Corpus-based study of hotel overviews on official websites: Language strategies for self-promotion

KONDO, Yukie

(Ritsumeikan University)

1. Introduction

Websites are an important platform for firms to communicate its service assurances with their consumers (Chen & Dhillon, 2003). Hotels are no exception and have their own official websites, where they publish official information and messages, as well as showcasing their features to attract potential guests. Although third-party travel agencies are a strong force in the market for finding and booking hotels, the latest research by the American Hotel & Lodging Association finds that 39% of travelers most often book their hotel rooms through a third-party website, and 29% of consumers report that they have been misled by third-party traveler resellers (The American Hotel & Lodging Association, 2018). Considering this fact, as Som and Blanckaert (2015) argue, official websites can serve as "a primary source of authentic information" (p. 126) and also a place for hotels to convey messages directly to consumers. In the case of English-language hotel websites, most websites have an overview section on the top-level page, where hotels briefly introduce themselves. *Figure 1* shows the top-level page of a 5-star hotel's official website; an overview section can be found below a large photo. Hotels are categorized into different grades by stars and accordingly differ in how they introduce and present themselves to potential guests depending on their grade. However, there have been no major empirical studies analyzing the overview texts of hotel websites to determine the language strategies hotels use to effectively communicate with potential guests. This study, therefore, explores the characteristics of the overview texts on the websites and the ways they differ according to hotel grades. The author complied a corpus of hotel overview texts taken from websites of hotels located in London; the analysis in this study is based on this corpus. Because websites are not only for advertising and marketing (McNeil & Riello, 2016) but are also the faces of the brand in digital space (Som & Blanckaert, 2015), analyzing the top-level overview texts will contribute to a better understanding of finding hotels' language strategies for promoting themselves to their potential guests on the web.

Figure 1. Example top-level page with an overview section from the website
          of a 5-star hotel

2. Corpus

The corpus for this study is composed of the texts of 124 overviews on the official websites of 3-star to 5-star hotels in London[1,2], collected between March and April, 2017. London was chosen because it was one of the largest cities where hotels in different grades were located, from affordable 3-star hotels to luxury 5-star hotels. Overview texts were defined as those texts containing general remarks about the hotel that were usually posted on the top-level page as shown in *Figure 1*. Titles or headings were also analyzed, if present. Table 1 shows the number of hotels and basic statistics of the tokens according to each grade. A hotel overview contained its hotel name at least once or more than once in most cases. As hotel names might affect the results of analyses, all hotel names in the corpus were replaced with "HOTELNAME."

Table 1

Number of hotels and basic statistics of the tokens in the corpus

Hotel grade	Number of hotels	Tokens	Average tokens	Minimum tokens	Maximum tokens
5-star	47	6289	132.55	21	348
4-star	66	10521	158.21	29	458
3-star	11	2414	219.45	31	659

3. Methods

Keywords for each grade were derived according to the value of log-likelihood using CasualConc[3] to examine the characteristics of words used in each grade. The BE06 Corpus[4] was used as a reference corpus. Co-occurrence network analysis and correspondence analysis were then conducted using KH Coder[5] to see how words were related to one another, as well as to each grade.

The co-occurrence network command of KH Coder creates a diagram that shows "the words with similar appearance patterns, i.e., with high degrees of co-occurrence, connected by lines (edges)" (Higuchi, 2017, p. 50). This diagram also shows "the association between words and variables/headings, in addition to the associations between words" (Higuchi, 2017, p. 50). Hotel grades were used as variables to examine whether words were used more exclusively in a certain grade or were more commonly used at different grades. To avoid the inclusion of too many words in the diagram, top sixty edges with strong Jaccard similarity coefficients and also words that occurred ten or more times in the corpus were used.

Correspondence analysis is "a method of data visualization that is applicable to cross-tabular data" (Greenacre, 2010, p. 613). It takes "the frequency of co-occurring features and converts them to distances" (Glynn, 2010, p. 1). Words are then plotted, revealing their relationship by their relative distance from each other (Glynn, 2010). The correspondence analysis command of KH Coder creates "a two-dimensional scatter diagram" that shows words with "a similar appearance pattern" (Higuchi, 2017, p. 44). Hotel grades were again used as valuables for this analysis. To avoid the inclusion of too many words in the diagram, top fifty words as filtered according to their chi-square value[6] that also occurred ten or more times in the corpus were used.

