A Study on Identity Construction of China-US Cross-border E-commerce Enterprises from the Perspective of Discourse-Historical Analysis

Wang Yayao

Xi'an International Studies University, Xi'an, Shaanxi, China, 710128

Abstract: *This study utilizes the 2022 annual reports of Alibaba and Amazon as its corpus and adopts a discourse analysis three-dimensional framework within the context of discourse history research. The research path begins with an examination of discourse strategy, specifically focusing on the implementation methods and language characteristics in constructing corporate identity in China and the United States. Through the analysis of referential strategies, declarative strategies, and perspective strategies, this paper elucidates the differences and distinctive features of discourse strategies used for identity construction in Chinese and American enterprises. Additionally, it aims to establish a comparative research framework for discourse strategies in identity construction between Chinese and foreign enterprises, offering insights for further exploration in this field.*

Keywords: discourse-historical analysis; identity construction; e-commerce enterprises

1. Introduction

Identity construction, a prominent topic in social studies, has gained traction with cross-disciplinary research, particularly in linguistics. Scholars explore how discourse shapes identity, departing from fixed notions. While much research focuses on national, teacher, and ethnic identities, corporate identity through discourse analysis remains understudied. Corporate discourse, vital for brand image and societal perception, influences resource acquisition and employee commitment, pivotal for enterprise development (Wu & Zhang, 2019).[10]

This article adopts a social constructionist view of identity and employs the discourse-historical approach to analyze the 2022 annual reports of Alibaba and Amazon, constructing a corpus for study. Using Wmatrix and business discourse analysis, we conduct a comparative analysis on macro themes, language strategies, and language techniques. The aim is to understand the discourse of corporate identity in Chinese and foreign cross-border e-commerce firms. The study offers insights for Chinese enterprises aiming to globalize and integrate into overseas markets. This article will address the following three questions: 1. What are the characteristics of the corporate identity themes of the cross-border e-commerce companies in China and the United States? 2. What are the language strategies used in the construction of corporate identity by cross-border e-commerce companies in China and the United States? 3. What are the specific language devices used in the construction of corporate identity by cross-border e-commerce companies in China and the United States?

2. Literature review

The study of corporate identity construction originated in the 1960s and was first proposed by Gordon Lippincott and Walter Margulies.[6] Corporate identity construction is a process in which enterprises establish internal and external company image and identity through various forms, such as symbols, language, and actions. Initially, domestic and foreign scholars utilized paradigms and methods from the fields of management, communication, and cross-cultural communication to conduct relevant research. With the rise and vigorous development of disciplines such as discourse analysis and pragmatics, traditional perspectives on identity began to undergo a discursive turn (Benwell & Stokoe 2006:4)[1]. Scholars at home and abroad began to pay attention to the interactive relationship between language and identity in discourse communication dynamic processes, focusing mainly on institutional identity in universities (Chen Jianping, 2016, 2017)[3][4], teacher identity construction in the classroom (Gan &Lin, 2023), corporate identity construction in annual reports (Wu & Zhang, 2019), and

corporate identity discourse construction strategies in speeches by corporate leaders (Huo, 2020), among others.

Research on corporate institutional identity is limited, mainly focusing on cross-country comparisons.[8] For instance, Coupland and Brown analyzed emails from Royal Dutch Shell's website to examine discourse strategies. Sun and Jiang studied metaphorical strategies in American and Chinese companies' mission statements[9]. Domestic scholars like Wang Lifei and Bu Han explored CEO speeches and reports from Chinese and American banks, while Chen Pinying and Shen Jie compared social responsibility reports. Wu Nan and Zhang Jingyuan analyzed discourse in Huawei and Apple's annual reports. These studies indicate that social constructionist discourse analysis effectively reveals the interaction between discourse and identity, offering new insights for identity construction research.

3. Research Framework and Corpus

3.1. Research Method

The rapid advancement of computer technology and the ongoing enhancement of corpus analysis tools have revolutionized linguistic analysis, shifting it from solely qualitative to a blend of qualitative and quantitative approaches.[5] This article utilizes Ruth Wodak's discourse-historical approach as its theoretical framework, coupled with the discourse research methods of Wu Nan and Zhang Jingyuan (2019). It introduces a three-tiered analysis framework—comprising macro themes, medium-level discourse strategies, and micro-level language representations—to examine and contrast the corporate identity construction of Alibaba and Amazon[10].

