

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

Elisa Claire Alemán Carreón <sup>1</sup>  · Hugo Alberto Mendoza España <sup>1</sup> · Hirofumi Nonaka <sup>1</sup> · Toru Hiraoka <sup>2</sup>

Received: date / Accepted: date

**Abstract** The Japanese spirit of hospitality and service *Omotenashi* is known worldwide for its excellence. Recent years show a steady increase in international tourists coming to Japan. Chinese tourists, especially, have been steadily increasing. However, before the shift that has brought a global perspective in recent years, most tourist behavior studies were biased for the Western world. Previous research shows that different cultural backgrounds result in different expectations and, arguably, different satisfaction factors. Knowing this, a cross-cultural study of differences between Chinese and Western cultures after the current boom in the Chinese economy in the high standard Japanese hospitality environment is fascinating. Will the top-grade hospitality of Japan influence both populations equally, or will their cultural differences set them apart? Will they be satisfied with the soft attributes like service or be more concerned, comparing soft attributes, such as service, with hard attributes like, such as location and facilities? We bring light to these questions and the differences in each population's satisfaction and dissatisfaction factors in, and studied different price ranges. Taking advantage of Web 2.0, we collected our data with web crawling, labeled a sample, applied Shannon's entropy to extract these factors automatically and then use

---

✉ Elisa Claire Alemán Carreón  
E-mail: elisa.claire.aleman.carreon@gmail.com, ORCID: 0000-0002-6437-0866  
Hugo Alberto Mendoza España  
E-mail: mendoza.espana@gmail.com  
Hirofumi Nonaka  
E-mail: nonaka@kjs.nagaokaut.ac.jp  
Toru Hiraoka  
E-mail: hiraoka@sun.ac.jp

<sup>1</sup> Nagaoka University of Technology, Nagaoka, Japan

<sup>2</sup> University of Nagasaki, Nagasaki, Japan

~~them in an SVM, and then used them~~ to classify a more extensive data set ~~with an SVM~~. We then used dependency parsing and ~~part-of-speech-part-of-speech~~ tagging to extract ~~which nouns the nouns that~~ were tied to ~~praising-positive~~ adjectives. We found that Chinese tourists are ~~less-concerned-with-hospitality~~ and ~~concerned~~ more with room quality than ~~Western-tourists~~. ~~The latter were delighted-hospitality, whereas, Western tourists are delighted more~~ by the staff behavior. We also found that ~~Chinese tourists are concerned with~~ the lack of a ~~Chinese-friendly-environment~~, and ~~Western customers are unsatisfied with dirty rooms or~~ ~~Chinese-friendly environment~~ for Chinese customers and the smell of cigarettes ~~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

## 1 Introduction

~~Japan-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 ~~for its hospitality-being the highest grade. The to be of the highest quality. Omotenashi, which describes the spirit of Japanese hospitality-is celebrated worldwide in a single Japanese word: Omotenashi. With~~, with roots in Japanese history and tea ceremony, ~~their hospitality-is famous around the world-is celebrated worldwide~~ (Alsheikh and Sato 2015). ~~Therefore~~ 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 ~~hold-have~~ different expectations (Engel et al. 1990). Thus, it could be theorized that their satisfaction factors should be different. ~~How will different cultures react and perceive hotels and their hospitality in this context? Our study attempts to bring light to this with two essential tourist populations that differ in culture to Japan: Chinese and Western tourists~~The Japanese tourist market is gradually becoming diverse because of multicultural tourist populations. This diversity means that the expectations when staying at a hotel will be varied. Therefore, hotel management must cater to these expectations to increase customer satisfaction, maintain a good reputation, and generate positive word-of-mouth.

In the last couple of decades, the Japanese economy has been more and more affected by an increase in inbound international tourism. There was a Year-on-Year Growth Rate of 19.3% in 2017, with a total of 28,691,073 inbound tourists that year. From this total, Our study focuses on Western and Chinese

tourists, considering their numbers. Chinese tourists accounted for 25.63% of the tourist population was mostly Asian (86.14%), and approximately a fourth of the total (25.63%) came from China. Western countries, counting English-speaking countries and Europe, make. On the other hand, Western countries accounted for 11.4% of the total, with a and 7.23% of the total being 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. From that, along with the number of researchers interested in this phenomenon has been increasing as well. studies on this phenomenon (Sun et al. 2017). With these and other multicultural tourist populations, the tourist market is more and more diverse. Diversity in customers' cultural backgrounds means that their expectations when staying at a hotel will also be varied. Therefore, Hotel management needs to cater to these needs and expectations to increase customer satisfaction, maintain a good reputation, and generate positive word-of-mouth.

However, many tourist behavior. Despite this, many tourist-behavior analyses have been performed with only involving Western subjects. As such, a gap in knowledge and discussion knowledge gap existed until recent decades. Those studies that do include

In studies involving Asian populations in their analysis most commonly study Chinese tourist behavior 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 that compare Asian to Western tourist behavior studies reporting comparisons between Asian and Western tourists' behaviors (e.g. Choi and Chu 2000) are commonly survey typically survey- or interview-based studies with, using small samples. These, while studies, although valid, can have their limitations. This gap in research 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 make 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 hotel lodging 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. For this, we used a data-driven approach to our analysis.

With Owing to the advent of Web 2.0 and customer review websites, researchers realized the benefits of online reviews for research, and their importance for 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). Consequentially, tourism

~~research also began to use~~ Consequently, information collected online is being  
 75 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 ~~;~~ machine  
~~learning ; and big data methodologies and machine learning technologies~~ can  
 80 ~~increase the number of manageable samples per study . The increase can be~~  
~~from the hundred samples manually analyzed by researchers to the~~ in a study  
~~from hundreds to~~ hundreds of thousands ~~automatically analyzed by machines.~~  
~~This technology . These technologies~~ can not only help confirm existing theo-  
 ries but also lead to finding new patterns and to knowledge discovery (Fayyad  
 et al. 1996).

85 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 ~~availability of enormous amounts of wide~~  
~~availability of~~ online reviews of Japanese hotels by both Mainland Chinese  
 tourists posting ~~in on~~ Ctrip and Western, English-speaking tourists ~~populations~~  
 90 ~~posting in~~ posting on TripAdvisor. ~~With this~~ Based on these data, we can con-  
 firm existing theories ~~about their differences in behavior and explore the data~~  
~~to regarding the differences in tourists' behavior and~~ discover factors that  
 could have been overlooked in the past. ~~To do this, we~~ We use machine learn-  
 ing to automatically classify ~~review sentences~~ sentences in the online reviews  
 95 as positive or negative opinions ~~of on~~ the hotel. We then perform a statisti-  
 cal extraction of the topics that ~~concerns most concern~~ the customers of each  
 population ~~the most~~.

## 2 Research objective

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

Our proposal includes ~~using large scale data from online hotel reviews in~~  
~~Chinese and English to study their differences in a statistical manner. In the~~  
~~past, survey-based studies have provided a theoretical background for a few~~  
 115 ~~specific tourist populations of a single culture or that travel with a single~~

purpose. These studies' short scope often leads to difficulties in observing cultural and language differences in a single study.

Our study attempts to uncover the difference in satisfaction and dissatisfaction factors between different cultures. These factors can become the focal point for improving the tourism and service industries and increasing customer satisfaction. Satisfied customers will then write more positive online reviews that will, in turn, increase sales and attract new customers.

### 3 Theoretical background and hypothesis development

#### 3.1 Japanese hospitality and service: *Omotenashi*

The spirit of Japanese hospitality, or *Omotenashi*, has roots in the countries history. However, country's history, and to this day, it is regarded as the highest standard (Ikeda 2013; Al-alsheikh and Sato 2015). There is even a famous phrase in customer service in Japan: *okyaku-sama wa kami-sama desu*, or translated meaning "The customer is god". Some. Some scholars say that *omotenashi* originated from the old Japanese art of the tea ceremony in the 16th century. However, other scholars found that its roots come from even earlier, 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. Today/Currently, 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 as to not fall behind in the competition. In Japanese businesses, hotels included/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 with different depths differently depending on the situation, where a light bow could be used to say "Please, allow me to guide you". Deep bows are also used to apologize for any inconvenience the customer could have faced, followed by a very respectful apology. In fact, despite Although the word *omotenashi* being can be translated directly as "hospitality", it includes both the concepts of hospitality and service (Kuboyama 2020). This hospitality culture permeates every kind 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 be a positive aspect that would be at the top of satisfaction for any customer 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 determine whether or not directly determine whether a hotel is practicing with the cultural standards of *omotenashi* directly. 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 ~~in general, which can be compared and compare them~~ to other places and business practices in the world. In summary, ~~our study considers~~ we consider the influence of the cultural aspect of *omotenashi* while analyzing the evaluations ~~of on~~ service and hospitality factors that are universal to all hotels in any country.

Therefore ~~we pose a research question for our study~~, 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 ~~comes from~~ is based on the Japanese culture. Different cultures interacting with it could ~~have~~ provide a different evaluation of it. ~~While some~~ 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 ~~above~~ 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?*

### 3.2 Customer satisfaction and dissatisfaction towards individual factors during hotel stay

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. Previous studies on the dimensions of culture that influence differences in expectations have been performed in the past Donthu and Yoo (1998). Western and Chinese customers ~~can~~ then have very different ~~satisfaction and dissatisfaction factors since they have different~~ backgrounds and cultures. These varying backgrounds 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. ~~These expectations will be there from the moment that they choose the hotel throughout their stay.~~ 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. ~~So~~ Therefore, we must measure their differences or similarities at their common ground as well.

However, 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, ~~suppose-if~~ a customer has a conscious expectation of a comfortable bed and a wide shower, and it is realized during their visit. ~~In that case~~, 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 ~~by-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.

Previous research has also focused more on soft attributes, with little focus on hard attributes, ~~like location or infrastructure, mostly focusing only-if~~ only focusing on facilities (e.g. Shanka and Taylor 2004; Choi and Chu 2001). However, hard factors, which are uncontrollable by the hotel staff, 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.



This leads to another couple of research questions:

**Research Question 2a:** *To what degree does 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.

### 3.3 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. ~~Explaining this increase, analyzed the number of studies related to Chinese tourists from 2001 to 2012 and found a steady increase until~~, with rapid growth in studies following the year 2007, ~~followed by the rapid growth of Chinese tourism studies. This increase in Chinese tourist behavior research resulted in several~~ (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 ~~less~~ ~~fewer~~ ~~are~~ the number of studies explicitly comparing Chinese and Western tourists. One example is a study by Choi and Chu (2000), ~~who~~ ~~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. ~~At the same time,~~ ~~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 ~~that American tourists were found~~ ~~American tourists~~ to be individualistic and motivated by novelty. ~~In contrast,~~ ~~while~~ Japanese tourists were collectivist and motivated by ~~prestige and family, with an escape from routine and increased knowledge as a common motivator~~ ~~increasing knowledge and escaping routine~~.

One thing to note with the above Asian vs. Western analyses is that they were performed before 2000 and ~~that they are not Chinese specific but study~~ ~~Asian people in general~~ ~~not Chinese specific~~. Meanwhile, the current Chinese ~~economy~~ ~~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 ~~by~~ ~~in~~ Hunt (1975), ~~that 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 U.S.A. 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, U.S.A. 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 U.S. tourists using text mining techniques and more massive datasets, although in a restaurant context.

