



A literature review of Corpus Linguistics Research in the Field of Fashion

Ke Zhou*

University of Jinan, Jinan, 250022, China

*Corresponding author. Email:1481171685@qq.com

Abstract. Fashion is a multifaceted phenomenon with significant cultural, social, and economic importance. To explore fashion, some researchers have adopted corpus linguistics methods to examine fashion language. However, no researcher has reviewed the corpus-based literature in this field. This study reviews the application of corpus linguistics in fashion and identifies the potential gaps in previous studies. Nineteen studies were selected from Google Scholar and Web of Science. The corpora, tools, and theories used in these studies were further analyzed. Results showed that fashion studies employing the corpus approach tend to use corpora from single sources and focus more on written discourse than spoken data. Furthermore, different corpus tools were widely utilized. Not limited to linguistics, theories from other disciplines were combined with corpus linguistics in fashion research. This review highlights the potential of a corpus-based approach in fashion-related studies. Meanwhile, the study also provides references for future research by pinpointing noteworthy gaps that warrant further exploration.

Keywords: corpus linguistics, fashion, literature review

1 Introduction

Fashion is a complex concept that involves art, culture, marketing, self-identity, and so forth, influencing every aspect of individual and collective lives. It can be seen as a shifting style of dress, an art form, and a multibillion-dollar worldwide business [1]. Besides being a symbol of socio-economic status, fashion is also an active force in shaping culture. As a medium of expression, fashion is an important tool for us to showcase uniqueness and establish identity. Fashion is a cultural practice closely related to how we define our identities as individuals and as members of social groups [2]. Since fashion has a complex and profound impact on society and individuals, research in the field of fashion is of great significance.

Corpus linguistics is a research method of analyzing spoken and written natural texts stored in an electronic format [3]. As “one of the fastest-growing methodologies in contemporary linguistics [4],” corpus linguistics has been applied in research across multiple fields such as fashion. Fashion language is the core of understanding

fashion. Studying fashion language offers insights into the impact of fashion on social change, cultural dissemination, and economic situations. Additionally, researching fashion language can help advance fashion education and provide inspiration for new designs. Corpora can provide a large and extensive amount of data for researchers conducting fashion studies from multiple perspectives, such as cross-cultural and historical perspectives. It can be seen that adopting corpus linguistics methods for studying fashion has many advantages and significant implications.

Researchers study different aspects of the fashion field using the corpus method. First, several studies focus on how to construct identity in fashion discourse, for fashion reflects personal choices as well as broader social, cultural, and economic contexts. Limatius, for instance, explored how plus-size women's identity is constructed by analyzing the terminology used in plus-size fashion blogs to describe body shape [5]. Since fashion consumption represents a huge commercial value in the market, several studies concentrated on fashion brand marketing and communication discourse. Belinda Crawford Camiciottoli analyzed the website and Facebook homepage texts of over 100 fashion brands, aiming to provide references for fashion brands to effectively convey their expected brand attributes through the Internet [6]. Furthermore, one of the main areas of concern in the previous studies is sustainable fashion. For example, Gajšt investigated the frequently occurring single words and multi-word terms in media narratives of fast fashion and sustainable fashion and, based on this, analyzed the common attitudes towards fast fashion and sustainable fashion in media narratives [7].

Multiple corpus-based studies have been conducted in the fashion field; however, no researcher has yet conducted an up-to-date systematic review of this research field. To fill this gap, this study reviews 19 corpus-based studies in the fashion field, aiming to develop an understanding of the research trends of corpus-based studies in fashion and try to provide directions for future research. More precisely, the research seeks to address the following questions.

1. What corpora and tools were used in previous studies?
2. From what theoretical perspective have researchers conducted corpus-based studies in the fashion field?

2 The Corpus-Based Studies in The Fashion Field

Fashion plays a crucial role in shaping collective life, culture, and society as a whole. Realizing the importance of fashion language, some scholars have conducted research on various forms of language related to the fashion industry based on corpus linguistics. Corpus-based studies in the fashion field have been selected and reviewed to discover the trend of corpus research in the fashion field and identify potential research gaps. First, studies were collected from Web of Science and Google Scholar in July 2024. Key terms “fashion,” “fashion language,” and “corpus-based,” “corpus,” “corpus-assisted,” and “corpus linguistics” were combined using Boolean operators. The search string used is as follows: (“fashion” OR “fashion language”) AND (“corpus-based” OR “corpus,” “corpus-assisted” OR “corpus linguistics”). A total of

27 articles were found in the search. Then, research publications that meet the following criteria are selected.

