



Differences in Chinese and Western tourists faced with Japanese hospitality: a natural language processing approach

Elisa Claire Alemán Carreón¹  · Hugo Alberto Mendoza España¹ · Hirofumi Nonaka¹ · Toru Hiraoka²

Received: 24 October 2019 / Revised: 16 April 2021 / Accepted: 17 May 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Since culture influences expectations, perceptions, and satisfaction, a cross-culture study is necessary to understand the differences between Japan's biggest tourist populations, Chinese and Western tourists. However, with ever-increasing customer populations, this is hard to accomplish without extensive customer base studies. There is a need for an automated method for identifying these expectations at a large scale. For this, we used a data-driven approach to our analysis. Our study analyzed their satisfaction factors comparing soft attributes, such as service, with hard attributes, such as location and facilities, and studied different price ranges. We collected hotel reviews and extracted keywords to classify the sentiment of sentences with an SVC. We then used dependency parsing and part-of-speech tagging to extract nouns tied to positive adjectives. We found that Chinese tourists consider room quality more than hospitality, whereas Westerners are delighted more by staff behavior. Furthermore, the lack of a Chinese-friendly environment for Chinese customers and cigarette smell for Western ones can be disappointing factors of their stay. As one of the first studies in the tourism field to use the high-standard Japanese hospitality environment for this analysis, our cross-cultural study contributes to both the theoretical understanding of satisfaction and suggests practical applications and strategies for hotel managers.

Keywords Sentiment analysis · Hotels and lodging · Text mining · Chinese · English · Satisfaction and dissatisfaction factors

✉ Elisa Claire Alemán Carreón
elisa.claire.aleman.carreon@gmail.com

Extended author information available on the last page of the article

1 Introduction

Inbound international tourism has been increasingly affecting Japanese economy (Jones et al. 2009). A year-on-year growth rate of 19.3% was observed in 2017, with 28,691,073 inbound tourists (Japan National Tourism Organization 2019).

Japan's hospitality has been known historically to be of the highest quality. *Omotenashi*, which describes the spirit of Japanese hospitality, with roots in Japanese history and tea ceremony, is celebrated worldwide (Al-alsheikh and Sato 2015). Consequently, it would stand to reason that tourists visiting Japan would have this hospitality as their first and foremost satisfaction factor. However, it is known that customers from different countries and cultures have different expectations (Engel et al. 1990). Thus, it could be theorized that their satisfaction factors should be different.

The Japanese tourist market is gradually becoming diverse because of multi-cultural tourist populations. This diversity means that the expectations when staying at a hotel will be varied. Cultural backgrounds have a decisive role in aspects of satisfaction and in the perceptions of quality (Mattila 1999; Winsted 1997), or behavioral intentions (Liu et al. 2001), such as the difference in Westerners and Asians in their willingness to pay more (Levy 2010). A difference in cultural background can also heavily influence customers' expectations, as well as their perceptions of quality, and the difference between these two is what expresses itself as satisfaction. This difference in expectations and perceptions of quality can be smaller or larger depending on the culture in reaction to the same service.

For a growing industry with increasing cultural diversity, it is essential to identify the cross-culture expectations of customers in order to provide the appropriate services, cater to these expectations to ensure and increase customer satisfaction, maintain a good reputation, and generate positive word-of-mouth.

In 2017, Chinese tourists accounted for 25.63% of the tourist population. On the other hand, Western countries accounted for 11.4% of the total, and 7.23% were countries where English is the official or the de facto national language (Japan National Tourism Organization 2019). The effect of Chinese tourists on international economies is increasing, along with the number of studies on this phenomenon, Sun et al. (2017). Despite this, many tourist-behavior analyses have been performed only involving Western subjects. Yet, it is known that Western and Asian customers are heavily differentiated (Levy 2010). As such, a knowledge gap existed until recent decades. Considering the numbers of inbound tourists in Japan and our team's language capabilities, our study focuses on Western and Chinese tourists.

In studies involving Asian populations in the analysis, Chinese-tourist behaviors have been evaluated most commonly (e.g. Liu et al. 2019; Chang et al. 2010; Dongyang et al. 2015). The few studies reporting comparisons between Asian and Western tourists' behaviors (e.g. Choi and Chu 2000) are typically survey- or interview-based, using small samples. These studies, although valid, can have limitations, namely, the scale and sampling. In the past, survey-based studies have provided a theoretical background for a few specific tourist populations of

a single culture or traveling with a single purpose. These studies' limited scope often leads to difficulties in observing cultural and language differences in a single study. This creates a need for large-scale cross-cultural studies for the increasing Asian and Western tourist populations. It could be said that Westerners account for a smaller portion of the tourist population compared to Asians. However, according to Choi and Chu (2000), Westerners are known as "long-haul" customers, spending more than 45% of their budget on hotels. In comparison, their Asian counterparts only spend 25% of their budget on hotels. Therefore, it is essential to study Asian and Western tourist populations, their differences, and the contrast with the existing literature results.

However, with ever-increasing customer populations, this is hard to accomplish without extensive studies of the customer base. There is a need for an automated method for identifying these expectations at a large scale. Our study intends to answer the need for such a methodology utilizing machine learning and natural language processing of large amounts of data. For this, we used a data-driven approach to our analysis, taking advantage of hotel review data. With this methodology, we explore the expectations and needs for the two most differing cultures currently interacting with the hospitality industry in Japan.

Owing to the advent of Web 2.0 and customer review websites, researchers realized the benefits of online reviews for research, sales (Ye et al. 2009; Basuroy et al. 2003), customer consideration (Vermeulen and Seegers 2009) and perception of services and products (Browning et al. 2013), among other effects of online interactions between customers (e.g. Xiang and Gretzel 2010; Ren and Hong 2019). Consequently, information collected online is being used in tourism research for data mining analysis, such as opinion mining (e.g. Hu et al. 2017), predicting hotel demand from online traffic (Yang et al. 2014), recommender systems (e.g. Loh et al. 2003), and more. Data mining and machine learning technologies can increase the number of manageable samples in a study from hundreds to hundreds of thousands. These technologies can not only help confirm existing theories but also lead to finding new patterns and to knowledge discovery (Fayyad et al. 1996).

In this study, we evaluate the satisfaction factors of two essential tourist populations that are culturally different from Japan: Chinese and Western tourists. We take advantage of the wide availability of online reviews of Japanese hotels by both Mainland Chinese tourists posting on *Ctrip* and Western, English-speaking tourists posting on *TripAdvisor*. Based on these data, we can confirm existing theories regarding the differences in tourists' behavior and discover factors that could have been overlooked in the past. We use machine learning to automatically classify sentences in the online reviews as positive or negative opinions on the hotel. We then perform a statistical extraction of the topics that most concern the customers of each population.

2 Research objective

With the knowledge that cultural background influences expectations in customers, which is the basis for satisfaction, it becomes important to know the difference in factors influencing satisfaction and dissatisfaction between the most differing and numerous tourist populations in a given area.

This study aims to determine the difference in factors influencing satisfaction and dissatisfaction between Chinese and English-speaking tourists in the context of high-grade hospitality of Japanese hotels across several price ranges. We use machine learning to classify the sentiment in texts and natural language processing to study commonly used word pairings. More importantly, we also intend to measure how hard and soft attributes influence customer groups' satisfaction and dissatisfaction. We define hard attributes as attributes relating to physical and environmental aspects, such as the hotel's facilities, location, infrastructure, and surrounding real estate. In contrast, soft attributes are the hotel's non-physical attributes related to services, staff, or management.

3 Theoretical background and hypothesis development

3.1 Cultural influence in expectation and satisfaction

Customer satisfaction in tourism has been analyzed since decades past, Hunt (1975) having defined customer satisfaction as the realization or overcoming of expectations towards the service. Oliver (1981) defined it as an emotional response to the provided services in retail and other contexts, and Oh and Parks (1996) reviewed the psychological processes of customer satisfaction for the hospitality industry. It is generally agreed upon that satisfaction and dissatisfaction stem from the individual expectations of the customer. As such, Engel et al. (1990) states that each customer's background, therefore, influences satisfaction and dissatisfaction. It can also be said that satisfaction stems from the perceptions of quality in comparison to these expectations.

These differences in customers' backgrounds can be summed up in cultural differences as well. In the past, satisfaction and perceived service quality have been found to be influenced by cultural differences (e.g. Mattila 1999; Winsted 1997). Service quality perceptions have been studied via measurements such as SERVQUAL (e.g. Armstrong et al. 1997).

Previous studies on the dimensions of culture that influence differences in expectations have been performed in the past as well (e.g. Mattila 2019; Levy 2010; Donthu and Yoo 1998), such as comparing individualism vs. collectivism, high context vs. low context, uncertainty avoidance, among other factors. While culture as a concept is difficult to quantify, some researchers have tried to use these and more dimensions to measure cultural differences, such as the six dimensions described by Hofstede (1984), or the nine dimensions of the GLOBE model (House et al. 1999).

These cultural dimensions are more differentiated in Western and Asian cultures (Levy 2010). Our study being located in Japan, it stands to reason that the differences in expectations between Western tourists and Asian tourists should be understood in order to provide a good service. However, even though geographically close, Japanese and Chinese cultures are both very different when it comes to customer service. This is why our study focuses on the difference between Chinese and Western customers in Japan. The contrasting cultural backgrounds between Chinese and Western customers will lead to varying expectations of the hotel services, the experiences they want to have while staying at a hotel, and the level of comfort that they will have. In turn, these different expectations will determine the distinct factors of satisfaction and dissatisfaction for each kind of customer and the order in which they prioritize them.

Because of their different origins, expectations, and cultures, it stands to reason Chinese and Western tourists could have completely different factors to one another. Therefore, it could be that some factors do not appear in the other reviews at all. For example, between different cultures, it can be that a single word can express some concept that would take more words in the other language. Therefore, we must measure their differences or similarities at their common ground as well.

3.2 Customer satisfaction and dissatisfaction towards individual factors during hotel stay

We reviewed the importance of expectations in the development of satisfaction and dissatisfaction and the influence that cultural backgrounds have in shaping these expectations. This is true for overall satisfaction for the service as a whole, as well as individual elements that contribute to satisfaction.

In this study, we study not overall customer satisfaction but the satisfaction and dissatisfaction that stem from individual-specific expectations, be they conscious or unconscious. For example, if a customer has a conscious expectation of a comfortable bed and a wide shower, and it is realized during their visit, they will be satisfied with this matter. However, suppose that same customer with a conscious expectation of a comfortable bed experienced loud noises at night. In that case, they can be dissatisfied with a different aspect, regardless of the satisfaction towards the bed. Then, the same customer might have packed their toiletries, thinking that the amenities might not include those. They can then be pleasantly surprised with good quality amenities and toiletries, satisfying an unconscious expectation. This definition of satisfaction does not allow us to examine overall customer satisfaction. However, it will allow us to examine the factors that a hotel can revise individually and how a population perceives them as a whole. In our study, we consider the definitions in Hunt (1975) that satisfaction is a realization of an expectation, and we posit that customers can have different expectations towards different service aspects. Therefore, in our study, we define satisfaction as the emotional response to the realization or overcoming of conscious or unconscious expectations towards an individual aspect or factor of a service. On the other hand, dissatisfaction is the emotional response

to the lack of a realization or under-performance of these conscious or unconscious expectations towards specific service aspects.

Studies on customer satisfaction (e.g. Truong and King 2009; Romão et al. 2014; Wu and Liang 2009) commonly use the Likert scale (Likert 1932) (e.g. 1 to 5 scale from strongly dissatisfied to strongly satisfied) to perform statistical analysis of which factors relate most to satisfaction on the same dimension as dissatisfaction (e.g. Chan et al. 2015; Choi and Chu 2000). The Likert scale's use leads to correlation analyses where one factor can lead to satisfaction, implying that the lack of it can lead to dissatisfaction. However, a binary distinction (satisfied or dissatisfied) could allow us to analyze the factors that correlate to satisfaction and explore factors that are solely linked to dissatisfaction. There are fewer examples of this approach, but studies have done this in the past (e.g. Zhou et al. 2014). This method can indeed decrease the extent to which we can analyze degrees of satisfaction or dissatisfaction. However, it has the benefit that it can be applied to a large sample of text data via automatic sentiment detection techniques using artificial intelligence.

3.3 Japanese hospitality and service: *Omotenashi*

The spirit of Japanese hospitality, or *Omotenashi*, has roots in the country's history, and to this day, it is regarded as the highest standard (Ikeda 2013; Al-alsheikh and Sato 2015). There is a famous phrase in customer service in Japan: *okyaku-sama wa kami-sama desu*, meaning "The customer is god." Some scholars say that *omotenashi* originated from the old Japanese art of the tea ceremony in the sixteenth century, while others found that it originates in the form of formal banquets in the 7th-century (Aishima et al. 2015). The practice of high standards in hospitality has survived throughout the years. Presently, it permeates all business practices in Japan, from the cheapest convenience stores to the most expensive ones. Manners, service, and respect towards the customer are taught to workers in their training. High standards are always followed to not fall behind in the competition. In Japanese businesses, including hotels, staff members are trained to speak in *sonkeigo*, or "respectful language," one of the most formal of the Japanese formality syntaxes. They are also trained to bow differently depending on the situation, where a light bow could be used to say "Please, allow me to guide you." Deep bows are used to apologize for any inconvenience the customer could have faced, followed by a very respectful apology. Although the word *omotenashi* can be translated directly as "hospitality," it includes both the concepts of hospitality and service (Kuboyama 2020). This hospitality culture permeates every type of business with customer interaction in Japan. A simple convenience shop could express all of these hospitality and service standards, which are not exclusive to hotels.

It stands to reason that this cultural aspect of hospitality would positively influence customer satisfaction. However, in many cases, other factors such as proximity to a convenience store, transport availability, or room quality might be more critical to a customer. In this study, we cannot directly determine whether a hotel is practicing the cultural standards of *omotenashi*. Instead, we consider it as a cultural factor that influences all businesses in Japan. We then observe the customers' evaluations

regarding service and hospitality factors and compare them to other places and business practices in the world. In summary, we consider the influence of the cultural aspect of *omotenashi* while analyzing the evaluations on service and hospitality factors that are universal to all hotels in any country.

Therefore, we pose the following research question:

Research Question 1a To what degree are Chinese and Western tourists satisfied with Japanese hospitality factors such as staff behavior or service?

However, Japanese hospitality is based on Japanese culture. Different cultures interacting with it could provide a different evaluation of it. Some might be impressed by it, whereas some might consider other factors more important to their stay in a hotel. This point leads us to a derivative of the aforementioned research question:

Research Question 1b Do Western and Chinese tourists have a different evaluation of Japanese hospitality factors such as staff behavior or service?

3.4 Customer expectations beyond service and hospitality

Staff behavior, hospitality and service, and therefore *Omotenashi*, are all soft attributes of a hotel. That is, they are non-physical attributes of the hotel, and as such, they are practical to change through changes in management. While it is important to know this, it is not known if the cultural differences between Chinese and Western tourists also influence other expectations and satisfaction factors, such as the hard factors of a hotel.

Hard factors are attributes uncontrollable by the hotel staff, which can play a part in the customers' choice behavior and satisfaction. Examples of these factors include the hotel's surroundings, location, language immersion of the country as a whole, or touristic destinations, and the hotel's integration with tours available nearby, among other factors.

Besides the facilities, many other aspects of the experience, expectation, and perception of the stay in a hotel can contribute to the overall satisfaction, as well as individual satisfactions and dissatisfactions. However, previous research focuses more on these soft attributes, with little focus on hard attributes, if only focusing on facilities (e.g. Shanka and Taylor 2004; Choi and Chu 2001). Because of this gap in knowledge, we decided to analyze the differences in cultures regarding both soft and hard attributes of a hotel.

This leads to two of our research questions:

Research Question 2a To what degree do satisfaction and dissatisfaction stem from hard and soft attributes of the hotel?

Research Question 2b How differently do Chinese and Western customers perceive hard and soft attributes of the hotel?

The resulting proportions of hard attributes to soft attributes for each population could measure how much the improvement of management in the hotel can increase future satisfaction in customers.

3.5 Chinese and Western tourist behavior

In the past, social science and tourism studies focused extensively on Western tourist behavior in other countries. Recently, however, with the rise of Chinese outbound tourism, both academic researchers and businesses have decided to study Chinese tourist behavior, with rapid growth in studies following the year 2007 (Sun et al. 2017). However, studies focusing on only the behavior of this subset of tourists are the majority. To this day, studies and analyses specifically comparing Asian and Western tourists are scarce, and even fewer are the number of studies explicitly comparing Chinese and Western tourists. One example is a study by Choi and Chu (2000), which found that Western tourists visiting Hong Kong are satisfied more with room quality, while Asians are satisfied with the value for money. Another study by Bauer et al. (1993) found that Westerners prefer hotel health facilities, while Asian tourists were more inclined to enjoy the Karaoke facilities of hotels. Both groups tend to have high expectations for the overall facilities. Another study done by Kim and Lee (2000) found American tourists to be individualistic and motivated by novelty, while Japanese tourists were collectivist and motivated by increasing knowledge and escaping routine.

One thing to note with the above Asian vs. Western analyses is that they were performed before 2000 and not Chinese-specific. Meanwhile, the current Chinese economic boom is increasing the influx of tourists of this nation. The resulting increase in marketing and the creation of guided tours for Chinese tourists could have created a difference in tourists' perceptions and expectations. In turn, if we follow the definition of satisfaction in Hunt (1975), the change in expectations could have influenced their satisfaction factors when traveling. Another note is that these studies were performed with questionnaires in places where it would be easy to locate tourists, i.e., airports. However, our study of online reviews takes the data that the hotel customers uploaded themselves. This data makes the analysis unique in exploring their behavior compared with Western tourists via factors that are not considered in most other studies. Furthermore, our study is unique in observing the customers in the specific environment of high-level hospitality in Japan.

More recent studies have surfaced as well. A cross-country study (Francesco and Roberta 2019) using posts from USA citizens, Italians, and Chinese tourists, determined using a text link analysis that customers from different countries indeed have a different perception and emphasis of a few predefined hotel attributes. According to their results, USA customers perceive cleanliness and quietness most positively. In contrast, Chinese customers perceive budget and restaurant above other attributes. Another couple of studies (Jia 2020; Huang 2017) analyze differences between Chinese and US tourists using text mining techniques and more massive datasets, although in a restaurant context.

These last three studies focus on the USA culture, whereas our study focuses on the Western culture. Another difference with our study is that of the context of the study. The first study (Francesco and Roberta 2019) was done within the context of tourists from three countries staying in hotels across the world. The second study chose restaurant reviews from the USA and Chinese tourists eating in three countries in Europe. The third study analyzed restaurants in Beijing.