These last three ~~articles focus on~~ studies focus on the U.S.A. culture, ~~while 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 ~~with~~ within the context of tourists from three countries staying in hotels across the world. The second ~~one study~~ chose restaurant reviews from the U.S.A. and Chinese tourists eating in three countries in Europe. The third ~~is analyzing restaurants in Beijing~~ 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, ~~do customers~~ customers could either hold their satisfaction to this hospitality regardless of their culture ~~, or are or value~~ other factors more ~~relevant to the customers?~~ 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 ~~different~~ countries. The general satisfaction score was noticed to be different ~~in 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 ~~tourists~~ 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 ~~could also help us analyze~~ analyzes the satisfaction and dissatisfaction factors cross-culturally and ~~compare~~ compares them with the existing literature.

Undoubtedly previous literature has examples of other cross-culture studies of tourist behavior and ~~to~~ 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. ~~This raises the question: in such an environment,  
will the customers be universally satisfied with this factor, or will they have  
differing views within their cultures?~~ 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.

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

	Bauer et.al (1993)	Choi and Chu (2000)	Kim and Lee (2000)	Huang (2017)	Francesco and Roberta (2019)	Jia (2020)	Our study
<b>Comparison objects</b>	Asians vs Westerns	Asians vs Westerns	Anglo-Americans vs Japanese	Chinese vs English-speakers	USA vs China vs Italy	Chinese vs US tourists	Chinese vs Westerners
<b>Study topic</b>	Changes in demand	Satisfaction Levels	Tourist Motivation	Dining experience of Roast Duck	Perception and Emphasis	Motivation and Satisfaction	<b>Satisfaction and Dissatisfaction</b>
<b>Geographical focus</b>	Asia Pacific region	Hong Kong	Global	Beijing	Multi-national	Multi-national	<b>Japan</b>
<b>Industry</b>	Hotels	Hotels	Tourism	Restaurant (Beijing Roast Duck)	Hotels	Restaurants	Hotels
<b>Study subjects</b>	Hotel managers	Hotel customers	Tourists arriving in airport	Diners online reviews	Hotel customers online reviews	Diners online reviews	Hotel customers online reviews
<b>Sample method</b>	surveys	surveys	survey	text mining	text mining	text mining	text mining
<b>Number of samples</b>	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)	<b>89,207 reviews</b> <b>(48,070 Chinese)</b> <b>(41,137 English)</b>
<b>Study method</b>	statistics	VARIMAX	MANOVA	Semantic Network Analysis	Text Link Analysis	Topic modeling (LDA)	<b>SVM, Dependency Parsing and POS tagging</b>
<b>Subject nationality</b>	Asians: China, Fiji, Hong Kong, Indonesia, Malaysia, Singapore, Taiwan, Guam, Tabiti, Thailand Westerners: Australia, New Zealand	Asians: China, Taiwan, Japan, South Korea, South-East Asia Westerners: North America, Europe, Australia, New Zealand	USA, Japan	English-speakers: U.K., U.S., Australia, New Zealand, Canada, Ireland Chinese-speakers: China	USA, China, Italy	USA, China	Chinese-speakers: China English-speakers: (U.K., U.S., Australia, New Zealand, Canada, Ireland)

### 3.4 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 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, ~~a never before seen~~ 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, ~~a critical 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).

~~The algorithm used~~ Our study used an algorithm for sentiment analysis ~~in our study is~~ called a Support Vector Machine ~~. It is a form of~~ (SVM), a super-

vised machine learning used for binary classification. ~~This means a sample of~~  
~~labeled training data is given to the algorithm to detect patterns in the data~~  
~~and use those patterns to establish a method for classifying other unlabeled~~  
~~data automatically.~~ 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 ~~data~~  
~~to know the correct task result template~~ 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 ~~without any~~  
~~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 ex-  
 ploring 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 interven-  
 tion from the researchers in making questionnaires, it can result in anything  
 that the researchers could not expect. These qualities determine why, ~~much~~  
~~like actual mining, 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 ~~being is that~~ we can discover knowl-  
 edge that could otherwise be missed by ~~looking at~~ observing a localized sample  
 rather than taking a holistic view of every ~~users'~~ user's opinion. In other words,  
 a machine algorithm can find the needles in a haystack that we did not know  
 were there ~~from taking by examining~~ small bundles of hay at a time.

In this study, ~~we can predict that several things might occur. Our data~~  
~~could show satisfaction and dissatisfaction factors that are universal, and it~~  
~~could also find strictly cultural factors. However, we expect that both of these~~  
~~options will present themselves. We can also assert that we could arrive at~~  
~~very similar results to previous literature if they are correct in their findings.~~  
~~However, we are using a database of several orders of magnitude larger. We can~~

also expect to discover patterns that researchers previously had not noticed because of the lack of questionnaire design and users' freedom to record their pleasures and grievances.

## 4 Methodology

We have extracted a large number of text reviews from a Chinese portal the site *Ctrip*, with mostly mainland Chinese users, and the travel site *TripAdvisor*. We then determined the most commonly used words that contribute 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 perform a Support Vector Machine (SVM) based 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 classified the sentences in the extracted reviews as emotionally positive or negative, using an optimized Support Vector Classifier (SVC). 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 other nouns used in the reviews the nouns they refer to. We split the dataset into price ranges to observe the satisfaction factors and their differences differences in keyword usage between lower-class and higher-class hotels. We observed the frequency of the terms in the dataset to extract the most utilized words in either review. We show an overview of this methodology in Figure 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. Soft attributes are non-physical aspects of the hotel, those that regard service, staff behavior, or hotel management, which can be changed by the staff or management. Hard attributes are related to physical and environmental aspects, such as facilities, infrastructure, and the hotel's surroundings, which are impossible to change by the staff alone.

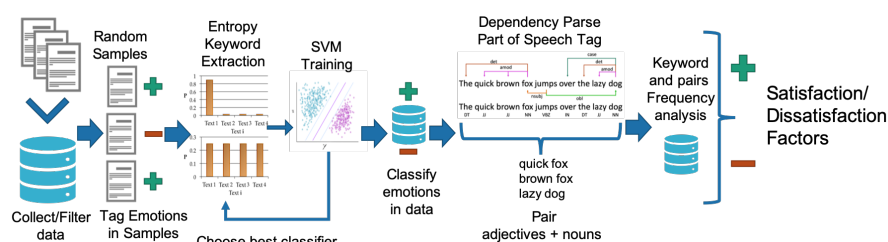


Fig. 1: Overview of the methodology to quantitatively rank satisfaction factors.



#### 4.1 Data collection

In the ~~data collection stage for Chinese reviews in Ctrip~~, Ctrip data collection, reviews from a total of 5774 ~~review pages of~~ 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 ~~from mainland China~~. 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, ~~we needed to make the data and comparisons we draw from each of these datasets fair. For that purpose~~ 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 ~~and cut off reviews outside of each other's date ranges~~. In addition, we selected only the hotels that had pricing information available. We extracted the lowest ~~price possible for a room or bed for one night, and the~~ 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 ~~of several classes~~, 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 ~~indirectly, giving us insight into the kind of service offered~~. After filtering, ~~we found that the number of the datasets contained 557~~ hotels in common ~~in the data collected was 557~~. The overlapping date range for reviews was from July 2014 to July 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. ~~After filtering the data, we found that the number of reviews was similar for both English and Chinese reviews, but that English reviews tend to be longer in general.~~

The price for a night in these hotels ranges from ~~low priced cheap~~ capsule hotels at 2000 yen per night to high-end hotels ~~188,000-188,000~~ yen a night ~~as at~~ the far ends of the bell curve. 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 Figure 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 to 2500 yen, 2500 to 5000 yen, 5000 to 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 to 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 to 20,000 yen, 20,000 to 30,000 yen). Further than that is more likely to be *ryokan* or high class resorts or five-star hotels (30,000 to 50,000 yen, 50,000 to 100,000 yen, 100,000 to 200,000 yen). Note that because of choosing the highest price per one night in each hotel, the cheapest two price ranges (0 to 2500 yen, 2500 to 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.

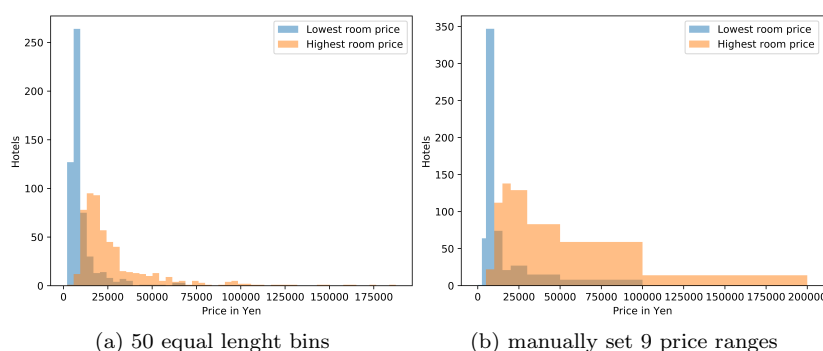


Fig. 2: Price for one night distribution, blue: lowest price, orange: highest price.

## 4.2 Text processing

~~Chinese text, unlike English, does not have spaces between each word to separate them. Besides, we also~~ 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 ~~the case of texts in English~~ 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 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 ~~with~~ 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 Figure 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

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 to 2500 yen	Hotels	0	0
	Reviews	0	0
2: 2500 to 5000 yen	Hotels	0	0
	Reviews	0	0
3: 5000 to 10,000 yen	Hotels	22	22
	Reviews	452	459
	Sentences	1,108	3,988
	Positive sentences	924	1,875
	Negative sentences	184	2,113
4: 10,000 to 15,000 yen	Hotels	112	112
	Reviews	2,176	2,865
	Sentences	4,240	24,107
	Positive sentences	3,566	11,619
	Negative sentences	674	12,488
5: 15,000 to 20,000 yen	Hotels	138	138
	Reviews	7,043	4,384
	Sentences	14,726	37,342
	Positive sentences	12,775	17,449
	Negative sentences	1,951	19,893
6: 20,000 to 30,000 yen	Hotels	129	129
	Reviews	11,845	13,772
	Sentences	24,413	115,830
	Positive sentences	21,068	55,381
	Negative sentences	3,345	60,449
7: 30,000 to 50,000 yen	Hotels	83	83
	Reviews	8,283	7,001
	Sentences	17,939	58,409
	Positive sentences	15,642	28,493
	Negative sentences	2,297	29,916
8: 50,000 to 100,000 yen	Hotels	59	59
	Reviews	16,670	9,646
	Sentences	36,255	81,940
	Positive sentences	31,638	38,217
	Negative sentences	4,617	43,723
9: 100,000 to 200,000 yen	Hotels	14	14
	Reviews	1,601	3,010
	Sentences	3,282	26,423
	Positive sentences	2,930	12,274
	Negative sentences	352	14,149

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 **entire** 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.

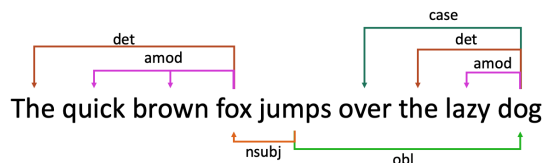


Fig. 3: Example of dependency parsing.

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 Figure 4. A Part of Speech is 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).



