



**UNIVERSIDAD DE MÁLAGA**  
**FACULTAD DE TURISMO**

Method for Automatically Extracting of  
Tourism Destination Image from Chinese  
Social Networks. The case of Spain

Metodología para la Extracción Automática de  
la Imagen del Destino en Redes Sociales  
Chino-Parlantes. El caso España.

**Fang-wei Chen**

Universidad de Málaga  
Facultad de Turismo,  
Málaga, España

2015



UNIVERSIDAD  
DE MÁLAGA

AUTOR: Fang-wei Chen

 <http://orcid.org/0000-0003-0538-2194>

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:

<http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Cualquier parte de esta obra se puede reproducir sin autorización pero con el reconocimiento y atribución de los autores.

No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): [riuma.uma.es](http://riuma.uma.es)

Method for Automatically Extracting of  
Tourism Destination Image from Chinese  
Social Networks. The case of Spain

Metodología para la Extracción Automática de  
la Imagen del Destino en Redes Sociales  
Chino-Parlantes. El caso España.

**Fang-wei Chen**

Tesis presentada como requisito parcial para optar al título de:

**Doctor en Dirección y Planificación del Turismo**

Directores:

Dra. Pilar Alarcón Urbistondo

Dr. Antonio Guevara Plaza

Universidad de Málaga

Facultad de Turismo,

Málaga, España

2015



UNIVERSIDAD DE MÁLAGA  
FACULTAD DE TURISMO

Dra. D. <sup>a</sup> PILAR ALARCÓN URBISTONDO, del Departamento de Economías y Administración de Empresa y Dr. D. ANTONIO GUEVARA PLAZA del Departamento de Lenguajes y Ciencias de Computación (Área de Lenguajes y Sistemas Informáticos) de la Universidad de de Málaga.

HACEN CONSTAR QUE:

D. <sup>a</sup> FANG-WEI CHEN, ha elaborado bajo nuestra dirección el trabajo de investigación correspondiente a su Tesis Doctoral titulada:

*Method for Automatically Extracting of Tourism Destination Image from Chinese Social Networks. The case of Spain.*

Revisado el presente trabajo estimamos que puede ser presentado al Tribunal que ha de juzgarlo.

Y para que conste a efectos de lo establecido en el artículo 11 del Real Decreto 56/2005, regulador de los estudios de Tercer Ciclo-Doctorado autorizamos la presentación de esta Tesis en la Universidad de Málaga.

Fdo.. Pilar Alarcón Urbistondo

Fdo.: Antonio Guevara Plaza

Málaga, a 15 de Noviembre de 2015

*A Iván y mis padres*

# Agradecimientos

Quiero agradecer a las personas que han participado directa o indirectamente en el desarrollo de esta Tesis. Sin vosotros, no hubiera sido posible de terminarla.

En primer lugar, quiero agradecer a mis directores, Pilar Alarcón Urbistondo y Antonio Guevara Plaza, por dedicar su tiempo y aportar sus conocimientos para poder llevar a cabo este gran trabajo. Quiero hacer mención especialmente, a Pilar Alarcón por su paciencia y dedicación absoluta. Ella nunca ha dudado en ayudarme, compartir sus experiencias personales tratándome como si fuera de la familia. Gracias a su simpatía y ánimos, me levanta siempre de nuevo en los momentos más difíciles durante el periodo de investigación. También a los compañeros Alfonso Cerezo, Plácido Sierra y José David Cisner por las asistencias administrativas y los intercambios de conocimientos.

En segundo lugar, quiero agradecer a la Dra. Wei-Lun Ku, a Wei-Fan Chen y todos investigadores y personal del laboratorio del Procesamiento de Lenguaje Chino en Sinica de Taiwán por ofrecerme la oportunidad de trabajar junto a este grupo humano con reconocimiento mundial en el procesamiento de lenguaje natural y aprender tanto de ellos. También a mi maestro de programación, Ting-hui Kao de Yahoo! en Taiwán, por enseñarme las técnicas del lenguaje Python en tan poco tiempo.

Por último, agradecer a mi familia el apoyo ofrecido durante todo el camino de la investigación. Sobre todo a Iván Montenegro por la ayuda, el cariño y compañía ofrecida durante este tiempo.

# Resumen

En los últimos años China nos ha mostrado su poder en el mercado mundial del turismo. Según informes publicados por la Organización Mundial del Turismo en 2014, apreciamos que China se ha convertido en el mercado con más tendencia y lidera el crecimiento del gasto en viajes al extranjero, alcanzando una cifra récord de 129,000 millones de dólares estadounidenses. Además, ocupa el primer lugar del ranking de los diez mercados emisores más importantes en el mundo (UNWTO, 2015).

No es de extrañar, por tanto que los destinos muestren un gran interés por este emisor. Este interés incluye desde conocer los hábitos de consumo del turista chino hasta la imagen y opinión que estos tienen sobre el destino. Este último caso es el que abordamos en este trabajo.

El estudio de la imagen de destino es un tema recurrente entre los investigadores turísticos que tradicionalmente emplean encuestas en las que se pregunta a los turistas por cada uno de los atributos que componen el destino. Sin embargo, la aparición de las redes sociales nos ha proporcionado una nueva fuente de información para el estudio de la imagen de destino.

Las opiniones que los turistas dan sobre los destinos nos permiten conocer qué imagen tienen de nosotros. Además esta información es espontánea, no dirigida por una encuesta. Así las opiniones encontradas ponen de manifiesto lo que realmente son aspectos que les preocupan a los turistas. Es decir, nos permite ver además de su valoración su peso. Por todo esto nos resulta interesante profundizar en el conocimiento de las opiniones de los turistas en las RRSS.

Sin embargo, revisando los estudios realizados en los últimos años, destacamos que aquellos relacionados con las RRSS y destino son escasos, principalmente por lo novedoso de esta línea de investigación. Las herramientas empleadas son herramientas pre-diseñadas para el análisis de contenido off-line. Es decir, los usuarios tienen que extraer en primer lugar los textos manualmente desde páginas web y después insertarlos en dichos programas. Esto supone un trabajo muy tedioso y el tiempo de obtención del resultado aumenta

considerablemente. Además, los usuarios tienen que crear una serie de palabras clave manualmente para que los sistemas analicen y busquen dentro de los textos insertados. Por estos motivos, los consideramos como herramientas semi-manuales. Estos métodos, además, son más proclives al sesgo ya que para trabajar necesitan que el investigador introduzca las palabras de búsqueda según su criterio.

Además, la mayoría de los trabajos recientes se limitan a estudiar textos en inglés. Tan solo hemos encontrado una investigación (Tse & Zhang, 2013) en la que se estudian textos en idioma chino y además se realiza el análisis de forma manual. La razón principal de esta escasez podría ser la dificultad de interpretar el idioma chino. Esta limitación proviene de la propia complejidad del lenguaje chino, pero también debido a que la mayoría de los programas actuales no están preparados para realizar un análisis semántico en texto chino. Es decir, los usuarios pese a poder insertar texto chino, no pueden obtener una lista con las palabras más frecuentes. La razón de esto es que dichos programas aíslan los términos a través de los espacios existentes entre unos y otros, y en un texto chino las palabras no se encuentran separadas mediante espacios.