On the other hand, our study focuses on Western culture, instead of a single Western country, and Chinese culture clashing with the hospitality environment in Japan, specifically. Japan's importance in this analysis comes from the unique environment of high-grade hospitality that the country presents. In this environment, customers could either hold their satisfaction to this hospitality regardless of their culture or value other factors more depending on their cultural differences. Our study measures this at a large scale across different hotels in Japan.

Other studies have gone further and studied people from many countries in their samples and performed a more universal and holistic (not cross-culture) analysis. Choi and Chu (2001) analyzed hotel guest satisfaction determinants in Hong Kong with surveys in English, Chinese and Japanese translations, with people from many countries in their sample. Choi and Chu (2001) found that staff service quality, room quality, and value for money were the top satisfaction determinants. As another example, Uzama (2012) produced a typology for foreigners coming to Japan for tourism, without making distinctions for their culture, but their motivation in traveling in Japan. In another study, Zhou et al. (2014) analyzed hotel satisfaction using English and Mandarin online reviews from guests staying in Hangzhou, China coming from many countries. The general satisfaction score was noticed to be different among those countries. However, a more in-depth cross-cultural analysis of the satisfaction factors was not performed. As a result of their research, Zhou et al. (2014) thus found that customers are universally satisfied by welcome extras, dining environments, and special food services.

Regarding Western tourist behavior, a few examples can tell us what to expect when analyzing our data. Kozak (2002) found that British and German tourists' satisfaction determinants while visiting Spain and Turkey were hygiene and cleanliness, hospitality, the availability of facilities and activities, and accommodation services. Shanka and Taylor (2004) found that English-speaking tourists in Perth, Australia were most satisfied with staff friendliness, the efficiency of check-in and check-out, restaurant and bar facilities, and lobby ambiance.

Regarding outbound Chinese tourists, academic studies about Chinese tourists have increased (Sun et al. 2017). Different researchers have found that Chinese tourist populations have several specific attributes. According to Ryan and Mo (2001) and their study of Chinese tourists in New Zealand, Chinese tourists prefer nature, cleanliness, and scenery in contrast to experiences and activities. Dongyang et al. (2015) studied Chinese tourists in the Kansai region of Japan and found that Chinese tourists are satisfied mostly with exploring the food culture of their destination, cleanliness, and staff. Studying Chinese tourists in Vietnam, Truong and King (2009) found that Chinese tourists are highly concerned with value for money. According to Liu et al. (2019), Chinese tourists tend to have harsher criticism compared with other international tourists. Moreover, as stated by Gao et al. (2017), who

analyzed different generations of Chinese tourists and their connection to nature while traveling, Chinese tourists prefer nature overall. However, the younger generations seem to do so less than their older counterparts.

Although the studies focusing only on Chinese or Western tourists have a narrow view, their theoretical contributions are valuable. We can see that depending on the study and the design of questionnaires and the destinations; the results can vary greatly. Not only that, but while there seems to be some overlap in most studies, some factors are completely ignored in one study but not in the other. Since our study uses data mining, each factor's definition is left for hotel customers to decide en masse via their reviews. This means that the factors will be selected through statistical methods alone instead of being defined by the questionnaire. Our method allows us to find factors that we would not have contemplated. It also avoids enforcing a factor on the mind of study subjects by presenting them with a question that they did not think of by themselves. This large variety of opinions in a well-sized sample, added to the automatic findings of statistical text analysis methods, gives our study an advantage compared to others with smaller samples. This study analyzes the satisfaction and dissatisfaction factors cross-culturally and compares them with the existing literature.

Undoubtedly previous literature has examples of other cross-culture studies of tourist behavior and may further highlight our study and its merits. A contrast is shown in Table 1. This table shows that older studies were conducted with surveys and had a different study topic. These are changes in demand (Bauer et al. 1993), tourist motivation (Kim and Lee 2000), and closer to our study, satisfaction levels (Choi and Chu 2000). However, our study topic is not the levels of satisfaction but the factors that drive it and dissatisfaction, which is overlooked in most studies. Newer studies with larger samples and similar methodologies have emerged, although two of these study restaurants instead of hotels (Jia 2020; Huang 2017). One important difference is the geographical focus of their studies. While Francesco and Roberta (2019), Jia (2020) and Huang (2017) have a multi-national focus, we instead focus on Japan. The focus on Japan is important because of its top rank in hospitality across all types of businesses. Our study brings light to the changes, or lack thereof, in different touristic environments where an attribute can be considered excellent. The number of samples in other text-mining studies is also smaller than ours in comparison. Apart from that, every study has a different text mining method.

3.6 Data mining, machine learning, knowledge discovery and sentiment analysis

In the current world, data is presented to us in larger and larger quantities. Today's data sizes were commonly only seen in very specialized large laboratories with supercomputers a couple of decades ago. However, they are now standard for market and managerial studies, independent university students, and any scientist connecting to the Internet. Such quantities of data are available to study now more than ever. Nevertheless, it would be impossible for researchers to parse all of this data by themselves. As Fayyad et al. (1996) summarizes, data by itself is unusable until it goes through a process of selection, preprocessing, transformation, mining, and

Table 1 Comparison between cross-culture or cross-country previous studies and our study

| | Bauer et al. (1993) | Choi and Chiu (2000) | Kim and Lee (2000) | Huang (2017) | Francesco and Roberta (2019) | Jia (2020) | Our study |
|---------------------|--|--|------------------------------------|---|---------------------------------|---|--|
| Comparison objects | Asians vs Westerns | Asians vs Westerners | Anglo-Americans vs Japanese | Chinese vs English-speakers | USA vs China vs Italy | Chinese vs US tourists | Chinese vs Westerners |
| Study topic | Changes in demand | Satisfaction Levels | Tourist Motivation | Dining experience of Roast Duck | Perception and Emphasis | Motivation and Satisfaction | Satisfaction and Dis-satisfaction |
| Geographical focus | Asia Pacific region | Hong Kong | Global | Beijing | Multi-national | Multi-national | Japan |
| Industry | Hotels | Hotels | Tourism | Restaurant (Beijing Roast Duck) | Hotels | Restaurants | Hotels |
| Study subjects | Hotel managers | Hotel customers | Tourists arriving in airport | Diners online reviews | Hotel customers | Diners online reviews | Hotel customers |
| Sample method | Surveys | Surveys | Survey | Text mining | Text mining | Text mining | Text mining |
| Number of samples | 185 surveys | 540 surveys | 165 Anglo-American 209 Japanese | 990 Chinese reviews 398 English reviews | 9000 reviews (3000 per country) | 2448 reviews (1360 Chinese) (1088 English) | 89,207 reviews (48,070 Chinese) (41,137 English) |
| Study method | statistics | VARIMAX | MANOVA | Semantic Network Analysis | Text Link Analysis | Topic modeling (LDA) | SVM, Dependency Parsing and POS tagging |
| Subject nationality | Asians: China, Fiji, Hong Kong, Indonesia, Malaysia, Singapore, Taiwan, Guam, Tahiti, Thailand | Asians: China, Taiwan, Japan, South Korea, South-East Asia | USA, Japan | English-speakers: UK, US, Australia, New Zealand, Canada, Ireland | USA, China, Italy | USA, China | Chinese-speakers: China English-speakers: (UK, US, Australia, New Zealand, Canada, Ireland) |
| Westerners: | Australia, New Zealand | America, Europe, Australia, New Zealand | | Chinese-speakers: China | | | |

evaluation. Only then can it be established as knowledge. With the tools available to us in the era of information science, algorithms can be used to detect patterns that would take researchers too long to recognize. These patterns can, later on, be evaluated to generate knowledge. This process is called Knowledge Discovery in Databases.

Now, there are, of course, many sources of numerical data to be explored. However, perhaps what is most available and interesting to managerial purposes is the resource of customers' opinions in text form. Since the introduction of Web 2.0, an unprecedented quantity of valuable information is posted to the Internet at a staggering speed. Text mining has then been proposed more than a decade ago to utilize this data (e.g. Rajman and Besançon 1998; Nahm and Mooney 2002). Using Natural Language Processing, one can parse language in a way that translates to numbers so that a computer can analyze it. Since then, text mining techniques have improved over the years. This has been used in the field of hospitality as well for many purposes, including satisfaction analysis from reviews (e.g. Berezina et al. 2016; Xu and Li 2016; Xiang et al. 2015; Hargreaves 2015; Balbi et al. 2018), social media's influence on travelers (e.g. Xiang and Gretzel 2010), review summarization (e.g. Hu et al. 2017), perceived value of reviews (e.g. Fang et al. 2016), and even predicting hotel demand using web traffic data (e.g. Yang et al. 2014).

More than only analyzing patterns within the text, researchers have found how to determine the sentiment behind a statement based on speech patterns, statistical patterns, and other methodologies. This method is called sentiment analysis or opinion mining. A precursor of this method was attempted decades ago (Stone et al. 1966). With sentiment analysis, one could use patterns in the text to determine whether a sentence was being said with a positive opinion, or a critical one. This methodology could even determine other ranges of emotions, depending on the thoroughness of the algorithm. Examples of sentiment analysis include ranking products through online reviews (e.g. Liu et al. 2017; Zhang et al. 2011), predicting political poll results through opinions in Twitter (O'Connor et al. 2010), and so on. In the hospitality field, it has been used to classify reviewers' opinions of hotels in online reviews (e.g. Kim et al. 2017; Al-Smadi et al. 2018).

Our study used an algorithm for sentiment analysis called a Support Vector Machine (SVM), a supervised machine learning used for binary classification. Machine learning is a general term used for algorithms that, when given data, will automatically use that data to "learn" from its patterns and apply them for improving upon a task. Learning machines can be supervised, as in our study, where the algorithm has manually labeled training data to detect patterns in it and use them to establish a method for classifying other unlabeled data automatically. Machine learning can also be unsupervised, where there is no pre-labeled data. In this latter case, the machine will analyze the structure and patterns of the data and perform a task based on its conclusions. Our study calls for a supervised machine since text analysis can be intricate. Many patterns might occur, but we are only interested in satisfaction and dissatisfaction labels. Consequently, we teach the machine through previously labeled text samples.

Machine learning and data mining are two fields with a significant overlap since they can use each other's methods to achieve the task at hand. Machine learning

methods focus on predicting new data based on known properties and patterns of the given data. Data mining, on the other hand, is discovering new information and new properties of the data. Our machine learning approach will learn the sentiment patterns of our sample texts showing satisfaction and dissatisfaction and using these to label the rest of the data. We are not exploring new patterns in the sentiment data. However, we are using sentiment predictions for knowledge discovery in our database. Thus, our study is a data mining experiment based on machine learning.

Because the methodology for finding patterns in the data is automatic and statistical, it is both reliable and unpredictable. Reliable in that the algorithm will find a pattern by its nature. Unpredictable in that since it has no intervention from the researchers in making questionnaires, it can result in anything that the researchers could not expect. These qualities determine why, similar to actual mining; data mining is mostly exploratory. One can never be sure that one will find a specific something. However, we can make predictions and estimates about finding knowledge and what kind of knowledge we can uncover. The exploration of large opinion datasets with these methods is essential. The reason is that we can discover knowledge that could otherwise be missed by observing a localized sample rather than taking a holistic view of every user's opinion. In other words, a machine algorithm can find the needles in a haystack that we did not know were there by examining small bundles of hay at a time.

4 Methodology

We extracted a large number of text reviews from the site *Ctrip*, with mostly mainland Chinese users, and the travel site *TripAdvisor*. We then determined the most commonly used words that relate to positive and negative opinions in a review. We did this using Shannon's entropy to extract keywords from their vocabulary. These positive and negative keywords allow us to train an optimized Support Vector Classifier (SVC) to perform a binary emotional classification of the reviews in large quantities, saving time and resources for the researchers. We then applied a dependency parsing to the reviews and a Part of Speech tagging (POS tagging) to observe the relationship between adjective keywords and the nouns they refer to. We split the dataset into price ranges to observe the differences in keyword usage between lower-class and higher-class hotels. We observed the frequency of the terms in the

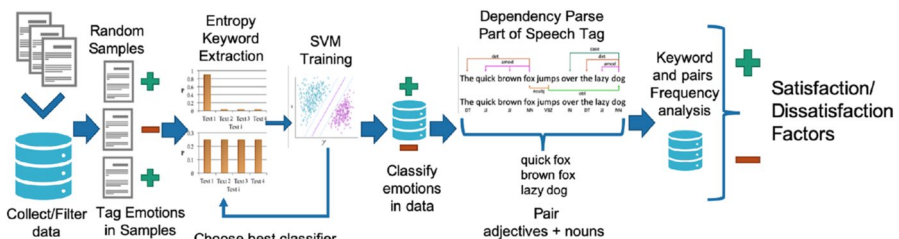


Fig. 1 Overview of the methodology to quantitatively rank satisfaction factors

dataset to extract the most utilized words in either review. We show an overview of this methodology in Fig. 1, which is an updated version of the methodology used by Alemán Carreón et al. (2018). Finally, we also observed if the satisfaction factors were soft or hard attributes of the hotel.

4.1 Data collection

In the *Ctrip* data collection, reviews from a total of 5774 hotels in Japan were collected. From these pages, we extracted a total of 245,919 reviews, from which 211,932 were detected to be standard Mandarin Chinese. Since a single review can have sentences with different sentiments, we separated sentences using punctuation marks. The Chinese reviews were comprised of 187,348 separate sentences.

In the *TripAdvisor* data collection, we collected data from 21,380 different hotels. In total, we collected 295,931 reviews, from which 295,503 were detected to be in English. Similarly to the Chinese data, we then separated these English reviews into 2,694,261 sentences using the *gensim* python library. For the language detection in both cases we used the *langdetect* python library.

However, to make the data comparisons fair, we filtered both databases only to contain reviews from hotels in both datasets, using their English names to do a search match. We also filtered them to be in the same date range. In addition, we selected only the hotels that had pricing information available. We extracted the lowest and highest price possible for one night as well. The difference in pricing can be from better room settings, such as double or twin rooms or suites, depending on the hotel. Regardless of the reason, we chose the highest-priced room since it can be an indirect indicator of the hotel's class. After filtering, the datasets contained 557 hotels in common. The overlapping date range for reviews was from July 2014 to 2017. Within these hotels, from *Ctrip* there was 48,070 reviews comprised of 101,963 sentences, and from *TripAdvisor* there was 41,137 reviews comprised of 348,039 sentences.

The price for a night in these hotels ranges from cheap capsule hotels at 2000 yen per night to high-end hotels 188,000 yen a night at the far ends of the bell curve.

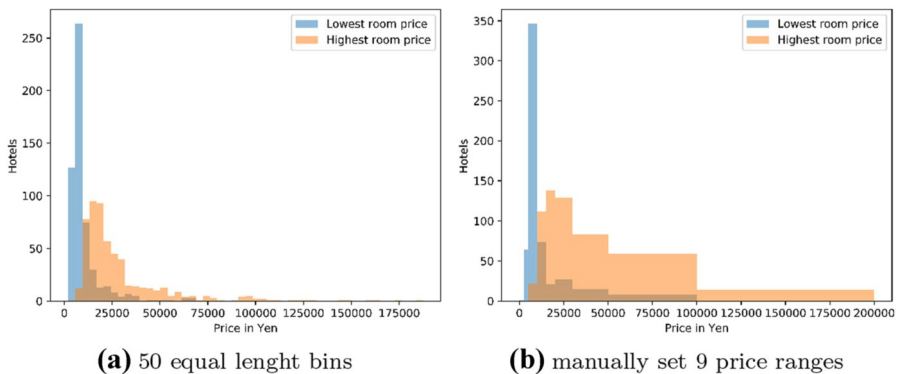


Fig. 2 Price for one night distribution, blue: lowest price, orange: highest price

Table 2 Collected data and structure after price range categorizing

| Price range | Data collected | Ctrip database | Tripadvisor database |
|-----------------------|--------------------|----------------|----------------------|
| 0: All Prices | Hotels | 557 | 557 |
| | Reviews | 48,070 | 41,137 |
| | Sentences | 101,963 | 348,039 |
| | Positive sentences | 88,543 | 165,308 |
| | Negative sentences | 13,420 | 182,731 |
| 1: 0–2500 yen | Hotels | 0 | 0 |
| | Reviews | 0 | 0 |
| 2: 2500–5000 yen | Hotels | 0 | 0 |
| | Reviews | 0 | 0 |
| 3: 5000–10,000 yen | Hotels | 22 | 22 |
| | Reviews | 452 | 459 |
| | Sentences | 1108 | 3,988 |
| | Positive sentences | 924 | 1875 |
| | Negative sentences | 184 | 2113 |
| 4: 10,000–15,000 yen | Hotels | 112 | 112 |
| | Reviews | 2176 | 2865 |
| | Sentences | 4240 | 24,107 |
| | Positive sentences | 3566 | 11,619 |
| | Negative sentences | 674 | 12,488 |
| 5: 15,000–20,000 yen | Hotels | 138 | 138 |
| | Reviews | 7043 | 4384 |
| | Sentences | 14,726 | 37,342 |
| | Positive sentences | 12,775 | 17,449 |
| | Negative sentences | 1951 | 19,893 |
| 6: 20,000–30,000 yen | Hotels | 129 | 129 |
| | Reviews | 11,845 | 13,772 |
| | Sentences | 24,413 | 115,830 |
| | Positive sentences | 21,068 | 55,381 |
| | Negative sentences | 3345 | 60,449 |
| 7: 30,000–50,000 yen | Hotels | 83 | 83 |
| | Reviews | 8283 | 7001 |
| | Sentences | 17,939 | 58,409 |
| | Positive sentences | 15,642 | 28,493 |
| | Negative sentences | 2297 | 29,916 |
| 8: 50,000–100,000 yen | Hotels | 59 | 59 |
| | Reviews | 16,670 | 9646 |
| | Sentences | 36,255 | 81,940 |
| | Positive sentences | 31,638 | 38,217 |
| | Negative sentences | 4617 | 43,723 |

Table 2 (continued)

| Price range | Data collected | Ctrip database | Tripadvisor database |
|------------------------|--------------------|----------------|----------------------|
| 9: 100,000–200,000 yen | Hotels | 14 | 14 |
| | Reviews | 1601 | 3010 |
| | Sentences | 3282 | 26,423 |
| | Positive sentences | 2930 | 12,274 |
| | Negative sentences | 352 | 14,149 |

Customers' expectations can vary greatly depending on the pricing of the hotel room they stay at. Therefore, we made observations on the distribution of pricing in our database's hotels and binned the data by price ranges, decided by consideration of the objective of stay. We show these distributions in Fig. 2. The structure of the data after division by price is shown in Table 2. This table also includes the results of emotional classification after applying our SVC, as explained in 4.3. The first three price ranges (0–2500 yen, 2500–5000 yen, 5000–10,000 yen) would correspond to low-class hotels or even hostels on the lower end and cheap business hotels on the higher end. Further on, there are business hotels in the next range (10,000–15,000 yen). After that, the stays could be at Japanese style *ryokan* when traveling in groups, high-class business hotels, luxury love hotels, or higher class hotels (15,000–20,000 yen, 20,000–30,000 yen). Further than that is more likely to be *ryokan* or high class resorts or five-star hotels (30,000–50,000 yen, 50,000–100,000 yen, 100,000–200,000 yen). Note that because of choosing the highest price per one night in each hotel, the cheapest two price ranges (0–2500 yen, 2500–5000 yen) are empty, despite some rooms being priced at 2000 yen per night. Because of this, other tables will omit these two price ranges.