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

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 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 (recommend)
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

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 (kilometer)
	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

#### 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  $\alpha$  for one class and  $\alpha'$  for the other, and then they are used in an **SVM-SVC** (Cortes and Vapnik 1995), optimizing keywords to select the best performing classifier using the  $F_1$ -measure (Powers 2011). The selected **SVM-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 **SVM-SVC**. C is a constant that affects the optimization process when minimizing the error of the separating hyperplane. Low values of C give some freedom of error, which minimizes false positives. ~~However, depending on the data, it can but can also~~ increase false negatives. Inversely, high C values will likely result in minimal false negatives but a possibility of false positives. **SVM-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

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

Keyword List	Classifier emotion	C	$F_1$ $\mu$	$F_1$ $\sigma$
Satisfaction keywords ( $\alpha = 2.75$ )	Satisfaction	2.5	0.91	0.01
Negative keywords ( $\alpha' = 3.75$ )	Dissatisfaction	0.5	0.67	0.11
<b>Combined</b> ( $\alpha = 2.75, \alpha' = 3.75$ )	<b>Satisfaction</b>	<b>0.5</b>	<b>0.95</b>	<b>0.01</b>

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

Keyword List	Classifier emotion	C	$F_1$ $\mu$	$F_1$ $\sigma$
Satisfaction keywords ( $\alpha = 1.5$ )	Satisfaction	1.75	0.82	0.02
Dissatisfaction keywords ( $\alpha' = 4.25$ )	Dissatisfaction	3	0.80	0.03
<b>Combined</b> ( $\alpha = 1.5, \alpha' = 4.25$ )	<b>Satisfaction</b>	<b>2</b>	<b>0.83</b>	<b>0.02</b>

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

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 **SVM-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 ~~the SVM-classification process's linear kernel~~ 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 ~~SVM~~ SVC learning algorithm, each data point classified incorrectly ~~causes a change in~~ 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  $\alpha$  to reduce the number of wrong classifications. This equation shows the last value of  $\alpha$  after the end of the cycle.

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

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

We tagged 159 Chinese sentences and 2357 English sentences as positive or negative for our training data. The entropy comparison factors  $\alpha$  and  $\alpha'$  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.

## 5 Data Analysis

### 5.1 Frequent keywords in differently priced hotels

~~To understand Chinese tourists and English-speaking tourists' satisfaction and dissatisfaction factors when lodging in Japan, we study both the frequency of the words they use. Following that, to know the relevance of a keyword as a preference for each group, we observed each entropy-based keyword's frequencies in our complete data set and in each price range. The frequency of the keywords in the database shows the level of priority it has for customers.~~

We observed the top 10 ~~words~~ satisfaction and dissatisfaction keywords with the highest frequencies ~~for keywords linked by entropy to satisfaction and dissatisfaction in~~ of emotionally positive and negative statements to study. The keywords are the quantitative rank of the needs of Chinese and ~~English speaking~~ 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 section 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.

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)	12892	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 to 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 to 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 to 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
6: 20,000 to 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 to 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 to 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 to 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
3: 5000 to 10,000 yen			annoying	220
			sense	220
	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
4: 10,000 to 15,000 yen			funny smell	3
			poor	3
			renovation	3
			carpet	2
	价格 (price)	98	dated	40
	一般 (general)	91	poor	29
	距离 (distance)	43	disappointing	26
	陈旧 (obsolete)	34	worst	24
5: 15,000 to 20,000 yen	地理 (geography)	31	uncomfortable	23
	老 (old)	30	cigarette	22
	中文 (Chinese)	26	pricey	22
			minor	21
			paper	19
			unable	19
	价格 (price)	296	poor	57
	一般 (general)	218	dated	41
6: 20,000 to 30,000 yen	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
7: 30,000 to 50,000 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
8: 50,000 to 100,000 yen			cigarette	79
			annoying	70
	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
9: 100,000 to 200,000 yen	老 (old)	45	dirty	51
	华人 (Chinese)	2	disappointing	50
			cigarette	46
			unable	43
	价格 (price)	561	pricey	163
	一般 (general)	510	dated	150
	中文 (Chinese)	337	disappointing	129
	地理 (geography)	239	poor	124
9: 100,000 to 200,000 yen	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
	价格 (price)	54	pricey	40
	一般 (general)	51	sense	34
9: 100,000 to 200,000 yen	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	15
			disappointment	14
			paper	14

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) : 12892 不错 酒店 (nice hotel) : 1462 不错 位置 (nice location) : 1426 不错 服务 (nice service) : 869 不错 环境 (nice environment) : 714 不错 (not bad) : 139 不错 酒店 (nice hotel) : 17 不错 位置 (nice location) : 16 不错 早餐 (nice breakfast) : 12 不错 服务 (nice service) : 8	大 (big) : 9844 大 房间 (big room) : 3197 大 床 (big bed) : 772 大 酒店 (big hotel) : 379 大 超市 (big supermarket) : 232 大 (big) : 76 大 房间 (big room) : 11 大 床 (big bed) : 10 大 超市 (big supermarket) : 5 大 商场 (big market) : 3 大 (big) : 348 大 房间 (big room) : 76 大 床 (big bed) : 30 大 社 (big club) : 26 大 空间 (big space) : 16 大 (big) : 1277 大 房间 (big room) : 316 大 床 (big bed) : 140 大 超市 (big supermarket) : 73 大 酒店 (big hotel) : 49 大 (big) : 2245 大 房间 (big room) : 680 大 床 (big bed) : 198 大 酒店 (big hotel) : 102 大 空间 (big space) : 64 大 (big) : 1913 大 房间 (big room) : 643 大 床 (big bed) : 141 大 超市 (big supermarket) : 74 大 酒店 (big hotel) : 66 大 (big) : 3670 大 房间 (big room) : 1340 大 床 (big bed) : 238 大 酒店 (big hotel) : 144 大 商场 (big market) : 88 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) : 6665 干净 房间 (clean room) : 1224 干净 酒店 (clean hotel) : 737 干净 卫生 (clean and hygiene) : 664 干净 环境 (clean environment) : 61 干净 (clean) : 114 干净 房间 (clean room) : 21 干净 酒店 (clean hotel) : 10 干净 卫生 (clean and hygiene) : 6 干净 总体 (clean overall) : 4 干净 (clean) : 455 干净 房间 (clean room) : 66 干净 卫生 (clean and hygiene) : 52 干净 酒店 (clean hotel) : 48 干净 打扫 (clean up) : 9 干净 (clean) : 1348 干净 房间 (clean room) : 234 干净 酒店 (clean hotel) : 161 干净 卫生 (clean and hygiene) : 92 干净 设施 (clean facilities) : 19 干净 (clean) : 1940 干净 房间 (clean room) : 360 干净 酒店 (clean hotel) : 203 干净 卫生 (clean and hygiene) : 137 干净 环境 (clean environment) : 21 干净 (clean) : 1159 干净 房间 (clean room) : 224 干净 酒店 (clean hotel) : 146 干净 卫生 (clean and hygiene) : 71 干净 环境 (clean environment) : 16 干净 (clean) : 1577 干净 房间 (clean room) : 310 干净 酒店 (clean hotel) : 161 干净 卫生 (clean and hygiene) : 101 干净 服务 (clean service) : 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 hygiene) : 5	近 (near) : 5181 近 酒店 (near hotel) : 453 近 桥 (near bridge) : 144 近 地铁站 (near subway station) : 122 近 站 (near station) : 108 近 (near) : 55 近 酒店 (near hotel) : 4 近 地铁 (near subway) : 2 近 (near) : 323 近 酒店 (near hotel) : 27 近 站 (near station) : 14 近 地铁 (near subway) : 12 近 车站 (near the station) : 10 近 (near) : 1016 近 酒店 (near hotel) : 82 近 站 (near station) : 35 近 地铁站 (near subway station) : 34 近 桥 (near bridge) : 29 近 (near) : 1433 近 酒店 (near hotel) : 164 近 地铁 (near subway) : 34 近 地铁站 (near subway station) : 31 近 车站 (near the station) : 27 近 (near) : 935 近 酒店 (near hotel) : 80 近 站 (near station) : 25 近 桥 (near bridge) : 20 近 山 (near mountain) : 12 近 (near) : 1354 近 酒店 (near hotel) : 88 近 桥 (near bridge) : 76 近 地铁站 (near subway station) : 35 近 铁路 (Kintetsu) : 24 近 (near) : 65 近 酒店 (near hotel) : 8 近 地铁站 (near subway station) : 3 近 市场 (near market) : 3	新设施 (new facility) : 363 新酒店 (new hotel) : 246 新装修 (new decoration) : 116 新房间 (new room) : 53 新 (new) : 2775 新设施 (new facility) : 9 新装修 (new decoration) : 2 新酒店 (new hotel) : 25 新装修 (new decoration) : 15 新房间 (new room) : 10 新 (new) : 517 新设施 (new facility) : 89 新酒店 (new hotel) : 51 新装修 (new decoration) : 24 新房间 (new room) : 10 新 (new) : 260 新设施 (new facility) : 63 新酒店 (new hotel) : 25 新装修 (new decoration) : 15 新房间 (new room) : 11 新 (new) : 1634 新设施 (new facility) : 141 新酒店 (new hotel) : 123 新装修 (new decoration) : 57 新盖 (new) : 22 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) : 3028 棒 酒店 (great hotel) : 463 棒 位置 (great position) : 218 棒 服务 (great service) : 168 棒 早餐 (great breakfast) : 164 棒 (great) : 11 棒 位置 (great position) : 2 棒 (great) : 73 棒 位置 (great position) : 6 棒 房间 (great room) : 3 棒 水平 (great level) : 3 棒 温泉 (great hot spring) : 3 棒 (great) : 241 棒 位置 (great position) : 43 棒 酒店 (great hotel) : 25 棒 服务 (great service) : 22 棒 早餐 (great breakfast) : 8 棒 (great) : 440 棒 酒店 (great hotel) : 51 棒 位置 (great position) : 45 棒 服务 (great service) : 23 棒 早餐 (great breakfast) : 20 棒 (great) : 448 棒 酒店 (great hotel) : 68 棒 位置 (great position) : 34 棒 服务 (great service) : 24 棒 早餐 (great breakfast) : 14 棒 (great) : 1626 棒 酒店 (great hotel) : 281 棒 早餐 (great breakfast) : 112 棒 位置 (great position) : 96 棒 服务 (great service) : 86 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
3: 5000 to 10,000 yen						
4: 10,000 to 15,000 yen						
5: 15,000 to 20,000 yen						
6: 20,000 to 30,000 yen						
7: 30,000 to 50,000 yen						
8: 50,000 to 100,000 yen						
9: 100,000 to 200,000 yen						