For both co-occurrence network analysis and correspondence analysis, words were drawn in a circle. The more frequently a word occurs, the larger the circle is. Numbers in the squares represent hotel grades.

4. Results

4.1 Keywords

Keywords are shown in Table 2. The top three keywords for both the 5-star and the 4-star hotels included "HOTELNAME," "hotel," and "London;" the top keywords for the 3-star hotels, however, included "Wembley" before "HOTELNAME." "Stadium" also came before "London," followed by "arena." Some of the 3-star hotels emphasized their location near "Wembley Stadium" and "Wembley Arena," which resulted in the prevalence of the keywords "stadium" and "arena." It is evident that while for the 5-star and the 4-star hotels, their hotel names and the hotels themselves were important elements, for the 3-star hotels, the landmarks near the hotels were perhaps more important than the hotels as such. The fourth-ranked keyword for the 5-star hotels was "luxury," and this was not included among the 4-star and the 3-star hotels' high-ranked keywords. Words keyed to more specific locations than simply "London" appeared for all hotel grades; "Mayfair" was common for the 5-star hotels, 'Kensington" for the 4-star hotels, and "Wembley," "stadium," and "arena" for the 3-star hotels. The keyword "locate" only appeared for the 4-star and the 3-star hotels, which suggests that location is a more important element for hotels in these grades. In short, higher-grade hotels tended to emphasize their grade and facilities (e.g., 5-star, 4-star, suite, and restaurant), while lower-grade hotels tended to place greater importance on their location.

Table 2

Keywords according to hotel grades

	5-star		4-star		3-star	
	Keywords	Log-likelihood	Keywords	Log-likelihood	Keywords	Log-likelihood
1	hotelname	1249.06	hotelname	1291.28	wembley	722.24
2	hotel	727.20	hotel	1161.94	hotelname	450.88
3	london	533.44	london	810.84	hotel	294.45
4	luxury	521.98	restaurant	325.93	stadium	210.05
5	suite	359.86	guest	274.53	london	180.30
6	mayfair	270.10	locate	244.41	arena	108.06
7	5-star	227.10	room	243.87	our	101.24
8	guest	195.30	kensington	233.00	locate	83.89
9	spa	164.99	our	196.36	guest	80.10
10	room	140.37	4-star	190.64	room	73.65

4.2 Co-occurrence network

Figure 2 shows the co-occurrence network diagram of the words used in hotel overviews. Thirty-nine words and sixty edges were drawn in the diagram, which revealed that there were clusters of words particular to a certain grade and also clusters that were associated with different grades.
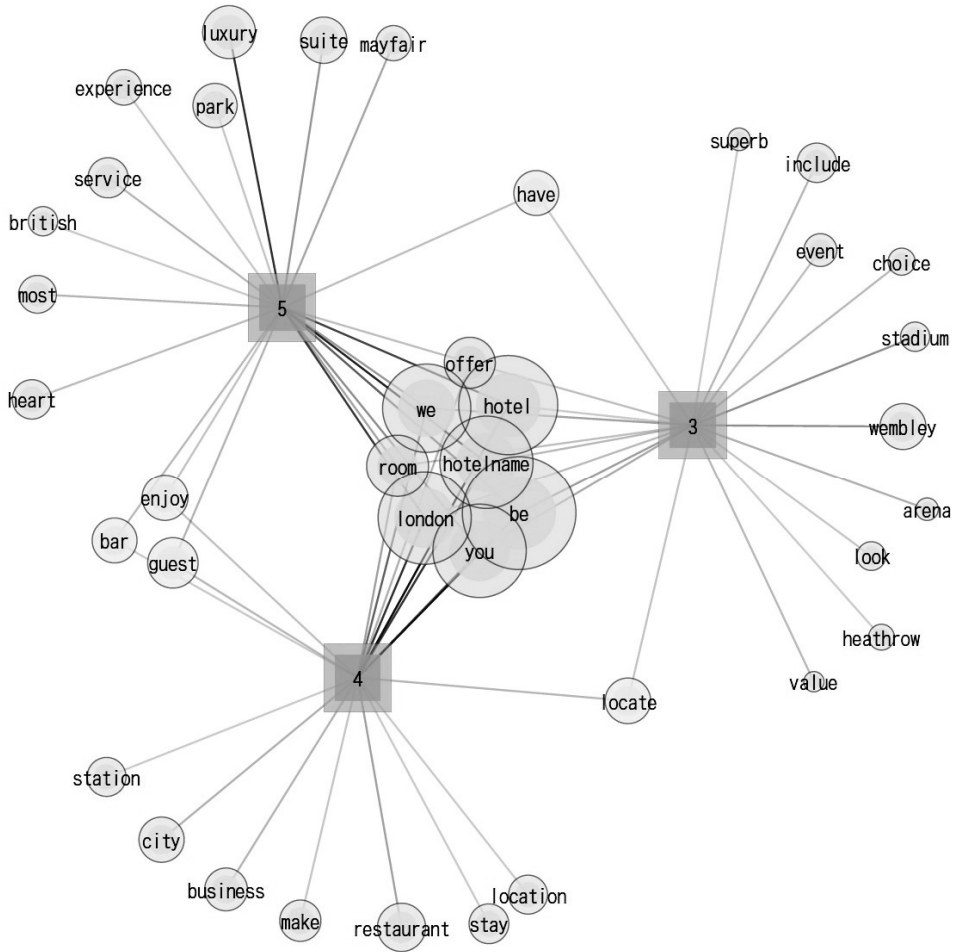

