Convergence and Divergence in Branding City Destinations: A Comparative Study of the Multimodal Discourse in Beijing and London Publicity Films

Qian Wang
Northwestern Polytechnical University, China

Dan Zhang
Hainan University, China

Abstract

City branding brings immense benefits for mega-cities in gaining international prestige in an increasingly competitive global arena. City publicity films, as an effective method for promoting the city through online dissemination, can reach and influence a wider audience. However, the deployment of different semiotic resources in the branding discourse in city publicity films remains under-explored, and in particular, the role of cultural attributes in the construction of meaning in the discourse of city branding through linguistic and nonverbal modalities remains unknown. This paper, drawing on theories of Systemic Functional Grammar and Visual Grammar, examines the multimodal discourse of publicity films of Beijing and London in terms of representational and interactive meanings achieved through various semiotic resources. It is found that, in verbal and visual discourse, both films share similarities regarding enhancing persuasiveness via emotional branding but exhibit differences regarding how to achieve persuasiveness through different semiotic resources that co-construct meaning. The Beijing publicity film blends functional and emotional values while the London publicity film is prone to being more functional. In addition, possible reasons for the differences observed are discussed.
1. Introduction

With booming digital technology, language is not the only “medium of communication” (Halliday, 1985, p. 222) to convey meaning. Instead, different semiotic resources, including language, image, sound, color, and other modes, can collaborate in the generation of meaning, thus changing discourse analysis, which restricts its study to pure language, into multimodal discourse analysis (henceforth MDA). This recognition indicates that the analysis of language in texts and other semiotic resources that operate independently contributes to understanding a multimodal text (Lim, 2007). Against this backdrop has emerged the burgeoning growth of the application of MDA in different genres—textbooks, advertisements, paintings, posters, etc.—as evidenced in numerous studies (e.g., O’Toole, 1999; O’Halloran, 2000; Thibault, 2000; Lemke, 2002; Kress, 2003; Norris, 2004; Pang, 2004; Royce & Bowcher, 2007; Bateman, 2008).

Due to the increased competition in the global arena, cities are currently striving to gain international prominence using favorable brand images to become important attractions for investors, new citizens, talents, and tourists from across the world. To meet this challenge, cities must seek strategic marketing techniques to ensure that their uniqueness is recognized as a distinctive brand (Riza et al., 2012). In fact, the concept of city branding stems from the principles and management of traditional brands developed for products (Yoo, Donthu, & Lee, 2000). Defined as the “purposeful symbolic embodiment of all information connected to a city in order to create associations around it” (Lucarelli & Berg, 2011, p. 21), city branding entails substantial economic, political, and cultural benefits (Kavaratzis, 2004; Balakrishnan, 2009). These benefits explain why recent decades have witnessed the exponential growth of the literature addressing this topic from the perspectives of city branding concept and measurement (e.g., Zavattaro, 2010; Boland, 2013), branding strategy (e.g., Peel & Lloyd, 2008; Brandt & de Mortanges, 2011; Kaplan et al., 2010; Ahmad et al., 2013), and branding culture and tourism (Murphy & Boyle, 2006; Northover, 2010; Giovanardi, 2011). In
a digital era where internet marketing has revolutionized communication efforts, the internet today breeds tremendous opportunities for enhancing a city’s image internationally (Dobers & Hallin, 2009) as a response to encouraging cities to be connected with the global and digitized community (Sassen, 2004). Therefore, scholarly attention has gradually shifted to online city branding, focusing on tourism marketing and destination branding via the internet (Law et al., 2010), the effectiveness of online distribution of tourism-related information (Gertner et al., 2007; Xiang & Gretzel, 2010), and the role of social media as a destination marketing strategy (Hays et al., 2013).

Among the available means of advertising cities, city promotional films, which integrate text, music, sound, animation and many other resources to present a city’s history, lifestyle, and cultural heritage, can best serve the function of “selling.” In particular, with the emergence of easily accessible online media, a city’s promotional film can have a global reach, if disseminated online, to provide viewers with virtual tours of a destination.

The multimodal analysis of branding discourse by virtue of publicity films aligns with a social semiotic perspective probing into discourse and communication, which maintains that linguistic and visual choices are deployed as semiotic resources to achieve particular semiotic identity performance (van Leeuwen, 2005; Machin & Mayr, 2012; Ng, 2018). Based on an investigation of the existing literature, there has been little interest in exploring how cultural attributes exert an influence on semiotic choices that constitute the branding discourse via the online publicity film as an important channel for city branding. Therefore, this study, drawing on theories of Systemic Functional Grammar (Halliday, 1978, 2000) and Visual Grammar (Kress & van Leeuwen, 1996, 2006), compares the multimodal discourse of publicity films from Beijing and London as two mega-cities located in the East and the West. By exploring the representational and interactive meanings achieved through various semiotic resources, this paper hopes to shed light on the role that cultural attributes play in the meaning construction of discourse about city branding through verbal and nonverbal modalities.