Teniendo en cuenta a la declaración anterior, esta Tesis tiene el objetivo principal de crear y proponer un método novedoso basándose en un análisis cualitativo para estudiar la imagen en las RRSS empleadas por turistas chinos sobre el destino donde viajan, en este caso España.

Para la persecución de este objetivo marcado no nos valen las técnicas usadas hasta ahora. Debemos diseñar una metodología más compleja con cinco objetivos específicos como se muestra a continuación:

1. Determinación de las redes sociales relevantes como sujetos de estudio.
2. Extracción y organización de los comentarios depurados.
3. Técnica para analizar sintácticamente los comentarios en chino.
4. Diseño de una estructura de la imagen de destino para determinar los atributos influyentes en los comentarios.
5. Detección de las opiniones procedentes de los sujetos de estudio.

Para el desarrollo de este método y para su contratación empírica necesitamos concretar un destino donde aplicarlo. En este caso hemos seleccionado España. La selección

de España como destino de prueba no es casual. Como se comentará más adelante se trata de un destino líder y muy variado en su gama de productos.

La contrastación empírica nos va a permitir conocer la imagen de España entre los turistas chino-parlantes. Este será nuestro segundo objetivo.

Este objetivo general, a su vez, implica plantearse:

1. los atributos que más intervienen en la construcción de la imagen de España entre los chino-parlantes,
2. la identificación de los atributos mejor valorados ,
3. la identificación de los atributos con valoraciones negativas y,
4. a su vez, nos permite identificar si, como expone la teoría, la imagen se construye con un modelo en el que las dimensiones están interrelacionadas entre ellas, no pudiéndose interpretar cada dimensión de forma independiente.

Todos ellos suman los cuatro objetivos específicos del segundo objetivo general.

Con estas premisas diseñamos un método de trabajo haciendo el uso del lenguaje de programación Python, ya que resulta más sencillo y flexible de programar y permite aplicarse a textos chinos, descartando la información inútil como los anuncios. Esto consigue evitar el consumo de tiempo y problemas de importación y exportación de información a la hora de cambiar programas. Con Python, los módulos pueden ser diseñados con funciones específicas y múltiples pasos, procesándose gracias a esto de forma secuencial. Es decir, Python nos permite adaptar la salida de datos en entrada de datos para realizar un procesamiento continuo.

Este método no consiste simplemente en la propuesta de una técnica sino de una rutina de trabajo dividida en distintas etapas. Cada una de las fases está diseñada para abordar un objetivo concreto que supone un escalón para alcanzar el objetivo final. Debemos mencionar que el diseño de las actividades de cada una de las fases no ha sido fácil y se ha topado con hitos o problemas que tuvieron que abordarse para obtener finalmente una propuesta metodológica congruente y consistente.

Para evitar posibles sesgos que pudieran derivarse a la hora de traducir texto chino a otro idioma, conservamos el idioma original en todos los procesos. Sólo los resultados finales que no contienen elementos subjetivos se encuentran traducidos al español.

Tras completar todo el proceso, conseguimos el primero objetivo y también nos encontramos con un resultado de la imagen real de España de los turistas chino-parlantes a través de los comentarios registrados en los dos foros que se han analizado, que son Backpacker.com y Qyer.com.

Observamos que los turistas se encuentran interesados en atributos relacionados con, en primer lugar, las infraestructuras generales, seguidas de las infraestructuras turísticas, el entorno natural, la cultura, la historia y el arte, los factores políticos y económicos, los recursos naturales, el entorno social y por último, el ocio y el entretenimiento turístico.

Como se muestra a continuación, encontramos también la percepción de la imagen con respecto a los atributos ordenados por número de observaciones en cada dimensión:

Atributos que reciben más opiniones positivas:

- Infraestructura general: carreteras e infraestructuras comerciales
- Infraestructura turística: información turística, excursiones y alojamiento
- Entorno natural: ciudades y pueblos, y paisajes
- Cultura, historia y arte: Festival, arte y costumbres
- Factores políticos y económicos: precios
- Recursos naturales: clima, montañas y reservas naturales
- Entorno social: comunicaciones
- Ocio y entretenimiento turístico: el deporte

Atributos que reciben opiniones más negativas:

- Infraestructura general: carreteras e infraestructuras comerciales
- Infraestructura turística: alojamiento, información turística y excursiones
- Entorno natural: ciudades y pueblos y paisajes
- Cultura, historia y arte: Festival, arte y costumbres
- Factores políticos y económicos: precios
- Recursos naturales: clima, lagos y montañas
- Entorno social: comunicaciones
- Ocio y entretenimiento turístico: deporte

Carreteras y Ciudades y Pueblos son los dos atributos comentados que aparecen con más frecuencia en estos dos foros. Además, estos dos atributos reciben la mayoría de las

críticas tanto positivas como negativas. Basándonos en esto, podemos interpretar que el transporte por carretera es el medio de transporte principal para este grupo de turistas y las actividades esenciales para ellos son visitar las ciudades y pueblos.

Sin embargo, también percibimos diferentes percepciones entre los usuarios Backpackers y Qyer. Los usuarios de Backpackers tienden a preferir las actividades relativas a arte y festival, publicar más opiniones sobre información turística y alojamiento y dependen más de las telecomunicaciones y el servicio de transporte que los usuarios Qyer. Por el contrario, los usuarios de Qyer prestan más atención a temas como: paisaje, precios, deportes, restaurantes y excursiones. Los atributos que reciben opiniones más negativas son: carreteras, festivales, alojamiento e información turística en el caso de Backpackers, y ciudades y pueblos, carreteras, excursión y paisajes en el caso de Qyer.

No obstante, se detectaron algunos errores de interpretación sentimental y ruidos en el análisis. En cuanto a los ruidos de interpretación, las revisiones de algunos comentarios que contienen la pareja de palabras como montaña\_puerto(山\_港灣) del atributo montaña o puerto\_hotel (港灣\_飯店) de alojamiento nos muestran opiniones negativas debido a otros elementos que aparecen en el mismo contenido. Este tipo de ruidos es la validación de que la imagen se construye con un modelo en el que las dimensiones están interrelacionadas entre sí, no pudiéndose interpretar cada dimensión de forma independiente. Ejemplos de errores de análisis son las parejas de palabras sección\_Dì (段\_第) del atributo carretera y viaje\_Yì (旅行-義) de información turística.

Para evaluar el desempeño de este método de extracción automática para los análisis de imagen de destino turístico, un cálculo de valor-F1 ha sido aplicado para calcular la precisión (precision) de la interpretación sentimental y la exhaustividad (recall) de este sistema del análisis. Por último, obtenemos el valor de precisión, exhaustividad y valor-F1 del Backpackers: 0,833, 0,5 y 0,624, y de Qyer 0,8, 0,47 y 0,592 respectivamente.

Después de llevar a cabo el análisis del lenguaje natural, los contenidos no estructurados han sido analizados para obtener una opinión general y temas de interés en relación con el destino a través de los comentarios de usuarios en las redes sociales.

En conclusión, los turistas chinos tienen una buena impresión con respecto a España. A menudo utilizan instalaciones de la carretera y comerciales, y se preocupan bastante por el tema de los robos. Sin embargo, no publican comentarios sobre el tema de la aglomeración, la

contaminación ni la congestión de tráfico. En general visitan ciudades y pueblos, participan en festivales locales y admiran las obras de arte. Aún así, el tema de transporte constituye el problema principal para el que tengan acceso a estos lugares. Además, están interesados en información relacionada con los costes de transporte, el precio del alojamiento, utilizando esta información para comparar precios antes de tomar decisiones.