- (1) They are written in English.
- (2) The corpus approach is used as one of the research methods
- (3) The research subject belongs to the fashion field

Ultimately, 19 articles were selected. These studies will be analyzed from the aspects of corpus used, topic concerned, tool employed, and theoretical perspective in this chapter.

2.1 Corpora Used

This analysis revealed the sources and usage of corpora for these 19 selected works of literature. As illustrated in Figure 2, the corpus used in these studies mainly come from three sources: magazines, brand websites, and social media. Fashion magazines, as the still active traditional media, account for the largest proportion. For instance, Pilyarchuk collected 1061 Vogue articles between July 2019 and June 2020 to build a corpus. Four studies’ corpora were composed of texts from social media platforms [8]. Given the growing significance of social media in shaping fashion discourse in contemporary society, fashion blogs, in particular, have emerged as the primary hub for the distribution of news and information about fashion. For example, in Limatius’s study, the corpus was composed of 20 fashion blogs and comments. [9] Also, the official websites of various fashion brands are the main battlefield for promoting and introducing products and expressing brand concepts.

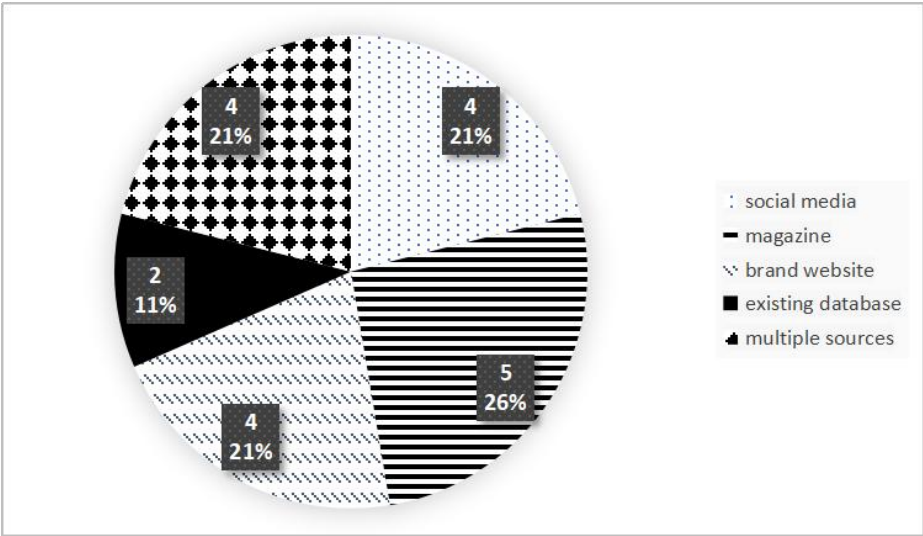
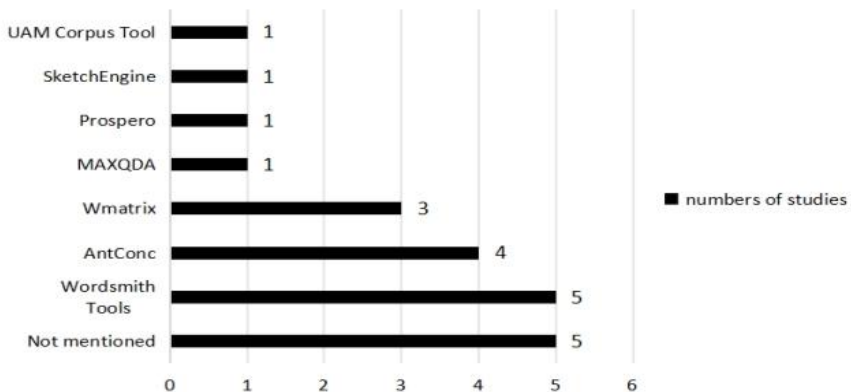


Fig. 1. The Sources of Corpus

In addition, 21% of studies used corpora composed of texts from different sources. Blanchet built a corpus by collecting 609 texts from the Factiva database and the websites and blogs of the fashion event and its exhibitor [10]. Fewer corpus is randomly sampled from large existing databases or corpus, such as NOW (News on the Web corpus), COCA (Corpus of Contemporary American English), and LexisNexis. The corpus used in Belinda's 2020 study was compiled from the LexisNexis database, a powerful online database that offers material from a wide range of news, business, legal sources, and well-known publications for academic research [11].