Figure 2. Co-occurrence network diagram of the words used in hotel overviews

The eight circles located in the middle are connected to all the grades with edges because they were used similarly in all the grades. The words in the circles include "HOTELNAME," "be," "London," "offer," "room," "you," "we," and "hotel." These words are considered to be the fundamental elements for London hotel overviews. Putting these words together creates a sentence, such as (1), containing the fundamental information that hotel overviews express.

(1)   We/HOTELNAME offer you hotel/room [in] London.

Of course, this is rather an intuitive conclusion, and the actual sentences used in hotel overviews were wider in variety, for example, in terms of what they offer, as in (2), (3), and (4). These are example sentences in which the word "offer" occurred. In (2) and (3), a particular service (free cancellation) or features (guestrooms and event spaces) are "offered," while in (4), an abstract expression (a fashionable and timeless base) is used.

(2)   On many of our rooms we offer free cancellation, so a change of plan is no problem as your booking can be cancelled or amended free of charge. (3-star)
(3)   […] and 4-star deluxe HOTELNAME offer a total of 281 guest rooms and various versatile event spaces, featuring state-of-the-art facilities and stylish decor. (4-star)
(4)   With world-famous restaurants and a stunning spa, we offer a fashionable and timeless base in the centre of the British capital. (5-star)

The common words between the 3-star and the 4-star hotels included "locate," while those common between the 4-star and the 5-star were "enjoy," "guest," and "bar." Thus, "locate" appears to be characteristic of lower-grade hotels, while the latter, of higher-grade hotels. This supports the idea that lower-grade hotels appeal to their guests by mentioning the location. Circles that are only connected to the 3-star hotels included specific destinations, such as "Wembly," "Heathrow," and "arena." They are places that can be conveniently reached from/to the hotel. On the contrary, higher-grade hotels appear to appeal to their guests through emphasizing the hotel's facilities, such as the bar.