El método propuesto nos permite entender en qué se encuentran interesados los turistas y cómo se forman opiniones sobre destinos turísticos. Este método no es simplemente una técnica para la extracción de datos, sino también es un flujo de trabajo compuesto de diferentes etapas. Se aplica una metodología cualitativa basada en técnicas de procesamiento del lenguaje natural para analizar los contenidos en chino en los medios sociales.

Este método de trabajo nos ahorra gran cantidad de tiempo en la visualización de cada comentario y la clasificación de palabras. Si queremos conocer más detalles, tan solo necesitamos contemplar comentarios relacionados con estas parejas de palabras en lugar de visualizar todos los comentarios. Incluso, podemos estudiar la nota individual de cada comentario para conocer los puntos débiles y fuertes de nuestros atributos.

A pesar de estas ventajas, hay que reconocer que este método tiene algunas limitaciones. Los módulos programados están diseñados exclusivamente para su uso con texto en chino. Por otra parte, los resultados del método deben ser supervisados por un investigador chino-parlante, a pesar de que los datos se procesan automáticamente. Además, y al igual que con otros métodos (Költringer y Dickinger, 2015), parte del contenido se pierde durante el procesamiento. Debido a que el método es una forma de procesamiento del lenguaje natural automático, los errores ortográficos y gramaticales complican la determinación de los atributos. Este es un problema general con el que los investigadores se enfrentan en este campo.

Por otra parte, nombre de la ciudad como "Barcelona" en un contexto chino natural, se pueden escribir "巴塞隆納, 巴賽羅那, 巴賽羅納, 巴賽... etc." o "Barcelona" sin traducir el idioma original fonéticamente. Las traducciones incoherentes de palabras como estas también afectan al rendimiento de este método.

Estas revelaciones sobre los errores o dificultades son pasos necesarios para tener éxito y ofrecen pistas para mejorar el rendimiento de este método de trabajo. También nos permite desarrollar una futura línea de investigación. Siempre es necesario tener un paso

previo que consiste en el perfeccionamiento de las bases de datos utilizadas para el análisis. Un corpus chino de los destinos turísticos mejoraría considerablemente el rendimiento del método propuesto. Como base para un futuro trabajo, recomendamos firmemente a los investigadores construir un diccionario de sinónimos de términos turísticos en idioma chino orientado a la investigación en el análisis de las opiniones de los destinos turísticos. Tal diccionario de sinónimos debe contener detalles de los atributos y productos relacionados con el turismo, junto con sus sinónimos, así como la web ontología para una comprensión más profunda e integrada de las opiniones con respecto a los destinos turísticos.

**Palabras clave:** Imagen de destino turístico, redes sociales, turistas chino-parlantes, procesamiento del lenguaje natural, análisis de contenido

## Abstract

In recent years, China has shown its influence on the world tourism market. According to a report published by the United Nations World Tourism Organization in 2014, China has become an important market, with increasingly growth in the number of tourists. Moreover, China leads growth in expenses from travelling abroad, reaching a record \$129,000 million US dollars. In addition, Chinese tourists continue to demonstrate the potential for consumption in the tourist sector, occupying the top place among the ten most important groups for outbound travel markets in the world (UNWTO, 2015).

Not surprisingly, tourism destinations have taken a serious interest in this market. This includes, for example, interest in understanding Chinese tourist habits in terms of consumption, and the opinions they have regarding particular destinations. Understanding the latter is the primary target in this study.

The study of tourism destinations image (TDI) is an appellant topic among researchers. Traditionally, surveys were utilized to ask tourists about certain features that apply to particular destinations. Nevertheless, the ubiquity of the social networks has provided a new source of information for studying tourists' images of a destination.

The opinions that tourists share in social networks allows us to better understand how they feel about the destination. In addition, this method of collecting opinions is advantageous. For instance, the information is spontaneous; it is not directed by a survey. Thus, we can understand not only the values attached to each aspect of a tourist destination, but also the importance that each of these aspects has for the tourist.

However, after reviewing recent researches with regard to the above mentioned matter, we find that there are few studies on the connection between social media and tourist destinations, owing to the novelty of this field. The tools employed in the majority of these studies comprise pre-designed software for offline content analysis. That is, users must first manually extract text from webpages before importing them into such programs. This task is both monotonous and time-consuming. In addition, users must provide a series of keywords, such that the system can match them with the input text. Methods that require researchers to

insert search words are prone to bias, because researchers tend to supply subjective criteria when selecting the input terms.

Furthermore, almost all studies carried out recently have been restricted to text in English. One of the reasons for preferring English text is the inherent difficulty in interpreting a language as complex as Chinese. Yet another reason is that the semantic analysis of Chinese text is unsupported by most standard content-analysis software. Specifically, these programs isolate words with reference to the spaces that exist between them, whereas in Chinese, spaces are not used to demarcate words. In conclusion, existing tourism-destination investigations of Chinese texts remain weak, and no approach is available to realize Chinese context semantically and automatically.

In view of this, the present thesis consists in introducing and providing a novel method based on qualitative analysis for studying the online images of Chinese tourists regarding the destinations to which they travel. To achieve this goal, we aim to create an automatic and continuous method that can process Chinese text, analyze large amounts of information quickly, and give appraisals regarding image opinions. The specific objectives of **this objective** are the following:

1. Determinating the relevant Chinese social networks as study subjects.
2. Correctly retrieving and organizing on-line comments without ads.
3. Performing a semantic analysis of Chinese text from social media.
4. Creating a structure to determinate influential attributes and feature extraction.
5. Analysing opinions and reviews.

As for the development of this method and its empirical engagement we need a destination to prove its viability, so that Spain has been selected. The choice of Spain as a test target is not accidental. As it will be discussed later, Spain is a leading destination and provides the diversity of tourism products.

This step will allow us to understand the image of Spain among the Chinese -speaking tourists. **This will be our second objective.** This second objective, in turn, is involved and consider to:

1. discover attributes most involved in building the image of Spain,
2. identify attributes with positive opinions,
3. identify attributes with negative opinions and,
4. identify if, as stated theory, the image is constructed with a model in which the dimensions are interrelated and not being able to interpret independently.

All the above mentioned are met four specific objectives of the second general objective.

We propose this method designed with the Python programming language, owing to its simplicity and flexibility, and we applied it to Chinese text. The proposed method discards irrelevant information such as ads and banners on websites. With Python, modules can be designed with specific roles and multiple steps, each proceeding sequentially. That is, Python allows us to use the output data from one step as the input data for another. This avoids the time-consuming task of importing and exporting data when using different software.

Each stage is designed to reach a specific objective, and each step builds upon the previous step. The development of each aspect is subject to challenges and problems or milestones that must be managed in order to derive a coherent and consistent methodological proposal.

To avoid possible biases that might result from translating Chinese into another language, we retain the original language during all processes. Only the final results that do not contain subjective elements are translated into Spanish.

In selecting the test object step, both traditional and simplified Chinese forums were chosen as the object of study. We chose Alexa, a global web-traffic and -ranking tool, to find out the most suitable samples at this stage. The candidate webpages should be from high-traffic websites for public non-commercial use, and we strove for neutrality during the collection process. After defining previous criteria, Backpackers.com.tw representing traditional Chinese was ranked highest. In a similar manner, we determined a study sample for simplified Chinese: Qyer.com.