4.2 Text processing

We needed to analyze the grammatical relationship between words, be it English or Chinese, to understand the connections between adjectives and nouns. For all these processes, we used the Stanford CoreNLP pipeline developed by the Natural Language Processing Group at Stanford University (Manning et al. 2014). In order to separate Chinese words for analysis, we used the Stanford Word Segmenter (Chang et al. 2008). In English texts, however, only using spaces is not enough to correctly collect concepts. The English language is full of variations and conjugations of words depending

Fig. 3 Example of dependency parsing

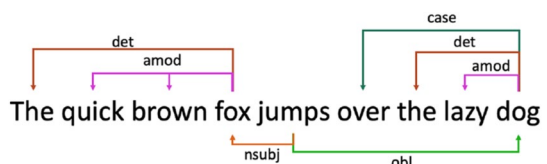


Fig. 4 Example of POS tagging with the Penn Treebank tags

The quick brown fox jumps over the lazy dog
 DT JJ JJ NN VBZ IN DT JJ NN

Table 3 Target parts of speech for extraction and pairing

| Language | POS Tag | Part of Speech | Examples |
|---------------------|----------------------------------|------------------------------|---------------------|
| Chinese target tags | NN | Noun (general) | 酒店 (hotel) |
| | VA | Predicative Adjective (verb) | 干净的 (clean) |
| | JJ | Noun modifier (adjectives) | 干净 (clean) |
| | VV | Verb (general) | 推荐 (recom- mend) |
| English target tags | NN | Noun (general) | Room |
| | NNS | Noun (plural) | Beds |
| | JJ | Adjective | Big |
| | JJS | Adjective (superlative) | Best |
| | JJR | Adjective (comparative) | Larger |
| | VB | Verb (base form) | Take |
| | VBP | Verb (single present) | Take |
| | VBN | Verb (past participle) | Taken |
| VBG | Verb (gerund/present participle) | Taking | |

on the context and tense. Thus, a better segmentation is achieved by using lemmatization, which returns each word's dictionary form. For this purpose, we used the *gensim* library for the English texts.

A dependency parser analyzes the grammatical structure, detecting connections between words, and describing the action and direction of those connections. We show an example of these dependencies in Fig. 3. This study uses the Stanford NLP Dependency Parser, as described by Chen and Manning (2014). A list of dependencies used by this parser is detailed by de Marneffe and Manning (2008). In more recent versions, they use an updated dependency tag list from Universal Dependencies (Zeman et al. 2018). In our study, this step was necessary to extract adjective modifiers and their subject. We did that by parsing the database and extracting instances of a few determined dependency codes. One of these dependency codes is “amod”, which stands for “adjectival modifier”. This is used when an adjective modifies a noun directly (e.g., A big apple). The other dependency code we used was “nsubj”, or nominal subject, the class's syntactic subject. We used this one for cases where the adjective is modifying the noun indirectly through other words (e.g., The apple is big). This dependency does not necessarily only include a combination of adjectives and nouns. However, it can also be connected with copular verbs, nouns, or other adjectives. We saw it necessary also to perform a Part of Speech (POS) tagging of these clauses.

A Part of Speech (POS) tagger is a program that assigns word tokens with tags identifying the part of speech. An example is shown in Fig. 4. A Part of Speech is

Table 4 Filtered out parts of speech to aid pairing

| Language | POS Tag | Part of Speech | Examples |
|------------------------|---------|-------------------------|---|
| Commonly filtered tags | DT | Determiner | a, an |
| | PN | Pronoun | I, you, they |
| | CD | Cardinal Number | 1, 2, 3, 4, 5 |
| | PU | Punctuation | .!? |
| Chinese filtered tags | DEV | Particle | 地 (Japan) (adverbial particle) |
| | NR | Noun (proper noun) | 日本 (Japan) |
| | M | Measure word | 个 (general classifier), 公里 (kilo-meter) |
| | SP | Sentence-final particle | 他 (he), 好 (good) |
| | IJ | Interjection | 啊 (ah) |
| English target tags | NNP | Noun (proper noun) | Japan |
| | PRP\$ | Possessive Pronoun | My, your, her, his |
| | WP | Wh-pronoun | What, who |

a category of lexical items that serve similar grammatical purposes, for example, nouns, adjectives, verbs, or conjunctions. In our study, we used the Stanford NLP POS tagger software, described by Toutanova and Manning (2000) and Toutanova et al. (2003), which uses the Penn Chinese Treebank tags (Xia 2000).

In this study, we were interested in identifying combinations of adjectives, some verbs, and nouns. We also needed to filter away bad combinations that were brought by the versatility of nominal subject dependencies. For this purpose, we identified the tags for nouns, verbs, and adjectives in Chinese and English, with the English tags being a bit more varied. What would be called adjectives in English corresponds more to stative verbs in Chinese, so we needed to extract those as well. We show a detailed description of the chosen tags in Table 3. We also show a detailed description of the tags we needed to filter. We selected these tags heuristically by observing commonly found undesired pairs in Table 4.

Once we had these adjective + noun or verb + noun pairs, we could determine what the customers referred to in their reviews. With what frequency they use those pairings positively or negatively.

Table 5 Best performing SVC 5-fold cross-validation Chinese text classifiers

| Keyword List | Classifier emotion | C | F_1 μ | F_1 σ |
|--|---------------------|------------|----------------|-------------------|
| Satisfaction keywords ($\alpha = 2.75$) | Satisfaction | 2.5 | 0.91 | 0.01 |
| Negative keywords ($\alpha' = 3.75$) | Dissatisfaction | 0.5 | 0.67 | 0.11 |
| Combined ($\alpha = 2.75$, $\alpha' = 3.75$) | Satisfaction | 0.5 | 0.95 | 0.01 |

Bold font shows the result of performance tests of several configurations of the classifiers. The boldened configuration was the one chosen for the classification process in each language

Table 6 Best performing SVC 10-fold cross-validation English text classifiers

| Keyword List | Classifier emotion | C | F_1 μ | F_1 σ |
|--|---------------------|----------|----------------|-------------------|
| Satisfaction keywords ($\alpha = 1.5$) | Satisfaction | 1.75 | 0.82 | 0.02 |
| Dissatisfaction keywords ($\alpha' = 4.25$) | Dissatisfaction | 3 | 0.80 | 0.03 |
| Combined ($\alpha = 1.5, \alpha' = 4.25$) | Satisfaction | 2 | 0.83 | 0.02 |

Bold font shows the result of performance tests of several configurations of the classifiers. The boldened configuration was the one chosen for the classification process in each language

Table 7 Examples of positive and negative sentences used for training SVM

| Language | Emotion | Sentences |
|----------|----------|--|
| Chinese | Positive | 酒店的服务很好和我住过的所有日本酒店一样各种隐形服务非常厉害 (translated as: "The service of the hotel is very good. All the services of the Japanese hotels I have stayed in are extremely good.") 有一个后门到地铁站非常近周边也算方便酒店服务和卫生都很好 (translated as: "There is a back door to the subway station very close to it. The surrounding area is also convenient hotel service and health are very good") |
| | Negative | 酒店旁边很荒凉连个便利店都要走很远 (translated as: "The hotel is very bleak, and you have to go very far to go to the nearest convenience store.") 唯一不足是价格太高 (translated as: "The only negative is that the price is too high.") |
| English | Positive | It was extremely clean, peaceful and the hotel Hosts made us feel super welcome Location is very good, close to a main road with a subway station, a bakery, a 7 eleven and a nice restaurant that is not too expensive but serves good food |
| | Negative | The only downside. Our room was labeled 'non-smoking' but our duvet reeked of smoke. A bit pricey though |

4.3 Sentiment analysis using a Support Vector Classifier

The sentiment analysis was performed using the methodology described by Alemán Carreón et al. (2018). Keywords are determined by a comparison of Shannon's entropy (Shannon 1948) between two classes by a factor of α for one class and α' for the other, and then they are used in an SVC (Cortes and Vapnik 1995), optimizing keywords to select the best performing classifier using the F_1 -measure (Powers 2011). The selected SVC keywords would then clearly represent the user driving factors leading to positive and negative emotions. We also performed experiments to choose the best value of the parameter C used in the SVC. C is a constant that affects the optimization process when minimizing the error of the separating hyper-plane. Low values of C give some freedom of error, which minimizes false positives but can also increase false negatives. Inversely, high C values will likely result in

minimal false negatives but a possibility of false positives. SVC performance results are displayed in Tables 5 and 6. Examples of tagged sentences are shown in Table 7.

Shannon's entropy can be used to observe the probability distribution of each word inside the corpus. A word included in many documents will have a high entropy value for that set of documents. Opposite to this, a word appearing in only one document will have an entropy value of zero.

An SVC is trained to classify data based on previously labeled data, generalizing the data's features by defining a separating $(p-1)$ -dimensional hyperplane in p -dimensional space. Each dimension is a feature of the data in this space. The separating hyperplane, along with the support vectors, divides the multi-dimensional space and minimizes classification error.

Our study used a linear kernel for the SVC, defined by the formula (1) below. Each training sentence is a data point, a row in the vector x . Each column represents a feature; in our case, the quantities of each of the keywords in that particular sentence. The labels of previously known classifications (1 for positive, 0 for negative) for each sentence comprise the $f(x)$ vector. The Weight Vector w is comprised of the influences each point has had in the training process to define the hyperplane angle. The bias coefficient b determines its position.

During the SVC learning algorithm, each data point classified incorrectly alters the weight vector to correctly classify new data. These changes to the weight vector are greater for features close to the separating hyperplane. These features have stronger changes because they needed to be taken into account to classify with a minimal error. Sequentially, the weight vector can be interpreted as a numerical representation of each feature's effect on each class's classification process. Below we show the formula for the weight vector w (2), where x is the training data and each vectorized sentence x_i in the data is labeled y_i . Each cycle of the algorithm alters the value of w by α to reduce the number of wrong classifications. This equation shows the last value of α after the end of the cycle.

$$f(x) = w^T x + b \quad (1)$$

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad (2)$$

We tagged 159 Chinese sentences and 2357 English sentences as positive or negative for our training data. The entropy comparison factors α and α' were tested from 1.25 to 6 in intervals of 0.25. We applied this SVC to classify the rest of our data collection. Subsequently, the positive and negative sentence counts shown in Table 2 result from applying our SVC for classification.

Table 8 English and Chinese comparison of the top 10 positive keywords

| Price range | Chinese keyword | Counts in Ctrip | English keyword | Counts in Tripadvisor |
|----------------------|----------------------|-----------------|-----------------|-----------------------|
| 0: All Prices | 不错 (not bad) | 12,892 | Good | 19148 |
| | 大 (big) | 9844 | Staff | 16289 |
| | 干净 (clean) | 6665 | Great | 16127 |
| | 交通 (traffic) | 6560 | Location | 11838 |
| | 早餐 (breakfast) | 5605 | Nice | 11615 |
| | 近 (near) | 5181 | Clean | 9064 |
| | 地铁 (subway) | 4321 | Helpful | 5846 |
| | 购物 (shopping) | 4101 | Excellent | 5661 |
| | 推荐 (recommend) | 3281 | Comfortable | 5625 |
| | 环境 (environment) | 3258 | Friendly | 5606 |
| 3: 5000–10,000 yen | 不错 (not bad) | 139 | Good | 206 |
| | 干净 (clean) | 114 | Staff | 181 |
| | 早餐 (breakfast) | 112 | Clean | 174 |
| | 大 (big) | 76 | Nice | 166 |
| | 交通 (traffic) | 72 | Great | 143 |
| | 地铁 (subway) | 66 | Location | 91 |
| | 近 (near) | 55 | Comfortable | 79 |
| | 地铁站 (subway station) | 51 | Helpful | 70 |
| | 远 (far) | 41 | Friendly | 64 |
| | 附近 (nearby) | 34 | Recommend | 59 |
| 4: 10,000–15,000 yen | 不错 (not bad) | 601 | Good | 1399 |
| | 干净 (clean) | 455 | Staff | 1165 |
| | 大 (big) | 348 | Great | 961 |
| | 近 (near) | 323 | Nice | 808 |
| | 早餐 (breakfast) | 270 | Location | 800 |
| | 卫生 (health) | 201 | Clean | 656 |
| | 交通 (traffic) | 196 | Excellent | 412 |
| | 地铁 (subway) | 164 | Friendly | 400 |
| | 远 (far) | 158 | Helpful | 393 |
| | 附近 (nearby) | 150 | Comfortable | 391 |
| 5: 15,000–20,000 yen | 不错 (not bad) | 1925 | Good | 2242 |
| | 干净 (clean) | 1348 | Staff | 1674 |
| | 大 (big) | 1277 | Great | 1414 |
| | 交通 (traffic) | 1058 | Clean | 1204 |
| | 近 (near) | 1016 | Nice | 1175 |
| | 地铁 (subway) | 801 | Location | 1109 |
| | 早餐 (breakfast) | 777 | Comfortable | 621 |
| | 地铁站 (subway station) | 639 | Friendly | 615 |
| | 附近 (nearby) | 572 | Free | 581 |
| | 购物 (shopping) | 516 | Helpful | 552 |

Table 8 (continued)

| Price range | Chinese keyword | Counts in Ctrip | English keyword | Counts in Tripadvisor |
|------------------------|-------------------|-----------------|-----------------|-----------------------|
| 6: 20,000–30,000 yen | 不错 (not bad) | 3110 | Good | 6550 |
| | 大 (big) | 2245 | Staff | 5348 |
| | 交通 (traffic) | 1990 | Great | 5074 |
| | 干净 (clean) | 1940 | Location | 4414 |
| | 近 (near) | 1433 | Nice | 3451 |
| | 地铁 (subway) | 1073 | Clean | 3364 |
| | 早餐 (breakfast) | 1007 | Shopping | 1992 |
| | 购物 (shopping) | 979 | Helpful | 1970 |
| | 周边 (surroundings) | 837 | Comfortable | 1941 |
| | 附近 (nearby) | 825 | Friendly | 1915 |
| 7: 30,000–50,000 yen | 不错 (not bad) | 2291 | Good | 3407 |
| | 大 (big) | 1913 | Staff | 2867 |
| | 干净 (clean) | 1159 | Great | 2620 |
| | 交通 (traffic) | 1105 | Location | 2186 |
| | 近 (near) | 935 | Nice | 2160 |
| | 早餐 (breakfast) | 846 | Clean | 1750 |
| | 推荐 (recommend) | 638 | Helpful | 1147 |
| | 购物 (shopping) | 636 | Train | 1040 |
| | 周边 (surroundings) | 552 | Subway | 1034 |
| | 环境 (environment) | 541 | Friendly | 1001 |
| 8: 50,000–100,000 yen | 不错 (not bad) | 4451 | Great | 4425 |
| | 大 (big) | 3670 | Good | 4350 |
| | 早餐 (breakfast) | 2422 | Staff | 3777 |
| | 交通 (traffic) | 2012 | Nice | 2991 |
| | 购物 (shopping) | 1764 | Location | 2439 |
| | 新 (new) | 1634 | Clean | 1655 |
| | 棒 (great) | 1626 | Excellent | 1555 |
| | 地铁 (subway) | 1604 | Helpful | 1313 |
| | 干净 (clean) | 1577 | Comfortable | 1246 |
| | 近 (near) | 1354 | Friendly | 1238 |
| 9: 100,000–200,000 yen | 不错 (not bad) | 375 | Great | 1488 |
| | 大 (big) | 315 | Staff | 1277 |
| | 棒 (great) | 189 | Good | 994 |
| | 早餐 (breakfast) | 171 | Nice | 864 |
| | 环境 (environment) | 157 | Location | 799 |
| | 交通 (traffic) | 127 | Excellent | 631 |
| | 选择 (select) | 112 | Beautiful | 455 |
| | 推荐 (recommend) | 109 | Large | 404 |
| | 赞 (awesome) | 101 | Helpful | 401 |
| | 购物 (shopping) | 98 | Wonderful | 372 |

Table 9 English and Chinese comparison of the top 10 negative keywords

| Price range | Chinese keyword | Counts in Ctrip | English keyword | Counts in Tripadvisor |
|----------------------|-----------------|-----------------|-----------------|-----------------------|
| 0: All Prices | 价格 (price) | 1838 | Pricey | 462 |
| | 一般 (general) | 1713 | Poor | 460 |
| | 中文 (Chinese) | 733 | Dated | 431 |
| | 地理 (geography) | 691 | Disappointing | 376 |
| | 距离 (distance) | 434 | Worst | 327 |
| | 陈旧 (obsolete) | 319 | Minor | 258 |
| | 老 (old) | 297 | Uncomfortable | 253 |
| | 华人 (Chinese) | 15 | Carpet | 240 |
| | | | Annoying | 220 |
| | | | Sense | 220 |
| 3: 5000–10,000 yen | 价格 (price) | 31 | Worst | 6 |
| | 一般 (general) | 28 | Walkway | 5 |
| | 距离 (distance) | 11 | Unable | 4 |
| | 地理 (geography) | 10 | Worse | 4 |
| | 中文 (Chinese) | 9 | Annoying | 3 |
| | 老 (old) | 2 | Dirty | 3 |
| | | | Funny smell | 3 |
| | | | Poor | 3 |
| | | | Renovation | 3 |
| | | | Carpet | 2 |
| 4: 10,000–15,000 yen | 价格 (price) | 98 | Dated | 40 |
| | 一般 (general) | 91 | Poor | 29 |
| | 距离 (distance) | 43 | Disappointing | 26 |
| | 陈旧 (obsolete) | 34 | Worst | 24 |
| | 地理 (geography) | 31 | Uncomfortable | 23 |
| | 老 (old) | 30 | Cigarette | 22 |
| | 中文 (Chinese) | 26 | Pricey | 22 |
| | | | Minor | 21 |
| | | | Paper | 19 |
| | | | Unable | 19 |
| 5: 15,000–20,000 yen | 价格 (price) | 296 | Poor | 57 |
| | 一般 (general) | 218 | Dated | 41 |
| | 地理 (geography) | 125 | Disappointing | 38 |
| | 中文 (Chinese) | 93 | Annoying | 36 |
| | 距离 (distance) | 84 | Worst | 36 |
| | 陈旧 (obsolete) | 43 | Cigarette | 31 |
| | 老 (old) | 26 | Rude | 28 |
| | 华人 (Chinese) | 3 | Uncomfortable | 26 |
| | | | Paper | 25 |
| | | | Pricey | 24 |