2. Theoretical Background

2.1 Systemic functional grammar
As claimed, multimodality involves combining the use of “various semiotic modes with a socio-cultural domain that yields a semiotic product or result in an event” (Kress & van Leeuwen, 2006, p. 59). This interpretation links multimodality with systemic functional grammar (henceforth SFG), which argues for the social semiotic nature of language (Hallidy, 1985). In SFG, the choices of a linguistic system organized by the experiential, the interpersonal, and the textual as three metafunctions of language are constrained by the field (what is talked about), the tenor (who is participating), and the mode (means of transmitting the message).

The ideational metafunction by which we understand reality is the function for construing human experience (Halliday, 2000). This metafunction relates to the “grammatical resources involved in construing the flux of experience” through the unit of the clause (Guijarro & Sanz, 2008, p. 1603). The realization of the experiential function is dependent on a semantic system to describe inner and outer experiences, so the role of a semantic system leads to the application of a transitivity system whose choices in the grammatical system are process types, participant types, and circumstances. Interpersonal function involves social relations and participation, depicting the role adopted by speakers in exchanging meanings or interacting with each other. During a specific situation of communication, three syntactic moods, organized as declarative, interrogative, and imperative, assist in expressing illocutionary forces such as statement, offer, question, and directive. In addition to mood, the modality that the speaker uses to convey his/her attitudes or evaluations is another important component of interpersonal meaning. In accordance with SFG, the interpersonal meaning is also embodied in the personal pronoun system, since the choice of different personal pronouns enables the speaker to express his or her attitudes toward something and the relationship among listeners (Halliday, 2000, p. 92).

2.2 Visual social semiotics
Building on Halliday’s social semiotics, Kress and van Leeuwen (1996, 2006) develop a descriptive framework of visual semiotic analysis for visual communication called visual grammar (henceforth VG). Similar to linguistic structures, visual processes within visual structures are also related to the roles of participants in a specific communicative context.

Narrative representation and conceptual representation constitute the
dimension of representational meaning. Narrative processes that present unfolding actions and events, processes of changes, and instantaneous spatial arrangement, are connected by vectors. In contrast to narrative processes, conceptual processes are static (no vectors) and do not involve the action or reaction of represented participants but represent them in terms of their “more generalized and more or less stable and timeless essence, in terms of class, structure, or meaning” (Kress & van Leeuwen, 2006, p. 267).

The interactive meaning in VG depicts the relations between image makers, represented participants, and image viewers, giving hints to viewers about the producer’s attitudes toward the represented objects, including people in images (Kress & van Leeuwen, 1996, pp. 52-53). This interactive meaning can be achieved by contact, social distance, perspective, and modality in visual mode (Kress & van Leeuwen, 1996, p. 67). Contact refers to the relationship established between represented participants and viewers or audiences, which can be realized by two acts—the “demand” act and the “offer” act (Kress & van Leeuwen, 1996, p. 113). The former relates to represented participants’ seeking attention or a relationship built up from viewers by virtue of a gaze or eye contact with them, the purpose of which is to obtain information; the latter refers to the represented participants’ offering information to viewers by not looking at them directly.

Social distance, as a second dimension of the interactive meanings of images, manifests the degree of intimacy established between the viewers and the represented participants through the framing size of the lens. It is determined by how close the represented participants depicted in images appear to the viewers, thereby resulting in feelings of intimacy or distance. This affinity and relationship can be revealed by a close-up shot implying intimacy, a medium shot denoting social distance, and a long shot signaling impersonal distance.

Perspective refers to the choices of angles that can reflect the attitude and affect how we understand the situation; these angles comprise horizontal and vertical angles. The horizontal perspective manifested with a frontal angle and an oblique angle connotes involvement and detachment from represented participants, respectively. The vertical perspective, realized through high, low, and eye-level angles, implies different levels of power between represented participants and viewers.

Modality refers to the degree of truth through which the real world is reflected. Objects may be filmed as if they were real by resorting to markers such as color,
contextualization, brightness, representation, and depth. Color saturation, color differentiation, and color modulation comprise three scales of color, which are graded from full to absence, from maximally diversified to monochrome, and from fully modulated to unmodulated, respectively. It is noteworthy that judging the reliability of color expression depends on different viewers, which characterizes the emotional factor of color (Kress & van Leeuwen, 2006, p. 163). Contextualization, another subset of modality, involves the image background, graded from detail to absence.

In short, the ideational and interpersonal meaning in SFG and their realization, as well as the representational and interactive meaning in VG and their realization, comprises the theoretical foundations of the present study. As a typical multimodal discourse, verbal and visual modes prevail in a city publicity film in which SFG will be used to analyze texts in verbal discourse, and VG will be used to analyze images in visual discourse.

2.3 Data description

Both Beijing and London are metropolises characterized by concentrations of economic, political, and cultural activities. Hence, the data used for analysis come from the two cities’ 2014 publicity films on YouTube ranked according to popularity, as indicated by the number of viewers: one is the international version of the Beijing Publicity Film (henceforth BPF) produced by Ke Zhaolei, and the other is the London Publicity Film (henceforth LPF) produced by Louis Theodossiou. Narrated by an adult male in Chinese with English subtitles, the BPF lasts 271 seconds while the LPF lasts 238 seconds, with the voice-over of a young girl.