After determining study samples, Spain was selected as a travel destination to evaluate the proposed method since it is one of the three most visited countries, accompanied with France and the United States (UNWTO, 2014). Moreover, Spain proved to be a destination with an enormous variety of opinions.

For this reason, in Information extraction step, we selected content related to Spain from Backpackers and Qyer for data extraction. The URLs for these two sources are as follows: (<http://www.backpackers.com.tw/forum/forumdisplay.php?f=117>) and (<http://bbs.qyer.com/forum-18-1.html>) for Backpackers and Qyer, respectively. Owing to the fact that social-media content is continuously changing, we downloaded all content from these URLs in order to focus the working range, limiting it to a particular time—a necessary preliminary. Two different Python modules are created in order to retrieve information from each of these websites, inasmuch as these webpages show the utilization of different Chinese characters (traditional and simplified characters) and the distinct structure and design.

Finally, in the Backpackers forum, we collected a total of 3,655 threads containing 16,371 comments related to Spain. From Qyer, we extracted 10,088 threads with 31,339 comments related to Spain. In both cases, the comments were collected from the date that the first comment was posted (April 13th, 2006 from the Backpackers forum, and March 7th, 2004 from the Qyer forum) until August 1st, 2014. The data extracted from each of these two forums had been stored in BSON format and arranged in the following fields:

- For data from Backpackers: Username, source, gender, title, content, thanks, date, \_id
- For data from Qyer: Username, content, title, experience, source, date, \_id

The field source is the place where this review is restored. Regarding the field thanks, it consists of two numbers, the number that other users think this comment is useful and the number of posts that the same author publishes. These two numbers are separated by a forward slash. As for the field date, it means the post date and is ordered by year, month, date, hour, and minute in brackets according to the local time setting of each forum. The field \_id is an automatic system-generated ID to store comments. With regard to experience, this field shows us the frequency of usage of users. The different classifications of two forums are due to the information provided on each webpage. In this process, we aim to retrieve all information demonstrating on target webs in case that we need more than on-line comments.

Before analysing the text from the forums, the content must be parsed. Since, Chinese has neither a phonetic alphabet nor spacing between words. Undoubtedly, this makes the detection of TDI attributes difficult. Moreover, punctuation is often missing, and poor grammar is common in these forums. Therefore, in Syntactic analysis step, we compared two existing specific parsing tools for the grammatical analysis of Chinese text and chose one of

them, which can reach our goal. Finally, we opted to use the Chinese Knowledge and Information Processing system (CKIP), developed by the Academia Sinica in Taiwan.

The results of this step segment sentences into words, adding part-of-speech tagging (POS tagging) in parentheses after each term that provide us a method of understanding the exact meanings of contexts and the possibility of detecting keywords for image analysis.

After carrying out the previous steps, in Determining attributes and feature extraction step, we face a difficult issue that is to identify the tourism-related elements mentioned by users on social media. An important prerequisite for doing so is a thesaurus of tourism-related words that can be used to configure a search for the relevant attributes. Such terms include ‘transport,’ ‘accommodation,’ ‘restaurants,’ etc. However, there is no existing TDI thesaurus. To resolve this problem, we combine theory and practice. By merging Beerli & Martín’s TDI dimensions (2004) with a tourism thesaurus, built by the United Nations World Tourism Organization (2001), we obtain a basic structure with 8 dimensions as class and 51 attributes as sub-class for TDI study use. The following are the detail of these elements.

- Natural resources: nature reserves, flora, countryside, deserts, mountains, lakes, beaches, weather and countryside.
- General infrastructure: health services, commercial infrastructures, airports, bridges, transportation, telecommunication, ports and roads.
- Tourist infrastructure: excursions, accommodation, tourist information, bar and discotheques, restaurant and tourist centres.
- Tourist leisure and recreation: shopping, nightlife, casinos, zoos, sports and theme parks.
- Cultural, history and art: handicraft, gastronomy, festival, art, customs, concerts, monuments, museums and religion.
- Political and economic factors: politics, prices, economy and safety.
- Natural environment: congestion, overcrowding, pollution, city and towns, cleanliness and scenery.
- Social environment: communication, local people and quality of life.

Still, this structure in English is unable to be performed on Chinese texts. To resolve this issue, we adopted Extended-HowNet system (E-HowNet) to define the word meanings. Providing English definition, E-HowNet is a lexical database for Chinese language

knowledge, developed by the Academia Sinica in Taiwan. Via the database of E-HowNet, we managed to obtain a list of synonyms for each attribute. All these synonyms in the lists were turned into keywords to make a set of search requests. In total there are 384 keywords in Chinese.

After establishing keywords for search requests, we must search for the features of these terms in order to filter the tokens that are associated with these terms. That is, for example, tourists consistently discussed the price, location, and categories when they commented on hotels. In this example, price, location and categories are hotels' features. In order to enhance the associated terms of keywords, we applied an algorithm — Point-wise Mutual Information (PMI) — to search for the features that pertain to the principal elements.

The PMI is a useful association measure that is employed to determine the relationship between two variables, and it is frequently applied in the field of linguistics for the purpose of extracting high-quality features (Popescu & Etzioni, 2007; Mata, 2011). When a PMI value for the pair of term  $x$  and term  $y$  is higher than the others, this means that the term  $x$  is strongly associated with the term  $y$ . This identifies features of the pre-selected attributes. That is, it determines the words that are closely related to a particular tourism attribute in the forums. This step involves finding associated words of each selected keyword and arranging these associated terms into pairs. The set of word-pairs obtained from this process are significantly large. Moreover, the existing criteria do not establish the exact threshold for inclusion given a certain PMI value. Therefore, we developed a Delphi study with experts from the field of tourism and tourist behaviour in social networks to decide the number of word-pairs that should be retained. As a consequence, we preserved the top five highest PMI value couples for each pre-selected keywords that were commonly agreed by experts, in order to create the list of word-pairs.

Finally, we inputted word-pair lists and extracted all the published content that contains inserted pairs. The retrieved contents were stored with the name of word-pair with the order of extracting in text files. Each file was stored in a folder named by attribute name to which the word-pair corresponds. The results from this processing step show us not only features belonging to each attribute, but also the importance of each attribute according to the amount of extracted comments.

The last step is Opinion analysis. Before analysing tourist opinions, we must identify those particular words that express an opinion. We could simply employ an existing Chinese

opinion-word list. However, such words do not exhaust the range of factors that determine the opinions in posted content, and they do not fully consider the range of constructing sentences. That is, different sentence structures and collocated vocabularies may result in different meanings. Ku & Chen (2010) demonstrated that an analysis of an opinion within a sentence is more ambiguous than an analysis of an opinion within a context, because, in a context, opinion words point to the interests of the author. Therefore, analysing opinions are especially complicated and requires considerable resources and time.

To achieve the aim pursued, we adopted the system of Chinese opinion-extraction system for opinionated information, known as CopeOpi. Developed from news sources, CopeOpi is a web-based system capable of extracting from documents words and phrases that express sentiments. Furthermore, CopeOpi determines the polarity and the strength of the extracted opinions. The system evaluates the relations between items by tracking words that express opinions and collocating the terms within the analysed text. There are three parameters that affect the output result: sentimental words, opinion operators, and negation operators. The best performance of this system achieves the f-measure 0.68 and 0.91 precision in the traditional Chinese opinion analysis tests. As to the best performance of the simplified Chinese opinion analysis, this system success to reach the f-measure 0.85 and 0.96. That is to say, this system correctly distinguishes between the positive, neutral and negative sentiments expressed by writers in more than 70% in traditional Chinese cases and 97% in simplified cases.