2.2 Tools Employed

Researchers have a variety of tool choices, but a few tools appear relatively frequently. According to Figure 3, five studies did not mention the specific tool employed. Wordsmith Tools are the most frequently used corpus tool, followed by AntConc. Then, Wmatrix was employed in 3 studies. Fewer studies utilized SketchEngine and UAM Corpus Tool in data analysis. In addition, MAXQDA, which is not a specialized corpus analysis tool, has also been in a study. MAXQDA is a tool for arranging, evaluating, displaying, and visualizing data from mixed and qualitative methods.

**Fig. 2.** Instruments Employed in the Analyzed Studies

These corpus tools all included basic analysis tools—word lists, keyword analysis, and KWIC, which can satisfy most of the research needs. However, when there is a need for semantic analysis in research, Wmatrix will be selected because it is equipped with a semantic annotation tool that automatically places each lexical item in one of more than 200 pre-established semantic domains. Crawford Camiciottoli processes corpus components by using the semantic tagger of Wmatrix to identify

their key semantic domains for observing the changing trends in the usage of Hyphenated Phrasal Expressions in Vogue [11].

2.3 Theoretical Perspectives

Most studies have conducted corpus-based research in the fashion field from different theoretical perspectives. Table 1 demonstrates that 13 studies employed theoretical perspectives, while six studies did not use any theory as a theoretical lens. Among the 13 studies, linguistic theories such as conceptual metaphor theory and multimodal discourse analysis were applied in eight studies. Additionally, theories from different fields were employed. For instance, the actor-network theory used by Blanchet in research is a widely used theory in the field of social sciences [12]. Crawford Camiciottoli takes the Brand Personality Scale, which belongs to the field of marketing and consumer psychology, as a Theoretical construct and applies it to investigate how fashion brands utilize web-based communications to convey brand personality [6].

Table 1. Theoretical Perspectives of previous studies

No.	Author/Year	Theoretical Perspective
1	Limatius (2019)	Community of Practice
2	Crawford Camiciottoli (2015)	None
3	Pristian (2016)	Multimodality Theory, Internet Linguistic Theory and Advertisement Language Theory
4	Crawford Camiciottoli (2020)	None
5	Limatius (2018)	None
6	Cao & Shi (2014)	Intertextuality
7	Crawford Camiciottoli (2018)	Dimensions of Brand Personality
8	Wells, V et al. (2021)	Legitimacy theory
9	Pilyarchuk (2024)	Conceptual Metaphor Theory and Critical Discursive Psychology
10	Cardoso et al. (2021)	Corpus Linguistics and Cognitive Linguistics
11	Ortolani (2024),.	Multidimensional Analysis
12	Gajšt (2024)	Discourse Analysis
13	Карпова (2023)	Multimodal Discourse Analysis

14	Mercuri (2023)	Pragma dialectics and Argumentum Model of Topics
15	Ho (2019)	The Appraisal Framework
16	Crawford Camiciottoli (2019)	None
17	Kömürcü (2011)	None
18	Cucchi and Piotti (2016)	None
19	Blanchet (2017)	Actor–Network Theory

3 DISSCUSSION

In this chapter, these 19 articles will be critically reviewed in terms of the following aspects: diversity and representativeness of corpora, robustness of methodologies and theoretical frameworks.

3.1 Diversity and Representativeness of Corpora

Diversity and representativeness should be the primary consideration in corpus design. In corpus-based studies of fashion, paying attention to the diversity of sources is crucial for achieving comprehensive, unbiased, and insightful results. This is because ideas, trends, and opinions about fashion are disseminated through various channels, including magazines, blogs, social media, etc. Also, fashion is a creative endeavor that constantly introduces new designs, materials, and technologies. These innovations continuously impact the evolution of fashion language. Acknowledging the intricate and diverse nature of language in fashion, Ortolani conducts a linguistic analysis of an extensive and varied collection of written and spoken texts in this domain [12]. A wide range of texts, including academic publications, fiction, Instagram posts, magazine articles, newspaper articles, tutorials, TV fashion shows, YouTube vlogs, and more, were used by Ortolani in his corpus of 2937 texts. The aim is to identify the primary linguistic characteristics and the variation parameters between written and spoken registers. Among these nineteen articles, the data collected in the 15 studies comes from only one source. This finding reveals a lack of diversity in the corpora used in most studies. For instance, Kömürcü collected data only from one fashion magazine for studying color-related expressions in fashion discourse, which led to its data's diversity not being guaranteed [13].