The verbal data in the two films are transcribed manually and thoroughly evaluated for accuracy, and the visual data gathered are segmented to obtain static images (O’Halloran, 2004) using Storm Media software. The obtained data amount to 141 words (12 sentences) and 97 pictures in the Beijing Publicity Film and 223 words (21 sentences) and 70 pictures in the London Publicity Film.

3. Findings and Discussions

3.1 Verbal discourse analysis: Ideational and interpersonal meanings

As contended, it is the language that empowers us to conceptualize and describe
our inner and outer experience (Halliday, 1985). The transitivity system distinguishes between outer (the processes of the external world) and inner (the processes of consciousness) experience; the former refers to material processes and the latter to mental processes. Another grammatical category that generalizes and relates one fragment of experience to another by classifying and identifying is called the relational process. Material, mental, and relational processes are the three main types of processes in the English transitivity system (Halliday & Matthiessen, 2014).

3.1.1 Transitivity

Table 1. Types of processes in the BPF and LPF

<table>
<thead>
<tr>
<th>Processes</th>
<th>BPF</th>
<th>LPF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage</td>
</tr>
<tr>
<td>Material</td>
<td>2</td>
<td>17%</td>
</tr>
<tr>
<td>Mental</td>
<td>6</td>
<td>50%</td>
</tr>
<tr>
<td>Relational</td>
<td>3</td>
<td>33%</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>100%</td>
</tr>
</tbody>
</table>

The analysis of transitivity in the verbal discourse of the BPF and LPF shows that the mental and relational processes exhibit similarly high frequency in the two films, accounting for 50% and 33% and for 52% and 34%, respectively. Recurrent mental processes (feel, discover in the BPF; trust, inspire in the LPF) represent the sources of cognition (e.g., feel) and desire (e.g., discover) for the described textual reality. The nature of a promotional film that intends to elicit the viewer’s emotional response may explain the high frequency of the mental processes observed. With references to the relational processes used to define the attributes of the cities and to quantify a speaker’s feelings (be in the BPF and LPF), the descriptive function of the city’s great charm has been fulfilled.

3.1.2 Mood

The basic types of moods for interpersonal exchanges can be represented by three syntactic forms: commands are usually expressed by imperative sentences, statements by declarative sentences, and questions by interrogative sentences. Offerings can
be expressed by any of those three forms. The following table demonstrates the distribution of mood structure in the two films.

Table 2. Distribution of mood structures

<table>
<thead>
<tr>
<th>Films</th>
<th>Mode Structure</th>
<th>Declarative</th>
<th>Imperative</th>
<th>Interrogative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPF</td>
<td>Frequency</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>0.92</td>
<td>0.08</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>LPF</td>
<td>Frequency</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

As illustrated in the table above, both of the city publicity films are inclined to use declaratives: 92% of the verbal discourse in the BPF and all of the verbal discourse in the LPF are declaratives. The imperative is seldom used and interrogative sentences never occur. Declaratives are characterized by brevity, through which enough information can be offered to target viewers to fulfill the purpose of publicity films. Only one imperative sentence appears in the BPF, inviting people to pay a visit.

Some snippets of the verbal discourse from the BPF include the following:

(1) One day, I realized that, even though I was born in Beijing, it feels strange yet familiar.

This is the opening sentence that serves to introduce Beijing by way of a story narration. The narrator intends to change people’s perception of China as an old place by directing the audience’s attention to how this city is being rebuilt in a contemporary way.

(2) It is classic and modern, calm and peaceful, yet high-flying and energetic.

This declarative appears when the narrator concludes with distinctive features of modern Beijing. The information offered aims to persuade the audience to believe that, currently, China is a nation not isolated from the rest of the world but integrated into modernity while still retaining its cultural features.
(3) Walking in the Hutongs, I can still feel the aroma of Beijing.

To further illustrate the city’s traditional culture, this declarative invites the audience to appreciate the noticeable charm of what characterizes Beijing’s traditional era. The Hutongs (a type of narrow street or alley commonly associated with northern Chinese cities, especially Beijing) not only embody a certain period of history, but also the cordial lifestyle of the people who live there.

(4) All of you who come to Beijing, let us discover together.

This is the only imperative sentence in the film, and it brings it to an end. The narrator here uses the imperative to invite people from different corners of the world to visit Beijing, fulfilling the purpose of a publicity film.

Taken together, it has been observed that the declarative mood occurs most frequently, which corresponds to the purpose of an invitation. The interpersonal meaning of the relationship between the represented participants and viewers is realized via presenting information about this city and describing what it looks like.

By comparison, analysis of the mood in the verbal discourse of the London Publicity Film shows a similar preference for choosing declaratives to offer information, as shown in the following examples.

(1) We beg for opportunities; this is why we reside here.

This declarative helps the audience capture the idea that London is a place full of opportunities.

(2) We put our trust in London, and London stills our hearts.

In this sentence, the speaker describes the credible core qualities of London, i.e., making people living there feel calm and peaceful in a fast-paced society.