Once all the contents associated with the word-pairs have been assessed, we can figure the global opinion of word-pairs. This global opinion is demonstrated by number of positive, neutral, and negative comments that each word-pair receives and is stored in a CSV file (a comma-separated values file), named with the attribute name to which word-pairs correspond. Through this file, we can detect with which feature a particular attribute has a worse rating. In addition to a global view of attribute and its features, independent results of comments and score detail of each review are also available. This information provides us Chinese tourists' real thinking about a particular attribute of destination images.

After completing the entire processing, we met a result with a real tourism destination image from Chinese tourist via on-line reviews. For example, users from these two forums are interested in attributes relative to 1. General infrastructure, 2. Tourist infrastructure, 3. Natural environment, 4. Cultural, history and art, 5. Political and economic factors, 6. Natural

resources, 7. Social environment and 8. Tourist leisure and recreation in order of preference. As shown below, we found also the perception of image with regard to attributes ordered by number of comments in each dimension:

Attribute receives more positive opinions

- General infrastructure: roads and commercial infrastructures
- Tourist infrastructure: tourist information, excursions and accommodation
- Natural environment: cities, towns and scenery
- Cultural, history and art: festival, art and customs
- Political and economic factors: prices
- Natural resources: weather, mountains and nature reserves
- Social environment: communication
- Tourist leisure and recreation: sport

Attribute receives more negative opinions

- General infrastructure: roads and commercial infrastructures
- Tourist infrastructure: accommodation, tourist information and excursions
- Natural environment: cities, towns and scenery
- Cultural, history and art: festival, art and customs
- Political and economic factors: prices
- Natural resources: weather, lakes and mountains
- Social environment: communication
- Tourist leisure and recreation: sport

Roads and cities towns are two attributes mentioned frequently in these two forums. Also, these two attributes receive the most criticism both positive and negative. Based on this, we can interpret that road transportation is the main means of transportation for Chinese speaking tourists and the essential activities for them are visiting cities and towns.

However, we also noticed different perceptions between Backpackers users and Qyer users. Backpackers users tend to prefer activities regarding art and festival, post more reviews about tourist information and accommodation and depend more on telecommunication and transportation service than Qyer users. On the contrary, Qyer users pay more attention on topics such as: scenery, prices, sports, restaurants and excursions. Attributes receive more

negative opinions are: road, festivals, accommodation and tourist information in the case of Backpackers, cities, towns, roads, excursions and scenery in the case of Qyer.

In order to know the main reason these attributes receive more negative reviews, first, we selected all attributes with a high percentage of negative comments, and then we studied those comments which receive a higher negative score. After reviewing those negative comments, we found the original problems which lead to bad image. Here we provide some examples. The word-pair `alleyway_guesthouse` (巷口\_巷子) from road attribute receives 75% review with negative opinions because of the frequent robberies occurred in the alleyway. As to the 67% negative comments receiving by the word-pair `information_tradition` (資料\_風俗), these negative opinions come from reviews that recommend understanding local custom and not only looking for basic tourism information.

However, we also detected some errors of sentimental interpretation and errors of parsing. As regards errors of sentimental interpretation, the revisions of certain contents containing word-pairs, such as `mountain_harbour` (山\_港灣) from mountain, `harbour_hotel` (港灣\_飯店) from accommodation, show us these word-pairs receive negative opinions due to other attributes in the same content. This is validation that image is constructed with a model in which the dimensions are interrelated and not being able to interpret independently. Examples from parsing errors are, for example, the word-pairs `setion_Dì` (段\_第) from road and `travel_Yì` (旅行-義) from tourist information.

To evaluate the performance of this automatically extracting method for tourism destination image analyses, an F-measure was applied to calculate the precision and recall of the processing regarding the step 4 and 5 (Determining attributes and feature extraction and Opinion analysis). In this examination, recall and precision are weighted equally. Finally, we got the value of precision, recall and F1 from Backpackers: 0.833, 0.5 and 0.624 respectively. Values from Qyer are 0.8, 0.47 and 0.592.

After processing the natural language analysis, unstructured Chinese content is analyzed to derive the general sentiment and relevant topics regarding the destination through user comments on social media. The proposed method allows us to understand what tourists are interested in and how they form opinions about tourist destinations. This method is not simply a technique for extracting data; it is also a workflow with different stages. A

qualitative methodology based on natural language processing techniques is applied to analyze Chinese content on social media.

With the proposed method, we were able to work with massive Chinese data to correctly identify tourism attributes according to parameters that were determined a priori. To study a particular attribute, we merely need to enter the word as a parameter. The objective destination can be changed as required.

In conclusion, Chinese tourists have a good impression regarding Spain. They often use road and commercial facilities, but worry about the issue of robbery. However, they do not have comments about the issue of overcrowding, pollution either traffic congestion. They visit cities and small towns, take part in local festivals and admire artwork. Still, transportation issue is the main problem for them to access these places. Also, they are interested in information regarding transportation costs and accommodation price, and are used to compare these prices before making decision.

Apart from the acquisition of a global image, this dissertation also puts theory of the destination image to the test. The results show that the tokens of word-pairs frequently match words of attributes for example, the word-pair `hotel_cleanliness`. The word `hotel` is one of the attributes from Tourist infrastructure dimension and `cleanliness` is an attribute of the dimension Natural environment. Also, there are occasionally interpretations of opinions from some word-pairs in involving other attributes of different dimensions. Following above explorations we can verify the theory of Tasci, Gartner, & Cavusgil (2007), that is “a destination image is an interactive system of thoughts, opinions, feeling, visualizations, and intentions toward a destination (p. 200)” and conclude that it requires a comprehensive study of all image components to define and understand a global tourism destination image.

Moreover, because it is automatic and sequential, our proposed technique is much faster than manual or semi-automatic techniques, which require inputting data to different programs at each step, and it furthermore decreases the propensity for subjective bias. In addition, the proposed method eliminates the need to read posted content and classify various words. For a detailed analysis, we can simply track down reviews with negative scores rather than reading all of the comments related to the word-pairs. Furthermore, in assessing the opinions of tourists, the strong and weak points related to each aspect of tourism can be determined.

Despite these advantages, we should recognize that the method has some limitations. The programmed modules are designed exclusively for use with Chinese text. Moreover, the results of the method must be inspected by a Chinese-speaking researcher, even though the data is processed automatically. In addition, and likewise with other methods (Költringer & Dickinger, 2015), some content is missed during processing. Because the method is a form of automatic natural-language processing, spelling and grammatical errors complicate the determination of attributes. This is an entrenched problem that researchers face in this field. There are many synonyms and alternative meanings to the tokens in an experimental tourism thesaurus, which we translated directly into standard Chinese from the original Spanish. Moreover, there are countless ways of misspelling them. A city name like ‘Barcelona’ is a clear example of this. In a natural Chinese context, when authors refer to this foreign city, they may type “巴塞隆納, 巴賽羅那, 巴賽羅納, 巴賽...etc.” or “Barcelona” without translating the original language phonetically. Such factors can result in the loss of data, and they can decrease the accuracy of the results. This an intractable problem related to the field in general, rather than our method in particular.