“Representativeness refers to the extent to which a sample includes the full range of variability in a population.” [14] Diversity and representativeness are interdependent. Collecting data from multiple sources, regions, and times can help cover different fashion expressions and perspectives, thereby increasing the representativeness of the corpus. And Limatius only collected fashion blogs from a private Facebook group in the UK to study how plus-sized women establish identity

in fashion discourse [5]. This corpus contains English data and can only represent the situation of plus-sized blogs in the UK. A corpus that over-represents certain sources, such as high-end fashion magazines or Western-centric fashion blogs, may lead to biased results that do not accurately reflect global fashion discourse. A representative corpus can help avoid such skewed outcomes.

3.2 Robustness of Tools

The widespread use of corpus tools can highly improve the robustness of methods used in fashion studies by ensuring systematic, reliable, and comprehensive analysis. Corpus is a large collection of data, so without relying on corpus software tools to process it, analyzing these data will become extremely difficult, consuming a lot of manpower and time. Furthermore, corpus software often includes features for documenting the analytical process so that other researchers can verify results by using the same software settings to reproduce the analysis, which contributes to transparency and the robustness of the research. Five articles do not mention the tools they use, which makes it difficult to reproduce their data analysis process and reduces transparency. Last, these tools have advanced analytical capabilities to provide quantitative metrics like word frequencies and collocation that add a layer of objectivity to the analysis, improving the robustness of the conclusions.

3.3 Theoretical Frameworks

It is important to have a theoretical perspective for corpus-based research in the fashion field. A theoretical perspective can provide a framework that guides the research design, data collection, analysis, and interpretation of results. For example, it can help researchers formulate specific and relevant research questions and hypotheses, ensuring that the study is focused and coherent. Moreover, theoretical frameworks inform the selection of appropriate corpus analysis methods and tools, ensuring that they align with the research objectives. In the absence of a theoretical framework, the methodology may adopt an ad hoc approach, with analytical techniques selected without a clear justification. Also, theoretical perspectives from different disciplines can be integrated to provide a more comprehensive understanding of fashion discourse. They allow researchers to draw on a broader range of theories and concepts, enriching the analysis and interpretation of findings. Several of these 19 articles combine corpus linguistics with theories from other disciplines as their theoretical structure. The Theoretical lens employed in Limatius's 2019 study belonged to sociology, known as the community of practice [5]. The findings of corpus-based research can contribute to theory building in the fashion field by validating, challenging, or extending existing theories. For example, Ho adapted the appreciation system fit to analyze the genre of luxury fashion promotional texts based on the original one proposed by Martin and White [15].

3.4 Research Gaps and Future Directions

This review revealed several gaps in the chosen research. Many studies rely on a limited range of data sources, such as fashion magazines, websites, or blogs, neglecting diverse platforms like podcasts and video content. This resulted in the verbal language of fashion being understudied. Only two of these 19 articles involve spoken discourse, and most articles only focus on written content. To fill this research gap, future researchers should pay more attention to spoken discourse of fashion and incorporate data from various formats, including text, audio, and video.

Most research only focuses on Western fashion discourses, underrepresenting non-Western perspectives and multicultural influences. Only three of the studies have touched upon non-Western fashion discourse, conducting cross-cultural comparative studies between Western and non-Western fashion discourse. Pristian compared the language used by Muslim fashion websites in India, Malaysia, the UK, and the US [17]. Also, Cao, Shi and Ho have established two corpora, Chinese and English, for comparative analysis of fashion advertising and luxury goods promotion discourse, respectively [18]. It can thus be suggested that cross-cultural comparative studies should be more conducted to understand differences and similarities in fashion discourse across regions.

Finally, the contextual use of fashion language in different registers and genres is not sufficiently explored. Most studies only focus on fashion discourse in a specific context, such as magazines, advertisements, and social media. However, no research has compared fashion language in different contexts. Studying how fashion language varies across different contexts can be a future direction of corpus-based study in fashion.