As shown in Table 2, all the clauses in the LPF are declaratives, which convey information about London’s characteristics and qualities, building the bond between what is represented in the film and the viewing audience.
3.1.3 Modality
Modality refers to the intermediate ground of meaning between the positive and negative polarity, consisting of modalization and modulation (Halliday, 2000). If the speaker is providing information through utterance, the modality relates to how valid the information is in terms of probability (how likely it is; e.g., possible, likely, probable, certain, sure, etc.) or usuality (how frequent it is; e.g., sometimes, often, usually, always, etc.). If the speaker talks about goods-and-services, the modality relates to how confident the speaker can be in the eventual success of the speech act. Halliday endows modal expressions with different values, in terms of high, median, and low, to reflect speakers’ attitudes and degree of politeness. In commands, the modal space concerns the degree of obligation of the person to carry out the commands, while in offers it regards the degree of the inclination of the speaker to fulfill the offer.

Table 3. Distribution of modality structures

<table>
<thead>
<tr>
<th>Type</th>
<th>BPF</th>
<th>LPF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage</td>
</tr>
<tr>
<td>Modalization</td>
<td>1</td>
<td>50%</td>
</tr>
<tr>
<td>Probability</td>
<td>1</td>
<td>25%</td>
</tr>
<tr>
<td>Usuality</td>
<td>1</td>
<td>25%</td>
</tr>
<tr>
<td>Obligation</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Inclination</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

As illustrated in Table 3, similar patterns of employing more expressions of modalization than modulation to offer information emerge in the BPF and LPF. Additionally, no obligation expression of modulation occurs in both films. Moreover, the expression of probability is preferred over other subsets, with 50% and 66.7% in the BPF and LPF respectively. However, modalization is realized through the high and median value of modality in the BPF, in contrast with preferences over the
median and low value of modality in the LPF. Furthermore, the distribution of usuality demonstrates differences. The BPF chooses both the high and median value of modality, whereas the LPF tends to deploy the low value modality to express usuality. It is noteworthy that the inclination occurs once in the LPF, with no counterpart in the BPF.

Let us consider an example from the Beijing Publicity Film: *Apart from what everyone knows, I can always discover new things in Beijing*. The word “can” suggests a probability that speaks about the validity and certainty of propositions between polarities to provide the audience with some freedom and choices, but with expected responses from viewers. In this sentence, the narrator has succeeded in building up connections with the audience by expressing his feelings. The word “can” joined with “always” indicates a high frequency in expressing the meaning of usuality, denoting the truthfulness and reliability of the information that Beijing is a place where numerous unknown excitements are waiting to be explored. Thus, the message is sent out as a lure for potential visitors, intending to persuade viewers by emphasizing the narrator’s feelings.

Some snippets of choice of modality structure in the London Publicity Film include the following:

1. The capital is a beautiful thing, and you'll never forget the sound.
2. The city will inspire you more than you will ever know.

In LPF, “will” occurs five times, where four of these instances are expressions of probability and one is of inclination. Furthermore, “never” functions as an expression of usuality. In these sentences, except for the “will” as used in “you will ever know” in sentence 2, which functions as inclination, the narrator’s choice of the median modal verb “will” showcases the confidence and confirmation of the probability of people’s memory of London. The choice of “will” in these examples leaves room for the audience to make a decision, thus making the information more suggestive. The inclination revealed by “will” in example 2 transmits the idea that the audience will be inspired by London in some way, regardless of how much one could expect to experience. “Never” in example 1 enhances the emphatic tone of what this capital city can do to impress the audience.
3.1.4 Personal pronouns
The specific usage of personal pronouns can uncover interpersonal and attitudinal ideologies toward the relationship concealed in discourse (Halliday, 2000). The first person pronoun “I” signals the addressee or the speaker. The plural form “we,” typically used in an inclusive sense, ideologically reflects a shared stance or opinions between speaker and addressee(s). However, the second person pronoun “you” reveals a certain distance between the speaker and listener, which may embody power (Stephan, 1992).

Table 4. Use of personal pronouns in the BPF and LPF

<table>
<thead>
<tr>
<th>Type</th>
<th>BPF</th>
<th>LPF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage</td>
</tr>
<tr>
<td>I</td>
<td>4</td>
<td>44%</td>
</tr>
<tr>
<td>we</td>
<td>3</td>
<td>34%</td>
</tr>
<tr>
<td>my</td>
<td>1</td>
<td>11%</td>
</tr>
<tr>
<td>us</td>
<td>1</td>
<td>11%</td>
</tr>
<tr>
<td>you</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 4 shows that in the BPF, the frequency of using first person pronouns, including the singular, plural, and object forms, to a great extent outnumbers what is observed in the LPF, which prefers the second personal pronoun “you” instead. By articulating the narrator’s feelings and seeking bonds with viewers using first person pronouns (e.g., “we” and “us”), the BFP tries to shorten the distance to viewers to influence the audience in a more subtle and implicit fashion. In contrast, the LPF is inclined to highlight that London will impress “you” as an audience, suggesting more objective and explicit messages.