Another point we should address is the matter of synonyms and calculation of PMI value. In the step of Determining attributes and feature extraction, we created the first destination image structure in Chinese for extraction destination features and opinion values. Notwithstanding, in the process of attribute assignment, we encountered obstacles related to synonyms of attributes, which contain only one letter or character. In many cases, a single Chinese character, such as “莊” changes its meaning when attached to another character. For example, when we add the character "布" (means cloth) before "莊", the set of these two forms a word which means fabric (布莊); if we add the character "農" (means agriculture) in front of "莊" then takes the meaning farm (農莊); and if we "村" before "莊" the two together mean town (村莊). Because of this, we strongly recommend caution in assigning the words of one character as attributes.

Apart from this problem, we detected some misinterpretation of opinion especially in long comments that contain multiple attributes. It is assumed that the comments of these two forums are classified by subject, however, in some cases contents could be involved in many issues, such as notes of the traveller. This type of comment is often very long and be associated with many attributes. We calculate the value of PMI based on posts rather than sentences, and therefore the features they get a high value of PMI is not always in the same

sentence with his attribute. Occasionally, we found attribute and its characteristic in separating paragraphs. These give word-pairs the possibility of being affected by other attributes or characteristics. That is why, in some cases we see that CopeOpi detects a comment that contains a word-pair negative, however, when we pass this comment to Chinese speakers, no one considers it a negative opinion. For future study, we recommend that investigator calculate the value of PMI based on phrases or paragraphs to examine whether through this change can improve the accuracy of the methodology of this thesis.

Moreover, sometimes the high value of PMI occurred when there is a no frequent word. Words that appear in few comments but always co-occur with attributes may have a high PMI value. To do this, we recommend consulting word frequency before creating word-pairs with attributes in future study.

These revelations about mistakes are necessary steps to succeed and offer clues to improving the performance of this method of work. It also allows us to develop a future line of research. There is always a preliminary step that involves refining the databases used for parsing. A sufficiently adequate Chinese corpus of tourism destinations and terms would considerably improve the performance of the proposed method. As a basis for future work, we strongly urge scholars to build a Chinese tourism thesaurus oriented toward research in analysing opinions of tourism destinations. Such a thesaurus should contain details of tourism-related attributes and products, along with their synonyms, as well as web ontology for a deep and integrated understanding of opinions with regard to tourism destinations.

**Keywords:** Tourism destination image, social networks, Chinese speaking tourists, natural language processing, content analysis

# Índices

# Índice de Contenido

AGRADECIMIENTOS.....	II
RESUMEN .....	III
ABSTRACT.....	X
ÍNDICE DE FIGURAS .....	XXVII
ÍNDICE DE TABLAS .....	XXIX
LISTA DE ABREVIATURAS.....	XXX
INTRODUCCIÓN .....	1
INTRODUCTION .....	7
<b>1. MARCO TEÓRICO.....</b>	<b>12</b>
<b>1.1. IMAGEN DE DESTINO TURÍSTICO .....</b>	<b>12</b>
<b>1.1.1. CONCEPTO DE LA IMAGEN DE DESTINO TURÍSTICO .....</b>	<b>14</b>
<i>1.1.1.1. LAS DIMENSIONES DE LA IMAGEN DE DESTINO.....</i>	<i>16</i>
<i>1.1.1.2. FACTORES QUE INFLUYEN EN LA FORMACIÓN DE LA IMAGEN DE DESTINO EN LA MENTE DEL CONSUMIDOR.....</i>	<i>21</i>
<b>1.1.2. ESTADO DE LA CUESTIÓN EN LA INVESTIGACIÓN SOBRE LA IMAGEN DE DESTINO.....</b>	<b>24</b>
<i>1.1.2.1. TEMAS MÁS INVESTIGADOS.....</i>	<i>25</i>
<i>1.1.2.2. METODOLOGÍAS, VARIABLES Y ESCALAS MÁS USADAS.....</i>	<i>29</i>
<i>1.1.2.3. CONCLUSIÓN DE LAS INVESTIGACIONES EN TORNO A LA IMAGEN DE DESTINO EN LA ÚLTIMA DÉCADA.....</i>	<i>33</i>
<b>1.2. REDES SOCIALES Y EL TURISMO .....</b>	<b>34</b>
<b>1.2.1. CONCEPTO DE LA RED SOCIAL.....</b>	<b>35</b>
<i>1.2.1.1. LA CLASIFICACIÓN DE LAS REDES SOCIALES VIRTUAL.....</i>	<i>37</i>
<i>1.2.1.2. IMPACTO DE LAS REDES SOCIALES VIRTUALES SOBRE LA INDUSTRIA TURÍSTICA .....</i>	<i>39</i>
<b>1.2.2. REVISIÓN CONCEPTUAL DE LA RED SOCIAL VIRTUAL CON EL TURISMO .....</b>	<b>41</b>
<b>1.3. ESTADO DE LA CUESTIÓN EN LA INVESTIGACIÓN DE REDES SOCIALES E IMAGEN DE DESTINO .....</b>	<b>43</b>

<b>1.4.</b>	<b>EL MERCADO CHINO-PARLANTE Y EL USO DE LAS RRSS</b> .....	<b>45</b>
<b>2.</b>	<b>OBJETIVOS</b> .....	<b>49</b>
<b>2.1.</b>	<b>OBJETIVOS GENERALES Y ESPECÍFICOS</b> .....	<b>50</b>
<b>3.</b>	<b>MODELO Y PROPUESTA DE LA EXTRACCIÓN AUTOMÁTICA PARA EL ANÁLISIS DE LA IMAGEN DE DESTINO</b> .....	<b>54</b>
<b>3.1.</b>	<b>PLANEAMIENTO DEL TRABAJO</b> .....	<b>55</b>
<b>3.2.</b>	<b>SELECCIÓN DE OBJETO DE ENSAYO</b> .....	<b>57</b>
<b>3.2.1.</b>	<b>OBJETIVO DE LA ETAPA</b> .....	<b>57</b>
<b>3.2.2.</b>	<b>HERRAMIENTA APLICADA</b> .....	<b>57</b>
<b>3.2.3.</b>	<b>RESULTADO DE ESTA ETAPA</b> .....	<b>58</b>
<b>3.3.</b>	<b>EXTRACCIÓN DE LA INFORMACIÓN</b> .....	<b>59</b>
<b>3.3.1.</b>	<b>OBJETIVO DE LA ETAPA</b> .....	<b>59</b>
<b>3.3.2.</b>	<b>HERRAMIENTA APLICADA Y PROCESOS</b> .....	<b>60</b>
<b>3.3.2.1.</b>	<i>EL CASO DE BACKPACKERS</i> .....	<i>60</i>
<b>3.3.2.2.</b>	<i>EL CASO DE QYER</i> .....	<i>62</i>
<b>3.3.3.</b>	<b>RESULTADO DE ESTA ETAPA</b> .....	<b>64</b>
<b>3.4.</b>	<b>REALIZACIÓN DEL ANÁLISIS SINTÁCTICO</b> .....	<b>66</b>
<b>3.4.1.</b>	<b>OBJETIVO DE LA ETAPA</b> .....	<b>66</b>
<b>3.4.2.</b>	<b>SELECCIÓN DE LA HERRAMIENTA</b> .....	<b>67</b>
<b>3.4.2.1.</b>	<i>CKIP (CHINESE KNOWLEDGE AND INFORMATION PROCESSING)</i> .....	<i>67</i>
<b>3.4.2.2.</b>	<i>STANFORD PARSER</i> .....	<i>67</i>
<b>3.4.2.3.</b>	<i>COMPARACIÓN DE LAS DOS HERRAMIENTAS</i> .....	<i>68</i>
<b>3.4.3.</b>	<b>RESULTADO DE ESTA ETAPA</b> .....	<b>70</b>
<b>3.5.</b>	<b>EXTRACCIÓN DE CARACTERÍSTICAS</b> .....	<b>71</b>
<b>3.5.1.</b>	<b>OBJETIVO Y TAREAS DE LA ETAPA</b> .....	<b>71</b>
<b>3.5.2.</b>	<b>HERRAMIENTA APLICADA</b> .....	<b>71</b>
<b>3.5.2.1.</b>	<i>LAS DIMENSIONES DE LA IMAGEN</i> .....	<i>72</i>

