



# A Study on the Influence of Linguistic Styles of Reviews on Sales Performance in Second-Hand Transactions

Yini Liu<sup>1,a\*</sup>, Hao Zhang<sup>2,b</sup>

<sup>1</sup>School of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, Jiangsu, China

<sup>2</sup>School of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, Jiangsu, China

<sup>a\*</sup> Corresponding Author Email: lyn\_vanilla@163.com

<sup>b</sup> Email: zhanghao0216@nuaa.edu.cn

**Abstract.** With the change of people's consumption philosophy, second-hand consumption has become more and more popular. However, how to solve the information asymmetry between buyers and sellers in the second-hand market has become a key concern for second-hand platforms. Reviews play an important role in consumer decision, and the linguistic styles reflected in reviews have different persuasive effects on users. This study explores the relationship between reviews and sales performance from the perspective of review linguistic style based on a second-hand book platform, using text mining methods to process unstructured review texts and multiple linear regression for empirical analysis. Conclusions are both specific and intimate styles are positively related with sales volume, while intimate style is also positively related with sales speed.

**Keywords:** second-hand transaction, review, linguistic style, sales performance

## 1 Introduction

With the increasing recognition of low-carbon living, second-hand consumption as a form of sharing economy is becoming a new norm. Many second-hand trading platforms have emerged one after another. They build bridges between buyers and sellers, improving the efficiency of idle resource sharing and reuse, so as to promote the development of sustainable socio-economic.

In second-hand transactions, sellers have more information and thus become the more advantageous party. Therefore, how to break the barrier of information asymmetry between buyers and sellers is the focus of second-hand platforms. Some platforms such as Xianyu, Zhuanzhuan, and Duozhuayu have made attempts from review mechanism to provide feedback channels for users. This can reduce sellers' moral risks and help users make comprehensive judgments of target product's quality.

Important role played by reviews in consumer decision has been much confirmed<sup>[10]</sup>. Existing studies have shown review characteristics (e.g., quantity, rating), reviewer characteristics (e.g., credit rating, expertise) and product types have significant effects

© The Author(s) 2023

S. Yacob et al. (eds.), *Proceedings of the 2023 7th International Seminar on Education, Management and Social Sciences (ISEMSS 2023)*, Advances in Social Science, Education and Humanities Research 779, [https://doi.org/10.2991/978-2-38476-126-5\\_114](https://doi.org/10.2991/978-2-38476-126-5_114)

on sales performance<sup>[5][9][11][14]</sup>. However, most of studies are based on statistical features of reviews, while in real consumption process, users pay more attention to the review text<sup>[1]</sup>. In addition to literal content, “how to say” is also important. Linguistic style of text has potential impacts on recipients. Human communication theory suggests communication style can enhance relationship and credibility of interlocutor.

However, most studies focus on first-hand platforms, and less attention is paid to second-hand. Differences exist between them. In addition to sales volume, second-hand platforms also focus on efficiency of product flow, which means the speed from sold to purchased again. Second-hand consumers are generally more price sensitive and environmentally conscious<sup>[3][4][13]</sup>. First-hand products are consistent in quality, while second-hand products have uneven quality and varying appearance due to different degrees of use. It is hard to be truly reflected in description, so users can obtain through reviews.

Therefore, understanding the relationship between reviews and consumer behavior in second-hand markets has certain theoretical value and practical significance. This study examines from the perspective of reviews textual characteristics based on a domestic second-hand book platform. Main questions to be explored is how different linguistic styles of reviews affect sales performance.

## 2 Research Hypotheses and Model Construction

Linguistic style refers to the way in which information is written. Language psychology considers that persuasive linguistic styles include specificity and intimacy<sup>[2][12]</sup>.

Specific linguistic style measured by the proportion of prepositions and quantifiers<sup>[7][8]</sup> refers to a higher degree of description certainty. Use of these words in review make it refer to specific details more often, thus processing faster by users through non-semantic system<sup>[6]</sup>. Therefore, this study concludes that an increase in specificity of review linguistic style has a positive impact on sales performance. This study further subdivides sales performance into financial indicator (sales volume) and dynamic efficiency indicator (sales speed). Hypotheses 1a and 1b are formulated.