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

Price range	一般 (general)	陈旧 (obsolete)	老 (old)
<b>0: All Prices</b>	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	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
<b>3: 5000 to 10,000 yen</b>			
<b>4: 10,000 to 15,000 yen</b>	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	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
<b>5: 15,000 to 20,000 yen</b>	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	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
<b>6: 20,000 to 30,000 yen</b>	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) : 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
<b>7: 30,000 to 50,000 yen</b>	一般 (general) : 311 一般设施 (general facilities) : 23 一般服务 (general service) : 22 一般早餐 (average breakfast) : 19 一般酒店 (average hotel) : 15	陈旧 (obsolete) : 71 陈旧设施 (obsolete facilities) : 43 陈旧设备 (obsolete equipment) : 5 陈旧房间 (outdated room) : 3	老 (old) : 45 老酒店 (old hotel) : 11 老设施 (old facility) : 7 老店 (old shop) : 3 老房间 (old room) : 2
<b>8: 50,000 to 100,000</b>	一般 (general) : 510 一般服务 (general service) : 39 一般设施 (general facilities) : 32 一般早餐 (average breakfast) : 30 一般酒店 (average hotel) : 25	陈旧 (obsolete) : 90 陈旧设施 (obsolete facilities) : 53 陈旧房间 (outdated room) : 5 陈旧感觉 (Stale feeling) : 2	老 (old) : 134 老酒店 (old hotel) : 34 老设施 (old facility) : 26 老装修 (old decoration) : 9 老店 (old shop) : 7
<b>9: 100,000 to 200,000</b>	一般 (general) : 51 一般服务 (general service) : 7 一般早餐 (average breakfast) : 5 一般位置 (general location) : 2 一般房间 (average room) : 2	陈旧 (obsolete) : 6 陈旧设施 (obsolete facilities) : 4	老 (old) : 5 老设施 (old facility) : 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
<b>0: All Prices</b>	good : 19148	clean : 9064	comfortable : 5625	helpful : 5846	free : 4318	large : 4104	friendly : 5606	great : 16127
	good location : 1985	clean room : 3596	comfortable bed : 1919	helpful staff : 2927	free wifi : 773	large room : 1256	friendly staff : 3819	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
<b>3: 5000 to 10,000 yen</b>	good hotel : 874	clean everything : 200	comfortable hotel : 238	helpful service : 74	free bus : 225	large room : 192	friendly person : 63	great hotel : 802
	good : 206	clean : 174	comfortable : 79	helpful : 70	free : 35	large : 31	friendly : 64	great : 143
	good location : 30	clean room : 55	comfortable bed : 21	helpful staff : 36	free wifi : 10	large room : 7	friendly staff : 53	great location : 21
	good value : 19	clean bathroom : 14	comfortable room : 9		free tea : 4	large area : 2	friendly everyone : 2	great view : 14
<b>4: 10,000 to 15,000 yen</b>	good english : 10	clean place : 12	comfortable futon : 8		free raman : 2	large size : 2	friendly service : 2	great place : 13
	good place : 7	clean hotel : 6	comfortable stay : 3		free toothbrush : 2			great experience : 5
	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
<b>5: 15,000 to 20,000 yen</b>	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	large room : 12	friendly hotel : 7	great hotel : 43
	good service : 67	clean everything : 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
<b>6: 20,000 to 30,000 yen</b>	good location : 242	clean room : 440	comfortable bed : 219	helpful staff : 301	free wifi : 109	large room : 85	friendly staff : 444	great location : 109
	good hotel : 116	clean hotel : 133	comfortable room : 99	helpful desk : 11	free shuttle : 35	large suitcase : 18	friendly hotel : 12	great view : 81
	good breakfast : 113	clean bathroom : 38	comfortable stay : 30	helpful concierge : 9	free bus : 30	large room : 18	friendly service : 8	great hotel : 68
	good service : 108	clean everything : 26	comfortable hotel : 20	helpful reception : 5	free breakfast : 27	large hotel : 17	friendly most : 7	great place : 61
<b>7: 30,000 to 50,000 yen</b>	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 concierge : 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	large room : 81	friendly person : 21	great hotel : 232
<b>8: 50,000 to 100,000 yen</b>	good breakfast : 303	clean everything : 77	comfortable hotel : 82	helpful receptionist : 17	free drink : 38	large bed : 43	friendly hotel : 19	great view : 220
	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	large room : 32	friendly service : 24	great view : 162
<b>9: 100,000 to 200,000 yen</b>	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
	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
<b>9: 100,000 to 200,000 yen</b>	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	large room : 38	friendly person : 12	great hotel : 241
	good : 994	clean : 261	comfortable : 347	helpful : 401	free : 240	large : 404	friendly : 370	great : 1488
<b>9: 100,000 to 200,000 yen</b>	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 poor service : 55 poor breakfast : 41 poor quality : 27 poor english : 24 poor : 3	dated : 431 outdated : 128 outdated room : 20 outdated hotel : 10 outdated bathroom : 7	worst : 327 worst hotel : 43 worst experience : 18 worst part : 15 worst service : 10 worst : 6 worst room : 2	dirty : 188 dirty carpet : 34 dirty room : 23 not dirty : 7 dirty bathroom : 6 dirty : 3	uncomfortable : 253 uncomfortable bed : 63 uncomfortable pillow : 20 uncomfortable mattress : 8 uncomfortable night : 8 uncomfortable : 2
3: 5000 to 10,000 yen					
4: 10,000 to 15,000 yen	poor : 29 poor breakfast : 3 poor service : 3 poor conditioning : 2 poor view : 2	dated : 40 outdated : 11 outdated decor : 2 outdated room : 2	worst : 24 worst hotel : 4 worst experience : 2	dirty : 11 dirty floor : 2	uncomfortable : 23 uncomfortable bed : 4 not uncomfortable : 2 uncomfortable night : 2 uncomfortable pillow : 2
5: 15,000 to 20,000 yen	poor : 57 poor service : 10 poor breakfast : 6 poor hotel : 5 poor experience : 3	dated : 41 outdated : 8	worst : 36 worst hotel : 8 worst experience : 3 worst part : 2 worst service : 2	dirty : 14 dirty room : 2	uncomfortable : 26 uncomfortable bed : 7 uncomfortable pillow : 2
6: 20,000 to 30,000 yen	poor : 136 poor breakfast : 15 poor service : 14 poor english : 9 poor quality : 9	dated : 131 outdated : 31 outdated room : 6 outdated hotel : 2	worst : 86 worst hotel : 11 worst part : 7 worst breakfast : 5 worst experience : 5	dirty : 67 dirty room : 10 dirty carpet : 8 dirty bathroom : 3 dirty chair : 2	uncomfortable : 103 uncomfortable bed : 24 uncomfortable pillow : 11 uncomfortable night : 4 uncomfortable experience : 3
7: 30,000 to 50,000 yen	poor : 92 poor service : 8 poor breakfast : 7 poor english : 7 poor connection : 5	dated : 65 outdated : 17 outdated hotel : 4 outdated bathroom : 2 outdated decor : 2	worst : 64 worst hotel : 10 worst room : 3 worst service : 3 worst part : 2	dirty : 51 dirty carpet : 11 dirty room : 7 dirty clothe : 2 dirty luggage : 2	uncomfortable : 55 uncomfortable bed : 20 uncomfortable mattress : 6 uncomfortable pillow : 5 uncomfortable room : 5
8: 50,000 to 100,000 yen	poor : 124 poor service : 16 poor breakfast : 9 poor quality : 9 poor english : 6	dated : 150 outdated : 58 outdated room : 9 outdated furniture : 6 outdated hotel : 4	worst : 98 worst hotel : 9 worst experience : 5 worst part : 3	dirty : 36 dirty carpet : 12 dirty room : 3 dirty cup : 2 dirty rug : 2	uncomfortable : 33 uncomfortable bed : 7
9: 100,000 to 200,000 yen	poor : 19 poor service : 4 poor choice : 2 poor experience : 2	dated : 3 outdated : 2	worst : 12 worst experience : 2	dirty : 6	uncomfortable : 8 little uncomfortable : 2



### 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 ~~regarding~~ 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 ~~bedsheets~~ bed sheets or curtains, and other physical attributes that are part of the service and not ~~of~~ the hotel’s physical structure are ~~also~~ considered soft attributes. Thus, we can observe the top 10 satisfaction and dissatisfaction keywords and determine whether they are soft or hard attributes.

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 ~~is describing~~ described unchangeable physical factors by the staff or management, we consider them hard. If the word ~~implies~~ 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 ~~4~~ four adjective and noun pairings used in the entire dataset. ~~We counted the percentage of usage and counted the usage percentage~~ in each context. If it ~~is was~~ not clear from the word or the pairing alone, we ~~declare~~ 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 Figure 5. For the English keywords, see Figure 6.

## 6 Results

### 6.1 ~~Experiment~~ Experimental results and ~~answering~~ answers to research questions

Our research questions were ~~about two things. In related to two issues. Based on~~ research questions 1a and 1b, ~~we decide that~~ the objective of this study ~~to determine was to determine the differences in~~ how Chinese and Western tourists ~~interact with the omotenashi culture influenced~~ perceive Japanese hotels, whose hospitality and service ~~in Japan, and how are they different in~~

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
Negative Keywords	赞	100% undefined
	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	75% hard, 25% soft
	华人	100% soft

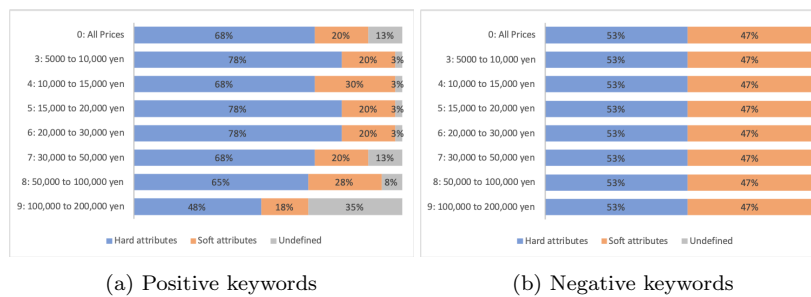


Fig. 5: Hard and soft attributes from the top Chinese keywords for all price ranges

their perceptions in this matter. We observed are influenced by the omotenashi culture.

Observing the top-ranking positive factors for Chinese tourists across different price ranges in Table 8, and specifically the word “不错(not bad)” and its pairings in Table 10. From these observations, we can infer keywords in Chinese reviews, as shown in Tables 8 and Table 10, it was revealed that, while service, cleanliness, and breakfast are were praised in most hotels, location is usually placed above it in importance on the location was more important when observing the pairings. When we see the rest of the factors Hard attributes were abundant lower on the list, we see that the list is more populated with hard attributes like location and transportation availability across different

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

price ranges. From the negative keyword usages lists. The negative keywords in Table 9 ,there are complaints about the indicate that a lack of a Chinese friendly environment . However, most complaints are also 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. Nevertheless, the most complained about aspect is the price of the hotel. Surprisingly, However, most complaints were about the hotel's price, which included all of the price ranges have this negative keyword at the top of the list, suggesting that it is the main concern to ; 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

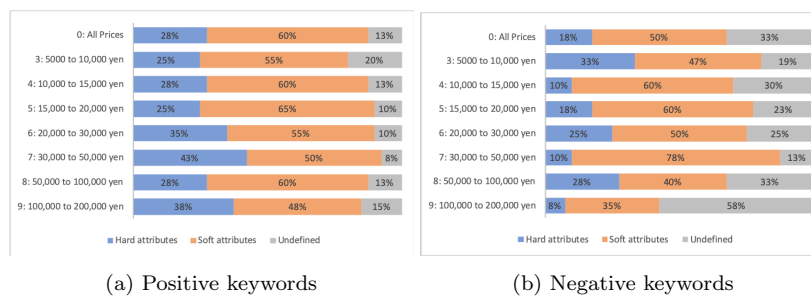


Fig. 6: Hard and soft attributes from the top English keywords for all price ranges

“poor” and its pairings in Table 13, we see that it is also service-related concepts that the Western tourists are disappointed with. ~~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.~~

In contrast, the English-speaking customers are easily upset by any lack of cleanliness and smoke smell from cigarettes. Surprisingly, the cigarette smell is an issue even in the middle to high-class hotels above 30,000 yen per night. However, above 50,000 yen per night, this problem seems to disappear from the list of top 10 concerns. Old and dated buildings seem to be a concern for both populations. On the positive side, for all price ranges considered, English-speaking tourists value staff friendliness over room quality when considering their satisfaction. In contrast, Chinese tourists consider location and transportation more often.