3.5.2.2.	<i>EL TESAURO TURÍSTICO DE OMT</i> .....	73
3.5.2.3.	<i>E-HOWNET (EXTENDED-HOWNET)</i> .....	75
3.5.2.4.	<i>PMI</i> .....	78
<b>3.5.3.</b>	<b>RESULTADO DE ESTA ETAPA</b> .....	<b>80</b>
<b>3.6.</b>	<b>ANÁLISIS DEL SENTIMIENTO</b> .....	<b>82</b>
<b>3.6.1.</b>	<b>OBJETIVO DE LA ETAPA</b> .....	<b>82</b>
<b>3.6.2.</b>	<b>HERRAMIENTA APLICADA</b> .....	<b>83</b>
<b>3.6.3.</b>	<b>RESULTADO DE ESTA ETAPA</b> .....	<b>85</b>
<b>3.7.</b>	<b>RESULTADO OBTENIDO CON EL MODELO DE EXTRACCIÓN PROPUESTO</b> ...	<b>89</b>
<b>4.</b>	<b>CONTRASTACIÓN EMPÍRICA DEL MODELO PROPUESTO</b> .....	<b>92</b>
<b>4.1.</b>	<b>PLANTEAMIENTO DEL TRABAJO</b> .....	<b>93</b>
<b>4.2.</b>	<b>RESULTADO DEL TRABAJO EMPÍRICO</b> .....	<b>95</b>
<b>4.2.1.</b>	<b>VISUALIZACIÓN DE LOS COMENTARIOS EN BACKPACKERS</b> .....	<b>102</b>
4.2.1.1.	<i>LA DIMENSIÓN DE INFRAESTRUCTURA GENERAL</i> .....	104
4.2.1.2.	<i>LA DIMENSIÓN RECURSOS NATURALES</i> .....	110
4.2.1.3.	<i>LA DIMENSIÓN INFRAESTRUCTURA TURÍSTICA</i> .....	115
4.2.1.4.	<i>LA DIMENSIÓN DE OCIO Y ENTRETENIMIENTO TURÍSTICO</i> .....	119
4.2.1.5.	<i>LA DIMENSIÓN DE CULTURA, HISTORIA Y ARTE</i> .....	122
4.2.1.6.	<i>LA DIMENSIÓN DE FACTORES POLÍTICOS Y ECONÓMICOS</i> .....	126
4.2.1.7.	<i>LA DIMENSIÓN DE ENTORNO NATURAL</i> .....	128
4.2.1.8.	<i>LA DIMENSIÓN DE ENTORNO SOCIAL</i> .....	130
<b>4.2.2.</b>	<b>VISUALIZACIÓN DE LOS COMENTARIOS EN QYER</b> .....	<b>132</b>
4.2.2.1.	<i>LA DIMENSIÓN DE INFRAESTRUCTURA GENERAL</i> .....	134
4.2.2.2.	<i>LA DIMENSIÓN DE RECURSOS NATURALES</i> .....	138
4.2.2.3.	<i>LA DIMENSIÓN DE INFRAESTRUCTURA TURÍSTICA</i> .....	142
4.2.2.4.	<i>LA DIMENSIÓN DE OCIO Y ENTRETENIMIENTO TURÍSTICO</i> .....	146

4.2.2.5.	<i>LA DIMENSIÓN DE CULTURA, HISTORIA Y ARTE</i> .....	149
4.2.2.6.	<i>LA DIMENSIÓN DE FACTORES POLÍTICOS Y ECONÓMICOS</i> .....	153
4.2.2.7.	<i>LA DIMENSIÓN DE ENTORNO NATURAL</i> .....	155
4.2.2.8.	<i>LA DIMENSIÓN DE ENTORNO SOCIAL</i> .....	157
<b>4.3.</b>	<b>EVALUACIÓN DE LAS TÉCNICAS APLICADAS</b> .....	<b>158</b>
<b>4.4.</b>	<b>IMAGEN DE ESPAÑA SEGÚN LOS COMENTARIOS EN LAS RRSS CHINO PARLANTES</b> .....	<b>161</b>
<b>5.</b>	<b>CONCLUSIONES, LIMITACIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN</b>	<b>165</b>
<b>5.</b>	<b>CONCLUSIONS, LIMITATIONS AND FUTURE INVESTIGATIONS</b> .....	<b>171</b>
	<b>BIBLIOGRAFÍA</b> .....	<b>176</b>
	<b>ANEXO</b> .....	<b>193</b>
<b>A. 1.</b>	<b>LAS DIMENSIONES DE IMAGEN DE DESTINO</b> .....	<b>193</b>
<b>A. 2.</b>	<b>TABLA DE LA REVISIÓN DE LAS VARIABLES EMPLEADAS</b> .....	<b>194</b>
<b>A. 3.</b>	<b>ETIQUETA USADO POR CKIP</b> .....	<b>196</b>
<b>A. 4.</b>	<b>STANFORD PARSER POST TAGS</b> .....	<b>198</b>
<b>A. 5.</b>	<b>GLOSARIO DE TÉRMINOS PARA LA MEDICIÓN DE LA IDT</b> .....	<b>199</b>
<b>A. 6.</b>	<b>PAREJAS DE PALABRAS OBTENIDAS CON SUS RESPECTIVOS VALORES DE PMI</b> .....	<b>200</b>
	<b>A. 6. 1. BACKPACKERS</b> .....	<b>200</b>
	<b>A. 6. 2. QYER</b> .....	<b>208</b>
<b>A. 7.</b>	<b>LAGO_RUTA (湖泊_路線 2.TXT)</b> .....	<b>216</b>