H1a: Specificity of reviews linguistic style is positively related with sales volume

H1b: Specificity of reviews linguistic style is positively related with sales speed

Intimate linguistic style is known as low level of psychological distance. Psychological distance refers to the degree to which reviewers remove themselves from the discussing topic and is measured by first person pronouns and negative emotional words<sup>[8]</sup>. More first person pronouns in reviews indicates that reviewer is more focused on the self, while more negative emotional words in reviews decreases the perception of trust<sup>[12]</sup>. They all increase the psychological distance between review reader and reviewer. Therefore, this study concludes that low psychological distance, high intimacy style has a positive effect on sales performance. Hypotheses 2a and 2b are formulated.

H2a: Intimacy of reviews linguistic style is positively related with sales volume

H2b: Intimacy of reviews linguistic style is positively related with sales speed

The research model is shown in the Fig 1 below:

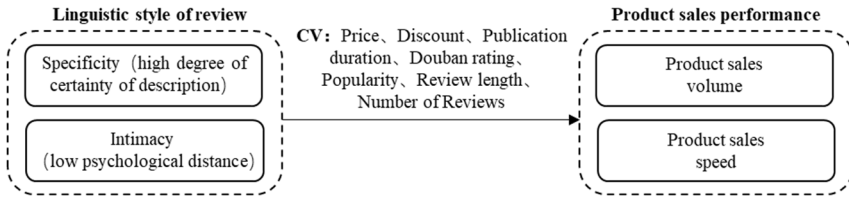


Fig. 1. Research Model: Relationship between Reviews Linguistic Style and Sales Performance

### 3 Data Collection and Processing

#### 3.1 Data sources and sample selection

This study selects a domestic second-hand book trading platform as the object. This platform is one of the earliest online second-hand book platforms. Its representative operating model is C2B2C, which means goods are recovered from users and sold after unified sterilization, refurbishment, encapsulation and pricing, solving credit and efficiency problems in second-hand books transactions through standardization.

In March 2023, we collected information from this second-hand book platform in categories of “Chinese literature”, “foreign literature”, “humanities and social sciences” and “art and design”. Data included basic information (e.g., title, price, publication date, Douban rating, sales volume, transaction time, etc.) and review text. Number of Douban ratings for each book used to measure popularity is collected from Douban.

Initial collection of second-hand book data amounted to 4172, of which 2039 had reviews. This study excluded books published before 1995 to better measure the impact of reviews on sales performance. Because these books belong to antique books, which have high academic research, collection and investment value and are often sold at a premium due to scarcity. So they are not in the scope of this study. Final filtered data contained 1207 books with 14991 reviews, with an average of 12.42 reviews per book.

#### 3.2 Text preprocessing

Data extracted from review texts is mainly their linguistic style features in this study. Before conducting text analysis, text format needs to be standardized. Then text segmentation, lexical annotation and word frequency statistics are performed. Word segmentation is the division of consecutive sentences in a text into separate words. This study mainly uses Python to call jieba.pseg. Python provides jieba word segmentation interface, which can be called by installing jieba library directly. It is one of mainstream word segmentation tools with good accuracy. Basic principle of jieba word segmentation is generating directed acyclic graph with the help of word graph scanning method, which reflects all words that may be composed of words in a text sentence, and the scanning process is based on prefix dictionaries, which is very efficient. It uses the dynamic programming method to find path with the highest probability in directed graph to find the maximum cut combination based on word frequency. In order to improve the accuracy of this tool, this study adds proper nouns such as book names and

author names to the original user dictionary, and then some proprietary words are manually added based on word segmentation results and debugged several times to improve accuracy. Finally, the lexical properties are marked and the word frequency is counted.

### 3.3 Variable measure

Dependent variables include book sales volume and sales speed. Sales volume data is measured by “number of people who bought” in the page. Sales speed is measured by the average number of days to sell.

In terms of independent variables, after text segmentation, lexical annotation and word frequency statistics, referring to the study of Theokary[8], number of prepositions and quantifiers were used to measure specificity of reviews, and number of first person pronouns and negative emotional words were used to measure intimacy.

In order to control for potential heterogeneity effects so that the hypothesis can be rigorously tested, this study used book price, discount, publication duration, Douban rating, popularity, review length, and number of reviews as control variables.