The research tool Wmatrix, developed by Professor Paul Rayson at Lancaster University in the UK (Paul Rayson, 2008), is employed for statistical analysis of corpora. Wmatrix functions to tag and compile word lists, offering both syntactic and semantic tagging capabilities. Syntactic tagging involves POS (part of speech) tagging using the Claws tagger, boasting an accuracy rate of 96% to 97%. Semantic tagging utilizes the USAS tagger with an accuracy rate of 92%. Wmatrix identifies key semantic domains within texts, revealing underlying semantic meanings.[7]

3.2. Data Collection

In order to conduct this research, we downloaded the 2022 annual reports from both Alibaba and Amazon official websites to construct a Chinese E-commerce Corporate Corpus (CECC) and an American E-commerce Corporate Corpus (AECC). CECC contains 179,544 tokens and 13,012 types, while AECC contains 43,552 tokens and 6,306 types.

4. Three-Dimensional Discourse Analysis

4.1. Analysis of Macro-Theme

Using CECC and AECC as observation corpora and assisted by the corpus analysis tool Wmatrix to automatically identify prominent thematic semantic domains, we used BNC CG Business corpus (a sub-corpus of the British National Corpus) as a reference corpus to generate thematic semantic domain diagrams (as shown in Tables 1 and 2).

Observing Tables 1 and 2, we can see that the 2022 annual reports of both Alibaba and Amazon share the following semantic domains: N1 (numbers); I2.2 (business: selling); I1.1 (money and pay); I1 (money: generally); A1.8+ (inclusion); A2.2 (cause&effect/ connection); I1.3 (money: cost and price); I2.1 (business: generally); X2.2 (knowledge); A9- (giving); Y2 (information technology and computing); S8+ (help); A2.1+ (change); G2.1 (law and order), totaling 14 domains. This indicates that the corporate identities constructed by Alibaba and Amazon in their 2022 annual reports have some commonalities. The 14 shared thematic semantic domains describe the common features of the two companies in terms of enterprise services, management approaches, and enterprise nature.

Example 1: We seek to offer our customers low prices , fast and free delivery , easy-to-use functionality , and timely customer *service*.(Amazon)

Example 2: Alongside the continued development of its logistics infrastructure , Cainiao has also been continuously strengthening its customer *service* capabilities . (Alibaba)

The semantic domain S8+ (help) is used to describe "corporate services", with high-frequency words such as "help", "support", and "service". Both companies construct themselves as service providers with a sense of corporate social responsibility by describing their corporate services (see Example 1 and Example 2).

Example 3: Our Business suffers when we are unsuccessful in making , *integrating* , and maintaining acquisitions and investment. We have acquired and invested in a number of companies , and we may in the future acquire or invest in or enter into joint ventures with additional companies. (Amazon)

Example 4: In response to these big and impactful changes , our guiding principle has been be consumption , cloud computing , and confident , be flexible and be ourselves *globalization* as the immovable pillars.(Alibaba)

The semantic domain A1.8+ (inclusion) contains high-frequency words such as "comprehensive", "integration" and "globalization" and is ranked ninth in CECC and sixth in AECC. This indicates that both companies value teamwork and international cooperation and strive to build an inclusive and open image through participating in international cooperation and communication.

Based on Tables 1 and 2, the unique thematic semantic domains in AECC and CECC can be identified. AECC has the semantic domains I1.2 (money: debt), O2 (objects generally), A1.5.2-(useless), T1.3 (time: period), X6 (deciding), E4.2+ (content), etc. These semantic domains, which include words such as "debt, products, liability, estimate, fulfillment", indicate that Amazon values technological innovation, leadership management, and economic benefits, and pays attention to planning and risk management.

CECC has the semantic domains G1.1 (government), G2.2+ (ethical), S7.1- (no power), S7.1+ (in power), L2 (living creature: animals; birds,etc.), N5.1 (part). It can be inferred that this company pays more attention to its business achievements and reputation and enhances its corporate image and competitiveness through the support and cooperation of related academic fields and institutions.

Table 1 :Partial thematic semantic domains in CECC (Top 20).