3.2 Visual discourse analysis: Representational and interactive meanings
For dynamic multimodal discourse, in fact, visual images are the primary carriers of the discourse’s core meaning to render information.
### 3.2.1 Representational meaning

In line with what is presented in the BPF and LPF, three umbrella categories, encompassing humans, inanimate objects, and spatial representations, and seven subcategories can be identified: 1) People and local residents (e.g., locals, everyday life, recreation); 2) infrastructures (e.g., skyscrapers, transportation); 3) fashion (e.g., clothes, accessories, shopping centers); 4) local cuisine and dining (e.g., food, drinks, restaurants); 5) cultural icons (e.g., cultural heritage, traditional events); 6) tourist attractions (e.g., historic buildings, scenic spots, works of art); and 7) natural landscapes.

<table>
<thead>
<tr>
<th>Category</th>
<th>BPF</th>
<th>LPF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage</td>
</tr>
<tr>
<td>human</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>27</td>
</tr>
<tr>
<td>objects</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>spatial representation</td>
<td>Total</td>
<td>97</td>
</tr>
</tbody>
</table>

Figure 1. Comparison of represented participants in the BPF and LPF
The differences in these two films lie in that the BPF chooses represented participants in terms of people (17%), inanimate objects (58%), and spatial representations (25%), whereas the LPF only involves inanimate objects (37%) and spatial representations (63%), without any human participants. Moreover, the BPF tends to describe cultural icons (e.g., a Chinese knot, calligraphy, tea culture), while the LPF is inclined to demonstrate infrastructures, tourist attractions, and landscapes. To be more specific, in the BPF, human activities are instantiated in the scenes involving serving Chinese tea, a Peking Opera performance, and ordinary people’s lives, especially young people’s lives. For inanimate objects, the transportation system, cultural icons, and food and cuisine are described in a great detail. However, in the LPF, for inanimate objects, only infrastructure, such as skyscrapers, and fashion are identified using recognizable scenic spots and images of landscapes, and these are depicted for spatial representations.

Moreover, we found that narrative images (67%) predominate in the BPF, whereas conceptual images (85%) prevail in the LPF. The BPF uses actional and reactional images (image 1 describes a girl playing the guitar and image 2 suggests the enjoyment of tea) to inject the film with vitality. As shown in image 3, which depicts a girl’s hand touching colorful bubbles, the symbolic meaning of Beijing as a place with different colors waiting to be felt and experienced is revealed in this conceptual process.

By comparison, landmarks or modern buildings filmed using conceptual processes dominate in the LPF, as instantiated in image 4. This image reflects a classification process, with the London Eye being more salient than other buildings, conveying a panoramic night view of London. Image 5 utilizes the shining red light as represented participant, which could possibly embody hopes in London.
3.2.2 Interactive meaning

- Contact

In the Beijing Publicity Film, there are 16 scenes involving human participants, in which only 4 demand information enacted from the viewers through eye contact, as illustrated in image 6. The construct of “demand” signals the engagement of viewers by virtue of a direct invitation. However, in the London Publicity Film, the director chooses to portray London by exhibiting London’s scenery rather than involving any human participant, as illustrated in image 7.

In the BPF, image 6 depicts an elegant lady dressed in a Chinese cheongsam introducing the Chinese tea ceremony with a smile. She makes direct eye contact with the audience to elicit their response and interaction, thereby sending an invitational message. By contrast, image 7, as one of London’s renowned landmarks, offers information about London’s beauty and magnificent architecture but with no intention of communicating with the viewers.

Therefore, both publicity films resort to offering images to introduce cities in a more direct fashion, especially London, which is in keeping with the main feature of publicity films, namely, to offer as much information to the viewers as possible. However, the Beijing film more purposely engages the audience in short-term interaction to send an invitation in a slightly warmer way, compared to London’s Promotional Film, which centers more on tourist attractions that highlight sightseeing experiences.
- Social Distance

As revealed in the BFP, the more frequent use of medium shots and close-up shots than long shots enhances the engagement of viewers. By contrast, there are more long shots and medium shots in the LPF, in which no animate participant is engaged; instead, it has a rich display of magnificent buildings. Thus, it can be concluded that the BPF mainly construes a personal and social distance, while the LPF scores high in social and impersonal relationship.

Table 6. Distribution of social distance

<table>
<thead>
<tr>
<th>Type</th>
<th>BPF</th>
<th></th>
<th>LPF</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage</td>
<td>Number</td>
<td>Percentage</td>
</tr>
<tr>
<td>Close-up shot</td>
<td>43</td>
<td>44%</td>
<td>16</td>
<td>23%</td>
</tr>
<tr>
<td>Medium shot</td>
<td>36</td>
<td>37%</td>
<td>17</td>
<td>24%</td>
</tr>
<tr>
<td>Long shot</td>
<td>18</td>
<td>19%</td>
<td>37</td>
<td>53%</td>
</tr>
<tr>
<td>Total</td>
<td>97</td>
<td>100%</td>
<td>70</td>
<td>100%</td>
</tr>
</tbody>
</table>

As exemplified in image 8, the close-up shot that depicts a girl smelling the hotpot establishes an intimacy between the viewers and the represented participant through bringing viewers into the scenario. This approach can stimulate viewers’ desire to sample this Beijing specialty. Images 9 and 10 show the Imperial Palace filmed from long shots moving toward close-up shots, creating an effective visual impact by converting frames, with whose help the majestic royal building as a whole and with the details of exquisite carved designs are presented.
By comparison, images 11 and 12 showcase the clock in One Canada Square, the iconic landmark in London, with a medium and then a close-up shot that highlight the fast-moving hands of the clock. The conversion of the frames from medium to near embodies the concept “time is money,” which accords exactly with the theme of the skyscraper. The long shot instantiated in image 13 indicating impersonal distance showcases the attraction of London Bridge being united with its surroundings. Taking these differences into account, the BPF endeavors to forge closer relations with the viewer, whereas the LPF seems to influence the viewer more objectively within an appropriate distance.