4 Conclusions

This review examined what corpora, tools, and theories are used in previous corpus-based studies of fashion. The results show that corpus-based exploration of the fashion field encompasses a great variety of topics. Corpus tools are widely used by most researchers to increase the reliability of data. In addition, several studies have made interdisciplinary attempts in terms of theoretical frameworks. Despite these advantages, the review identifies several gaps that warrant further exploration. Most studies analyze spoken discourse and non-English texts, which leads to lack of diversity in terms of data sources. Another issue is that current research lacks comparative studies on fashion language in different contexts.

The study lays a foundation for future corpus-based fashion studies by providing a systematic review of existing research. Specifically, this study provides a guide on the usage of corpora, tools, and theories. To be more robust and inclusive, future research should enrich the sources of data, demonstrate complete presentation of the data analysis process, and pay more attention to the application of theoretical frameworks. At the same time, the understanding gained here should help to identify future research directions. For example, cross-cultural perspectives on fashion study and research on oral fashion texts are both potential directions discovered in the review.

Some limitations need to be noted regarding the present study. First, the research sample size is relatively small because it relies mostly on two data sources, which may impact the validity, reliability, and generalizability of the findings. Moreover, limited in language, this study lacks attention to corpus-based study in non-English fashion language. The current study only examined corpora sources, tools, and theories from previous studies without considering other aspects such as themes, data filtering methods, and analytical techniques.

Future research could increase sample size by conducting a more comprehensive search to include a larger number of studies and incorporating studies focusing on fashion language in non-western contexts. Also, it is recommended that further research broaden the scope of analysis.

References

1. Kennedy, A. Stoehrer, E. B., Calderin, J. (2013) *Fashion design, referenced: A visual guide to the history, language, and practice of fashion*. Rockport Publishers, Los Angeles.
2. Craik, J. (2024). *Fashion: the key concepts*. Bloomsbury Publishing, London.
3. Kübler, N., Aston, G. (2010). Using corpora in translation. In Schmitt, N., Rodgers, M. P. (Eds.), *The Routledge handbook of corpus linguistics*. pp. 501-515. Routledge, London.
4. Gries, S. T. (2009). What is corpus linguistics?. *Language and linguistics compass*, 3(5), 1225-1241.
5. Limatius, H. (2019). “I’ ma fat bird and I just don’ t care” : a corpus-based analysis of body descriptors in plus-size fashion blogs. *Discourse, Context & Media*, 31, 100316.
6. Crawford Camiciottoli, B. (2018). The nuances of brand personality: a corpus-assisted linguistic analysis of web-based communications of fashion brands. *Hermes*, 57, 141-159.
7. Gajšt, N. (2024). Sustainable fashion is “in” and fast fashion is “out” – a corpus-driven analysis of media narrative regarding the fashion Industry’ s (un-)sustainability. *Challenges in the Turbulent Economic Environment and Organizations’ Sustainable Development*.
8. Pilyarchuk, K. (2024). In/exclusion in fashion discourse: are we in or out?. *Discourse & Society*, 0(0), 1-9.
9. Limatius, H. (2018). Fat, curvy or plus-size? A corpus-linguistic approach to identity construction in plus-size fashion blogs. *Search of Meaning: Literary, Linguistic, and Translational Approaches to Communication*. Tampere, Finland: Tampere University, 12-38.
10. Blanchet, V. (2017). ‘ We make markets ’ : The role of the Ethical Fashion Show in categorising the ethical fashion. *Recherche et Applications en Marketing (English Edition)*, 32(2), 26-45.
11. Crawford Camiciottoli, B. (2020). Hyphenated phrasal expressions in fashion journalism: a diachronic corpus-assisted study of Vogue magazine. *Lingue Culture Mediazioni*, 7(2), 1-22.
12. Ortolani, K. O. (2024). The language of fashion from a multidimensional perspective. In S. Araújo, (Ed.), *Portugal Digital Humanities Looking at the World: Exploring Innovative Approaches and Contributions to Society* (pp. 75-85). Cham: Springer Nature Switzerland.
13. Kömürçü, S. (2023). *Color terminology in the written fashion discourse* [Master’ s thesis, The University of Freiburg].

14. Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing*, 8(4), 243-257.
15. Ho, N. K. M. (2019). Evaluation in English and Chinese marketing communications: An adaptation of the Appraisal framework for the genre of luxury fashion promotional texts [Doctoral dissertation, Heriot-Watt University].

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.