Table 9 (continued)

| Price range | Chinese keyword | Counts in Ctrip | English keyword | Counts in Tripadvisor |
|------------------------|-----------------|-----------------|-----------------|-----------------------|
| 6: 20,000–30,000 yen | 一般 (general) | 504 | Poor | 136 |
| | 价格 (price) | 472 | Dated | 131 |
| | 地理 (geography) | 164 | Pricey | 120 |
| | 中文 (Chinese) | 155 | Disappointing | 112 |
| | 距离 (distance) | 116 | Uncomfortable | 103 |
| | 陈旧 (obsolete) | 75 | Minor | 93 |
| | 老 (old) | 55 | Smallest | 88 |
| | 华人 (Chinese) | 2 | Worst | 86 |
| | | | Cigarette | 79 |
| | | | Annoying | 70 |
| 7: 30,000–50,000 yen | 价格 (price) | 326 | Poor | 92 |
| | 一般 (general) | 311 | Pricey | 92 |
| | 地理 (geography) | 110 | Dated | 65 |
| | 中文 (Chinese) | 94 | Worst | 64 |
| | 陈旧 (obsolete) | 71 | Carpet | 55 |
| | 距离 (distance) | 68 | Uncomfortable | 55 |
| | 老 (old) | 45 | Dirty | 51 |
| | 华人 (Chinese) | 2 | Disappointing | 50 |
| | | | Cigarette | 46 |
| | | | Unable | 43 |
| 8: 50,000–100,000 yen | 价格 (price) | 561 | Pricey | 163 |
| | 一般 (general) | 510 | Dated | 150 |
| | 中文 (Chinese) | 337 | Disappointing | 129 |
| | 地理 (geography) | 239 | Poor | 124 |
| | 老 (old) | 134 | Worst | 98 |
| | 距离 (distance) | 97 | Walkway | 82 |
| | 陈旧 (obsolete) | 90 | Carpet | 71 |
| | 华人 (Chinese) | 8 | Minor | 63 |
| | | | Sense | 63 |
| | | | Outdated | 58 |
| 9: 100,000–200,000 yen | 价格 (price) | 54 | Pricey | 40 |
| | 一般 (general) | 51 | Sense | 34 |
| | 中文 (Chinese) | 19 | Minor | 33 |
| | 距离 (distance) | 15 | Lighting | 20 |
| | 地理 (geography) | 12 | Disappointing | 19 |
| | 陈旧 (obsolete) | 6 | Poor | 19 |
| | 老 (old) | 5 | Annoying | 16 |
| | | | | Mixed |
| | | | Disappointment | 14 |
| | | | Paper | 14 |

5 Data analysis

5.1 Frequent keywords in differently priced hotels

We observed the top 10 satisfaction and dissatisfaction keywords with the highest frequencies of emotionally positive and negative statements to study. The keywords are the quantitative rank of the needs of Chinese and English-speaking customers. We show the top 10 positive keywords for each price range comparing English and Chinese in Table 8. For the negative keywords, we show the results in Table 9.

We can observe that the most used keywords for most price ranges in the same language are similar, with a few changes in priority for the keywords involved. For example, in Chinese, we can see that the customers praise cleanliness first in cheaper hotels, whereas the size of the room or bed is praised more in hotels of higher class. Another example is that in negative English reviews, complaints about price appear only after 10,000 yen hotels. After this, it climbs in importance following the increase in the hotel's price.

5.2 Frequently used adjectives and their pairs

Some keywords in these lists are adjectives, such as the word “大 (big)” mentioned before. To understand those, we performed the dependency parsing and part of speech tagging explained in Sect. 4.2. While many of these connections, we only considered the top 4 used keyword connections per adjective per price range. We show the most used Chinese adjectives in positive keywords in Table 10, and for negative Chinese adjective keywords in Table 11. Similarly, for English adjectives used in positive sentences we show the most common examples in Table 12, and for adjectives used in negative sentences in Table 13.

5.3 Determining hard and soft attribute usage

To further understand the differences in satisfaction and dissatisfaction in Chinese and Western customers of Japanese hotels, we classified these factors as either hard or soft attributes of a hotel. We define hard attributes as matters regarding the hotel's physical or environmental aspects, such as facilities, location, or infrastructure. Some of these aspects would be impractical for the hotel to change, such as its surroundings and location. Others can be expensive to change, such as matters requiring construction costs, which are possible but would require significant infrastructure investment. On the other hand, soft attributes are the non-physical attributes of the hotel service and staff behavior that are practical to change through management. For example, the hotel's services or the cleanliness of the rooms are soft attributes. For our purposes, amenities, clean or good quality bed sheets or curtains, and other physical attributes that are part of the service and not the hotel's physical structure are considered soft attributes. Thus, we can observe the top 10 satisfaction and dissatisfaction keywords and determine whether they are soft or hard attributes.

Table 10 Top 4 words related to the mainly used adjectives in positive Chinese texts

| Price range | 不错 (not bad) | 大 (big) | 干净 (clean) | 近 (near) | 新 (new) | 棒 (great) |
|----------------------|------------------------------|----------------------------|--------------------------------|---------------------------------|---------------------------|----------------------------|
| 0: All Prices | 不错 (not bad): 12,892 | 大 (big): 9844 | 干净 (clean): 6665 | 近 (near): 5181 | 新 (new): 2775 | 棒 (great): 3028 |
| | 不错酒店 (nice hotel): 1462 | 大房间 (big room): 3197 | 干净房间 (clean room): 1224 | 近酒店 (near hotel): 453 | 新设施 (new facility): 363 | 棒酒店 (great hotel): 463 |
| | 不错位置 (nice location): 1426 | 大床 (big bed): 772 | 干净酒店 (clean hotel): 737 | 近桥 (near bridge): 144 | 新酒店 (new hotel): 246 | 棒位置 (great position): 218 |
| | 不错服务 (nice service): 869 | 大酒店 (big hotel): 379 | 干净卫生 (clean and hygienic): 464 | 近地铁站 (near subway station): 122 | 精装修 (new decoration): 116 | 棒服务 (great service): 168 |
| | 不错环境 (nice environment): 714 | 大超市 (big supermarket): 232 | 干净环境 (clean environment): 61 | 近站 (near station): 108 | 新房间 (new room): 53 | 棒早餐 (great breakfast): 164 |
| | 不错 (not bad): 139 | 大 (big): 76 | 干净 (clean): 114 | 近 (near): 55 | 棒 (great): 11 | 棒位置 (great position): 2 |
| | 不错酒店 (nice hotel): 17 | 大房间 (big room): 11 | 干净房间 (clean room): 21 | 近酒店 (near hotel): 4 | 棒位置 (great position): 2 | |
| | 不错位置 (nice location): 16 | 大床 (big bed): 10 | 干净酒店 (clean hotel): 10 | 近地铁 (near subway way): 2 | | |
| | 不错早餐 (nice breakfast): 12 | 大超市 (big supermarket): 5 | 干净卫生 (clean and hygienic): 6 | | | |
| | 不错服务 (nice service): 8 | 大商场 (big market): 3 | 干净总体 (clean overall): 4 | | | |
| 4: 10,000–15,000 yen | 不错 (not bad): 601 | 大 (big): 348 | 干净 (clean): 455 | 近 (near): 323 | 新 (new): 37 | 棒 (great): 73 |
| | 不错位置 (nice location): 72 | 大房间 (big room): 76 | 干净房间 (clean room): 66 | 近酒店 (near hotel): 27 | 新设施 (new facility): 9 | 棒位置 (great position): 6 |
| | 不错酒店 (nice hotel): 37 | 大床 (big bed): 30 | 干净卫生 (clean and hygienic): 52 | 近站 (near station): 14 | 精装修 (new decoration): 2 | 棒房间 (great room): 3 |
| | 不错服务 (nice service): 34 | 大社 (big club): 26 | 干净酒店 (clean hotel): 48 | 近地铁 (near subway way): 12 | 新酒店 (new hotel): 2 | 棒水平 (great level): 3 |
| | 不错早餐 (nice breakfast): 26 | 大空间 (big space): 16 | 干净打扫 (clean up): 9 | 近车站 (near station): 10 | | 棒温泉 (great hot spring): 3 |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

Table 10 (continued)

| Price range | 不错 (not bad) | 大 (big) | 干净 (clean) | 近 (near) | 新 (new) | 棒 (great) |
|----------------------|------------------------------|---------------------------|--------------------------------|--------------------------------|--------------------------|---------------------------|
| 5: 15,000–20,000 yen | 不错 (not bad): 1925 | 大 (big): 1277 | 干净 (clean): 1348 | 近 (near): 1016 | 新 (new): 234 | 棒 (great): 241 |
| | 不错位置 (nice location): 207 | 大房间 (big room): 316 | 干净房间 (clean room): 234 | 近酒店 (near hotel): 82 | 新设施 (new facility): 47 | 棒位置 (great position): 33 |
| | 不错酒店 (nice hotel): 168 | 大床 (big bed): 140 | 干净酒店 (clean hotel): 161 | 近站 (near station): 35 | 新酒店 (new hotel): 25 | 棒酒店 (great hotel): 25 |
| | 不错服务 (nice service): 131 | 大超市 (big supermarket): 73 | 干净卫生 (clean and hygienic): 92 | 近地铁站 (near subway station): 34 | 新装修 (new decoration): 15 | 棒服务 (great service): 22 |
| | 不错早餐 (nice breakfast): 109 | 大酒店 (big hotel): 49 | 干净设施 (clean facilities): 19 | 近桥 (near bridge): 29 | 新房间 (new room): 10 | 棒早餐 (great breakfast): 8 |
| 6: 20,000–30,000 yen | 不错 (not bad): 3110 | 大 (big): 2245 | 干净 (clean): 1940 | 近 (near): 1433 | 新 (new): 517 | 棒 (great): 440 |
| | 不错位置 (nice location): 409 | 大房间 (big room): 680 | 干净房间 (clean room): 360 | 近酒店 (near hotel): 164 | 新设施 (new facility): 89 | 棒酒店 (great hotel): 51 |
| | 不错酒店 (nice hotel): 326 | 大床 (big bed): 198 | 干净酒店 (clean hotel): 203 | 近地铁 (near subway): 34 | 新酒店 (new hotel): 51 | 棒位置 (great position): 45 |
| | 不错服务 (nice service): 206 | 大酒店 (big hotel): 102 | 干净卫生 (clean and hygienic): 137 | 近地铁站 (near subway station): 31 | 新装修 (new decoration): 24 | 棒服务 (great service): 23 |
| | 不错环境 (nice environment): 183 | 大空间 (big space): 64 | 干净环境 (clean environment): 21 | 近车站 (near the station): 27 | 新房间 (new room): 10 | 棒早餐 (great breakfast): 20 |

Table 10 (continued)

| Price range | 不错 (not bad) | 大 (big) | 干净 (clean) | 近 (near) | 新 (new) | 棒 (great) |
|-----------------------|---|--------------------------------------|--|---------------------------------------|--|---|
| 7: 30,000–50,000 yen | 不错 (not bad): 2291 不错位置 (nice location): 277 | 大 (big): 1913 大房间 (big room): 643 | 干净 (clean): 1159 干净房间 (clean room): 224 | 近 (near): 935 近酒店 (near hotel): 80 | 新 (new): 260 新设施 (new facility): 63 | 棒 (great): 448 棒酒店 (great hotel): 68 |
| | 不错酒店 (nice hotel): 274 | 大床 (big bed): 141 | 干净酒店 (clean hotel): 146 | 近站 (near station): 24 | 新酒店 (new hotel): 25 | 棒位置 (great position): 34 |
| | 不错服务 (nice service): 140 | 大超市 (big supermarket): 74 | 干净卫生 (clean and hygienic): 71 | 近桥 (near bridge): 20 | 新装修 (new decoration): 15 | 棒服务 (great service): 24 |
| | 不错环境 (nice environment): 140 | 大酒店 (big hotel): 66 | 干净环境 (clean environment): 16 | 近山 (near mountain): 12 | 新房间 (new room): 11 | 棒早餐 (great breakfast): 14 |
| 8: 50,000–100,000 yen | 不错 (not bad): 4451 | 大 (big): 3670 | 干净 (clean): 1577 | 近 (near): 1354 | 新 (new): 1634 | 棒 (great): 1626 |
| | 不错酒店 (nice hotel): 587 | 大房间 (big room): 1340 | 干净房间 (clean room): 310 | 近酒店 (near hotel): 88 | 新设施 (new facility): 141 | 棒酒店 (great hotel): 281 |
| | 不错位置 (nice location): 415 | 大床 (big bed): 238 | 干净酒店 (clean hotel): 161 | 近桥 (near bridge): 76 | 新酒店 (new hotel): 123 | 棒早餐 (great breakfast): 112 |
| | 不错服务 (nice service): 328 | 大酒店 (big hotel): 144 | 干净卫生 (clean and hygienic): 101 | 近地铁站 (near subway station): 35 | 新装修 (new decoration): 57 | 棒位置 (great position): 96 |
| | 不错早餐 (nice breakfast): 251 | 大商场 (big market): 88 | 干净服务 (clean service): 13 | 近铁 (Kintetsu): 24 | 新高 (new): 22 | 棒服务 (great service): 86 |

Table 10 (continued)

| Price range | 不错 (not bad) | 大 (big) | 干净 (clean) | 近 (near) | 新 (new) | 棒 (great) |
|------------------------|--|--|---|--|--|--|
| 9: 100,000–200,000 yen | 不错 (not bad): 375 不错酒店 (nice hotel): 53 不错位置 (nice location): 30 不错环境 (nice environment): 27 不错服务 (nice service): 22 | 大 (big): 315 大房间 (big room): 131 大面积 (large area): 19 大床 (big bed): 15 大卫生间 (big toilet): 13 | 干净 (clean): 72 干净房间 (clean room): 9 干净酒店 (clean hotel): 8 干净卫生 (clean and hygienic): 5 | 近 (near): 65 近酒店 (near hotel): 8 近地铁站 (near subway station): 3 近市场 (near market): 3 | 新 (new): 77 新酒店 (new hotel): 19 新设施 (new facility): 13 新装修 (new decoration): 3 新位置 (new location): 2 | 棒 (great): 189 棒酒店 (great hotel): 36 棒体验 (great experience): 10 棒服务 (great service): 10 棒早餐 (great breakfast): 8 |

Table 11 Top 4 words related to the mainly used adjectives in negative texts

| Price range | 一般 (general) | 陈旧 (obsolete) | 老 (old) |
|----------------------|--|---|--|
| 0: All Prices | 一般 (general): 1713 一般设施 (general facilities): 137 一般服务 (general service): 115 一般酒店 (average hotel): 106 一般早餐 (average breakfast): 97 一般 (general): 28 一般设施 (general facilities): 5 一般早餐 (average breakfast): 3 一般味道 (general taste): 2 一般效果 (general effect): 2 一般 (general): 91 一般设施 (general facilities): 10 一般位置 (general location): 8 一般酒店 (average hotel): 6 一般早餐 (average breakfast): 5 一般 (general): 218 一般设施 (general facilities): 23 一般酒店 (average hotel): 21 一般早餐 (average breakfast): 14 一般卫生 (general hygiene): 8 一般 (general): 504 一般设施 (general facilities): 42 一般酒店 (average hotel): 37 一般服务 (general service): 34 一般早餐 (average breakfast): 21 | 陈旧 (obsolete): 319 陈旧设施 (obsolete facilities): 184 陈旧设备 (obsolete equipment): 18 陈旧房间 (outdated room): 10 陈旧酒店 (outdated hotel): 10 | 老 (old): 297 老酒店 (old hotel): 74 老设施 (old facility): 58 老店 (old shop): 15 老装修 (old decoration): 11 老 (old): 2 |
| 3: 5000–10,000 yen | | | |
| 4: 10,000–15,000 yen | | 陈旧 (obsolete): 34 陈旧设施 (obsolete facilities): 17 陈旧家具 (obsolete furniture): 2 陈旧设备 (obsolete equipment): 2 | 老 (old): 30 老酒店 (old hotel): 8 老设施 (old facility): 7 老建筑 (old building): 3 |
| 5: 15,000–20,000 yen | | 陈旧 (obsolete): 43 陈旧设施 (obsolete facilities): 25 陈旧设备 (obsolete equipment): 3 陈旧酒店 (outdated hotel): 2 | 老 (old): 26 老酒店 (old hotel): 11 老设施 (old facility): 7 老外观 (old appearance): 2 |
| 6: 20,000–30,000 yen | | 陈旧 (obsolete): 75 陈旧设施 (obsolete facilities): 42 陈旧设备 (obsolete equipment): 7 陈旧装修 (old decoration): 3 陈旧酒店 (outdated hotel): 2 | 老 (old): 55 老酒店 (old hotel): 9 老设施 (old facility): 8 老店 (old shop): 3 老房间 (old room): 3 |