- Perspectives

The following table shows the distribution of the perspectives in the publicity films of the two cities.

Table 7. Distribution of perspectives

<table>
<thead>
<tr>
<th></th>
<th>Horizontal</th>
<th>Vertical</th>
<th>BPF</th>
<th>LPF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage</td>
<td>Number</td>
<td>Percentage</td>
</tr>
<tr>
<td>Front</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>8</td>
<td>8%</td>
<td>16</td>
<td>23%</td>
</tr>
<tr>
<td>Low</td>
<td>6</td>
<td>7%</td>
<td>8</td>
<td>11%</td>
</tr>
<tr>
<td>Eye-level</td>
<td>34</td>
<td>35%</td>
<td>22</td>
<td>31%</td>
</tr>
<tr>
<td>Oblique</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>17</td>
<td>18%</td>
<td>18</td>
<td>26%</td>
</tr>
<tr>
<td>Low</td>
<td>5</td>
<td>5%</td>
<td>2</td>
<td>3%</td>
</tr>
<tr>
<td>Eye-level</td>
<td>27</td>
<td>27%</td>
<td>4</td>
<td>6%</td>
</tr>
<tr>
<td>Total</td>
<td>97</td>
<td>100%</td>
<td>70</td>
<td>100%</td>
</tr>
</tbody>
</table>
In the BPF, images 14 and 16 are clear instantiations from the front angle, whereas images 15 and 17 are filmed from an oblique angle, denoting different levels of engagement of viewers. The majesty of the Imperial Palace is demonstrated in image 14 by virtue of the low angle, whereas playing with grasshoppers, one of the interesting entertaining activities Beijing People used to enjoy is revealed in image 15 via a high angle. Touring tricycles in the streets of Beijing suggest what common people’s daily life looks like, eliciting shared feelings of working hard for a better life. An oblique angle in image 17 functions to convey the message of harmony between human and nature.

Likewise, a low angle is used to evoke viewers’ admiration when presenting magnificent buildings, as seen in image 18. However, an eye-level angle and front angle would be chosen to invite viewers to become part of the scenario, as what is seen in image 19 is Oxford Street, the world-famous shopping area, which greets and embraces international visitors.
In line with what is presented in Table 7, we find that both front and oblique angles are preferred in the BPF, while the front angle is favored in the LPF. Comparatively speaking, a high angle (49%) occurs more frequently in the LPF than in the BPF (26%). In addition, both films are inclined to convey equality with the viewers by their use of a large proportion of eye-level angle shots. In particular, in the BPF, when an intimate relationship is construed, the viewer is either empowered or positioned equally to feel being respected.

- **Modality**

Image 20 from the Beijing Publicity Film and image 21 from the London Publicity Film, though they exhibit different modalities in color, intend to grasp the viewer’s attention with meaning construction to either “energize or calm people” (Kress & van Leeuwen, 2002, p. 349). The former shows a higher modality in terms of color saturation, color differentiation, and color modulation with red and black, while the latter displays tranquility with the color blue. Pictures of higher modality usually have a visual impact on the audience, which in turn influences their emotions. The two images shown here serve different purposes: the red Chinese knot, as an auspicious cultural symbol, adds a festive feel of excitement; by contrast, image 21 depicts a different facet of London in peace at dusk, quite different from its bustling life in daytime.

Interestingly, a message of hope is conveyed similarly in the two films by making the background prominent in contextualization. In image 22, the male as a represented
participant from the Beijing Publicity Film highlights the background, especially the sun rising from the mountains, which symbolizes hope in the citizens’ hearts. Likewise, in image 23, the right light shining as a represented participant denotes hope in London as well.

A close examination of the two films reveals that the Beijing film plays around a higher modality of color by using bright colors such as red, yellow, and orange, in contrast to that of the LPF with its dominant colors of blue and gray due to time-lapse cinematography featuring fast-moving images to create the illusion of motion. Moreover, the BPF uses more abundant contextualization, suggesting that the Chinese Publicity Film has more substantial content and well-made details.

To summarize, in accordance with the analysis of representational meanings in the BPF, Beijing not only celebrates its rich cultural heritage as a historic city by displaying iconic buildings and cultural symbols, but also promotes itself as a regional and global hub of business and economy with images of skyscrapers in financial centers and of developed transportation systems. This approach can be conducive to representing Beijing as a mega-city characterized by tradition and modernity. For the interactive meaning realized in the BPF, Beijing also endeavors to use an emotional appeal to elicit viewers’ feedback through a gaze or more personal relationships.