"Have" is the only words connected to the 3-star and also the 5-star hotels. (5) and (6) are the example sentences using the word "have."
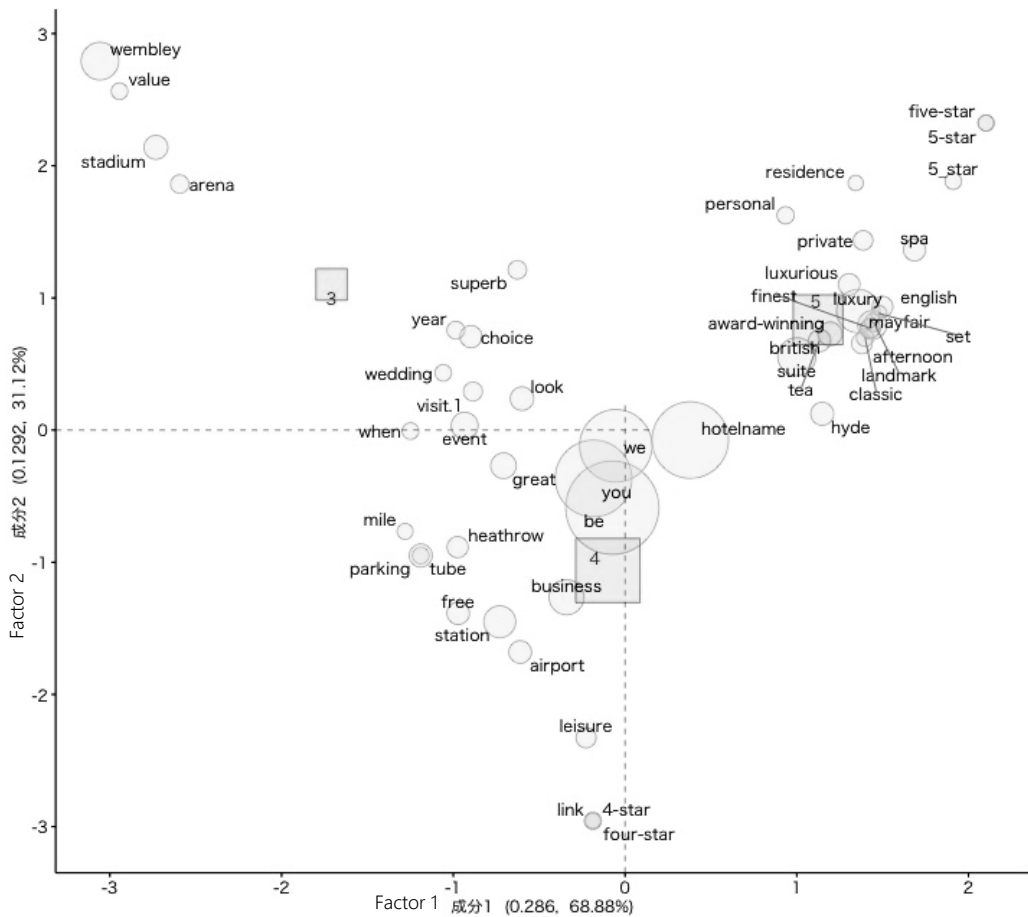
(5)   …our newly renovated guest suites and 5-star amenities will ensure that you have everything you need for a one-of-a-kind experience. (5-star)
(6)   Whether you're here on business or pleasure we have all of the necessary amenities to provide you an enjoyable experience. (3-star)

Both of these sentences from the 3-star and the 5-star overviews mention what is needed/necessary; however, the sentence for the 5-star hotel stimulates the reader's imagination, using open-ended, non-specific phrases such as "everything you need" or "one-of-a-kind experience," which of course will vary depending on who reads it. The sentence from the 3-star hotel, on the contrary, while referring to necessary amenities,

also reminds readers of specific items or activities, such as business or pleasure.

4.3 Correspondence analysis

The scatter diagram drawn by correspondence analysis is shown in *Figure 3*. Factor 1 on the horizontal axis accounts for 68.88% of the variance, and factor 2 on the vertical axis, for 31.12%. Factor 1 is the primary factor explaining the characteristics of the corpus, and from the left, words for the 3-star, 4-star, and 5-star overviews are observed. This suggests that the texts do exhibit tendencies according to hotel grades.

Figure 3. Scatter diagram of the correspondence analysis for words used in hotel overviews

Factor 1 can be considered as the quality that hotels have. At the extreme for the 5-star and the 4-star hotels, specific words indicating their star rating (e.g., five-star, four-star) are observed. For the 5-star hotels, in particular, there are also clusters of words indicating qualities such as "luxury," "luxurious," "award-winning," "finest,"
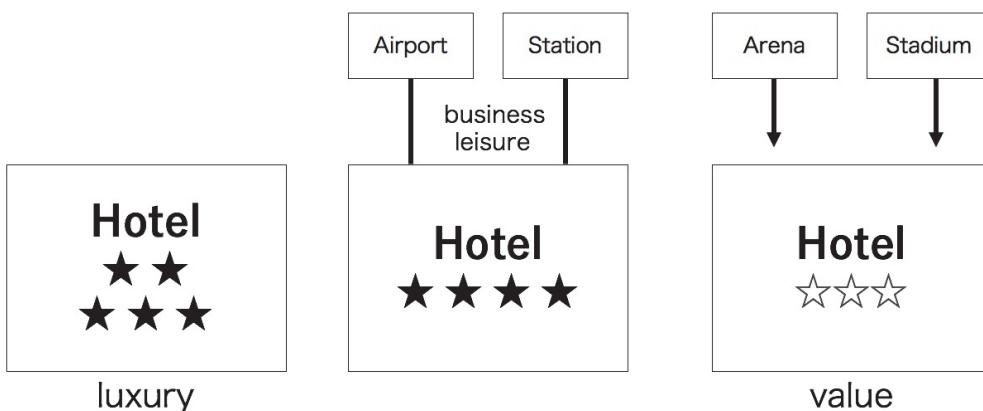
"personal," and "private." For the 3-star hotels, there are no words indicating the hotel grade, but the word "value" does appear.