~~We also can~~ 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, ~~while it appears~~ 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 yen to 200,000 yen per night. Meanwhile, the term “shopping” ~~only appears in the~~ appeared solely in the top 10 positive keywords list for English speakers who stayed in rooms priced 20,000 to 30,000 yen per night ~~top 10 positive keywords list for English-speakers.~~

In our research questions 2a and 2b, we ponder how customers of both cultural backgrounds evaluate hard and soft attributes of the hotel and how they differ in those evaluations. Here we define hard attributes as those relating ~~to~~ 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 physical, structural, or environmental aspects. These are often impossible or impractical to change by the hotel management and staff, such as facilities, infrastructure, the surroundings, view, or location convenience. On the other hand, hotel staff and management can change soft attributes, for example, by improving the hotel's services via training or hiring specialized staff, improving the quantity or quality of amenities, bedsheets, or general cleanliness 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 found-discovered that Chinese tourists are mostly positively reacting more-mostly positively react to the hotel's hard attributes. There is a slightly hard leaning (53%) concern with hard attributes in negative sentences, albeit this is-, albeit the negative evaluations are more uniform than the positive evaluations. English-speaking tourists, on-, with a tendency of 53 % towards hard attributes. On the other hand, are both positively and negatively English-speaking tourists were more responsive to soft attributes, either positively or negatively. In the case of negative keywords, English-speaking tourists are overwhelmingly more concerned with they were more concerned about the hotel's soft attributes dissatisfaction somehow.

One factor that both populations have-had in common is that, when perceiving the hotel negatively, the “老 (old)”, “dated”, “outdated”, “dated.” “outdated.” or “陈旧 (obsolete)” aspects of the room or the hotel are-being criticized across-, surprisingly, were surprisingly criticized across most price ranges. This is, however, 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 are-satisfied mostly by Japanese hotels'-were mainly satisfied by big and clean spaces. From the adjectival pairings that we in Japanese hotels. The adjectival pairings extracted with dependency parsing and POS tagging in Table 10, we can observe that mostly they mean-(Table 10) imply big and clean rooms. Other mentions are also included big markets nearby or a big bed. We can observe that across Across different price ranges, the usage of the word “大 (big)” increases as the hotel increases in price. However, we can see that they still react positively in a significant manner in cheaper hotels. Inspecting 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 can see that there are- we notice that there were also many references to the public bathing facilities in the hotel. We can also see them mentioned as Such references were also implied by a word pairing “棒 温泉 (great hot spring)”.

In Japan, there ~~is what is called~~ are the so-called “銭湯 (*sentō*)”~~;~~ which are artificially ~~made~~ constructed public bathing facilities, ~~on occasions~~ including saunas and baths with unique qualities. On the other hand, there are natural hot springs, called “温泉 (*onsen*)”~~. They can either be bathing in the~~ natural source of the water or using the hot springs. ~~However, they are interchangeable if natural hot spring water is used~~ in artificially made tiled bath facilities. It is a Japanese custom ~~and culture~~ that all customers ~~use the facilities after cleaning first~~ clean themselves in a shower and ~~go into the baths without any clothes. It can afterward use the baths nude.~~ It could be a cultural shock for many tourists ~~;~~ but this is ~~but~~ a fundamental attraction for many others.

~~However,~~ Chinese customers are satisfied with the size of the room or ~~the bed~~ is a hard attribute. ~~Without considering rebuilding the hotel bed; however,~~ it is not trivial to ~~improve on.~~ ~~On the other hand~~ change this. ~~In contrast,~~ cleanliness is mostly ~~relating~~ 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 then variably 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 ~~is~~ was a mention of criticizing the “一般 卫生 (general hygiene)” of the hotel, although it ~~is~~ was an uncommon pair. Therefore, we can assert that cleanliness ~~is~~ was an important soft attribute for Chinese customers ~~and that they are mostly pleased with their expectations being met,~~ and they were mostly pleased when their expectations were fulfilled.

~~One key component we found in Chinese customer satisfaction soft factors is~~ 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, ~~like~~ such as “food” or “eating,” ~~which were lower ranking~~ and were lower-ranking. In contrast, the word “早餐 (breakfast)” ~~refers possibly to its inclusion in~~ referring to the hotel commodities, was frequently used in positive texts compared to other food-related words ~~.~~ ~~The word “早餐 (breakfast)” is also observed~~ across all price ranges, ~~although albeit~~ at different priorities in each of them. ~~However, we assert that it is~~ For this reason, we regard it as an important factor. ~~Observing word pairs from~~ From the word pairs of the positive Chinese keywords in Table 10, we can also ~~see~~ note that “不错 (not bad)” is paired ~~as~~ with “不错 早餐 (nice breakfast)” in four of the seven price ranges with reviews available as part of the top ~~4~~ four pairings. It is only slightly lower ~~on~~ in other categories, although it is not ~~shown on~~ depicted in the table. Thus, we consider that a recommended strategy for hotel management is to invest in the inclusion or ~~betterment~~ 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 ~~can see that~~ observed at least three words ~~relate directly were related~~ to staff friendliness and services ~~being “staff,” “helpful in the general database: “staff,” “helpful,” and “friendliness” in the general database.”~~ The word “staff” is the ~~second most frequently used word highest-ranked of these three, ranking second~~ for satisfied customers across most price ranges ~~and only third in one of them.~~ ~~Adding to that, “helpful” and “friendly” follow it lower in the list in most price~~ ~~ranges.~~ The word “good” ~~is mostly about~~ mainly refers to the location, the service, breakfast, or English availability in Table 12. ~~Like~~ Similar to Chinese customers, Western customers also ~~seem seemed~~ to enjoy the included breakfasts regarding their satisfaction keyword pairings. However, the relevant word does not appear ~~directly~~ in the top 10 list ~~as in directly, in contrast to~~ their Chinese counterparts. The words “helpful” and “friendly” are mostly paired with “staff,” ~~“concierge,” “desk,”~~ “concierge,” “desk,” and “service.” ~~When we look at.~~ By considering the negative keyword ~~“poor”~~ “poor” and its pairings in Table 13, we ~~see that it is also realized once again that Western tourists were disappointed with service-related concepts that the Western tourists are disappointed with when they react concepts and reacted~~ negatively.

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

An interesting detail of the keyword ranking is that the word “comfortable” ~~is was~~ was high on the satisfaction factors, and “uncomfortable” was high on the dissatisfaction factors. The words ~~are were~~ paired with nouns like ~~“bed,” or “room,” “pillow” or “mattress,” generally referring such as “bed,” “room,” “pillow,” and “mattress,” when they generally referred to their sleep conditions in the hotel. It seems that Western tourists are highly sensitive to comfort levels in the hotels and whether it reaches 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.~~

~~While less high~~ Albeit lower in priority, the price range of 15,000 to 20,000 yen hotels also ~~mentions includes~~ includes “free” as one of the top 10 positive keywords, ~~paired mostly with “wifi,” mainly paired with “Wi-Fi.”~~ This price range ~~is mostly for corresponds to~~ corresponds to business hotels, where ~~we infer users would be~~



~~expecting users would expect~~ this feature the most. ~~Western tourists are highly sensitive to comfort levels in the hotels and whether it reaches their expectations.~~

#### 930 6.4 Tobacco, ~~what's that~~ an unpleasant smell ~~?in the room~~

A concern for Western tourists was uncleanliness and the smell of ~~tobacco~~ cigarettes in their room, which can be ~~considered a soft attribute. Tobacco was found not only as a standalone word~~ regarded as soft attributes. Cigarette smell was an issue even in the middle- and high-class hotels, of which the rooms were  
 935 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 as  
in word pairs in Table 13. ~~We can find other related word pairs such as "funny smell". Upon manual inspection of.~~ By manually inspecting a sample of re-  
 940 views with this keyword, we found-noticed that the room was often advertised as non-smoking, ~~yet,~~ however, the smell permeated the room and curtains. Another common complaint was that there were no ~~non-smoking facilities available at all in the first place~~ nonsmoking facilities available. The smell of smoke can completely ruin some customers' stay ~~and give a bad impression to~~  
 945 ~~review writers,~~ leading to bad reviews, thereby lowering the number of future customers.

~~However, in comparison, Chinese customers seem~~ In contrast, Chinese customers seemed not to be bothered by this ~~at all. We consulted studies involving the use of tobacco in different countries. Previous research states~~  
 950 Previous research has stated that 49–60 % of Chinese men (and 2.0–2.8 % of women) currently smoke or ~~have smoked before~~ smoked in the past. This was ~~taken derived~~ from a sample of 170,000 Chinese adults in ~~2013-2014~~ 2013-2014, which is high compared to many English-speaking countries (Zhang et al. 2019; World Health Organization 2015).

955 Japan has a polarized view on the topic of smoking. ~~Despite being~~ Although it has one of the world's largest tobacco markets, ~~its 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 ~~urged~~ suggested and not mandatory to ~~have lift~~ smoking restrictions in restaurants,  
 960 bars, hotels, and public areas. ~~However, many~~ Many places have designated smoking rooms ~~are available~~ to keep the smoke in an enclosed area and avoid bothering others. ~~Despite this~~

Nevertheless, businesses, especially those who cater to certain customers, ~~will generally be discouraged from having~~ are generally discouraged by smoking  
 965 restrictions if they want to ~~keep maintain~~ their clientele. ~~If Japanese hotels want to~~ To cater to all kinds of customers, including Western and Asian ~~alike, they,~~ Japanese hotels must provide spaces without tobacco smell. ~~After all, even if it~~ Even if the smoke does not bother a few customers, the lack of ~~smell~~ would 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 ~~transport, and nearby shops' availability were observed~~ 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~~ 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 ~~signal that while it was not the priority for either of them, the~~ signify that the hotel's location ~~is was~~ a secondary but still important point ~~in the hotel's for their~~ satisfaction. However, since this is a hard attribute, ~~unchangeable to the hotel's management,~~ it is not often considered in the literature. ~~Upon inspection of~~ By examining examples from the data, we ~~found-recognized~~ that most customers were satisfied if the hotel was near ~~to~~ at least two ~~other subjects of the following facilities:~~ subway, train, and convenience stores.

Japan is a country with a peculiar public ~~transport system. The rush~~ hour makes for a subway filled to the brim with people in suits making their commute transportation system. During rush hour, the subway is crowded with commuters, and trains and subway stations ~~in Tokyo~~ create a confusing public ~~transport transportation~~ map for a visitor in Tokyo. Buses are also available, ~~although albeit~~ less used than ~~the~~ rail systems in ~~the big metropolitan~~ cities. These three ~~are unusually means of transportation are usually~~ affordable in price. ~~Then there are the more expensive transports~~ There are more expensive means, such as the bullet train *shinkansen* for traveling across the country ; and taxis. ~~Taxis in Japan are a luxury. The latter is a luxury in Japan~~ compared to other countries. In ~~less developed countries, a taxi is the cheap method of~~ transport of choice. In Japan, taxis ~~are made to~~ provide a high-quality experience with a matching price. ~~This means that for tourists~~ Therefore, for people under a budget, subway availability and maps or GPS applications ~~and,~~ 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 ~~;~~ on the other hand, are also famous worldwide. ~~Japanese convenience stores are a haven for the traveler in need. It offers anything 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, ~~international ATMs,~~ among other things and international ATMs. If some trouble ~~occurred~~ occurs, or a traveler forgot to pack a particular item, it is ~~almost sure most certain~~ that they can find it.

Therefore, considering that both ~~transport transportation~~ systems and nearby shops are points of interest for Chinese and Western tourists, ~~Japanese~~

hotels have to carefully choose their location from the moment they are constructed. While not a top priority, this is a universal factor for both customer groups, and it can be an instant way to generate positive reviews and perhaps offering guide maps and information about these as an appeal point could result in greater satisfaction.

## 7 Discussion

Below we explore the possible interactions with Japanese hospitality, the differences between Chinese and Western tourists, the possible cause for them, how they vary across different price ranges, and what they imply for the industry. We also discuss the differences between the hotel's hard and soft attributes and how they contribute to customers' satisfaction.