Table 11 (continued)

| Price range | 一般 (general) | 陈旧 (obsolete) | 老 (old) |
|-----------------------|-------------------------------|--------------------------------|-------------------------|
| 7: 30,000–50,000 yen | 一般 (general): 311 | 陈旧 (obsolete): 71 | 老 (old): 45 |
| | 一般设施 (general facilities): 23 | 陈旧设施 (obsolete facilities): 43 | 老酒店 (old hotel): 11 |
| | 一般服务 (general service): 22 | 陈旧设备 (obsolete equipment): 5 | 老设施 (old facility): 7 |
| | 一般早餐 (average breakfast): 19 | 陈旧房间 (outdated room): 3 | 老店 (old shop): 3 |
| | 一般酒店 (average hotel): 15 | | 老房间 (old room): 2 |
| | 一般 (general): 510 | 陈旧 (obsolete): 90 | 老 (old): 134 |
| | 一般服务 (general service): 39 | 陈旧设施 (obsolete facilities): 53 | 老酒店 (old hotel): 34 |
| | 一般设施 (general facilities): 32 | 陈旧房间 (outdated room): 5 | 老设施 (old facility): 26 |
| | 一般早餐 (average breakfast): 30 | 陈旧感觉 (Stale feeling): 2 | 老装修 (old decoration): 9 |
| | 一般酒店 (average hotel): 25 | | 老店 (old shop): 7 |
| 8: 50,000–100,000 | 一般 (general): 51 | 陈旧 (obsolete): 6 | 老 (old): 5 |
| | 一般服务 (general service): 7 | 陈旧设施 (obsolete facilities): 4 | 老设施 (old facility): 2 |
| | 一般早餐 (average breakfast): 5 | | |
| | 一般位置 (general location): 2 | | |
| 9: 100,000 to 200,000 | 一般房间 (average room): 2 | | |

Table 12 Top 4 words related to the mainly used adjectives in positive English texts

| Price range | Good | Clean | Comfortable | Helpful | Free | Large | Friendly | Great |
|----------------------|-------------------------------------|---------------------------------|--|--------------------------------------|------------------------------|---------------------------------|--|---------------------------------------|
| 0: All Prices | Good: 19,148 Good location: 1985 | Clean: 9064 Clean room: 3596 | Comfortable: 5625 Comfortable bed: 1919 | Helpful: 5846 Helpful staff: 2927 | Free: 4318 Free wifi: 773 | Large: 4104 Large room: 1256 | Friendly: 5606 Friendly staff: 3819 | Great: 16,127 Great location: 2313 |
| | Good service: 1042 | Clean hotel: 969 | Comfortable room: 1098 | Helpful concierge: 304 | Free shuttle: 286 | Large hotel: 268 | Friendly service: 169 | Great view: 1099 |
| | Good breakfast: 942 | Clean bathroom: 282 | Comfortable stay: 272 | Helpful desk: 110 | Free drink: 234 | Large bathroom: 202 | Friendly hotel: 73 | Great service: 841 |
| | Good hotel: 874 | Clean everything: 200 | Comfortable hotel: 238 | Helpful service: 74 | Free bus: 225 | Larger room: 192 | Friendly person: 63 | Great hotel: 802 |
| 3: 5000–10,000 yen | Good: 206 Good location: 30 | Clean: 174 Clean room: 55 | Comfortable: 79 Comfortable bed: 21 | Helpful: 70 Helpful staff: 36 | Free: 35 Free wifi: 10 | Large: 31 Large room: 7 | Friendly: 64 Friendly staff: 53 | Great: 143 Great location: 21 |
| | Good value: 19 | Clean bathroom: 14 | Comfortable room: 9 | Free tea: 4 | Free every- one: 2 | Large area: 2 | Friendly every- one: 2 | Great view: 14 |
| | Good english: 10 | Clean place: 12 | Comfortable futon: 8 | Free raman: 2 | Friendly service: 2 | Large size: 2 | Friendly service: 2 | Great place: 13 |
| | Good place: 7 | Clean hotel: 6 | Comfortable stay: 3 | Free toothbrush: 2 | Great experi- ence: 5 | | | Great experi- ence: 5 |
| 4: 10,000–15,000 yen | Good: 1399 | Clean: 656 | Comfortable: 391 | Helpful: 393 | Free: 271 | Large: 250 | Friendly: 400 | Great: 961 |
| | Good location: 159 | Clean room: 247 | Comfortable bed: 123 | Helpful staff: 206 | Free wifi: 53 | Large room: 84 | Friendly staff: 292 | Great location: 158 |
| | Good breakfast: 87 | Clean hotel: 74 | Comfortable room: 90 | Helpful concierge: 20 | Free breakfast: 15 | Large bathroom: 20 | Friendly service: 15 | Great service: 51 |
| | Good hotel: 71 | Clean bathroom: 20 | Comfortable hotel: 26 | Helpful desk: 10 | Free service: 12 | Larger room: 12 | Friendly hotel: 7 | Great hotel: 43 |

Table 12 (continued)

| Price range | Good | Clean | Comfortable | Helpful | Free | Large | Friendly | Great |
|--------------------------|------------------------|---------------------------|--------------------------|-------------------------------|-----------------------|-----------------------|-------------------------|------------------------|
| 5: 15,000– 20,000 yen | Good service: 67 | Clean every- thing: 14 | Comfortable stay: 20 | Helpful service: 4 | Free drink: 11 | Large hotel: 10 | Friendly person: 6 | Great place: 35 |
| | Good: 2242 | Clean: 1204 | Comfortable: 621 | Helpful: 552 | Free: 581 | Large: 349 | Friendly: 615 | Great: 1414 |
| | Good location: 242 | Clean room: 440 | Comfortable bed: 219 | Helpful staff: 301 | Free wifi: 109 | Large room: 85 | Friendly staff: 444 | Great location: 199 |
| | Good hotel: 116 | Clean hotel: 133 | Comfortable room: 99 | Helpful desk: 11 | Free shuttle: 35 | Large suitcase: 18 | Friendly hotel: 12 | Great view: 81 |
| 6: 20,000– 30,000 yen | Good breakfast: 113 | Clean bathroom: 38 | Comfortable stay: 30 | Helpful con- cierge: 9 | Free bus: 30 | Larger room: 18 | Friendly service: 8 | Great hotel: 68 |
| | Good service: 108 | Clean every- thing: 26 | Comfortable hotel: 20 | Helpful recep- tion: 5 | Free breakfast: 27 | Large hotel: 17 | Friendly most: 7 | Great place: 61 |
| | Good: 6550 | Clean: 3364 | Comfortable: 1941 | Helpful: 1970 | Free: 1186 | Large: 1257 | Friendly: 1915 | Great: 5074 |
| | Good location: 703 | Clean room: 1379 | Comfortable bed: 658 | Helpful staff: 1019 | Free wifi: 269 | Large room: 329 | Friendly staff: 1311 | Great location: 881 |
| | Good service: 331 | Clean hotel: 379 | Comfortable room: 359 | Helpful con- cierge: 79 | Free breakfast: 68 | Large hotel: 87 | Friendly service: 51 | Great service: 249 |
| | Good english: 304 | Clean bathroom: 95 | Comfortable stay: 100 | Helpful desk: 42 | Free coffee: 57 | Larger room: 81 | Friendly person: 21 | Great hotel: 232 |
| | Good breakfast: 303 | Clean every- thing: 77 | Comfortable hotel: 82 | Helpful recep- tionist: 17 | Free drink: 38 | Large bed: 43 | Friendly hotel: 19 | Great view: 220 |

Table 12 (continued)

| Price range | Good | Clean | Comfortable | Helpful | Free | Large | Friendly | Great |
|-----------------------|---------------------|----------------------|-----------------------|------------------------|-------------------|--------------------|----------------------|---------------------|
| 7: 30,000–50,000 yen | Good: 3407 | Clean: 1750 | Comfortable: 1000 | Helpful: 1147 | Free: 933 | Large: 580 | Friendly: 1001 | Great: 2620 |
| | Good location: 380 | Clean room: 725 | Comfortable bed: 345 | Helpful staff: 607 | Free drink: 145 | Large room: 174 | Friendly staff: 715 | Great location: 393 |
| | Good breakfast: 191 | Clean hotel: 197 | Comfortable room: 193 | Helpful concierge: 53 | Free wifi: 129 | Larger room: 32 | Friendly service: 24 | Great view: 162 |
| | Good service: 182 | Clean bathroom: 61 | Comfortable hotel: 49 | Helpful service: 20 | Free coffee: 45 | Large hotel: 30 | Friendly hotel: 13 | Great hotel: 134 |
| | Good english: 155 | Clean everything: 36 | Comfortable stay: 47 | Helpful desk: 17 | Free bus: 38 | Large bed: 28 | Friendly person: 13 | Great service: 114 |
| 8: 50,000–100,000 yen | Good: 4350 | Clean: 1655 | Comfortable: 1246 | Helpful: 1313 | Free: 1072 | Large: 1233 | Friendly: 1238 | Great: 4425 |
| | Good location: 406 | Clean room: 648 | Comfortable bed: 425 | Helpful staff: 589 | Free shuttle: 181 | Large room: 442 | Friendly staff: 810 | Great location: 506 |
| | Good service: 296 | Clean hotel: 156 | Comfortable room: 266 | Helpful concierge: 108 | Free wifi: 172 | Large hotel: 109 | Friendly service: 51 | Great view: 436 |
| | Good hotel: 196 | Clean bathroom: 48 | Comfortable stay: 56 | Helpful service: 28 | Free bus: 127 | Large bathroom: 58 | Friendly hotel: 20 | Great service: 267 |
| | Good breakfast: 191 | Cleanliness: 40 | Comfortable hotel: 51 | Helpful desk: 26 | Free service: 65 | Larger room: 38 | Friendly person: 12 | Great hotel: 241 |

Table 12 (continued)

| Price range | Good | Clean | Comfortable | Helpful | Free | Large | Friendly | Great |
|-----------------------|--------------------|-----------------|-----------------------|-----------------------|--------------------|--------------------|----------------------|---------------------|
| 9,100,000–200,000 yen | Good: 994 | Clean: 261 | Comfortable: 347 | Helpful: 401 | Free: 240 | Large: 404 | Friendly: 370 | Great: 1488 |
| | Good location: 65 | Clean room: 102 | Comfortable bed: 128 | Helpful staff: 169 | Free wifi: 31 | Large room: 135 | Friendly staff: 194 | Great location: 155 |
| | Good service: 56 | Clean hotel: 24 | Comfortable room: 82 | Helpful concierge: 35 | Free breakfast: 19 | Large bathroom: 38 | Friendly service: 18 | Great view: 155 |
| | Good breakfast: 53 | Cleanliness: 8 | Comfortable stay: 16 | Helpful everyone: 7 | Free drink: 16 | Large hotel: 15 | Friendly everyone: 7 | Great service: 101 |
| | Good hotel: 40 | Clean place: 7 | Comfortable hotel: 10 | Helpful team: 5 | Free bus: 14 | Large bed: 12 | Friendly person: 4 | Great hotel: 80 |

Table 13 Top 4 words related to the mainly used adjectives in negative English texts

| Price range | Poor | Dated | Worst | Dirty | Uncomfortable |
|----------------------|----------------------|----------------------|----------------------|-------------------|-----------------------------|
| 0: All Prices | Poor: 460 | Dated: 431 | Worst: 327 | Dirty: 188 | Uncomfortable: 253 |
| | Poor service: 55 | Outdated: 128 | Worst hotel: 43 | Dirty carpet: 34 | uncomfortable bed: 63 |
| | Poor breakfast: 41 | Outdated room: 20 | Worst experience: 18 | Dirty room: 23 | Uncomfortable pillow: 20 |
| | Poor quality: 27 | Outdated hotel: 10 | Worst part: 15 | Not dirty: 7 | Uncomfortable mattress: 8 |
| | Poor english: 24 | Outdated bathroom: 7 | Worst service: 10 | Dirty bathroom: 6 | Uncomfortable night: 8 |
| | Poor: 3 | | Worst: 6 | Dirty: 3 | Uncomfortable: 2 |
| 3: 5000–10,000 yen | | | Worst room: 2 | | |
| | Poor: 29 | Dated: 40 | Worst: 24 | Dirty: 11 | Uncomfortable: 23 |
| | Poor breakfast: 3 | Outdated: 11 | Worst hotel: 4 | Dirty floor: 2 | Uncomfortable bed: 4 |
| | Poor service: 3 | Outdated decor: 2 | Worst experience: 2 | | Not uncomfortable: 2 |
| | Poor conditioning: 2 | Outdated room: 2 | | | Uncomfortable night: 2 |
| 5: 15,000–20,000 yen | Poor view: 2 | | | | Uncomfortable pillow: 2 |
| | Poor: 57 | Dated: 41 | Worst: 36 | Dirty: 14 | Uncomfortable: 26 |
| | Poor service: 10 | Outdated: 8 | Worst hotel: 8 | Dirty room: 2 | Uncomfortable bed: 7 |
| | Poor breakfast: 6 | | Worst experience: 3 | | Uncomfortable pillow: 2 |
| | Poor hotel: 5 | | Worst part: 2 | | |
| | Poor experience: 3 | | Worst service: 2 | | |
| 6: 20,000–30,000 yen | Poor: 136 | Dated: 131 | Worst: 86 | Dirty: 67 | Uncomfortable: 103 |
| | Poor breakfast: 15 | outdated: 31 | worst hotel: 11 | dirty room: 10 | Uncomfortable bed: 24 |
| | Poor service: 14 | Outdated room: 6 | Worst part: 7 | Dirty carpet: 8 | Uncomfortable pillow: 11 |
| | Poor english: 9 | Outdated hotel: 2 | Worst breakfast: 5 | Dirty bathroom: 3 | Uncomfortable night: 4 |
| | Poor quality: 9 | | Worst experience: 5 | Dirty chair: 2 | Uncomfortable experience: 3 |

Table 13 (continued)

| Price range | Poor | Dated | Worst | Dirty | Uncomfortable |
|------------------------|--------------------|-----------------------|---------------------|------------------|---------------------------|
| 7: 30,000–50,000 yen | Poor: 92 | Dated: 65 | Worst: 64 | Dirty: 51 | Uncomfortable: 55 |
| | Poor service: 8 | Outdated: 17 | Worst hotel: 10 | Dirty carpet: 11 | Uncomfortable bed: 20 |
| | Poor breakfast: 7 | Outdated hotel: 4 | Worst room: 3 | Dirty room: 7 | Uncomfortable mattress: 6 |
| | Poor english: 7 | Outdated bathroom: 2 | Worst service: 3 | Dirty clothe: 2 | Uncomfortable pillow: 5 |
| | Poor connection: 5 | Outdated decor: 2 | Worst part: 2 | Dirty luggage: 2 | Uncomfortable room: 5 |
| | Poor: 124 | Dated: 150 | Worst: 98 | Dirty: 36 | Uncomfortable: 33 |
| | Poor service: 16 | Outdated: 58 | Worst hotel: 9 | Dirty carpet: 12 | Uncomfortable bed: 7 |
| 8: 50,000–100,000 yen | Poor breakfast: 9 | Outdated room: 9 | Worst experience: 5 | Dirty room: 3 | |
| | Poor quality: 9 | Outdated furniture: 6 | Worst part: 3 | Dirty cup: 2 | |
| | Poor english: 6 | Outdated hotel: 4 | | Dirty rug: 2 | |
| | Poor: 19 | Dated: 3 | Worst: 12 | Dirty: 6 | Uncomfortable: 8 |
| | Poor service: 4 | Outdated: 2 | Worst experience: 2 | | Little uncomfortable: 2 |
| | Poor choice: 2 | | | | |
| | Poor experience: 2 | | | | |
| 9: 100,000–200,000 yen | | | | | |

Table 14 Determination of hard and soft attributes for Chinese keywords

| Keyword Emotion | Keyword | Attribute Category |
|-------------------|----------------|-----------------------------------|
| Positive Keywords | 不错 | 50% hard, 25% soft, 25% undefined |
| | 大 | 100% hard |
| | 干净 | 25% hard, 75% soft |
| | 早餐 | 100% soft |
| | 交通 | 100% hard |
| | 棒 | 25% hard, 50% soft, 25% undefined |
| | 近 | 100% hard |
| | 购物 | 100% hard |
| | 环境 | 100% hard |
| | 地铁 | 100% hard |
| | 卫生 | 100% soft |
| | 新 | 50% hard, 25% soft, 25% undefined |
| | 推荐 | 100% undefined |
| | 选择 | 100% undefined |
| | 地铁站 | 100% hard |
| | 远 | 100% hard |
| | 附近 | 100% hard |
| 周边 | 100% hard | |
| 赞 | 100% undefined | |
| Negative Keywords | 价格 | 100% soft |
| | 一般 | 50% hard, 50% soft |
| | 中文 | 100% soft |
| | 距离 | 100% hard |
| | 地理 | 100% hard |
| | 陈旧 | 100% hard |
| | 老 华人 | 75% hard, 25% soft 100% soft |

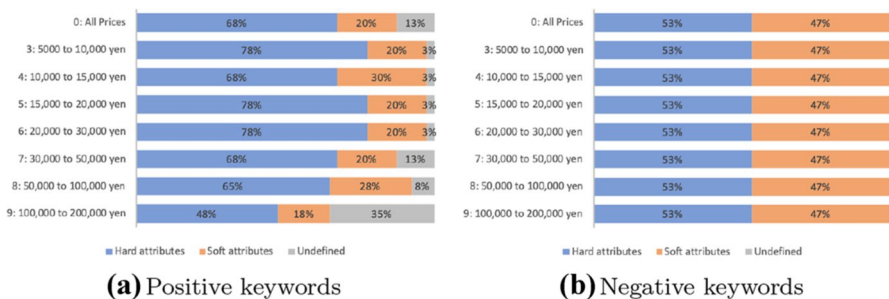


Fig. 5 Hard and soft attributes from the top Chinese keywords for all price ranges

Table 15 Determination of hard and soft attributes for English keywords

| Keyword Emotion | Keyword | Attribute Category | |
|-------------------|-------------------------|-----------------------------------|-----------------------------------|
| Positive Keywords | Good | 25% hard, 50% soft, 25% undefined | |
| | Great | 50% hard, 25% soft, 25% undefined | |
| | Staff | 100% soft | |
| | Clean | 100% soft | |
| | Location | 100% hard | |
| | Nice | 50% hard, 25% soft, 25% undefined | |
| | Excellent | 25% hard, 50% soft, 25% undefined | |
| | Helpful | 100% soft | |
| | Comfortable | 25% hard, 50% soft, 25% undefined | |
| | Shopping | 100% hard | |
| | Beautiful | 25% hard, 75% soft | |
| | Friendly | 100% soft | |
| | Train | 100% hard | |
| | Large | 100% hard | |
| | Free | 100% soft | |
| | Subway | 100% hard | |
| | Recommend | 100% undefined | |
| | Wonderful | 50% soft, 50% undefined | |
| | Negative Keywords | Pricey | 100% soft |
| | | Worst | 25% hard, 50% soft, 25% undefined |
| Dated | | 75% hard, 25% undefined | |
| Poor | | 100% soft | |
| Walkway | | 100% hard | |
| Sense | | 100% undefined | |
| Unable | | 100% soft | |
| Disappointing | | 50% soft, 50% undefined | |
| Minor | | 100% undefined | |
| Worse | | 100% undefined | |
| Annoying | | 75% hard, 25% undefined | |
| Lighting | | 100% soft | |
| Uncomfortable | | 100% soft | |
| Carpet | | 100% soft | |
| Dirty | | 75% soft, 25% undefined | |
| Cigarette | | 100% soft | |
| Funny smell | | 100% soft | |
| Rude | | 100% soft | |
| Smallest | | 75% hard, 25% undefined | |
| Mixed | | 100% undefined | |
| Renovation | 100% hard | | |
| Paper | 100% undefined | | |
| Disappointment | 100% undefined | | |
| Outdated | 75% hard, 25% undefined | | |

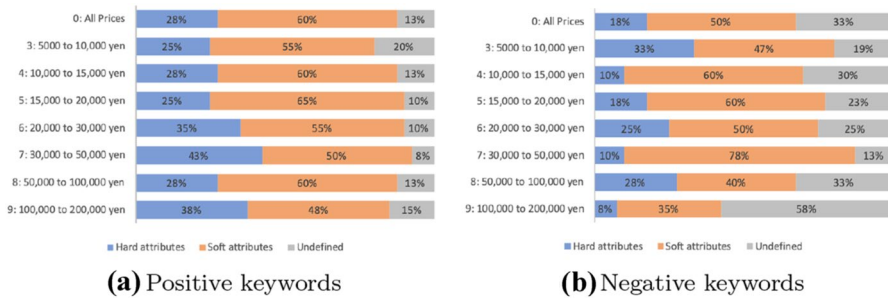


Fig. 6 Hard and soft attributes from the top English keywords for all price ranges

We manually labeled each language's top keywords into either hard or soft by considering how the word would be used when writing a review. If the word described unchangeable physical factors by the staff or management, we consider them hard. If the word implied an issue that could be solved or managed by the hotel staff or management, we consider it soft. For adjectives, we looked at the top four adjective and noun pairings used in the entire dataset and counted the usage percentage in each context. If it was not clear from the word or the pairing alone, we declared it undefined. Then, we added the counts of these words in each category. A single word with no pairing is always deemed 100% in the category it corresponds to. We add the partial percentages for each category when an adjective includes various contexts. The interpretation of these keywords is shown in the Tables 14 and 15. We can see the summarized results for the hard and soft percentages of positive and negative Chinese keywords in Fig. 5. For the English keywords, see Fig. 6.