Item	O1	%1	O2	%2	LL	LogRatio	
I2.1	3610	2.01	427	0.30 +	2163.16	2.73	Business: Generally
I1.1	3363	1.87	353	0.25 +	2147.56	2.90	Money and pay
G2.1	2215	1.23	116	0.08 +	1837.76	3.91	Law and order
N1	5579	3.11	1437	1.02 +	1716.46	1.61	Numbers
I2.2	3411	1.90	725	0.51 +	1307.41	1.89	Business: Selling
A2.2	2545	1.42	434	0.31 +	1191.31	2.20	Cause&Effect/Connection
I1	1710	0.95	254	0.18 +	887.97	2.40	Money generally
G1.1	1484	0.83	243	0.17 +	717.26	2.26	Government
A1.8+	1116	0.62	139	0.10 +	649.09	2.66	Inclusion
G2.2+	840	0.47	91	0.06 +	527.82	2.86	Ethical
S8+	2063	1.15	612	0.43 +	520.54	1.41	Helping
A2.1+	1591	0.89	443	0.31 +	440.77	1.50	Change
A9-	1560	0.87	451	0.32 +	409.28	1.44	Giving
Y2	1058	0.59	242	0.17 +	375.04	1.78	Information technology and computing
I1.3	1106	0.62	266	0.19 +	370.18	1.71	Money: Cost and price
S7.1-	495	0.28	37	0.03 +	366.35	3.39	No power
S7.1+	2201	1.23	831	0.59 +	356.14	1.06	In power
L2	480	0.27	37	0.03 +	351.14	3.35	Living creatures: animals, birds, etc.
N5.1-	835	0.47	167	0.12 +	339.86	1.97	Part
X2.2	306	0.17	7	0.00 +	299.43	5.10	Knowledge

Table 2: Partial thematic semantic domains in AECC (Top 20).

Item	O1	%1	O2	%2	LL	LogRatio	
N1	1952	4.48	1437	1.02 +	1793.63	2.14	Numbers
I2.2	1281	2.94	725	0.51 +	1466.64	2.52	Business: Selling
I1.1	944	2.17	353	0.25 +	1399.02	3.12	Money and pay
I1	618	1.42	254	0.18 +	870.18	2.98	Money generally
I1.2	507	1.16	189	0.13 +	752.60	3.12	Money: Debts
A1.8+	414	0.95	139	0.10 +	647.43	3.27	Inclusion
A2.2	633	1.45	434	0.31 +	620.64	2.24	Cause&Effect/Connection
I1.3	489	1.12	266	0.19 +	576.24	2.57	Money: Cost and price
G2.1	342	0.79	116	0.08 +	532.23	3.26	Law and order
I2.1	532	1.22	427	0.30 +	448.95	2.01	Business: Generally
O2	571	1.31	642	0.45 +	317.80	1.53	Objects generally
A1.5.2-	102	0.23	5	0.00 +	257.02	6.05	Useless
T1.3	718	1.65	1078	0.76 +	237.35	1.11	Time: Period
X6	81	0.19	0	0.00 +	234.05	9.04	Deciding
E4.2+	130	0.30	35	0.02 +	223.93	3.59	Content
X2.2	90	0.21	7	0.00 +	213.53	5.38	Knowledge
A9-	374	0.86	451	0.32 +	186.75	1.43	Giving
Y2	260	0.60	242	0.17 +	186.16	1.80	Information technology and computing
S8+	448	1.03	612	0.43 +	179.67	1.25	Helping
A2.1+	359	0.82	443	0.31 +	172.61	1.39	Change

4.2. Analysis of Mid-level Discourse Strategy

By comparing the Chinese E-commerce Corpus (CECC) and the American E-commerce Corpus (AECC), we observed that enterprise reports from both countries extensively utilize referential, predicative, and perspective-taking strategies in the discourse-historical analysis method for corporate promotion.

4.2.1. Referential Strategy

Referential strategies are crucial for shaping Chinese identities. Alibaba uses its corporate name "Alibaba" for self-reference, whereas Amazon favors adjectival possessive pronouns like "our" and

first-person pronouns such as "we". These differences highlight diverse group identities. American e-commerce firms' personalized language fosters an intra-group connection with readers, reducing the psychological distance from consumers. Conversely, Chinese e-commerce companies adopt a third-person perspective, aiming to establish an authoritative extra-group identity, independent of specific readership groups.

Example 6: And, *we* made important adjustments in our investment decisions and the way in which well invent moving forward ......(Amazon)

Example 7: Today, *Alibaba* has multiple consumer-facing brands in many regions and countries around the world.....(Alibaba)

4.2.2. Predicative Strategy

Predicative strategy conveys the author's stance. Both companies in the report use the modal expression "will" in their discourse, indicating intention or action plans, falling under speech acts of commitment.