### 7.1 Western and Chinese tourists in the Japanese hospitality environment

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

Our results indicate that out of the two, Western tourists are the most more satisfied with soft attributes ; such as friendly and helpful staff in Japan than Chinese tourists. As explained earlier in this paper, Japan is famous well known for its customer service all over the world. Respectful language and bowing are not exclusive to high priced high-priced hotels or businesses. These can even be found; these are met in convenience stores . The level of hospitality in even the cheapest of convenience stores as well. Even in the cheapest convenience store, the level of hospitality is starkly different from Westerner experiences. While it could be a culture shock to some, it is mostly seen positively. After all, the Japanese staff respectfully treats all customers. However, for some customers, this could be the best way they have been treated until that moment. Now, in higher priced Western culture and perhaps unexpected. In higher-priced hotels, the adjectives used to praise the service go ranged from normal descriptors like "good" to higher levels of praise like "wonderful staff", "wonderful experience", "excellent service", and "excellent staff". We can also see that . Furthermore, Kozak (2002) and Shanka and Taylor (2004) have also found proven that hospitality and staff friendliness is a vital determinant in are two determinants of Western tourists' satisfaction.

However, ~~we can see from~~ the negative English keywords ~~that a big indicate~~ that a large part of the dissatisfaction with Japanese hotels ~~stems stemmed~~ from a lack of hygiene and room cleanliness. Although Chinese customers ~~only had~~ had solely positive keywords about cleanliness, English-speaking customers ~~have found deemed~~ many places unacceptable to their standards. ~~This is particularly true at hotels~~, 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 ~~stench and general~~ dirty smell and lack of general hygiene. Kozak (2002) also ~~found proved~~ that hygiene and cleanliness were essential satisfaction determinants for Western tourists. However, in the previous literature, this was linked merely to satisfaction. In ~~comparison contrast~~, our research ~~uncovered that words relating to cleanliness are revealed that words related to cleanliness were~~ mostly linked to dissatisfaction. ~~Westerners could be said to have~~ We could assert that ~~Westerners had~~ a high standard of room cleanliness ~~when~~ compared to their Chinese counterparts.

According to previous research, ~~we can see that~~ Western tourists are already inclined to appreciate hospitality for their satisfaction. When presented with Japanese hospitality, this expectation is met and overcome. In contrast, ~~we can see from our results that Chinese tourists had less focus on according to our results, Chinese tourists were more concerned about room quality rather than~~ hospitality, staff, or service ~~and were more concerned with room quality~~. However, when analyzing the word pairs for “不错 (not bad)” and for “棒 (great)”, we can see that they ~~do~~ praise staff, service, and breakfast. ~~Observing~~ By observing the percentage of hard to soft attributes in Figure 5, however, we ~~know discover~~ that Chinese customers ~~are satisfied more were~~ more satisfied with hard attributes, ~~compared to the Western tourists who seem 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 ~~that stems from this derived~~ is less than that if it was expected. ~~On the other hand, we have the phenomenon of~~ In contrast, some tourists report a “nice surprise”: ~~When when~~ an unknown need is unexpectedly met, there is more satisfaction. It is necessary to note the difference between these two ~~phenomenons reactions~~. The “nice surprise” reaction fulfills a need unexpectedly. Perhaps the hospitality grade in Japan does not fulfill a ~~high enough need need high enough~~ for the Chinese population, thereby resulting in less satisfaction. For greater satisfaction, ~~the existence of a need being met is necessary a need must be met~~. However, the word “not bad” is at the top of the list ~~at in~~ most price ranges, and one of the uses is related to service. Thus, we cannot ~~say that they are conclude that they were~~ not satisfied with ~~this matter. Rather, they hold the service. Instead, they held~~ other factors at a higher priority, ~~considering; thus,~~ the keyword frequency ~~is was~~ higher for other pairings.

Another possibility ~~presents itself occurs~~ when we observe the Chinese tourists’ dissatisfaction factors. Chinese tourists may have expectations about

the Chinese visitors' ~~their~~ treatment that are not being met, even in this ~~high standard~~ high-standard hospitality environment. ~~Japan is known worldwide for their hospitality, but they are also known historically to be monolingual and have~~ 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 ~~taking shape, 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 just~~ the treatment towards Chinese people ~~could be examples~~ It is natural to be dissatisfied since traveling in a strange land without knowing the language can be a daunting experience. 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. ~~It seems like this is a problem that is not singular~~ 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 ~~brought us closer to attempted to determine~~ the answer. ~~On the one hand, there is a possibility~~ It is possible that Chinese customers ~~did have~~ had high-grade hospitality and ~~did not get were~~ equally satisfied with Westerners. In that case, it appears that the difference in perception stems from a psychological source. ~~Expectation ; expectation~~ leads to satisfaction and a lack of expectation results in lesser satisfaction. ~~On the other hand, there~~ 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 ~~from our results~~ which of these ~~could be the case. One thing is clear for hotel managers, however. Competing two is most likely from our results. However, competing~~ in hospitality and service ~~does include includes~~ language services, especially in the international tourism industry. Better multilingual support can only improve ~~that already high the hospitality~~ standard in Japan. Considering that most of the tourists in Japan come from other countries in Asia, ~~this is an endeavor that truly can bring benefits to their investment~~ multilingual support is beneficial. Proposals for this endeavor include hiring ~~Chinese speaking Chinese speaking~~ staff, preparing pamphlets in Chinese, or ~~have having~~ a translator application readily available with staff trained in interacting through an electronic translator.

## 7.2 Hard vs. soft satisfaction factors

As ~~we~~ stated in section ~~3.23.2~~, previous research ~~is focused mostly~~ 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. ~~Examples of hard attributes include the hotel's~~

location relative to public transport and shops, language immersion of the country, noise pollution, or weather. Because our study left 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 ~~to be decided statistically~~ were decided statistically in our study via customers' online reviews, we can see the importance of ~~those 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~~ in contrast, only 20 % are soft factors. The rates are similar for most price ranges ~~excepting~~ except the highest-priced hotels, ~~where 35% of the keywords are undefined~~. However, ~~the soft attributes are still similar at 18%~~. However, two of these ~~managerial words~~ soft attributes are all concentrated at the top of the list (“不错 (not bad)”, “干净 (clean)”), ~~plus and~~ the adjective pairs ~~relating~~ related to soft attributes of “不错 (not bad)” ~~which are are also~~ at the top in most price ranges ~~as well~~. Chinese tourists ~~could may~~ expect spaciousness and cleanliness when coming to Japan. ~~That expectation could be caused by~~ The expectation may be due to reputation, previous experiences, or cultural backgrounds. ~~Some scholars argue that different cultures have different room size perceptions~~. Although the study subjects are German and South Korean, the study presents the results as differences influenced by Asian and Western cultures. We argue that one country is not representative of others' cultures, so there can be differences between South Korea and China in room size perception. However, an interesting point appears. It could be that a different room size perception affects the satisfaction of Chinese tourists in contrast with Westerners. Westerners only start placing a priority on praising room size as the price of the hotel goes up. We can 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 ~~other few~~ factors found in previous literature could be linked to the keyword “环境 (environment or surroundings)” as well. This keyword ~~is present in~~ was found for hotels priced at more than 20,000 yen per night.

In ~~comparison~~ contrast, English speakers are mostly satisfied with the ~~hotel's~~ 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 to 20,000 yen per night ~~price range~~. This price range corresponds to ~~affordable business hotels, which corresponds to~~, for example. English-speaking customers ~~also have soft attributes at the top of their list~~. The exception, affordable business hotels. The exception to this is the hard attribute that is the ~~hotel's~~ hotels' location, which is consistently around the middle of the top 10 lists for all price ranges. ~~If one considers both Chinese and Western tourists' satisfaction, a hotel can improve to attract more customers in the future. If it was the other~~



way around, and the satisfaction was related more with hard attributes overall for 1020 both cultures, hotels would have to compete solely on their location.

For both customer groups, the main reason for dissatisfaction ~~is was~~ pricing, which can be interpreted as a concern about value for money. However, ~~it is interesting to note that while~~ English-speaking customers ~~complain about price with a lower rank~~ complained less about the price in lower-priced hotels. In contrast, ~~the~~ Chinese customers consistently have had “价格 (price)” as a ~~top the first~~ or second-most concern across all price ranges. A paper studying study on Chinese tourists found that they had this concern (Truong and King 2009). However, our results indicate that this ~~is less of a cultural attribute in Japanese hotels and~~ has more to do with the pricing of hotels overall. ~~The tourists coming to Japan could be both experienced travelers or first-time travelers. However, the fact is that their expectation of the price for hotels was lower than what they found in Japan 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 x 1 meters cost around ~~3,000 to 6,000~~ 3000 to 6000 yen per night. Bigger business hotel rooms are relatively expensive, ranging from ~~5,000-5000~~ to 12,000 yen per night. For comparison, hotels in the USA with a similar quality can ~~be charge~~ half the price.

Around half of the dissatisfaction factors for both Chinese and Western customers are caused by issues that could be ~~solved with improved management.~~ ~~The previous improved; this~~ is true for all price ranges. ~~Of course, the~~ The improvements could be staff training (perhaps in language), hiring professional cleaning services for rooms with cigarette smoke smells, or improving the bedding. ~~All of these options; however, these considerations~~ can be costly. However, ~~this paper provides a good guideline for which factors to consider first and which ones will be best suited to each customer group. Hotels can use the price range categorization in order to choose the appropriate strategy as well. However, once the hotel's location and construction are set for Chinese customers, not much else can be done to satisfy them; only a few changes can be made to satisfy Chinese customers~~ further. As mentioned ~~before previously~~, Chinese language availability is ~~another a~~ soft attribute that can be improved with staff and training investment.

~~On the other hand,~~ Western tourists are ~~all around dissatisfied with mostly~~ mainly dissatisfied with soft attributes. ~~They show this by having a low~~ 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 % ~~at most~~ in the price range ~~from of~~ 30,000 to 50,000 yen per night in a hotel. ~~The room for improvement for~~ Improvement scope for Western tourists is more extensive than ~~that for~~ their Chinese counterparts. As such, it presents a ~~bigger larger~~ investment opportunity. ~~As mentioned earlier in this paper, Westerners are known as “long-haul” customers, spending more than 45% of their budget on hotel lodging. On the other hand, their Asian counterparts only spend 25% of their budget on hotels. With bigger returns on managerial improvements, it seems like we can recommend investing in improving attributes that dissatisfy~~

Western customers, such as cleanliness and removing tobacco smell. Making more hotel facilities tobacco-free and deodorizing the rooms can be a low-cost investment that could increase returns many times over.

However, the opposite argument could also be made that Chinese customers provide a more significant number of customers, even though they tend to spend less on lodging. Attracting a large number of Chinese customers can be a viable strategy for hotels. However, as mentioned before, they tend to focus more on hard attributes, leaving language barrier-breaking as one of the few strategies to accomplish this.

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 preference of hard or soft attributes.

In , subjects from 10 different countries were compared in their expectations of service quality and analyzed through the lens of Hofstede's typology of culture . That study states that although culture has no one 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 versus 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 the fifth dimension that Hofstede proposed, which was added afterward with the alternate name *Confucian dynamic*. Academics with a Chinese background criticized Hofstede for being misinformed on the philosophical aspects of Confucianism, as well as being a difficult dimension to measure . Other models, such as the GLOBE model, also take issue with some of Hofstede's dimensions and replace them with others, making a total of 9 dimensions . The *masculinity-femininity* dimension, for example, is proposed to be instead 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. Gender stereotypes such as these have also been the subject of critique for Hofstede's model. Our study agrees with these critiques and, therefore, will avoid considering these for our discussion.

The backgrounds of collectivism in China and individualism in Western countries have been studied before . These backgrounds and 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 these differences in expectations exist .

For our purposes in 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. Perhaps 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 gives less importance to these concepts in their lives.