It is claimed that, similar to product brands, functional and emotional needs are two constituents that a city could satisfy (Morgan, 1999), but “the attributes that satisfy those needs need to be orchestrated into the city’s unique proposition” (Kavaratzis, 2004, p. 66). The last decade has witnessed a shift from strategies of city branding underlining functional attractions such as infrastructure, local services, and natural landscapes, to promoting emotional elements such as culture, people, and lifestyle (De Chernatony, 2006; Fetscherin, 2010) to sell the distinctive aura of a tourist destination. In this sense, the BPF seems to incorporate both functional and emotional values. Nevertheless, the LPF primarily accentuates sightseeing experiences by presenting scenic spots, which tend to be tinted with a more functional value. Time-lapse cinematography narrates London’s stories via fast-moving scenes, depicting the vicissitudes of this city in a diachronic sense.

3.3 Why similar and different?

3.3.1 Convergence: Shared strategy of emotional branding

Human cognition, to a great extent, relies on the experience that derives from having
a body with different sensorimotor capacities (Rosch, 1991) and thus being shaped by recurring bodily experience or embodiment (Gibbs, 2006). This phenomenon may explain why people’s understanding of the world is sometimes universal, regardless of different cultural contexts, since people across cultures share basic cognitive processes as demonstrated in similar ways of categorization, causal analysis, and inference (Nisbett, 2003).

As we know, tourist destinations seek to promote themselves as places that are friendly, enjoyable, and unique to compete with other destinations for a limited pool of tourism. Publicity films can be a significant resource and visual gateway to attract potential visitors, especially those who are geographically remote, to acquire the information they may need and, in the meantime, boost a city’s image. In verbal discourse, the frequent use of mental and relational processes regarding transitivity succeeds in eliciting positive affective emotional responses by virtue of describing a narrator’s experience with the city and depicting the city’s charm. Furthermore, declaratives in mood as a means of offering information facilitate the persuasive function. In visual discourse, more numerous images also accord with the ultimate purpose of a publicity film. Moreover, a large number of eye-level angle shots is used to create a sense of equality for the target viewers to evoke positive feelings of respect. All of these approaches serve the communicative function of a promotional film, but also and above all the “emotional branding” purpose (Morgan, 2004). Emotional branding derives from the idea that it is the emotional aspect—for example, positive feelings aroused or the personality or identity of tourist destinations—that distinguishes them (Blain, 2005; Ekinci, 2006). Thus, to establish emotional ties with visitors, thereby influencing potential tourist’ decisions, becomes critical (Mohammed et al., 2018). It thus becomes apparent that both of the films studied here resort to mental and relational processes to evoke viewers’ emotional response by defining cities’ uniqueness, especially when homogenization—“the practice of selling places may even generate sameness, blandness” (Philo & Kearns, 1993, p. 21)—has long been recognized in tourism.

3.3.2 Divergence: Cultural attributes
Culture is defined as “the collective programming of the mind that distinguishes the members of one group or category of people from another” (Hofstede, 2005, p. 39). It is argued that culture is of vital importance in recognizing a specific cultural
background in which different psychological cognitive styles are embedded (Shweder, 1991) to understand the cultural variations observed. Easterners tend to view the world in a more “holistic” way, whereas people from the west are more “analytic” (Nisbett, 2003, p. 76). Briefly explained, the holistic way of thinking is characterized by an emphasis on context and the situations in which people, objects, and relationships between objects and people are involved. By contrast, western people have a preference for salient objects and a rule-based categorization of objects, focusing on “a single dimension or aspect and a tendency to disentangle phenomena from the contexts in which they are embedded” (Grossmann & Na, 2014, p. 8). This difference could explain why, in terms of what was chosen to be filmed, the BPF tends to group people, objects, and the natural environment together to create harmony, whereas in the LPF, people are considered an insignificant element, while inanimate objects are regarded as more salient.

Hofstede’s cultural typology is one of the most valid differentiators that describe the interplay between the influence of culture on a society and society members’ behaviors (Bochner, 1994; Simon, 2000; Grossmann & Kross, 2010). The Individualism and Collectivism (IDV) index defines the relations between members and the groups to which they belong. An individualistic society construes individuals as distinguishable from their social affiliations and as separate entities, thus aligning them with fulfilling personal needs. However, collectivism blurs the distinction between individual and group, focusing on mutual dependence, the interests of the group, and relationship-building as hallmarks (Hofstede, 2005). Thus, the Chinese culture, which falls under that category of collectivism, is characterized by an emphasis on relationships with a group and on cherishing tradition and cultural heritage that bind members as a group together. However, British culture, which is characterized by individualism, is more task-oriented rather than relation-driven, and the “other world” brought about by detachment aligns with their individual preferences as independent persons. Therefore, in verbal discourse, preferences for the median and low value of modality in the LPF compared with the high and median value of modality detected in the BPF denote that the British director is apt to make the information offered more suggestive, leaving space for individual choices. In addition, in visual discourse, eye contact and more frequent close-up shots intending to build up intimate relationships with viewers are more frequent in the BPF, echoing an “in-groupers” collectivist preference. Moreover, cultural icons such as tea culture,
Chinese knots, opera, etc., as distinctive cultural symbols are in full display in the BPF, strengthening the ties of Chinese cultural identity. It is noteworthy that though “the rate at which particular personal pronouns are used in a language, for example, use of first-person pronouns index individualism,” which asserts one’s viewpoint (Uz, 2014, p. 1671), and this must be carefully interpreted in context. As demonstrated in this case under discussion, the narrator in the BPF definitely prefers the first-person pronoun, the purpose of which is to exert influence on viewers’ perceptions and emotions by way of conveying his intense feelings about Beijing. This approach forms a striking contrast with the narrator’s choice of the second-person pronoun “you” in the LPF, which highlights feelings being brought to audiences, and as a matter of fact signals that collectivist culture believes that personal feelings have the power to influence others with information credibility if a relationship is solid (Lau, et al., 2004).