Example 8: We *will* continue to learn from both our successes and our failures. We *will* make bold rather than timid investment decisions.......(Amazon)

Example 9: In the future , Alibaba *will* push technology boundaries even......(Alibaba)

4.2.3. Perspective-taking Strategy

Perspective-taking strategy refers to the method of demonstrating one's perspective and viewpoints. The perspective-taking strategy is often analyzed using the method of quotation, including direct and indirect quotations.

Alibaba's report uses many direct quotations, while Amazon has almost no direct quotations in the entire text, which is a significant difference. Amazon uses indirect quotations in the report. It can be seen that both Chinese and American cross-border e-commerce companies in the report constructed responsible corporate identities through perspective-taking strategies.

4.3. Analysis of Micro-language Feature

Word frequency analysis is an important aspect of corpus linguistics research. By counting the frequency of important words in the discourse and analyzing them, we can interpret the latent meaning of the discourse to the greatest extent possible. The Wmatrix tool was used to generate a word frequency table from the highest to the lowest frequencies, with functional words removed for the analysis.

Table 3: Frequently used terms in the AECC and CECC.

Number	CECC Corpus		AECC Corpus	
	Frequency	Word	Frequency	Word
1	820	business	178	sales
2	820	service	177	cash
3	664	PRC	172	net
4	625	equity	165	tax
5	547	company	162	services
6	487	group	161	operating
7	478	China	143	financial
8	466	financial	141	cost
9	435	revenue	129	business
10	419	fiscal	120	customer

From Table 3, it can be seen that Alibaba uses more national-related terms (such as PRC, China) in their report, while Amazon does not have this characteristic. Amazon emphasizes its commercial activities. In its annual report, it frequently uses terms such as "sales", "services", "cost", "business", and "customers", indicating that the company emphasizes its sales business, provided services, costs, commercial activities, and customer focus in the annual report. The use of high-frequency terms such as "financial", "cost", and "services" shows that Amazon's foundation for sustainable growth actually includes highly standardized management and a sound risk management system.

5. Conclusion

Discourse practice and identity have an interactive relationship.[2] Discourse and representation are important dimensions of identity construction research. This paper analyzed the 2022 annual reports of Alibaba and Amazon using critical linguistics and discourse-historical analysis. The study found that the discourse identities constructed by the two countries' enterprises have both similarities and differences. This study provides a foundation for further analysis of discourse identity construction strategies of enterprises and comparative studies of discourse identity construction models of domestic and foreign enterprises.

References

[1] Benwell, B., & Stokoe, E. (2006). Discourse and identity. Edinburgh University Press.

[2] Chai, G., & Han, H. (2017). An Interpersonal Pragmatic Study of Relational Identity in Government Business Promotion Discourse. Foreign Language Teaching, (1), 49-54.

[3] Chen, J. P. (2016). A Comparative Study on Institutional Identity Discourse Construction in Chinese and Foreign Universities. Chinese Language Learning, (4), 29-39.

[4] Chen, J. P. (2017). A Comparative Study of Discourse Construction Strategies of Institutional Identity in Chinese, British, and American Universities. Modern Foreign Languages, (1), 24-36.

[5] Chen, X. R. (2014). Identity Research from the Perspective of Pragmatics: Key Issues and Major Paths. Modern Foreign Languages, (5), 702-710.

[6] Lippincott, G., & Margulies, W. (1960). The Corporate Image. Journal of Business Identity Construction, 5(2), 123-137.

[7] Rayson, P. (2008). From key words to key semantic domains. International Journal of Corpus Linguistics, 13(4), 519-549.

[8] Shi, X. S., & Shan, X. H. (2019). Exploring the Cross-cultural Adaptation Level of Chinese Enterprise Websites Based on Corpus Analysis. Chinese Language Learning, (2), 71-80.

[9] Sun, Y., & Jiang, J. L. (2014). Metaphor Use in Chinese and American Corporate Mission Statements: A Cognitive Sociolinguistic Analysis. English for Specific Purposes, 33(2), 4-14.

[10] Wu, N., & Zhang, J. Y. (2019). Discourse Construction Strategies of Institutional Identity in Chinese and American Enterprises. Modern Foreign Languages, (2), 220-230.