6 Results

6.1 Experimental results and answers to research questions

Our research questions were related to two issues. Based on research questions 1a and 1b, the objective of this study was to determine the differences in how Chinese and Western tourists perceive Japanese hotels, whose hospitality and service are influenced by the *omotenashi* culture.

Observing the top-ranking positive keywords in Chinese reviews, as shown in Tables 8 and 10, it was revealed that, while service, cleanliness, and breakfast were praised in most hotels, the location was more important when observing the pairings. Hard attributes were abundant lower on the lists. The negative keywords in Table 9 indicate that a lack of a Chinese-friendly environment was perceived, although there were more complaints about hard attributes such as the building's age and the distance from other convenient spots. However, most complaints were about the hotel's price, which included all of the price ranges; therefore, the price was the primary concern for Chinese customers with different travel purposes.

On the other hand, the word “staff” is the second or third in the lists of satisfaction factors in English-written reviews in all the price ranges. This word is followed by a few other keywords lower in the top 10 list, such as “helpful” or “friendly”. When we look at the pairings of the top-ranked keyword “good” in Table 12, we find that customers mostly praise the location, service, breakfast, or English availability. When we look at the negative keyword “poor” and its pairings in Table 13, we see that it is also service-related concepts that the Western tourists are disappointed with.

We can also observe some keywords that are not considered by their counterparts. For example, English-speaking customers mentioned tobacco smell in many reviews. However, it was not statistically identified as a problem for their Chinese counterparts. On the other hand, although they appear in both English and Chinese lists, references to “购物 (shopping)” are more common in the Chinese lists across hotels of 15,000–200,000 yen per night. Meanwhile, the term “shopping” appeared solely in the top 10 positive keywords list for English speakers who stayed in rooms priced 20,000–30,000 yen per night.

With these results, we can observe that both Chinese and English-speaking tourists in Japan have different priorities. However, both populations consider the hotel’s location and transport availability (subways and trains) nearby as secondary but still essential points in their satisfaction with a hotel. The Chinese customers are primarily satisfied with the room quality in spaciousness and cleanliness and the service of breakfast.

For research questions 2a and 2b, we considered how customers of both cultural backgrounds evaluated the hard and soft attributes of hotels. Our study discovered that Chinese tourists mostly positively react to the hotel’s hard attributes, albeit the negative evaluations are more uniform than the positive evaluations, with a tendency of 53% towards hard attributes. On the other hand, English-speaking tourists were more responsive to soft attributes, either positively or negatively. In the case of negative keywords, they were more concerned about the hotel’s soft attributes.

One factor that both populations had in common is that, when perceiving the hotel negatively, the “老 (old),” “dated,” “outdated,” or “陈旧 (obsolete)” aspects of the room or the hotel were surprisingly criticized across most price ranges. However, this is a hard attribute and is unlikely to change for most hotels.

6.2 Chinese tourists: a big and clean space

We found that mainland Chinese tourists were mainly satisfied by big and clean spaces in Japanese hotels. The adjectival pairings extracted with dependency parsing and POS tagging (Table 10) imply big and clean rooms. Other mentions included big markets nearby or a big bed. Across different price ranges, the usage of the word “大 (big)” increased with the increasing price of the hotel. When inspecting closer by taking random samples of the pairs of “大空间 (big space)” or “大面积 (large area),” we notice that there were also many references to the public bathing facilities

in the hotel. Such references were also implied by a word pairing “棒 温泉 (great hot spring).”

In Japan, there are the so-called “銭湯 (*sentō*),” which are artificially constructed public bathing facilities, including saunas and baths with unique qualities. On the other hand, there are natural hot springs, called “温泉 (*onsen*).” However, they are interchangeable if natural hot spring water is used in artificially made tiled bath facilities. It is a Japanese custom that all customers first clean themselves in a shower and afterward use the baths nude. It could be a cultural shock for many tourists but a fundamental attraction for many others.

Chinese customers are satisfied with the size of the room or bed; however, it is not trivial to change this. In contrast, cleanliness is mostly related to soft attributes when we observe its adjectival pairings. We can observe pairs such as “干净 房间 (clean room)” at the top rank of all price ranges and thereupon “干净 酒店 (clean hotel),” “干净 总体 (clean overall),” “干净 环境 (clean environment),” and “干净 设施 (clean facilities),” among other examples. In negative reviews, there was a mention of criticizing the “一般 卫生 (general hygiene)” of the hotel, although it was an uncommon pair. Therefore, we can assert that cleanliness was an important soft attribute for Chinese customers, and they were mostly pleased when their expectations were fulfilled.

A key soft satisfaction factor was the inclusion of breakfast within the hotel. While other food-related words were extracted, most of them were general, such as “food” or “eating,” and were lower-ranking. In contrast, the word “早餐 (breakfast),” referring to the hotel commodities, was frequently used in positive texts compared to other food-related words across all price ranges, albeit at different priorities in each of them. For this reason, we regard it as an important factor. From the word pairs of the positive Chinese keywords in Table 10, we can also note that “不错 (not bad)” is paired with “不错 早餐 (nice breakfast)” in four of the seven price ranges with reviews available as part of the top four pairings. It is only slightly lower in other categories, although it is not depicted in the table. Thus, we consider that a recommended strategy for hotel management is to invest in the inclusion or improvement of hotel breakfast to increase good reviews.

6.3 Western tourists: a friendly face and absolutely clean

From the satisfaction factors of English-speaking tourists, we observed at least three words were related to staff friendliness and services in the general database: “staff,” “helpful,” and “friendliness.” The word “staff” is the highest-ranked of these three, ranking second for satisfied customers across most price ranges and only third in one of them. The word “good” mainly refers to the location, service, breakfast, or English availability in Table 12. Similar to Chinese customers, Western customers also seemed to enjoy the included breakfasts regarding their satisfaction keyword pairings. However, the relevant word does not appear in the top 10 list directly, in contrast to their Chinese counterparts. The words “helpful” and “friendly” are mostly paired with “staff,” “conciierge,” “desk,” and “service.” By considering the negative

keyword “poor” and its pairings in Table 13, we realized once again that Western tourists were disappointed with service-related concepts and reacted negatively.

Another soft attribute that is high on the list for most of the price ranges is the word “clean”, so we examined its word pairings. Customers largely praised “clean rooms” and “clean bathrooms” and also referred to the hotel in general. When observing the negative keyword frequencies for English speakers, we can find words such as “dirty” and “carpet” as well as word pairings such as “dirty carpet,” “dirty room,” and “dirty bathroom.” Along with complaints about off-putting smells, we could conclude that Western tourists had high expectations about cleanliness when traveling in Japan.

An interesting detail of the keyword ranking is that the word “comfortable” was high on the satisfaction factors, and “uncomfortable” was high on the dissatisfaction factors. The words were paired with nouns such as “bed,” “room,” “pillow,” and “mattress,” when they generally referred to their sleep conditions in the hotel. It seems that Western tourists were particularly sensitive about the hotels’ comfort levels and whether they reached their expectations. The ranking for the negative keyword “uncomfortable” is similar across most price ranges except the two most expensive ones, where this keyword disappears from the top 10 list.

Albeit lower in priority, the price range of 15,000–20,000 yen hotels also includes “free” as one of the top 10 positive keywords, mainly paired with “Wi-Fi.” This price range corresponds to business hotels, where users would expect this feature the most.

6.4 Tobacco, an unpleasant smell in the room

A concern for Western tourists was uncleanliness and the smell of cigarettes in their room, which can be regarded as soft attributes. Cigarette smell was an issue even in the middle- and high-class hotels, of which the rooms were priced at more than 30,000 yen per night. For hotels with rooms priced above 50,000 yen per night, however, this problem seemed to disappear from the list of top 10 concerns. Tobacco was referenced singularly as “cigarette”, but also in word pairs in Table 13 as “funny smell.” By manually inspecting a sample of reviews with this keyword, we noticed that the room was often advertised as non-smoking; however, the smell permeated the room and curtains. Another common complaint was that there were no non-smoking facilities available. The smell of smoke can completely ruin some customers’ stay, leading to bad reviews, thereby lowering the number of future customers.

In contrast, Chinese customers seemed not to be bothered by this. Previous research has stated that 49–60% of Chinese men (and 2.0–2.8% of women) currently smoke or smoked in the past. This was derived from a sample of 170,000 Chinese adults in 2013–2014, which is high compared to many English-speaking countries (Zhang et al. 2019; World Health Organization 2015).

Japan has a polarized view on the topic of smoking. Although it has one of the world’s largest tobacco markets, tobacco use has decreased in recent years. Smoking in public spaces is prohibited in some wards of Tokyo (namely Chiyoda, Shinjuku, and Shibuya). However, it is generally only suggested and not mandatory to

lift smoking restrictions in restaurants, bars, hotels, and public areas. Many places have designated smoking rooms to keep the smoke in an enclosed area and avoid bothering others.

Nevertheless, businesses, especially those who cater to certain customers, are generally discouraged by smoking restrictions if they want to maintain their clientele. To cater to all kinds of customers, including Western and Asian, Japanese hotels must provide spaces without tobacco smell. Even if the smoke does not bother a few customers, the lack of such a smell will make it an appropriate space for all customers.

6.5 Location, location, location

The hotel's location, closeness to the subway and public transportation, and availability of nearby shops proved to be of importance to both Chinese and English-speaking tourists. In positive word pairings in Tables 10 and 12, we can find pairs such as “不错 位置 (nice location),” “近 地铁站 (near subway station),” “近 地铁 (near subway)” in Chinese texts and “good location,” “great location,” and “great view” as well as single keywords “location” and “shopping” for English speakers, and “交通 (traffic),” “购物 (shopping),” “地铁 (subway),” and “环境 (environment or surroundings)” for Chinese speakers. All of these keywords and their location in each population's priorities across the price ranges signify that the hotel's location was a secondary but still important point for their satisfaction. However, since this is a hard attribute, it is not often considered in the literature. By examining examples from the data, we recognized that most customers were satisfied if the hotel was near at least two of the following facilities: subway, train, and convenience stores.

Japan is a country with a peculiar public transportation system. During rush hour, the subway is crowded with commuters, and trains and subway stations create a confusing public transportation map for a visitor in Tokyo. Buses are also available, albeit less used than rail systems in metropolitan cities. These three means of transportation are usually affordable in price. There are more expensive means, such as the bullet train *shinkansen* for traveling across the country and taxis. The latter is a luxury in Japan compared to other countries. In Japan, taxis provide a high-quality experience with a matching price. Therefore, for people under a budget, subway availability and maps or GPS applications, as well as a plan to travel the city, are of utmost necessity for tourists, using taxis only as a last resort.

Japanese convenience stores are also famous worldwide because they offer a wide range of services and products, from drinks and snacks to full meals, copy and scanning machines, alcohol, cleaning supplies, personal hygiene items, underwear, towels, and international ATMs. If some trouble occurs, or a traveler forgot to pack a particular item, it is most certain that they can find it.

Therefore, considering that both transportation systems and nearby shops are points of interest for Chinese and Western tourists, and perhaps offering guide maps and information about these as an appeal point could result in greater satisfaction.

7 Discussion

7.1 Western and Chinese tourists in the Japanese hospitality environment

To date, scholars have been correcting our historical bias towards the West. Studies have determined that different cultural backgrounds lead to different expectations, which influence tourists' satisfaction. In other words, tourists of a particular culture have different leading satisfaction factors across different destinations. However, Japan presents a particular environment; the spirit of hospitality and service, *omotenashi*, which is considered to be of the highest standard across the world. Our study explores whether such an environment can affect different cultures equally or whether it is attractive only to certain cultures.

Our results indicate that Western tourists are more satisfied with soft attributes than Chinese tourists. As explained earlier in this paper, Japan is well known for its customer service. Respectful language and bowing are not exclusive to high-priced hotels or businesses; these are met in convenience stores as well. Even in the cheapest convenience store, the level of hospitality is starkly different from Western culture and perhaps unexpected. In higher-priced hotels, the adjectives used to praise the service ranged from normal descriptors like "good" to higher levels of praise like "wonderful staff," "wonderful experience," "excellent service," and "excellent staff." Furthermore, Kozak (2002) and Shanka and Taylor (2004) have also proven that hospitality and staff friendliness are two determinants of Western tourists' satisfaction.

However, the negative English keywords indicate that a large part of the dissatisfaction with Japanese hotels stemmed from a lack of hygiene and room cleanliness. Although Chinese customers had solely positive keywords about cleanliness, English-speaking customers deemed many places unacceptable to their standards, particularly hotels with rooms priced below 50,000 yen per night. The most common complaint regarding cleanliness was about the carpet, followed by complaints about cigarette smell and lack of general hygiene. Kozak (2002) also proved that hygiene and cleanliness were essential satisfaction determinants for Western tourists. However, in the previous literature, this was linked merely to satisfaction. In contrast, our research revealed that words related to cleanliness were mostly linked to dissatisfaction. We could assert that Westerners had a high standard of room cleanliness compared to their Chinese counterparts.

According to previous research, Western tourists are already inclined to appreciate hospitality for their satisfaction. When presented with Japanese hospitality, this expectation is met and overcome. In contrast, according to our results, Chinese tourists were more concerned about room quality rather than hospitality, staff, or service. However, when analyzing the word pairs for "不错 (not bad)" and "棒 (great)," we can see that they praise staff, service, and breakfast. By observing the percentage of hard to soft attributes in Fig. 5, however, we discover that Chinese customers were more satisfied with hard attributes compared to Western tourists, who seemed to be meeting more than their expectations.

It could be considered that Chinese culture does not expect high-level service initially. When an expectation that is not held is met, the satisfaction derived is less than that if it was expected. In contrast, some tourists report a “nice surprise”: when an unknown need is unexpectedly met, there is more satisfaction. It is necessary to note the difference between these two reactions. The “nice surprise” reaction fulfills a need unexpectedly. Perhaps the hospitality grade in Japan does not fulfill a need high enough for the Chinese population, thereby resulting in less satisfaction. For greater satisfaction, a need must be met. However, the word “not bad” is at the top of the list in most price ranges, and one of the uses is related to service. Thus, we cannot conclude that they were not satisfied with the service. Instead, they held other factors at a higher priority; thus, the keyword frequency was higher for other pairings.

Another possibility occurs when we observe the Chinese tourists’ dissatisfaction factors. Chinese tourists may have expectations about their treatment that are not being met, even in this high-standard hospitality environment. This could be because Japan is monolingual and has a relatively large language barrier to tourists (Heinrich 2012; Coulmas and Watanabe 2002). While the Japanese effort to accommodate English speakers is slowly developing, efforts for Chinese accommodations can be lagging. Chinese language pamphlets and Chinese texts on instructions for the hotel room and its appliances and features (e.g., T.V. channels, Wi-Fi setup, etc.), or the treatment towards Chinese people, could be examples of these accommodations. Ryan and Mo (2001) also found that communication difficulty was one of the main reasons Chinese customers would state for not visiting again. However, this issue is not exclusive to Japan.

Our initial question was whether the environment of high-grade hospitality would affect both cultures equally. This study attempted to determine the answer. It is possible that Chinese customers had high-grade hospitality and were equally satisfied with Westerners. In that case, it appears that the difference in perception stems from a psychological source; expectation leads to satisfaction and a lack of expectation results in lesser satisfaction. There is also a possibility that Chinese customers are not receiving the highest grade of hospitality because of cultural friction between Japan and China.

It is unclear which of these two is most likely from our results. However, competing in hospitality and service includes language services, especially in the international tourism industry. Better multilingual support can only improve the hospitality standard in Japan. Considering that most of the tourists in Japan come from other countries in Asia, multilingual support is beneficial. Proposals for this endeavor include hiring Chinese-speaking staff, preparing pamphlets in Chinese, or having a translator application readily available with staff trained in interacting through an electronic translator.

7.2 Hard vs. soft satisfaction factors

As stated in Sect. 3.2, previous research has mostly focused on the hotel’s soft attributes and their influence on customer satisfaction (e.g., Shanka and Taylor 2004; Choi and Chu 2001). Examples of soft attributes include staff behavior,

commodities, amenities, and appliances that can be improved within the hotel. However, hard attributes are not usually analyzed in satisfaction studies. It is important to consider both kinds of attributes. If the satisfaction was based on soft attributes, a hotel can improve its services to attract more customers in the future. Otherwise, if the satisfaction was related more to hard attributes overall, hotels should be built considering the location while minimizing other costs. Because the satisfaction factors were decided statistically in our study via customers' online reviews, we can see the importance of the hard or soft attributes in their priorities.

