Smelser (Amsterdam: Elsevier).
Heisler, T., Vincent, D., and Bergeron, A. (2003). Evaluative metadiscursive comments and face-work in conversational discourse. J. Pragmat. 35, 1613–1631.
Hidi, S., and Boscolo, P. (eds) (2006). Writing and Motivation. Leiden: Brill.
Hilt, L., Vandenkerkhove, R., and Daelemans, W. (2020). Linguistic accommodation in teenagers’ social media writing: convergence
Parhankangas, A., and Renko, M. (2017). Linguistic style and crowdfunding success among social and commercial entrepreneurs. J. Bus. Vent. 32, 215–236. doi: 10.1016/j.jbusvent.2016.11.001

Pasupathi, M. (2007). Telling and the remembered self: linguistic differences in memories for previously disclosed and previously undisclosed events. Memory 15, 258–270. doi: 10.1080/09658210701256456

Pavlou, P. A., and Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: implications for trust building, price premiums, and seller differentiation. Inform. Syst. Res. 17, 392–414.

Pennebaker, J. W. (2011). The Secret Life of Pronouns: What Our Words Say About Us. New York: Bloomsbury Press.

Pennebaker, J. W. (2015). The Development and Psychometric Properties of LIWC 2015. Austin: University of Austin.

Pennebaker, J. W., Mehler, M. R., and Niederhoffer, K. G. (2003). Psychological aspects of natural language use: our words, our selves. Ann. Rev. Psychol. 54, 547–577. doi: 10.1146/annurev.psych.54.101101.094007

Poggi, I., D’Errico, F., and Vincze, L. (2013). Comments by words, face and body. J. Multi. User Int. 7, 67–78.

Pecht, K. (2000). Patterns of Stance in English. Flagstaff, AZ: Northern Arizona University.

Pecht, K. (2003). Stance moods in spoken english: evidentiality and affect in british and american conversation. Text Talk 23, 239–257. doi: 10.1515/text.2003.010

Ren, W. (2018). Mitigation in chinese online consumer reviews. Dis. Context Med. 26, 5–12. doi: 10.1016/j.dcm.2018.01.001

Romero, D. M., Swaab, R. I., Uzzi, B., and Galinsky, A. D. (2015). Mimicry is presidential: linguistic style matching in presidential debates and improved polling numbers. Personlals. Soc. Psychol. Bull. 41, 1311–1319. doi: 10.1177/014616721591168

Salehan, M., and Kim, D. J. (2016). Predicting the performance of online consumer reviews: a sentiment mining approach to big data analytics. Dec. Support Syst. 81, 30–40. doi: 10.1016/j.dss.2015.10.006

Semo, E., and Culpeper, J. (2011). "Stylistics," in Pragmatics in Practice. Handbooks of Pragmatics, eds J.-O. Östman and J. Verschueren (Amsterdam: John Benjamins).

Schindler, R. M., and Bickart, B. (2012). Perceived helpfulness of online consumer reviews: the role of message content and style. J. Consumer Behav. 11, 234–243. doi: 10.1002/cb.1372

Schuckert, M., Liu, X., and Law, R. (2015). Hospitality and tourism online reviews: recent trends and future directions. J. Travel Tourism Mark. 32, 608–621. doi: 10.1080/10548408.2014.933154

Simmons, R. A., Gordon, P. C., and Chambliss, D. L. (2005). Pronouns in marital interaction: what do “you” and “I” say about marital health? Psychol. Sci. 16, 932–936. doi: 10.1111/j.1467-9280.2005.01639.x

Srivastava, V., and Kalfo, A. D. (2019). Enhancing the helpfulness of online consumer reviews: the role of latent (content) factors. J. Int. Mark. 48, 33–50. doi: 10.1016/j.intmar.2018.12.003

Steenkamp, J. B. E., and Ter Hofstede, F. (2002). International market segmentation: issues and perspectives. Int. J. Res. Mark. 19, 185–213. doi: 10.1016/S0167-8116(02)00076-9

Swaab, R. I., Maddux, W. W., and Sinaceur, M. (2011). Early words that work: when and how virtual linguistic mimicry facilitates negotiation outcomes. J. Exp. Soc. Psychol. 47, 616–621. doi: 10.1016/j.jesp.2011.01.005

Taucszik, Y. R., and Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. J. Lang. Soc. Psychol. 29, 24–54. doi: 10.1177/0269927209351676

Teso, E., Olmedilla, M., Martínez-Torres, M. R., and Toral, S. L. (2018). Application of text mining techniques to the analysis of discourse in eWOM communications from a gender perspective. Technol. Forecast. Soc. Change 129, 131–142. doi: 10.1016/j.techfore.2017.12.018

Tseng, M. Y., and Zhang, G. (2020). Perceptions of and attitudes toward elastic language in online health communication in chinese. Lingua 233:102750. doi: 10.1016/j.lingua.2019.102750

Wang, F., and Karimi, S. (2019). This product works well (for me): the impact of first-person singular pronouns on online review helpfulness. J. Bus. Res. 104, 283–294. doi: 10.1016/j.jbusres.2019.07.028

Willemens, L. M., Neijens, P. C., Bronner, F., and De Ridder, J. A. (2011). Highly recommended! the content characteristics and perceived usefulness of online consumer reviews. J. Comput. Med. Commun. 17, 19–38. doi: 10.1111/j.1083-6101.2011.01551.x

World Health Organization (2010). Global Recommendations on Physical Activity for Health: Geneva: World Health Organization.

Wu, L., Shen, H., Fan, A., and Mattila, A. S. (2017). The impact of language style on consumers’ reactions to online reviews. Tourism Manage. 59, 590–596. doi: 10.1016/j.tourman.2016.09.006

Yin, D., Bond, S. D., and Zhang, H. (2017). Keep your cool or let it out: nonlinear effects of expressed arousal on perceptions of consumer reviews. J. Mark. Res. 54, 447–463. doi: 10.1509/jmr.13.0379

Yu, G. R. (2012). Questionnaire preparation and dimension construction of clothing brand evaluation. Silk 4, 65–69.

Zhang, Y. Z. (2019). The weight calculation of garment aesthetic evaluation and how virtual linguistic mimicry facilitates negotiation outcomes. J. Mark. Res. 47, 201601. doi: 10.1509/jmr.13.0379

Zwaan, R. A., and Radvansky, G. A. (1998). Situation models in language comprehension and memory. Psychol. Bull. 123:162. doi: 10.1037/0033-2909.123.2.162

Conflict of Interest: QW was employed by ByteDance Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Mu, Li, Noble and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.