The avoidance of uncertainty, as another cultural dimension, is defined as “the extent to which the members of a culture feel threatened by the ambiguous or unknown” (Hofstede, 2005, p. 86). Compared with Britain, China scores lower in this index (Hofstede, 2005) and thus falls into the lower degree of the avoidance of uncertainty. Chinese people tolerate uncertainty and ambiguity. That phenomenon is why the LPF is likely to express itself directly by deploying more frontal angle shots, to be straightforward and explicit, while the BPF tends to show itself much more implicitly using both front and oblique angles as a safer way of transmitting information. It is worth noting that front angle shots suggest “involvement” (Kress & van Leeuwen, 2006), but as discussed earlier, the LPF merely intends to depict inanimate objects such as skyscrapers, shopping centers, or landscapes in an impressive and persuasive manner by filming them from front angles to make the film more functional.

Taken together, we understand that, aside from shared bodily experiences coming from an interaction with the world that makes people develop similar cognitive mechanisms, cultural attributes do have a substantial role to play in shaping different nations’ perceptions of the world. In this sense, our findings are consistent with those studies that recognized cultural mediators in carriers of meaning realized through verbal and visual semiotic resources (Baldry, 2000; Baldry & Thibault, 2006; Hagan, 2007; Matthiessen, 2007). There is no doubt that differences in cinematography and aesthetics between the East and the West should also be taken into consideration.
4. Conclusion

The main contribution of this study illuminates the similar attributes shared by cities’ promotional films and how these films differ in their multimodal branding discourse for constructing a city’s image by drawing on linguistic analysis and providing a new lens to unmask preferences of cities in narrating their stories within a cultural context. In verbal and visual discourse, both the BPF and LPF share similarities in terms of fulfilling the communicative function of a promotional film whose persuasiveness would be greatly enhanced via emotional branding. However, these films exhibit differences regarding how to achieve persuasiveness through different semiotic resources that co-construct meaning. Based on the different layers of analyses in the multimodal branding discourse of these two films, it can be seen that the Beijing publicity film blends functional and emotional values, while the London publicity film is prone to be more functional. Nevertheless, due to space limitations, this paper fails to examine the textual and compositional meanings and collaborations of the different modalities in the meaning construction of the discourse. We envisage that future work on these aspects may provide new insights into the interplay between cultural attributes and the deployment of semiotics in city branding discourse across cultures.

Faced with regional, national, and global competition, Beijing and London, as two mega-cities in the East and the West, both seek new ways to promote themselves to galvanize economic development and to attract tourism, investment, and talent. In promotional films, Beijing is framed as a city with tradition and modernity, whereas London is branding itself as a city of a modern and cosmopolitan character. There is no denying that a publicity film, as one of the city branding channels, can serve to communicate the city’s competitive advantage on the one hand and to set cities apart from the fierce global competition on the other hand, based on the construction of a symbolic image that can be transmitted both locally and internationally. Therefore, as the sum of promotional and communicative activities, cities’ publicity film-makers should ask questions such as “What is this city endowed with?”, “What would make this city different?”, and “Who is the intended audience of this film?” to retain a unique and distinguishable city image to entice visitors or investors. In this sense, strategies for the best trajectory to distinguish a city’s characteristics are vital in
communicating a city’s image and identity. More importantly, city branding should be disseminated online, taking full advantage of social media in particular as a more effective, low-cost platform, thus enhancing “world-citiness” in global space and networked time.

Note
1 This study is supported by the Fundamental Research Funds for the Central Universities of NWPU 2019, the Teaching Reform Project of NWPU 2018, and the Higher Education Research Fund of NWPU (G2018KY0207).

References
Boland, P. (2013). Sexing up the city in the international beauty contest: The performative nature of spatial planning and the fictive spectacle of place branding. The Town Planning Review, 84, 251-274.


Tourism Management, 30, 857-66.

(Copy editing: Curtis Harrison)

About the authors
Qian Wang (wangq@nwpu.edu.cn) is an associate professor of English at the School of Foreign Studies, Northwestern Polytechnical University. She is currently a doctoral candidate in the Faculty of Humanities, The Hong Kong Polytechnic University. Her research interests include academic discourse, cognitive linguistics and language
education.

Dan Zhang (180533@hainu.edu.cn) is a lecturer at the Institute of Tropical Agriculture and Forestry, Hainan University. She is currently a doctoral candidate in the Faculty of Humanities, The Hong Kong Polytechnic University. Her research interests include critical discourse analysis, media discourse and corpus linguistics.