Figure 5 shows that, in regards to Chinese customer satisfaction, in general, 68% of the top 10 keywords are hard factors; in contrast, only 20% are soft factors. The rates are similar for most price ranges except the highest-priced hotels. However, two of these soft attributes are all concentrated at the top of the list (“不错 (not bad),” “干净 (clean)”), and the adjective pairs related to soft attributes of “不错 (not bad)” are also at the top in most price ranges. Chinese tourists may expect spaciousness and cleanliness when coming to Japan. The expectation may be due to reputation, previous experiences, or cultural backgrounds. We can compare these results with previous literature, where traveling Chinese tourists choose their destination based on several factors, including cleanliness, nature, architecture, and scenery (Ryan and Mo 2001). These factors found in previous literature could be linked to the keyword “环境 (environment or surroundings)” as well. This keyword was found for hotels priced at more than 20,000 yen per night.

In contrast, English speakers are mostly satisfied with the hotels' soft attributes. Figure 6 shows that soft attributes are above 48% in all price ranges, the highest being 65% in the price range of 15,000–20,000 yen per night, which corresponds to, for example, affordable business hotels. The exception to this is the hard attribute that is the hotels' location, which is consistently around the middle of the top 10 lists for all price ranges.

For both customer groups, the main reason for dissatisfaction was pricing, which can be interpreted as a concern about value for money. However, English-speaking customers complained less about the price in lower-priced hotels. In contrast, Chinese customers consistently had “价格 (price)” as the first or second-most concern across all price ranges. A study on Chinese tourists found that they had this concern (Truong and King 2009). However, our results indicate that this has more to do with the pricing of hotels in Japan than with Chinese culture. In general, Japan is an expensive place to visit, thereby impacting this placement in the ranking. Space is scarce in Japan, and capsule hotels with cramped spaces of 2×1 m cost around 3000–6000 yen per night. Bigger business hotel rooms are relatively expensive, ranging from 5000 to 12,000 yen per night. For comparison, hotels in the USA with a similar quality can charge half the price.

Around half of the dissatisfaction factors for both Chinese and Western customers are caused by issues that could be improved; this is true for all price ranges. The improvements could be staff training (perhaps in language), hiring professional cleaning services for rooms with cigarette smoke smells, or improving the bedding; however, these considerations can be costly. However, once the hotel's location and construction are set, only a few changes can be made to satisfy Chinese customers

further. As mentioned previously, Chinese language availability is a soft attribute that can be improved with staff and training investment.

Western tourists are mainly dissatisfied with soft attributes. This is revealed by a low satisfaction level of 35% in the highest price range where undefined factors are the majority and a maximum of 78% in the price range of 30,000–50,000 yen per night in a hotel. Improvement scope for Western tourists is more extensive than that for their Chinese counterparts. As such, it presents a larger investment opportunity.

7.3 Satisfaction across different price ranges

In previous sections of this paper, we mentioned the differences reflected in hotel price ranges. The most visible change across differently priced hotels is the change in voice when describing satisfaction. We noticed this by observing the adjective-noun pairs and finding pairs with different adjectives for the same nouns. For example, in English, words describing nouns such as “location” or “hotel” are “good” or “nice” in lower-priced hotels. In contrast, the adjectives that pair with the same nouns for higher-priced hotels are “wonderful” and “excellent.” In Chinese, the change ranges from “不错 (not bad)” to “棒 (great)” or “赞 (awesome).” We can infer that the level of satisfaction is high and influences how customers write their reviews. Regarding the negative keywords, however, the change ranges from “annoying” or “disappointing” to “worst.”

In this paper, we follow the definition of satisfaction by Hunt (1975), where meeting or exceeding expectations produces satisfaction. Conversely, the failure of meeting expectations causes dissatisfaction. We can assume that a customer that pays more for a higher-class experience has higher expectations. For example, in a highly-priced hotel, any lack of cleanliness can lead to disappointment. In the case of English-speaking customers in the 30,000–50,000 yen per night price range, cigarette smell is particularly disappointing. However, we consistently see customers with high expectations for high-class hotels reacting even more positively when satisfied. In the positive case, expectations appear to be exceeded in most cases, judging from their reactions.

We argue that these are two different kinds of expectations: logical and emotional. In the first case, customers are determined that the service must not fall below a specific standard; for example, they can be disappointed with unhygienic rooms or cigarette smell. In contrast, in the second case, customers have a vague idea of having a positive experience but do not measure it against any standard. For example, they expect a pleasant customer service experience or a hospitable treatment by the staff at a high-class hotel. Regardless of their knowledge in advance, positive emotions offer them a perception of exceeded expectations and high satisfaction. Thus, hospitality and service enhance the experience of the customers.

There are interesting differences between Chinese and English-speaking tourists in their satisfaction to differently priced hotels. For example, Chinese tourists have “购物 (shopping)” as a top keyword in all the price ranges. In contrast, English-speaking tourists mention it only as a top keyword in the 20,000–30,000 yen price range. It is widely known in Japan that many Chinese tourists visit Japan for

shopping. Tsujimoto (2017) analyzed the souvenir purchasing behavior of Chinese tourists in Japan and showed that common products besides food and drink are: electronics, cameras, cosmetics, and medicine, among *souvenir* items representative of the culture or places that they visit (Japan Tourism Agency 2014). Furthermore, Chinese tourists' choice to shop in Japan is more related to the quality of the items rather than their relation to the tourist attractions. Our results suggested that Western tourists were engaging more in tourist attractions rather than shopping activities compared to Chinese tourists.

Another interesting difference is that English-speaking tourists start using negative keywords about the hotel's price only if it concerns hotels of 15,000 yen or more; thereafter, the more expensive the hotel, the higher the ranking. In contrast, for Chinese customers, this keyword is the top keyword across all price ranges. Previous research suggests that value for money is a key concern for Chinese and Asian tourists (Choi and Chu 2000, 2001; Truong and King 2009), whereas Western customers are more concerned about hospitality (Kozak 2002).

While some attributes' value changes depending on the hotel's price range, some other attributes remain constant for each culture's customers. For example, appreciation for staff from English-speaking tourists is ranked close to the top satisfaction factor in all the price ranges. Satisfaction for cleanliness by both cultures constantly remains part of the top 10 keywords, except for the most expensive one, where other keywords replace keywords related to satisfaction or cleanliness in the ranking; however, they remain still high on the list. Chinese tourists have a high ranking for the word “早餐 (breakfast)” across all price ranges as well. As discussed in Sect. 6.5, transportation and location are also important for hotels of all classes and prices. While the ranking of attributes might differ between price ranges, hard and soft attribute proportions also appear to be constant within a 13% margin of error per attribute. This suggests that, from a cultural aspect, the customers have a particular bias to consider some attributes more than others.

7.4 Cross-culture analysis of expectations and satisfaction

The basic premise of this study is that different cultures lead to different expectations and satisfaction factors. This premise also plays a role in the differentiation between the preferences of hard or soft attributes.

In Donthu and Yoo (1998), subjects from 10 different countries were compared with respect to their expectations of service quality and analyzed based on Hofstede's typology of culture (Hofstede 1984). The previous study states that, although culture has no specific index, five dimensions of culture can be used to analyze or categorize a country in comparison to others. These are *power distance*, *uncertainty avoidance*, *individualism—collectivism*, *masculinity—femininity*, and *long-term—short-term orientation*. In each of these dimensions, at least one element of service expectations was found to be significantly different for countries grouped under contrasting attributes (e.g., individualistic countries vs. collectivist countries, high uncertainty avoidance countries vs. low uncertainty avoidance countries).

However, Hofstede's typology has received criticism from academics, particularly for the fifth dimension that Hofstede proposed, which was later added with the alternative name *Confucian dynamic*. Academics with a Chinese background criticized Hofstede for being misinformed on the philosophical aspects of Confucianism as well as considering a difficult dimension to measure (Fang 2003). Other models, such as the GLOBE model, also consider some of Hofstede's dimensions and replace them with others, making a total of nine dimensions (House et al. 1999). The *masculinity–femininity* dimension, for example, is proposed to be instead of two dimensions: *gender egalitarianism* and *assertiveness*. This addition of dimensions avoids assuming that assertiveness is either masculine or feminine, which stems from outdated gender stereotypes. Such gender stereotypes have also been the subject of critique on Hofstede's model (Jeknić 2014). We agree with these critiques and thus avoid considering such stereotypes in our discussion.

For our purposes of contrasting Western vs. Chinese satisfaction stemming from expectations, these dimensions could explain why Chinese customers are generally satisfied more often with hard factors while Westerners are satisfied or dissatisfied with soft factors.

The backgrounds of collectivism in China and individualism in Western countries have been studied previously (Gao et al. 2017; Kim and Lee 2000). These backgrounds as well as the differences in these cultural dimensions could be the underlying cause for differences in expectations. Regardless of the cause, however, measures in the past have proven that such differences exist (Armstrong et al. 1997).

The cultural background of Chinese tourists emphasizes their surroundings and their place in nature and the environment. Chinese historical backgrounds of Confucianism, Taoism, and Buddhism permeate the thought processes of Chinese populations. However, scholars argue that the changes in generations and their economic and recent history attaches less importance to these concepts in their lives (Gao et al. 2017). Nevertheless, one could argue a Chinese cultural attribute emphasizes that the environment and the location affect satisfaction rather than the treatment they receive.

A more anthropocentric and individualistic Western culture could correlate more of their expectations and priorities to the treatment in social circumstances rather than the environment. According to Donthu and Yoo (1998), highly individualistic customers, in contrast to collectivist customers, have a higher expectation of empathy and assurance from the provider, which are aspects of service, a soft attribute of a hotel.

Among other dimensions in both models, we can consider uncertainty avoidance. Customers of high uncertainty avoidance carefully plan their travel and thus have higher expectations towards service. In contrast, customers of lower uncertainty avoidance do not take risks in their decisions and thus face less disappointment with different expectations. However, according to Xiumei and Jinying (2011), the difference between China and the USA in uncertainty avoidance is not clear when measuring with the Hofstede typology and the GLOBE typology. While the USA is not representative of Western society, uncertainty avoidance may not cause the difference in hard-soft attribute satisfaction between Chinese and Western cultures. Differences in another factor, power distance, were also noted when using Hofstede's

method compared to the GLOBAL method; therefore, power distance was not considered for comparison.

7.5 Implications for hotel managers

Our study reached two important conclusions: one about hospitality and cultural differences and another about managerial decisions towards two different populations. Overall, Chinese tourists did not attach much importance to hospitality and service factors. Instead, they focused on the hard attributes of a hotel. In particular, they were not satisfied with hospitality as much as Western tourists were; otherwise, they felt that basic language and communication needs were not met; thereby, they were not much satisfied. Western tourists were highly satisfied with Japanese hospitality and preferred soft attributes to hard ones.

The other conclusion is that managerial decisions could mostly benefit Western tourists, except for language improvements and breakfast inclusion could satisfy both groups. As mentioned earlier in this paper, Westerners are “long-haul” customers, spending more of their budget on lodging than Asian tourists (Choi and Chu 2000). With bigger returns on managerial improvements, we recommend investing in improving attributes that dissatisfy Western customers, such as cleanliness and removing tobacco smell. In addition, breaking the language barrier is one of the few strategies to satisfy both groups. Recently, Japan has been facing an increase in Chinese students as well as students of Western universities. Hiring students as part-time workers could increase the language services of a hotel.

To satisfy both customer types, hotel managers need to invest in cleanliness, deodorizing, and making hotel rooms tobacco-free. It could also be recommended to invest in breakfast inclusion and multilingual services and staff preparedness to deal with Chinese and English speakers. Western tourists were also observed to have high comfort standards, which could be managerially improved for better reviews. Perhaps it could be suggested to perform surveys of the bedding that is most comfortable for Western tourists. However, not all hotels can invest in all of these factors simultaneously. Our results suggest that satisfying cleanliness needs could satisfy both customer types. We suggest investing in making the facilities tobacco-free. Our results are also divided by price ranges; thereby, a hotel manager could consider which analysis suits their hotel the most. Hard attributes are difficult to change; however, improvements in service can be made to accompany these attributes. For example, transportation guides for foreigners that might not know the area could increase satisfaction.

The managers must consider their business model for implementing the next strategy. One option could be attracting more Chinese customers with their observed low budgeting. Another could be attracting more big-budget Western customers. For example, investing more in cleanliness could improve Western customers looking for high-quality lodging satisfaction, even for an increased price per night. On the other hand, hotels might be deemed costly by Chinese customers wherever such an investment is made.

8 Limitations and future work

In this study, we analyzed keywords based on whether they appeared on satisfied reviews or dissatisfied ones. Following that, we attempted to understand these words' context by using a dependency parser and observing the related nouns. However, a limitation is that it analyzed solely the words directly related to each keyword and did not search for further connections. This means that if the words were used in combination with other keywords, we did not trace the effects of multiple contradicting statements. For example, in the sentence "The room is good, but the food is lacking," we extracted "good room" and "lacking food" but did not consider the fact that both occurred in the same sentence.

This study analyzed the differences in customers' expectations at different levels of hospitality and service factors by dividing our data into price ranges. However, in the same price range, for example, the highest one, we can find both a Western-style five-star resort and a high-end Japanese style *ryokan*. Services offered in these hotels are of high quality, albeit very different. Nevertheless, most of our database was focused on the middle range priced hotels, the services of which are comparably less varied.

An essential aspect of this study is that we focused on the satisfaction and dissatisfaction towards the expectations of individual aspects of the hotels. This gave us insight into the factors that hotel managers can consider. However, each customer's overall satisfaction was not measured since it would require methods that are out of the scope of this paper. Another limitation is that further typology analysis could not be made because of the nature of the data collected (for example, Chinese men and women of different ages or their Westerner counterparts).

In future work, we plan to investigate these topics further. We plan to extend our data to research different trends and regions of Japan, different kinds of hotels, and customers traveling alone or in groups, whether for fun or for work. Another point of interest in this study's future work is to use word clusters with similar meanings instead of single words.

9 Conclusion

In this study, we analyzed the differences in satisfaction and dissatisfaction between Chinese and English-speaking customers of Japanese hotels, particularly in the context of Japanese hospitality, *omotenashi*. We extracted keywords from their online reviews on *Ctrip* and *TripAdvisor* using Shannon's entropy calculations. We used these keywords for sentiment classification via an SVC. We then used dependency parsing and part of speech tagging to extract common pairs of adjectives and nouns as well as single words. We divided these data by sentiment and hotel price range (most expensive room/night).

We found that Western tourists were most satisfied with staff behavior, cleanliness, and other soft attributes. However, Chinese customers had other concerns for their satisfaction; they were more inclined to praise the room, location, and hotel's convenience. We found that the two cultures had different reactions to the hospitality

environment and the prices. Thus, we discussed two possible theories on why Chinese tourists responded differently from Westerners in the environment of *omotenashi*. One theory is that, although they were treated well, their experience was deteriorated by language or culture barriers. The second possible theory is that they reacted to hospitality differently since they did not have the same expectations. We theorized that a lack of expectations could result in lessened satisfaction than that to the same service if expected. On the other hand, even when they held high expectations in a high-priced hotel, Japanese hospitality exceeded Western tourists' expectations, judging by their vocabulary for expressing their satisfaction. We considered that Western tourists were more reactive to hospitality and service factors Chinese tourists.

Lastly, we measured the satisfaction and dissatisfaction factors, that is, a hotel's hard and soft attributes. Hard attributes are physical and environmental elements, and as such, are impractical elements to change. In contrast, soft attributes can be changed via management and staff by an improvement in services or amenities. We found that, for satisfaction, Western tourists favored soft attributes in contrast to Chinese tourists, who were more interested in the hard attributes of hotels across all the price ranges consistently. For dissatisfaction, Western tourists were also highly inclined to criticize soft attributes, such as cleanliness or cigarette smell in rooms. In contrast, Chinese tourists' dissatisfaction derived from both hard and soft attributes evenly.

One approach for hotel managers is to work to satisfy Chinese tourists more, who dedicate a lower percentage of their budget to hotels but are more numerous. They are less satisfied with soft attributes but have an identifiable method for improving satisfaction by lessening language barriers and providing a satisfactory breakfast. Another approach was focused on the cleanliness, comfort, and tobacco-free space expected by Western tourists. "Long-haul" Western tourists, who spend almost half of their budget on hotels with this strategy, were favored. Although Westerners are less in number than Chinese tourists, it could be proven that they have more substantial returns. This is because Chinese customers also favor cleanliness as a satisfaction factor, and both populations could be pleased.

Acknowledgements Acknowledgements During our research, we received the commentary and discussion by our dear colleagues necessary to understand particular cultural aspects that could influence the data's interpretation. We would like to show gratitude to Mr. Liangyuan Zhou, Ms. Min Fan, and Ms. Eerdengqiqige for this.

We would also like to show gratitude to Ms. Aleksandra Jajus, from whom we also received notes on the editing and commentary on the content of our manuscript.

Author Contributions HN and TH conceived of the presented idea. HN and EAC developed the theory. All members participated in planning the experiments. EAC and HME performed data collection, the experiments, and computations. EAC wrote the manuscript with support from HN and HME. EAC and HN participated in the discussion of results. T. Hiraoka supervised the project.

Funding This work was supported by the Japan Construction Information Center Foundation (JACIC).

Declarations

Conflict of interest None.