Factor 2 can be considered as the features using which hotels promote themselves to their potential guests. Clustered around the 5-star hotels are words referring to luxury features such as "spa" and "suite." Surrounding the 4-star hotels are words indicating location and transportation such as "link," "station," and "airport" as well as words referring to purposes such as "leisure" and "business." Between the 4-star and the 3-star hotels are words indicating location such as "parking," "tube," and "Heathrow." Those closer to the 3-star area include facilities convenient from/to the hotel such as "[Wembley] stadium" and "[Wembley] arena."

In summary, the words associated with the 5-star and the 3-star hotels had the most distance between them, while those associated with the 4-star hotels fell in the middle. For the 4-star and the 5-star hotels, the hotel grade was an important quality for marketing, while for the 3-star hotels, "value" took that role. For words associated with features, the 5-star and the 3-star hotels were closer together, while the 4-star hotels were distant. Facilities were an important factor for the 3-star and the 5-star hotels, but there was a difference: the 5-star hotels emphasized their own facilities, while the 3-star hotels mentioned the convenient facilities outside the hotel. For the 4-star hotels, facilities outside the hotels such as an airport and a station were important, but the focus was more on transportation.

5. Discussion

Hotel overviews briefly introduce the hotel on the top-level page and reach out to potential guests. Interestingly, the 5-star hotels describe themselves using words related to quality and mentioning their own facilities, while the 3-star hotels make an appeal by mentioning facilities convenient from/to the hotel. The 4-star hotels are in

Figure 4. Diagram showing how hotels promote themselves in overview texts

between them and mention what guests can do or where guests can go. The relationship between the different hotel grades and basic keywords are shown as a diagram in *Figure 4*.

The characteristics found from these corpus analyses indicate that the language strategies to promote a hotel on the websites differ depending on the hotel grades. Higher-grade hotels subjectively promote themselves by referring to their own facilities and qualities, while lower-grade hotels objectively promote themselves through reference to other entities; that is, by mentioning facilities outside the hotel or how it is convenient to get to the hotel through a stated means of transportation.

This method of self-promotion using social references can be applied to the "looking-glass self" proposed by Cooley (1902). Cooley (1902) observed that the looking-glass self, or social self, had the following three elements: first, "the imagination of our appearance to the other person," second, "the imagination of his judgment of that appearance," and third, "some sort of self-feeling, such as pride or mortification" (p. 189). Applying this to hotel overviews, hotels also imagine how guests view the hotels and imagine what the guests expect the hotels have to offer. Lower-grade hotels tend to fulfill the first two elements by describing their amenities as people might reasonably expect to find them. Higher-grade hotels, meanwhile, also satisfy the third element by imagining the judgment and expressing pride through use of elite words to describe themselves such as "luxury," "luxurious," "award-winning," "finest," "personal," or "private."

Extending Cooley's theory, Mead (1934) proposed "I" and "me" perspectives of the social self. According to these perspectives, "me" stands for "the organized set of attitudes of others which one himself assumes" (p. 258), whereas "I" stands for "something responding to a social situation" (p. 261) and "the answer that the individual makes to the attitude that others take toward him when he assumes an attitude toward them" (p. 261). If one is applied to the hotel overviews, lower-grade hotels put themselves in society and talk about themselves by indicating where they are socially and geographically located from the "me" perspective, while higher-grade hotels focus on expectations and respond to them from the "I" perspective.

Below is an example overview of a 5-star hotel, which is also shown in *Figure 1*.

Brown's Hotel is not only London's first hotel, but also one of the most illustrious in the British capital. Since opening on Albemarle Street in 1837, our distinguished hotel has hosted esteemed guests, from royals and presidents to world-famous authors. Add your name to the guest book and revel in the history of this iconic London hotel. Here, you can unwind in our serene spa, dine on award-winning cuisine and have a nightcap in the stylish Donovan Bar.