Nevertheless, one could argue that a Chinese cultural attribute emphasizes the environment and the place one is in towards satisfaction, rather than the way one is treated. According to previous research, Chinese tourists are collectivist, while Westerners are individualists. A more anthropocentric and individualistic Western culture could result in more of their expectations and priorities be related to how one is treated in social circumstances, rather than the environment one is in. According to , highly individualistic customers have a higher expectation of empathy and assurance from the provider than collectivist customers. Empathy and assurance from the provider are aspects of service, a soft attribute of a hotel.

Among other dimensions in both models, we can consider uncertainty avoidance. High uncertainty avoidance customers would carefully plan their travel and therefore have higher expectations towards service. On the other hand, lower uncertainty avoidance customers would have certain room for risk in their decisions, and therefore face less disappointment with different expectations. However, according to , the difference between China and the USA in uncertainty avoidance is not so clear when measuring with the Hofstede typology and the GLOBE typology. While the USA is not representative of Western society, this dimension might not be the one causing the difference in hard-soft attribute satisfaction between Chinese and Western cultures. Another factor, power distance, was also different when measured by Hofstede's method compared to the GLOBAL method, so we decided against making this comparison.

### 7.3 Satisfaction across different price ranges

In previous sections of this paper, we have mentioned the differences reflected in hotel price ranges. Nevertheless, it is interesting to discuss this further. The most visible change in satisfaction factors across differently priced hotels is the change in voice to describe their satisfaction with the same topics. We can know when describing satisfaction. We noticed this by observing the adjective + noun 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 more highly-priced higher-priced hotels are "wonderful" and "excellent". In Chinese, the change goes 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. However, when we look at Regarding the negative keywords, the change is however, the change ranges from "annoying" or "worst", to "disappointing". Here we can see how expectations influence satisfaction and dissatisfaction in different ways. disappointing" to "worst."

In this paper, we follow the definition of satisfaction by Hunt (1975), where meeting or exceeding expectations produces satisfaction. ~~Therefore, the lack~~ Conversely, the failure of meeting expectations ~~would cause dissatisfaction.~~ In the cases above, we can infer causes dissatisfaction. We can assume that a customer that pays more for a ~~higher class of~~ higher class experience has higher expectations. ~~This is true in dissatisfaction, where their expectation is higher in a more expensive hotel.~~ As such For example, in a highly-priced hotel, any lack of cleanliness can lead to disappointment ~~or outrage~~. In the case of English-speaking customers in the ~~30,000 to 50,000~~ 30,000 to 50,000 yen per night price range, cigarette smell is particularly disappointing. However, we consistently see customers with high expectations for high-class hotels ~~react~~ 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 ~~interactions with expectations.~~ We can observe logical expectations. Customers set a standard in their mind expectations: logical and emotional. In the first case, customers are determined that the service must not fall below ~~or be disappointed—~~ a specific standard; for example, ~~a customer being disappointed with dirty~~ they can be disappointed with unhygienic rooms or cigarette smell.

In contrast, ~~we can observe emotional expectations, where in the second~~ case, customers have a vague idea of having a positive experience. ~~However, they but~~ do not measure it against any standard. For example, ~~having they expect~~ a pleasant customer service experience or ~~being treated hospitably a hospitable treatment~~ by the staff at a high-class hotel. Regardless of their knowledge ~~beforehand of the service to be provided~~ in advance, positive emotions ~~give offer~~ them a perception of exceeded expectations and high satisfaction. ~~This is where~~ Thus, hospitality and service ~~come into play and enhances~~ enhance the experience of the customers.

There are interesting differences between Chinese and English-speaking tourists in their ~~change in satisfaction factors~~ satisfaction to differently priced hotels. For example, ~~we can observe that the~~ Chinese tourists have “购物 (shopping)” as a top keyword in all the price ranges. In contrast, English-speaking tourists ~~only mention it~~ mention it only as a top keyword in the 20,000 ~~to 30~~ 30,000 yen price range. It is ~~common knowledge~~ widely known in Japan that ~~Chinese tourists coming to Japan with the express intention of shopping are common~~ many Chinese tourists visit Japan for shopping. Tsujimoto (2017) analyzed the souvenir purchasing behavior of Chinese tourists in Japan. ~~The study shows and showed~~ that common products besides food and drink are: electronics, cameras, cosmetics, and medicine, among ~~other more traditional souvenir items~~ such as objects representative of the culture or places that they visit Japan Tourism Agency (2014). ~~There is an understanding that tourists choose~~ Furthermore, Chinese tourists’ choice to shop in Japan ~~has more to do with~~ is more related to the quality of the items rather than their relation to the tourist attractions. Our results ~~suggest~~ suggested that Western tourists ~~are~~ were engaging more in tourist attractions ~~in comparison~~.

~~with shopping activities~~; 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 ~~after~~ if it concerns hotels of 15,000 yen or more; ~~and it rises in its ranking~~; thereafter, the more expensive the hotel is, the higher the ranking. In contrast, ~~Chinese customers have this keyword as their~~ 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), ~~while~~ whereas Western customers are more concerned ~~with~~ about hospitality (Kozak 2002).

While some ~~aspects of satisfaction and dissatisfaction change attributes'~~ value changes depending on the hotel's price range, some other ~~factors stay mostly~~ 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 ~~stays~~ remains part of the top 10 keywords, except for the most expensive one, where other keywords ~~take that place~~ replace keywords related to satisfaction or cleanliness in the ranking. ~~However, it is~~; 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 section 6.5, ~~transport~~ 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 ~~at most~~ a 13 % margin of error per attribute; ~~often being lower~~. This suggests that ~~culturally~~, 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 ~~presents~~ reached two important conclusions: one about hospitality and cultural differences ~~;~~ and another about managerial decisions towards two different populations. ~~As a whole~~ Overall, Chinese tourists ~~are not showing the most satisfaction towards~~ did not attach much importance to hospitality and service factors. Instead, they ~~focus~~ focused on the hard attributes of a hotel. ~~Either they do not get as much satisfaction from hospitality~~ In particular, they were not satisfied with hospitality as much as Western tourists ~~or feel were; otherwise, they felt~~ that basic language and communication needs are not being met, so they receive a lesser impression. ~~On the other hand, Western tourists are elated~~ were not met; thereby, they were not much satisfied. Western tourists were highly satisfied with Japanese hospitality ~~;~~ preferring and preferred soft attributes to ~~hard-set ones, hard ones.~~

The other conclusion is that managerial decisions ~~will~~ could mostly benefit Western tourists, except ~~that Chinese for~~ language improvements and breakfast inclusion ~~can satisfy more Chinese customers. Japan is recently seeing~~ 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 ~~Western students of 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 ~~can be improved upon~~ managerially 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 ~~can~~ could satisfy both customer types. ~~A low-cost investment could be to make~~ We suggest investing in making the facilities tobacco-free. Our results are also divided by price ranges, ~~so;~~ thereby, a hotel manager ~~can~~ could consider which analysis suits their hotel the most.

While not manageable after a hotel has finished its construction, hard attributes are essential to consider for managers. As previously stated, transport systems and nearby shops are points of interest for both Chinese and Western tourists. Japanese hotel managers have to consider the location and surroundings before the hotel is constructed. A suggestion could be to purchase land and start the construction after public plans to make new subway lines are made. 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.

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

## 8 Limitations and Future Work

This paper is not without its limitations. We analyzed satisfaction and dissatisfaction in this study, we analyzed keywords based on whether they appeared on satisfied reviews or dissatisfied ones. Following that, we attempted to understand the context in which these words were used these words' context by using a dependency parser and observing the related nouns. However, the study is limited in that it only analyzes a limitation is that it analyzed solely the words directly related to each keyword and does not follow the upstream or downstream path down did not search for further connections. This means that if the words are 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 would extract "good food," we extracted "good room" and "lacking food", but do 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 Western-style five-star resort and a high-end Japanese-style Japanese style ryokan. Services offered in these hotels are very of high quality, although albeit very different. However Nevertheless, most of our database is was focused on the middle range priced hotels, which is the services of which are comparably less varied in service. However, there is still a divide between western and Japanese style hotels.

An essential aspect of this study is that we focus focused on the satisfaction and dissatisfaction towards the expectations of individual aspects of the hotels.

This ~~gives gave~~ us insight into ~~which factors can focus on by hotel managers in applying this knowledge~~the factors that hotel managers can consider. However, ~~our study was limited in that the overall satisfaction of each customer~~each customer's overall satisfaction was not measured. ~~This could be done by rating the volume of text used to describe satisfaction factors against the text volume used for dissatisfaction. However, this imposes a few difficulties since it would require methods~~ that are out of the scope of this study.

~~paper~~. Another limitation is that ~~a large portion of the Asian tourists coming to Japan is Taiwanese and Korean. We could not analyze these populations because our team members do not know those languages. Aside from 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 ~~further into this topic~~these topics further. We plan to extend our data to research ~~for different trends for different~~different trends and regions of Japan ~~and~~ different kinds of hotels ~~and between~~ 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, ~~our objective was to analyze~~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*. ~~To answer our research questions 1a and 1b, we~~We extracted keywords from their online reviews ~~uploaded to the portal sites 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 ~~commonly found~~common pairs of adjectives and nouns, as well as single words. We divided ~~this~~these data by sentiment and hotel price range, ~~considering the~~ (most expensive room ~~for one night~~ /night).

~~In the context of Japanese hospitality, we~~We found that Western tourists ~~had the most satisfaction~~were most satisfied with staff behavior, cleanliness, and other ~~attributes relating to the hotel's services and hospitality~~soft attributes. However, ~~we found that~~ Chinese customers had other concerns ~~than hospitality when studying their satisfaction, for their satisfaction; they~~were more inclined to praise the room, location, ~~or~~and hotel's convenience. We found that ~~both cultures have a different reaction to this hospitality environment. Both cultures have a different way of reacting to different prices. From this the two cultures had different reactions to the hospitality environment and the prices. Thus,~~ we discussed two possible theories on why Chinese tourists ~~respond differently to Westerners in this~~responded differently ~~from Westerners in the~~ environment of *omotenashi*. One theory is that ~~while they are being treated well and react positively, the environment is not compatible~~



with them because of, although they were treated well, their experience was deteriorated by language or culture barriers, which lessens their experience. The second possible theory is that they react differently to hospitality since they do reacted to hospitality differently since they did not have the same expectations to be satisfied in the same way. We theorize. We theorized that a lack of expectations could result in lessened satisfaction, even if than that to the same service is presented if expected. On the other hand, even when they hold held high expectations in a highly-priced hotel, Western tourists show that Japanese hospitality exceeds their high-priced hotel, Japanese hospitality exceeded Western tourists' expectations, judging by their vocabulary for expressing their satisfaction. We consider considered that Western tourists are were more reactive to hospitality and service factors than their Chinese counterparts Chinese tourists.

Lastly, we measured the satisfaction and dissatisfaction factors, referring to that is, a hotel's hard and soft attributes. 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. On the other hand, hard attributes are physical and impractical elements to change, such as the size of a room that has already been constructed, the location of a hotel, closeness to convenient spots, or elements out of the control of the hotel managers or amenities. We found that, for satisfaction, Western tourists favor soft attributes. In contrast, Chinese tourists are favored soft attributes in contrast to Chinese tourists, who were more interested in the hard attributes of hotels, consistently across price ranges across all the price ranges consistently. For dissatisfaction, Western tourists are were also highly inclined to criticize soft attributes, such as cleanliness or cigarette smell in rooms. In contrast, Chinese tourists' dissatisfaction comes evenly derived from both hard and soft attributes evenly.