References

- Aishima T, Sato Y et al (2015) The origin of Japanese Omotenashi in Man-yo-shu. *Bus Acc Rev* 16:103–122
- Al-alsheikh A, Sato Y (2015) Characteristics of the hospitality, Omotenashi in the traditional Japanese inn: a case study of Kagaya. *Bus Acc Rev* 16:123–142
- Al-Smadi M, Al-Ayyoub M, Jararweh Y, Qawasmeh O (2018) Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features. *Inf Process Manag*. <https://doi.org/10.1016/j.ipm.2018.01.006>
- Alemán Carreón EC, Nonaka H, Hiraoka T, Kumano M, Ito T, Hirota M (2018) Emotional contribution analysis of online reviews. In: *Proceedings of the 2018 International Conference on artificial life and robotics (ICAROB2018)*, Beppu, Japan, vol 23, pp 359–362. <https://doi.org/10.5954/ICAROB.2018.OS5-3>
- Armstrong RW, Mok C, Go FM, Chan A (1997) The importance of cross-cultural expectations in the measurement of service quality perceptions in the hotel industry. *Int J Hosp Manag* 16(2):181–190
- Balbi S, Misuraca M, Scepi G (2018) Combining different evaluation systems on social media for measuring user satisfaction. *Inf Process Manag* 54(4):674–685. <https://doi.org/10.1016/j.ipm.2018.04.009>
- Basuroy S, Chatterjee S, Ravid S (2003) How critical are critical reviews? the box office effects of film critics, star power, and budgets. *J Mark* 67(4):103–117. <https://doi.org/10.1509/jmkg.67.4.103.18692>
- Bauer T, Jago L, Wise B (1993) The changing demand for hotel facilities in the Asia Pacific region. *Int J Hosp Manag* 12(4):313–322. [https://doi.org/10.1016/0278-4319\(93\)90048-E](https://doi.org/10.1016/0278-4319(93)90048-E)
- Berezina K, Bilgihan A, Cobanoglu C, Okumus F (2016) Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. *J Hosp Mark Manag* 25(1):1–24. <https://doi.org/10.1080/19368623.2015.983631>
- Browning V, So KKF, Sparks B (2013) The influence of online reviews on consumers' attributions of service quality and control for service standards in hotels. *J Travel Tour Mark* 30(1–2):23–40. <https://doi.org/10.1080/10548408.2013.750971>
- Chan A, Hsu CH, Baum T (2015) The impact of tour service performance on tourist satisfaction and behavioral intentions: a study of Chinese tourists in Hong Kong. *J Travel Tour Mark* 32(1–2):18–33. <https://doi.org/10.1080/10548408.2014.986010>
- Chang P, Galley M, Manning C (2008) Optimizing Chinese word segmentation for machine translation performance. In: *Proceedings of the Third Workshop On Statistical Machine (Statmt'08)*, Columbus, Ohio, USA, pp 224–232. <http://nlp.stanford.edu/pubs/acl-wmt08-cws.pdf>. Accessed Apr 2019
- Chang R, Kivela J, Mak A (2010) Food preferences of Chinese tourists. *Ann Tour Res* 37(4):989–1011. <https://doi.org/10.1016/j.annals.2010.03.007>
- Chen D, Manning CD (2014) A fast and accurate dependency parser using neural networks. In: *Proceedings of the 2014 Conference on empirical methods in natural language processing (EMNLP)*, pp 740–750
- Choi TY, Chu R (2000) Levels of satisfaction among Asian and Western travellers. *Int J Qual Reliab Manag* 17(2):116–132. <https://doi.org/10.1108/02656710010304537>
- Choi TY, Chu R (2001) Determinants of hotel guests' satisfaction and repeat patronage in the Hong Kong hotel industry. *Int J Hosp Manag* 20(3):277–297. [https://doi.org/10.1016/S0278-4319\(01\)00006-8](https://doi.org/10.1016/S0278-4319(01)00006-8). <http://www.sciencedirect.com/science/article/pii/S0278431901000068>
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. <https://doi.org/10.1007/bf00994018>
- Coulmas F, Watanabe M (2002) Japans nascent multilingualism. *Contrib Sociol Lang* 87:249–274
- de Marneffe MC, Manning CD (2008) Stanford typed dependencies manual. Tech. rep., Stanford University. https://nlp.stanford.edu/software/dependencies_manual.pdf
- Dongyang Z, Mori T, Hayashi K et al (2015) A study on preferences and behavioral patterns of Chinese tourists in Kansai region, Japan. *Konan Econ Papers* 55(1–2):31–46. <https://doi.org/10.14990/00001507>
- Donthu N, Yoo B (1998) Cultural influences on service quality expectations. *J Serv Res* 1(2):178–186
- Engel J, Blackwell R, Miniard P (1990) *Consumer behavior*, 6th edn. Dryden Press, Hinsdale
- Fang T (2003) A critique of Hofstede's fifth national culture dimension. *Int J Cross Cult Manag* 3(3):347–368

- Fang B, Ye Q, Kucukusta D, Law R (2016) Analysis of the perceived value of online tourism reviews: influence of readability and reviewer characteristics. *Tour Manag* 52:498–506. <https://doi.org/10.1016/j.tourman.2015.07.018>
- Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. *AI Mag* 17(3):37–37. <https://doi.org/10.1609/aimag.v17i3.1230>. <https://www.aaai.org/ojs/index.php/aimagazine/article/download/1230/1131>
- Francesco G, Roberta G (2019) Cross-country analysis of perception and emphasis of hotel attributes. *Tour Manag* 74:24–42. <https://doi.org/10.1016/j.tourman.2019.02.011>. <http://www.sciencedirect.com/science/article/pii/S0261517719300408>
- Gao J, Zhang C, Huang Z (2017) Chinese tourists' views of nature and natural landscape interpretation: a generational perspective. *J Sustain Tour* 26(4):668–684. <https://doi.org/10.1080/09669582.2017.1377722>
- Hargreaves C (2015) Analysis of hotel guest satisfaction ratings and reviews: an application in Singapore. *Am J Mark Res* 1(4):208–214
- Heinrich P (2012) The making of monolingual Japan: language ideology and Japanese modernity, vol 146. In: Multilingual matters
- Hofstede G (1984) Culture's consequences: international differences in work-related values, vol 5. Sage, London
- House RJ, Hanges PJ, Ruiz-Quintanilla SA, Dorfman PW, Javidan M, Dickson M, Gupta V et al (1999) Cultural influences on leadership and organizations: Project GLOBE. *Adv Glob Leader* 1(2):171–233
- Hu YH, Chen YL, Chou HL (2017) Opinion mining from online hotel reviews—a text summarization approach. *Inf Process Manag* 53(2):436–449. <https://doi.org/10.1016/j.ipm.2016.12.002>
- Huang J (2017) The dining experience of Beijing roast duck: A comparative study of the Chinese and English online consumer reviews. *Int J Hosp Manag* 66:117–129. <https://doi.org/10.1016/j.ijhm.2017.07.003>. <http://www.sciencedirect.com/science/article/pii/S0278431917301251>
- Hunt JD (1975) Image as a factor in tourism development. *J Travel Res* 13(3):1–7
- Ikeda N (2013) Omotenashi: Japanese hospitality as the global standard. *Manag Serv Bus Jpn World Sci Chap* 9:145–154
- Japan National Tourism Organization (2019) Nationality/monthly foreign visitors to Japan (2003–2019). Tech. rep., Japan National Tourism Organization, https://www.jnto.go.jp/jpn/statistics/since2003_visitor_arrivals.pdf, (in Japanese)
- Japan Tourism Agency (2014) Consumption trend survey for foreigners visiting Japan
- Jeknić R (2014) Gender equality, young women and culture in the context of the "masculinity/femininity" as a dimension in Geert Hofstede's model of "national culture".. *Zbornik radova Pravng fakulteta u Splitu* 51(3):681–696
- Jia SS (2020) Motivation and satisfaction of Chinese and U.S. tourists in restaurants: a cross-cultural text mining of online reviews. *Tour Manag* 78:104071. <https://doi.org/10.1016/j.tourman.2019.104071>. <http://www.sciencedirect.com/science/article/pii/S0261517719302687>
- Jones T, Nagata S, Nakajima M, Masuyama K (2009) Prefectural branding in Japan—tourism, national parks and the Shinshu brand. *Place Brand Public Dipl* 5(3):192–201. <https://doi.org/10.1057/pb.2009.13>
- Kim C, Lee S (2000) Understanding the cultural differences in tourist motivation between Anglo-American and Japanese tourists. *J Travel Tour Mark* 9(1–2):153–170. https://doi.org/10.1300/J073v09n01_09
- Kim K, jung Park O, Yun S, Yun H (2017) What makes tourists feel negatively about tourism destinations? application of hybrid text mining methodology to smart destination management. *Technol Forecast Soc Change* 123:362–369. <https://doi.org/10.1016/j.techfore.2017.01.001>
- Kozak M (2002) Measuring tourist satisfaction with multiple destination attributes. *Tour Anal* 7(3–4):229–240. <https://doi.org/10.3727/108354203108750076> (12)
- Kuboyama T (2020) "Omotenashi" must comprise hospitality and service. In: *Serviceology for services*, Springer Singapore, pp 34–53. https://doi.org/10.1007/978-981-15-3118-7_3
- Levy SE (2010) The hospitality of the host: a cross-cultural examination of managerially facilitated consumer-to-consumer interactions. *Int J Hosp Manag* 29(2):319–327. <https://doi.org/10.1016/j.ijhm.2009.10.003>. <https://www.sciencedirect.com/science/article/pii/S027843190900111X>, special issue on Event Studies
- Likert R (1932) A technique for the measurement of attitudes. *Arch Psychol* 22:55

- Liu BSC, Furrer O, Sudharshan D (2001) The relationships between culture and behavioral intentions toward services. *J Serv Res* 4(2):118–129
- Liu Y, Bi JW, Fan ZP (2017) Ranking products through online reviews: a method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Inf Fusion* 36:149–161. <https://doi.org/10.1016/j.inffus.2016.11.012>
- Liu Y, Huang K, Bao J, Chen K (2019) Listen to the voices from home: an analysis of Chinese tourists' sentiments regarding Australian destinations. *Tour Manag* 71:337–347. <https://doi.org/10.1016/j.tourman.2018.10.004>. <http://www.sciencedirect.com/science/article/pii/S0261517718302395>
- Loh S, Lorenzi F, Saldaña R, Licthnow D (2003) A tourism recommender system based on collaboration and text analysis. *Inf Technol Tour* 6(3):157–165. <https://doi.org/10.3727/1098305031436980>
- Manning CD, Surdeanu M, Bauer J, Finkel J, Bethard SJ, McClosky D (2014) The Stanford CoreNLP natural language processing toolkit. In: Proceedings of the Association for Computational Linguistics (ACL) System Demonstrations, pp 55–60. <http://www.aclweb.org/anthology/P/P14/P14-5010>
- Mattila AS (1999) The role of culture and purchase motivation in service encounter evaluations. *J Serv Mark* 13:376–389
- Mattila AS (2019) A commentary on cross-cultural research in hospitality & tourism inquiry (invited paper for luminaries special issue of International Journal of Hospitality Management). *Int J Hosp Manag* 76:10–12. <https://doi.org/10.1016/j.ijhm.2018.06.007>. <https://www.sciencedirect.com/science/article/pii/S027843191830505X>, HOSPITALITY LUMINARIES ISSUE
- Nahm UY, Mooney RJ (2002) Text mining with information extraction. In: Proceedings of the AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases, Stanford CA, pp 60–67
- O'Connor B, Balasubramanyan R, Routledge B, Smith N (2010) From tweets to polls: Linking text sentiment to public opinion time series. In: Proceedings of the Fourth International AAAI Conference On Weblogs And Social Media, 11(1-2), p 122-129
- Oh H, Parks SC (1996) Customer satisfaction and service quality: a critical review of the literature and research implications for the hospitality industry. *Hosp Res J* 20(3):35–64
- Oliver RL (1981) Measurement and evaluation of satisfaction processes in retail settings. *J Retail* 57(3):25–48
- Powers D (2011) Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal Of Machine Learning Technologies* 2(1):37–63. http://www.flinders.edu.au/science_engineering/fms/School-CSEM/publications/tech_reps-research_artfcts/TRRA_2007.pdf
- Rajman M, Besançon R (1998) Text mining-knowledge extraction from unstructured textual data. In: Advances in data science and classification. Springer, Berlin, Heidelberg, pp 473–480
- Ren G, Hong T (2019) Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews. *Inf Process Manag* 56(4):1425–1438. <https://doi.org/10.1016/j.ipm.2018.04.003>
- Romão J, Neuts B, Nijkamp P, Shikida A (2014) Determinants of trip choice, satisfaction and loyalty in an eco-tourism destination: a modelling study on the Shiretoko Peninsula, Japan. *Ecol Econ* 107:195–205. <https://doi.org/10.1016/j.ecolecon.2014.07.019>
- Ryan C, Mo X (2001) Chinese visitors to New Zealand—demographics and perceptions. *J Vacat Mark* 8(1):13–27. <https://doi.org/10.1177/135676670200800103>
- Shanka T, Taylor R (2004) An investigation into the perceived importance of service and facility attributes to hotel satisfaction. *J Qual Assur Hosp Tour* 4(3–4):119–134. https://doi.org/10.1300/J162v04n03_08
- Shannon C (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):279–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Stone PJ, Dunphy DC, Smith MS (1966) The general inquirer: a computer approach to content analysis. MIT Press, Cambridge
- Sun Y, Wei Y, Zhang L (2017) International academic impact of Chinese tourism research: a review based on the analysis of SSCI tourism articles from 2001 to 2012. *Tour Manag* 58:245–252. <https://doi.org/10.1016/j.tourman.2016.03.008>
- Toutanova K, Manning C (2000) Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In: Proceedings of the 2000 Joint SIGDAT Conference EMNLP/VLC, 63-71, 2000
- Toutanova K, Klein D, Manning CD, Singer Y (2003) Feature-rich part-of-speech tagging with a cyclic dependency network. In: Proceedings of the 2003 conference of the North American chapter of the association for computational linguistics on human language technology-volume 1. Association for Computational Linguistics, pp 173–180

- Truong T, King B (2009) An evaluation of satisfaction levels among Chinese tourists in Vietnam. *Int J Tour Res* 11(6):521–535. <https://doi.org/10.1002/jtr.726>
- Tsujimoto N (2017) The purchasing behavior of Chinese tourists at popular visiting areas in Japan. *J Glob Tour Res* 2(2):99–104
- Uzama A (2012) Yokoso! Japan: classifying foreign tourists to Japan for market segmentation. *J Hospit Mark Manag* 21(2):132–154. <https://doi.org/10.1080/19368623.2011.615016>
- Vermeulen I, Seegers D (2009) Tried and tested: the impact of online hotel reviews on consumer consideration. *Tour Manag* 30:123–127. <https://doi.org/10.1016/j.tourman.2008.04.008>
- Winsted KF (1997) The service experience in two cultures: a behavioral perspective. *J Retail* 73(3):337–360
- World Health Organization (2015) WHO global report on trends in prevalence of tobacco smoking 2015. https://apps.who.int/iris/bitstream/handle/10665/156262/9789241564922_eng.pdf, ISBN: 978-9241564922
- Wu CHJ, Liang RD (2009) Effect of experiential value on customer satisfaction with service encounters in luxury-hotel restaurants. *Int J Hosp Manag* 28(4):586–593. <https://doi.org/10.1016/j.ijhm.2009.03.008>
- Xia F (2000) The Part-Of-Speech tagging guidelines for the Penn Chinese Treebank (3.0). Tech. Rep. IRCS Technical Reports Series. 38, Institute for Research in Cognitive Science, University of Pennsylvania. http://repository.upenn.edu/ircs_reports/38
- Xiang Z, Gretzel U (2010) Role of social media in online travel information search. *Tour Manag* 31(2):179–188. <https://doi.org/10.1016/j.tourman.2009.02.016>
- Xiang Z, Schwartz Z, Gerdes JH, Uysal M (2015) What can big data and text analytics tell us about hotel guest experience and satisfaction? *Int J Hosp Manag* 44:120–130. <https://doi.org/10.1016/j.ijhm.2014.10.013>
- Xiumei S, Jinying W (2011) Cultural distance between China and US across GLOBE model and Hofstede model. *Int Bus Manag* 2(1):11–17
- Xu X, Li Y (2016) The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: a text mining approach. *Int J Hosp Manag* 55:57–69. <https://doi.org/10.1016/j.ijhm.2016.03.003>
- Yang Y, Pan B, Song H (2014) Predicting hotel demand using destination marketing organization's web traffic data. *J Travel Res* 53(4):433–447. <https://doi.org/10.1177/0047287513500391>
- Ye Q, Law R, Gu B (2009) The impact of online user reviews on hotel room sales. *Int J Hosp Manag* 28(1):180–182. <https://doi.org/10.1016/j.ijhm.2008.06.011>
- Zeman D, Popel M, Straka M, Hajič J, Nivre J, Ginter F, Luotolahti J, Pyysalo S, Petrov S, Potthast M, Tyers F, Badmaeva E, Gökırmak M, Nedoluzhko A, Cinková S, Hajič J Jr, Hlaváčová J, Kettnerová V, Urešová Z, Kanerva J, Ojala S, Missilä A, Manning C, Schuster S, Reddy S, Taji D, Habash N, Leung H, de Marneffe MC, Sanguinetti M, Simi M, Kanayama H, de Paiva V, Droganova K, Martínez Alonso H, Çöltekin c, Sulubacak U, Uszkoreit H, Macketanz V, Burchardt A, Harris K, Marheinecke K, Rehm G, Kayadelen T, Attia M, Elkahky A, Yu Z, Pitler E, Lertpradit S, Mandl M, Kirchner J, Fernandez Alcalde H, Strnadová J, Banerjee E, Manurung R, Stella A, Shimada A, Kwak S, Mendonça G, Lando T, Nitisaroj R, Li J (2018) CoNLL 2018 shared task: Multilingual parsing from raw text to universal dependencies. In: Proceedings of the CoNLL 2018 Shared Task: Multilingual parsing from raw text to universal dependencies, pp 1–21
- Zhang H, Yu Z, Xu M, Shi Y (2011) Feature-level sentiment analysis for Chinese product reviews. In: Proceedings of the 2011 3rd International Conference on computer research and development, vol 2, pp 135–140. <https://doi.org/10.1109/iccrd.2011.5764099>
- Zhang M, Liu S, Yang L, Jiang Y, Huang Z, Zhao Z, Deng Q, Li Y, Zhou M, Wang L et al (2019) Prevalence of smoking and knowledge about the smoking hazards among 170,000 Chinese adults: a nationally representative survey in 2013–2014. *Nicotine Tob Res*. <https://doi.org/10.1093/ntr/ntz020>
- Zhou L, Ye S, Pearce PL, Wu MY (2014) Refreshing hotel satisfaction studies by reconfiguring customer review data. *Int J Hosp Manag* 38:1–10. <https://doi.org/10.1016/j.ijhm.2013.12.004>

Authors and Affiliations

Elisa Claire Alemán Carreón¹  · Hugo Alberto Mendoza España¹ ·
Hirofumi Nonaka¹ · Toru Hiraoka²

Hugo Alberto Mendoza España
mendoza.espana@gmail.com

Hirofumi Nonaka
nonaka@kjs.nagaokaut.ac.jp

Toru Hiraoka
hiraoka@sun.ac.jp

¹ Nagaoka University of Technology, Nagaoka, Japan

² University of Nagasaki, Nagasaki, Japan