Variables are summarized in the following Table 1 :

**Table 1.** Summary of Variables

Type	Variable Symbol	Definition	Variable Description
Dependent variables	Volume	Sales volume	“Number of people who bought” in the page
	Speed	Sales speed	Match “sell-buy” from transaction records and calculate the average number of days, which is taken as negative side
Independent variables	Specificity	Specificity	Proportion of prepositions and quantifiers
	Intimacy	Intimacy	Proportion of first person pronouns and negative emotional words, which is taken as negative side
Control variables	Price	Price	Product selling price shown in the page
	Discount	Discount	Discount rate for second-hand books shown in the page
	Year	Publication duration	Subtract the year of book publication from 2023
	Rating	Douban rating	Book's Douban rating
	Popularity	Popularity	Number of Douban ratings for books
	Length	Review Length	Total number of words in reviews
	Quantity	Number of Reviews	Number of reviews displayed on the page

## 4 Empirical Analysis

### 4.1 Descriptive statistics

As shown in the Table 2 below, the result of descriptive statistics indicates that standard deviations of sales volume (Volume), product price (Price), publication duration (Year), popularity (Popularity) and review length (Length) are too large. In order to

reduce heteroskedasticity effect, these variables were logarithmic transformed (based on natural logarithm e).

**Table 2.** Result of Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Max	Min
Volume	14991	1425.8	1760.2	14021	12
Speed	14991	0.88	0.343	2.551	0.009
Specificity	14991	0.079	0.052	0.291	0.002
Intimacy	14991	9.56	4.102	24.89	1
Price	14991	25.008	28.14	320.3	8
Discount	14991	0.751	0.075	0.834	0.324
Year	14991	2952.1	1885.6	9517	233
Rating	14991	8.14	0.893	9.8	4.6
Popularity	14991	18176	73683	773288	16
Length	14991	268.26	672.77	14224	1
Quantity	14991	4.858	3.628	10	1

### 4.2 Correlation analysis

Correlation coefficient  $r$  is used to measure the strength of this relationship and takes values in  $[-1,1]$ , the closer to 0, the lower the correlation between two variables. As shown in the Table 3 below, the result of correlation analysis indicates that the relationship between specificity of linguistic style and sales volume is significant and the correlation coefficient is greater than 0.3, which laterally indicates that hypothesis 1a may be valid. The relationship between intimacy and sales speed is significant and the correlation coefficient is greater than 0.3, which means that hypothesis 2b may be valid.

**Table 3.** Result of Correlation Analysis

	Specificity	Intimacy	Volume	Speed	Price	Discount	Year	Rating	Popularity	Length	Quantity
Specificity	1										
Intimacy	0.183*	1									
Volume	0.412***	0.269**	1								
Speed	0.198*	0.501***	0.287**	1							
Price	-0.034	0.036	0.054	0.078	1						
Discount	-0.067	0.037	0.083	0.087	0.145*	1					
Year	-0.039	0.037	0.046	0.067	0.382***	0.358***	1				
Rating	0.075	0.023	0.0784	0.084	0.172*	0.041	0.235**	1			

<b>Popularity</b>	0.191 *	0.091	0.29 4**	0.1 1*	0.048	0.166 *	0	0.19 4*	1		
<b>Length</b>	0.157 *	0.004	0.21 9**	0.1 1*	0.032	0.232 **	- 0.07 5	0.13 3*	0.587* **	1	
<b>Quantity</b>	0.139 *	0.002	0.3* **	0.0 81	0.051	- 0.276 **	0.03 3	0.17 8*	0.656* **	0.83 ***	1

### 4.3 Regression analysis

Hypothesis testing was performed by building a multiple regression model. A regression equation based on independent variables and control variables with respect to the dependent variable as sales volume is established as follows:

$$Sales_i = \varepsilon + \beta_1 Specificity_i + \beta_2 Intimacy_i + \beta_3 Price_i + \beta_4 Discount_i + \beta_5 Year_i + \beta_6 Rating_i + \beta_7 Popularity_i + \beta_8 Length_i + \beta_9 Quantity_i \quad (1)$$

As shown in the Table 4 below, significance of the model meets the requirement and VIF are less than 10, so there is no problem of multicollinearity, and the results can be used for hypothesis testing. According to the result, both specificity and intimacy of reviews linguistic styles are significantly and positively related to sales volume ( $\beta=0.347, p<0.001$ ;  $\beta=0.204, p<0.001$ ), supporting hypothesis 1a and hypothesis 2a. So specificity of linguistic style can increase users' purchase willingness. While close psychological distance and enhanced intimacy can increase purchase intentions.