One possible approach for hotel managers is to improve the satisfaction levels of Chinese tourists work to satisfy Chinese tourists more, who dedicate less a lower percentage of their budget to hotels but are more abundant in number numerous. They are less satisfied with soft attributes but have an identifiable method of for improving satisfaction by lessening language barriers and providing a satisfactory breakfast. Another approach we discussed was focusing was focused on the cleanliness and comfort that Western tourists expect and making the hotels, comfort, and tobacco-free. We favor "long-haul space expected by Western tourists. "Long-haul" Western tourists, who spend almost half of their budget on hotels with this strategy. While, were favored. Although Westerners are less in number than Chinese tourists, it could prove to 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. This paper provides results and discussion that can be utilized as a guideline for managerial decisions when considering Chinese and Western tourists in Japan. We can observe their stark differences and shared attributes.



**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.

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

Conflict of interest: none

## References

- Aishima T, Sato Y, et al. (2015) The origin of Japanese omotenashi in Man-yo-shu. *Business & accounting review* 16:103–122
- Al-alsheikh A, Sato Y (2015) Characteristics of the hospitality, omotenashi in the traditional Japanese inn: A case study of Kagaya. *Business & accounting review* 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. *Information Processing & Management* 56(2), DOI 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, DOI 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. *International Journal of Hospitality Management* 16(2):181–190
- Balbi S, Misuraca M, Scepti G (2018) Combining different evaluation systems on social media for measuring user satisfaction. *Information Processing & Management* 54(4):674–685, DOI 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. *Journal Of Marketing* 67(4):103–117, DOI 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. *International Journal of Hospitality Management* 12(4):313–322, DOI 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. *Journal of Hospitality Marketing & Management* 25(1):1–24, DOI 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. *Journal of Travel & Tourism Marketing* 30(1-2):23–40, DOI 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. *Journal of Travel & Tourism Marketing* 32(1-2):18–33, DOI 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, URL <http://nlp.stanford.edu/pubs/acl-wmt08-cws.pdf>
- Chang R, Kivela J, Mak A (2010) Food preferences of Chinese tourists. *Annals Of Tourism Research* 37(4):989–1011, DOI 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. *International Journal of Quality & Reliability Management* 17(2):116–132, DOI 10.1108/02656710010304537

- Choi TY, Chu R (2001) Determinants of hotel guests' satisfaction and repeat patronage in the Hong Kong hotel industry. *International Journal of Hospitality Management* 20(3):277–297, DOI 10.1016/S0278-4319(01)00006-8, URL <http://www.sciencedirect.com/science/article/pii/S0278431901000068>
- 1690 Cortes C, Vapnik V (1995) Support-vector networks. *Machine Learning* 20(3):273–297, DOI 10.1007/bf00994018
- Coulmas F, Watanabe M (2002) Japan's nascent multilingualism. *Contributions To The Sociology Of Language* 87:249–274
- 1695 Dongyang Z, Mori T, Hayashi K, et al. (2015) A study on preferences and behavioral patterns of Chinese tourists in Kansai region, Japan. *Konan Economic Papers* 55(1-2):31–46, DOI 10.14990/00001507
- Donthu N, Yoo B (1998) Cultural influences on service quality expectations. *Journal of service research* 1(2):178–186
- Engel J, Blackwell R, Miniard P (1990) *Consumer Behavior* (6th edition). Dryden Press, Hinsdale, Illinois, USA
- 1700 Fang B, Ye Q, Kucukusta D, Law R (2016) Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management* 52:498–506, DOI 10.1016/j.tourman.2015.07.018
- Fang T (2003) A critique of Hofstede's fifth national culture dimension. *International journal of cross cultural management* 3(3):347–368
- 1705 Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. *AI magazine* 17(3):37–37, DOI 10.1609/aimag.v17i3.1230, URL <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. *Tourism Management* 74:24 – 42, DOI 10.1016/j.tourman.2019.02.011, URL <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. *Journal of Sustainable Tourism* 26(4):668–684, DOI 10.1080/09669582.2017.1377722
- 1715 Hargreaves C (2015) Analysis of hotel guest satisfaction ratings and reviews: An application in Singapore. *American Journal Of Marketing Research* 1(4):208–214
- Heinrich P (2012) The making of monolingual Japan: Language ideology and Japanese modernity, vol 146. *Multilingual matters*
- Hofstede G (1984) *Culture's consequences: International differences in work-related values*, vol 5. Sage Publications
- 1720 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. *Advances in global leadership* 1(2):171–233
- Hu YH, Chen YL, Chou HL (2017) Opinion mining from online hotel reviews – a text summarization approach. *Information Processing & Management* 53(2):436–449, DOI 10.1016/j.ipm.2016.12.002
- 1725 Huang J (2017) The dining experience of Beijing roast duck: A comparative study of the Chinese and English online consumer reviews. *International Journal of Hospitality Management* 66:117 – 129, DOI 10.1016/j.ijhm.2017.07.003, URL <http://www.sciencedirect.com/science/article/pii/S0278431917301251>
- 1730 Hunt JD (1975) Image as a factor in tourism development. *Journal of travel research* 13(3):1–7
- Ikeda N (2013) Omotenashi: Japanese hospitality as the global standard. In: *Management Of Service Businesses In Japan*, World Scientific, chap 9, pp 145–154
- 1735 Japan National Tourism Organization (2019) Nationality / monthly foreign visitors to Japan (2003–2019). Tech. rep., Japan National Tourism Organization, URL [https://www.jnto.go.jp/jpn/statistics/since2003\\_visitor\\_arrivals.pdf](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 Pravnog fakulteta u Splitu* 51(3):681–696
- 1740 Jia SS (2020) Motivation and satisfaction of Chinese and U.S. tourists in restaurants: A cross-cultural text mining of online reviews. *Tourism Management* 78:104071, DOI 10.

- 1016/j.tourman.2019.104071, URL <http://www.sciencedirect.com/science/article/pii/S0261517719302687>
- 1745 Jones T, Nagata S, Nakajima M, Masuyama K (2009) Prefectural branding in Japan – tourism, national parks and the Shinshu brand. *Place Branding and Public Diplomacy* 5(3):192–201, DOI 10.1057/pb.2009.13
- Kim C, Lee S (2000) Understanding the cultural differences in tourist motivation between Anglo-American and Japanese tourists. *Journal of Travel & Tourism Marketing* 9(1-2):153–170, DOI 10.1300/J073v09n01.09
- 1750 Kim K, Joung 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. *Technological Forecasting and Social Change* 123:362–369, DOI 10.1016/j.techfore.2017.01.001
- 1755 Kozak M (2002) Measuring tourist satisfaction with multiple destination attributes. *Tourism Analysis* 7(3-4):229–240(12), DOI 10.3727/108354203108750076
- Kuboyama T (2020) “Omotenashi” must comprise hospitality and service. In: *Serviceology for Services*, Springer Singapore, pp 34–53, DOI 10.1007/978-981-15-3118-7\_3, URL [https://doi.org/10.1007/978-981-15-3118-7\\_3](https://doi.org/10.1007/978-981-15-3118-7_3)
- 1760 Likert R (1932) A technique for the measurement of attitudes. *Archives of psychology*
- Liu Y, Bi JW, Fan ZP (2017) Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion* 36:149–161, DOI 10.1016/j.inffus.2016.11.012
- 1765 Liu Y, Huang K, Bao J, Chen K (2019) Listen to the voices from home: An analysis of Chinese tourists’ sentiments regarding Australian destinations. *Tourism Management* 71:337 – 347, DOI 10.1016/j.tourman.2018.10.004, URL <http://www.sciencedirect.com/science/article/pii/S0261517718302395>
- Loh S, Lorenzi F, Saldaña R, Lichnow D (2003) A tourism recommender system based on collaboration and text analysis. *Information Technology & Tourism* 6(3):157–165, DOI 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, URL <http://www.aclweb.org/anthology/P/P14/P14-5010>
- 1775 de Marneffe MC, Manning CD (2008) Stanford typed dependencies manual. Tech. rep., Stanford University, URL [https://nlp.stanford.edu/software/dependencies\\_manual.pdf](https://nlp.stanford.edu/software/dependencies_manual.pdf)
- 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
- 1780 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. *Hospitality Research Journal* 20(3):35–64
- 1785 Oliver RL (1981) Measurement and evaluation of satisfaction processes in retail settings. *Journal of retailing*
- Powers D (2011) Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal Of Machine Learning Technologies* 2(1):37–63, URL [http://www.flinders.edu.au/science\\_engineering/fms/School-CSEM/publications/tech\\_reps-research\\_artfcts/TRRA\\_2007.pdf](http://www.flinders.edu.au/science_engineering/fms/School-CSEM/publications/tech_reps-research_artfcts/TRRA_2007.pdf)
- 1790 Rajman M, Besançon R (1998) Text mining-knowledge extraction from unstructured textual data. In: *Advances in data science and classification*, Springer, pp 473–480
- 1795 Ren G, Hong T (2019) Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews. *Information Processing & Management* 56(4):1425–1438, DOI 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. *Ecological Economics* 107:195–205, DOI 10.1016/j.ecolecon.2014.07.019
- 1800

- Ryan C, Mo X (2001) Chinese visitors to New Zealand - demographics and perceptions. *Journal Of Vacation Marketing* 8(1):13–27, DOI 10.1177/135676670200800103
- Shanka T, Taylor R (2004) An investigation into the perceived importance of service and facility attributes to hotel satisfaction. *Journal of Quality Assurance in Hospitality & Tourism* 4(3-4):119–134, DOI 10.1300/J162v04n03\_08
- Shannon C (1948) A mathematical theory of communication. *Bell System Technical Journal* 27(3):279–423, DOI 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
- 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. *Tourism Management* 58:245–252, DOI 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. *International Journal Of Tourism Research* 11(6):521–535, DOI 10.1002/jtr.726, URL <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jtr.726>
- Tsujimoto N (2017) The purchasing behavior of Chinese tourists at popular visiting areas in Japan. *Journal of Global Tourism Research* 2(2):99–104
- Uzama A (2012) Yokoso! Japan: Classifying foreign tourists to Japan for market segmentation. *Journal of Hospitality Marketing & Management* 21(2):132–154, DOI 10.1080/19368623.2011.615016
- Vermeulen I, Seegers D (2009) Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management* 30:123–127, DOI 10.1016/j.tourman.2008.04.008
- World Health Organization (2015) WHO global report on trends in prevalence of tobacco smoking 2015. URL [https://apps.who.int/iris/bitstream/handle/10665/156262/9789241564922\\_eng.pdf](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. *International Journal of Hospitality Management* 28(4):586–593, DOI 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, URL [http://repository.upenn.edu/ircs\\_reports/38](http://repository.upenn.edu/ircs_reports/38)
- Xiang Z, Gretzel U (2010) Role of social media in online travel information search. *Tourism Management* 31(2):179–188, DOI 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? *International Journal of Hospitality Management* 44:120–130, DOI 10.1016/j.ijhm.2014.10.013
- Xiumei S, Jinying W (2011) Cultural distance between China and US across GLOBE model and Hofstede model. *International Business and Management* 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. *International journal of hospitality management* 55:57–69, DOI 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. *Journal of Travel Research* 53(4):433–447, DOI 10.1177/0047287513500391
- Ye Q, Law R, Gu B (2009) The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management* 28(1):180–182, DOI 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, Drohanova 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, DOI 10.1109/icrd.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 & tobacco research: official journal of the Society for Research on Nicotine and Tobacco* ntz020, DOI 10.1093/ntr/ntz020
- Zhou L, Ye S, Pearce PL, Wu MY (2014) Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management* 38:1–10, DOI 10.1016/j.ijhm.2013.12.004