(5-star hotel, see also *Figure 1*.)

This overview uses social references to its history, location, distinguished visitors, and facilities such as the spa and the bar, and, at the same time, the description proudly responds to them. The hotel describes itself as "London's first" and also "one of the most illustrious." It boasts that "royals and presidents" and "world-famous authors" have stayed there, and therefore, it is a "distinguished" and "iconic" hotel. It has unique facilities such as a "serene spa" and the "Donovan Bar," which make it a place for guests to "unwind." This shows one way that a hotel can judge itself from the "me" perspective using social references, and simultaneously responds to the judgment from the "I" perspective. This appears to be a common strategy used by higher-grade hotels when introducing and promoting themselves in the overview text.

6. Conclusion

This study investigated how hotels use language strategically to promote themselves on their official websites. Three corpus analyses—keywords, co-occurrence network analysis, and correspondence analysis—demonstrated how the words adopted by each hotel grade varied. These differences were further discussed in light of Cooley's "looking-glass self" theory (1902) and Mead's "I" and "me" perspectives (1934). As a result, it was suggested that luxury hotels tend to speak from the "I" perspective responding to social expectations, while lower-grade hotels, the "me" perspective using social references to express themselves. Although these socio-psychological concepts were originally developed to discuss the development of children's selfhood and might not necessarily match how hotels express themselves on the web; they do shed some light on the hotels' language strategies. Further research needs to be carried out to investigate whether there are other factors than hotel grades that affect the language strategies. Nevertheless, the results obtained from this study will be meaningful for the luxury industries to learn how to promote themselves to their potential customers in a limited space on their websites. As the corpus used for this study was small, further studies with larger corpora are necessary to investigate whether the results obtained in this study can be generalized. It would also be beneficial to analyze similar texts from the websites in other luxury industries to investigate if there are any similarities in their language strategies.

Notes

[1] The AA Hotel Guide 2016 (AA Publishing, 2015) was used to find hotels according to hotel grades.

[2] This study did not use 1-star and 2-star hotels as the number of hotels was limited: zero hotel for 1-star and one hotel for 2-star. The number of hotels ranked from 3-star to 5- star was 128, but 4 hotels were excluded because their websites did not have a clear overview section.

[3] CasualConc is a freeware concordance program for Macintosh OS X created by Dr. Yasuhiro Imao of Osaka University. The program can be downloaded from https://sites.google.com/site/casualconc/

[4] The BE06 Corpus is "a one-million-word reference corpus of general written British English that was designed to be comparable to the Brown family of corpora" (Baker, 2009). https://doi.org/10.1075/ijcl.14.3.02bak

[5] KH Coder is a free software for quantitative content analysis or text mining created by Dr. Koichi Higuchi of Ritsumeikan University. This study used KH Coder Version. 3 Alpha. The software can be downloaded from http://khc.sourceforge.net/en/

[6] This can be done by selecting the "Filter words by chi-square value" option.

Acknowledgments

References

Chen, S. C., & Dhillon, G. S. (2003). Interpreting dimensions of consumer trust in E-commerce. *Information Technology and Management*, *4*, 303–318. http://doi.org/10.1023/a:1022962631249

Cooley, C. H. (1902). *Human nature and the social order*. New York: C. Scribner's Sons.

Glynn, D. (2010). Corpus-driven cognitive semantics introduction to the field. In D. Glynn & K. Fisher (Eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches.* (pp. 1–41). Berlin; New York: De Gruyter Mouton.

Greenacre, M. J. (2010). Correspondence analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(5), 613–619. http://doi.org/10.1002/wics.114

Higuchi, K. (2017). *KH Coder 3 reference manual*. Retrieved May 1, 2018, from http://khc. sourceforge.net/en

Mead, G. H. (1934). *Mind, self and society*. Chicago: University of Chicago Press.

McNeil, P., & Riello, G. (2016). *Luxury: A rich history*. Oxford: Oxford University Press.

Som, A., & Blanckaert, C. (2015). *The road to luxury: The evolution, markets, and strategies of luxury brand management*. Singapore: John Wiley & Sons.

The American Hotel & Lodging Association. (2018, May 3). New research shows consumers spend $5.2 billion annually in fraudulent and misleading hotel booking transactions. Retrieved May 5, 2018, from https://www.ahla.com/press-release/new-research-shows-consumers-spend-52-billion-annually-fraudulent-and-misleading