In terms of control variables, popularity is significantly and positively related with sales volume ( $\beta = 0.098, p < 0.05$ ), indicating that more popular products receive more purchases in the second-hand market. Number of reviews is also significantly and positively related with sales volume ( $\beta = 0.278, p < 0.001$ ), indicating that products receiving more feedback give users a stronger perception of trust to buy.

**Table 4.** Multiple Linear Regression Results (Dependent Variable: Sales Volume)

Results of linear regression analysis n=1207									
	Non-standardized coefficient		Standardization factor	t	P	VIF	R <sup>2</sup>	Adjustment of R <sup>2</sup>	F
	B	Standard Error	Beta						
<b>Constants</b>	4.337	0.8	-	5.423	0.000***	-			
<b>Specificity</b>	8.137	0.867	0.347	9.387	0.000***	1.053			
<b>Intimacy</b>	0.061	0.011	0.204	5.537	0.000***	1.05			
<b>Price</b>	-0.101	0.085	-0.048	1.194	0.233	1.241			
<b>Discount</b>	0.066	0.625	0.004	0.106	0.916	1.142	0.3	0.288	F=25.674 P=0.000***
<b>Year</b>	0.033	0.075	0.018	0.439	0.661	1.283			
<b>Rating</b>	0.05	0.055	0.036	0.9	0.369	1.26			
<b>Popularity</b>	0.06	0.031	0.098	1.964	0.050*	1.929			
<b>Length</b>	-0.06	0.047	-0.077	1.288	0.198	2.738			

<b>Quantity</b>	0.094	0.022	0.278	4.187	0.000***	3.401
<b>Dependent variable: Sales volume</b>						
<b>Note: ***, **, * represent 1%, 5%, 10% significance levels, respectively</b>						

A regression equation based on independent variables and control variables with respect to the dependent variable as sales speed is established as follows:

$$Speed_i = \varepsilon + \beta_1 Specificity_i + \beta_2 Intimacy_i + \beta_3 Price_i + \beta_4 Discount_i + \beta_5 Year_i + \beta_6 Rating_i + \beta_7 Popularity_i + \beta_8 Length_i + \beta_9 Quantity_i \quad (2)$$

As shown in the **Table 5** below, significance of the model meets the requirement and VIF are less than 10, so there is no problem of multicollinearity, and the results can be used for hypothesis testing. According to the result, intimacy of reviews linguistic style is significantly and positively related to sales speed ( $\beta = 0.514, p < 0.001$ ), supporting hypothesis 2b. This suggests that more intimate reviews can bring consumers psychologically closer to reviewers, thus allowing users to enhance speed of purchase decisions. In contrast, specificity does not have a significant relationship with sale speed although coefficient is positive, so hypothesis 1b is not tested. This suggests that more specificity can not contribute to speed of users' purchase decisions.

**Table 5.** Multiple Linear Regression Results (Dependent Variable: Sales Speed)

<b>Results of linear regression analysis n=1207</b>									
	Non-standardized coefficient		Standardization factor	t	P	VIF	R <sup>2</sup>	Adjustment of R <sup>2</sup>	F
	B	Standard Error	Beta						
<b>Constants</b>	0.554	0.225	-	2.46	0.014**	-			
<b>Specificity</b>	0.224	0.244	0.034	0.919	0.359	1.053			
<b>Intimacy</b>	0.043	0.003	0.514	13.887	0.000** *	1.05			
<b>Price</b>	0.017	0.024	0.029	0.722	0.47	1.241			
<b>Discount</b>	-0.223	0.176	-0.049	-1.269	0.205	1.142	0.297	0.285	F=25.316 P=0.000** *
<b>Year</b>	-0.027	0.021	-0.052	-1.266	0.206	1.283			
<b>Rating</b>	0.017	0.016	0.044	1.09	0.276	1.26			
<b>Popularity</b>	0.001	0.009	0.009	0.171	0.864	1.929			
<b>Length</b>	0.016	0.013	0.073	1.226	0.221	2.738			
<b>Quantity</b>	0	0.006	0	-0.007	0.994	3.401			
<b>Dependent variable: Speed</b>									
<b>Note: ***, **, * represent 1%, 5%, 10% significance levels, respectively</b>									

The summary results of the hypothesis testing are shown in the **Table 6** below:

**Table 6.** Summary of Hypothesis Testing Results

Hypothetical content	Result
H1a: Specificity of reviews linguistic style is positively related with sales volume	Support
H1b: Specificity of reviews linguistic style is positively related with sales speed	Not support
H2a: Intimacy of reviews linguistic style is positively related with sales volume	Support
H2b: Intimacy of reviews linguistic style is positively related with sales speed	Support

## 5 Conclusions

This study explores the relationship between reviews and sales performance from the perspective of linguistic style, based on a domestic second-hand book platform. This study uses text mining method to process the unstructured review texts and uses multiple linear regression to take empirically analysis. Conclusions obtained are as follows: both specific and intimate linguistic styles are positively related with sales volume, while intimate linguistic style is also positively related with sales speed.

Due to data limitations and limited proficiency, there are still some shortcomings. First, this study only involves one kind of second-hand products. In subsequent studies, data on other types of second-hand products can be supplemented to enrich the sample types and expand the applicability boundary. Second, conclusions drawn by text mining method are only inferences of correlations and cannot be taken as causal conclusions, so subsequent research can take experiments, questionnaires, and other means.

## References

- Chevalier, J. A., Mayzlin, D. (2006) The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3): 345-354. <https://doi.org/10.1509/jmkr.43.3.345>
- Dickert, S., Sagara, N., Slovic, P. (2011) Affective motivations to help others: A two-stage model of donation decisions. *Journal of Behavioral Decision Making*, 24(4): 361-376. <https://doi.org/10.1002/bdm.697>
- Ferraro, C., Sands, S., Brace-Govan, J. (2016) The role of fashionability in second-hand shopping motivations. *Journal of Retailing and Consumer Services*, 32: 262-268. <https://doi.org/10.1016/j.jretconser.2016.07.006>
- Guiot, D., Roux, D. (2010) A second-hand shoppers' motivation scale: Antecedents, consequences, and implications for retailers. *Journal of retailing*, 86(4): 355-371. <https://doi.org/10.1016/j.jretai.2010.08.002>
- Jiang, X. D. (2015) What Kind of Product Reviews Are Most Useful—— Research on the Influence of the Quantity and Text Features of Online Reviews on Their Usefulness. *Foreign Economics & Management*, (04): 41-55. doi:10.16538/j.cnki.fem.2015.04.004.
- Koh, Y., Lee, M., Kim, J., Yang, Y. (2020) Successful restaurant crowdfunding: the role of linguistic style. *International Journal of Contemporary Hospitality Management*, 32(10): 3051-3066. <https://doi.org/10.1108/IJCHM-02-2020-0159>



7. Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., Gorski, S. (2011) Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research*, 39(1): 19-37. <https://doi.org/10.1080/00909882.2010.536844>
8. Lee, M. T., Theokary, C. (2021) The superstar social media influencer: Exploiting linguistic style and emotional contagion over content?. *Journal of Business Research*, 132: 860-871. <https://doi.org/10.1016/j.jbusres.2020.11.014>
9. Li, A., Zhao, Z. J. (2019) A Review of Research on the Usefulness of Online Reviews. *Business Economics Research*, (06): 77-80. <https://kns.cnki.net/kcms/detail/10.1286.F.20190319.0858.042.html>
10. Liu, X., Lee, D., Srinivasan, K. (2019) Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *Journal of Marketing Research*, 56(6):918-943. <https://doi.org/10.1177/0022243719866690>
11. Sun, M. (2012) How does the variance of product ratings matter?. *Management Science*, 58(4): 696-707. <https://doi.org/10.1287/mnsc.1110.1458>
12. Toma, C. L., D'Angelo, J. D. (2015) Tell-tale words: Linguistic cues used to infer the expertise of online medical advice. *Journal of Language and Social Psychology*, 34(1): 25-45. <https://doi.org/10.1177/0261927X14554484>
13. Yan, R. N., Bae, S. Y., Xu, H. (2015) Second-hand clothing shopping among college students: the role of psychographic characteristics. *Young Consumers*, 16(1): 85-98. <https://doi.org/10.1108/YC-02-2014-00429>
14. Zhang, Y. H., Li, Z. W. (2016) Research on the Influencing Factors of the Usefulness of Online Reviews: Moderating Effects Based on Product Types. *Management Review*, (10): 123-132. doi:10.14120/j.cnki.cn11-5057/f.2016.10.011.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