# Índice de Figuras

FIGURA 1. TRES DIMENSIONES DE LA IDT DE ECHTNER Y RITCHIE .....	16
FIGURA 2. LA INTERRELACIÓN DE LOS COMPONENTES DE LA IMAGEN DE DESTINO .....	18
FIGURA 3. FACTORES QUE INFLUYEN LA EXPERIENCIA DE TURISTAS .....	20
FIGURA 4. LA FORMACIÓN DE LA IMAGEN .....	22
FIGURA 5. MODELO DE IMAGEN DE DESTINO Y SUS RELACIONES .....	24
FIGURA 6. CRECIMIENTO DEL NÚMERO DE USUARIOS DE INTERNET .....	34
FIGURA 7. PANORAMA DE LOS MEDIOS SOCIALES EN CHINA.....	46
FIGURA 8. ESQUEMA DE LAS RAMAS DE ESTUDIO .....	55
FIGURA 9. LA PROPUESTA DE LA METODOLOGÍA PARA EL ANÁLISIS DE LA IMAGEN DE DESTINO A TRAVÉS DE LA EXTRACCIÓN AUTOMÁTICA EN REDES SOCIALES .....	57
FIGURA 10. CAPTURAS DE PANTALLA DE LA WEB DE BACKPACKERS Y QYER.....	59
FIGURA 11. EL CÓDIGO HTML DE LA WEB BACKPACKERS.....	61
FIGURA 12. EL CÓDIGO HTML DE LA WEB QYER .....	63
FIGURA 13. EJEMPLO DE E-HOEÑET.....	75
FIGURA 14. LA ESTRUCTURA PARA EL PROCESAMIENTO DEL LENGUAJE NATURAL.....	81
FIGURA 15. FLUJO DEL TRABAJO.....	89
FIGURA 16. PESO DE CADA DIMENSIÓN .....	95
FIGURA 17. CANTIDAD DE COMENTARIOS RELACIONADOS CON LA DIMENSIÓN INFRAESTRUCTURA GENERAL.....	96
FIGURA 18. CANTIDAD DE COMENTARIOS RELACIONADOS CON LA DIMENSIÓN INFRAESTRUCTURA TURÍSTICA ....	96
FIGURA 19. CANTIDAD DE COMENTARIOS RELACIONADOS CON LA DIMENSIÓN ENTORNO NATURAL.....	97
FIGURA 20. CANTIDAD DE COMENTARIOS RELACIONADOS CON LA DIMENSIÓN CULTURA, HISTORIA Y ARTE .....	97
FIGURA 21. CANTIDAD DE COMENTARIOS RELACIONADOS CON LA DIMENSIÓN FACTOR POLÍTICO Y ECONÓMICO.....	98
FIGURA 22. CANTIDAD DE COMENTARIOS RELACIONADOS CON LA DIMENSIÓN RECURSO NATURAL .....	98
FIGURA 23. CANTIDAD DE COMENTARIOS RELACIONADOS CON LA DIMENSIÓN ENTORNO SOCIAL .....	99

FIGURA 24. CANTIDAD DE COMENTARIOS RELACIONADOS CON LA DIMENSIÓN OCIO Y ENTRETENIMIENTO TURÍSTICO .....	99
FIGURA 25. COMPARATIVA DE LA IMAGEN DE DESTINO A TRAVÉS DE LOS DOS FOROS .....	101
FIGURA 26. LAS INTERACCIONES ENTRE LOS ATRIBUTOS.....	162

# Índice de Tablas

TABLA 1. EJEMPLO DE CKIP .....	70
TABLA 2. EJEMPLO DE LA ADAPTACIÓN A LA TEORÍA DE MARTÍN Y BEERLI .....	73
TABLA 3. EJEMPLO DE LA ADAPTACIÓN DEL TESAURO DE OMT .....	74
TABLA 4. EJEMPLO DE LA ADAPTACIÓN DE E-HOWNET .....	76
TABLA 5. EJEMPLO DEL RESULTADO DE COPEOPI POR EL ATRIBUTO.....	85
TABLA 6. EJEMPLO DEL RESULTADO DE COPEOPI POR LAS PAREJAS DE PALABRAS.....	86
TABLA 7. RESULTADO DE BACKPACKERS.....	103
TABLA 8. RESULTADO DE QYER .....	133
TABLA 9. LOS COMPONENTES PARA EL CÁLCULO DEL VALOR- F.....	159

# Lista de Abreviaturas

<u>Abreviatura</u>	<u>Término</u>
<i>RRSS</i>	Redes sociales
<i>OGD</i>	Organización de gestión de destino
<i>IDT</i>	Imagen de destino turístico
<i>OMT</i>	Organización Mundial de Turismo
<i>TIC</i>	Tecnologías de la Información y la Comunicación
<i>DAFO</i>	Debilidad, amenazas, fortalezas y oportunidades

# INTRODUCCIÓN

# INTRODUCCIÓN

La potencia de los mercados emisores asiáticos, sobre todo China, llama bastante la atención y capta el interés de investigadores y de la industria turística. Según informes publicados de la Organización Mundial del Turismo (OMT, 2014) y de la ETC (European Travel Commission, 2014), China se ha convertido en el mercado con más tendencia en los últimos años y lidera el crecimiento del gasto en viajes al extranjero, alcanzando un récord de 129,000 millones de dólares estadounidenses y ocupando el primer lugar en el ranking de los diez mercados emisores con más de 98 millones de salidas en 2013, China ha seguido demostrando su potencial de consumir en sector turístico (OMT, 2015).

No es de extrañar, por tanto que los destinos, la industria y los investigadores muestren un gran interés por este emisor. Este interés, además, viene incrementado porque se trata de un turista con costumbres y valores muy diferentes a los turistas procedentes de mercados tradicionales, como pueden ser Alemania y Gran Bretaña. Así, para poder dar respuestas apropiadas a este emisor es necesario conocer desde los hábitos de consumo del turista chino hasta analizar sus posibles riesgos de crecimiento pasando por indentificar los segmentos más rentables o la imagen y opinión que estos tienen sobre el destino.

Este último caso es el que nos llama la atención como investigadores y, por ello, lo abordamos en este trabajo.

El estudio de la imagen de destino es un tema recurrente entre los investigadores turísticos. Ya en los años 70 aparecen con notable presencia de estudios referentes a la imagen del destino (IDT) (Hunt, 1975; Gunn C. , 1972; Goodrich, 1978; Mayo, 1973), concretamente ligados a temas relacionados con el marketing de la IDT. Desde entonces no ha hecho más que incrementarse el número de estudios sobre la IDT. Solo en 2014 se publicaron más de 270 artículos sobre este tema solo en revistas recogidas en el JCR.

Como parece lógico el enfoque, modelización y metodología sobre la imagen de destinos han ido evolucionando. Así por ejemplo se pasa de estudios sobre el papel y la influencia de la imagen de destino en el comportamiento y satisfacción de los viajeros (Chon, 1990; Echtner & Ritchie, 1993), a afinar la metodología con la aparición de análisis factorial, una prueba de *t* (t-test), análisis de medios, análisis discriminatorio, etc. y proponer modelos explicativos más complejos (Pike, 2002). Otro ejemplo de evolución en el estudio de la

imagen lo encontramos en las técnicas cualitativas y cuantitativas aplicadas. Así, en los primeros estudios se aplican las técnicas cuantitativas mientras que en la actualidad se aplican técnicas cualitativas como el análisis de contenido.

Tradicionalmente el estudio de la imagen se aborda con encuestas en las que se preguntan a los turistas por cada uno de los atributos que componen el destino. Sin embargo, la aparición de las redes sociales (RRSS) nos ha proporcionado una nueva e interesante fuente de información para el estudio de la imagen de destino (Tse & Zhang, 2013; Xiang & Gretzel, 2010).

La opinión que los turistas comparten en las RRSS nos permite conocer la imagen que tienen sobre el destino. Además esta forma de conocer las opiniones de los turistas conlleva ventajas. Por ejemplo, esta información es espontánea no dirigida por una encuesta. De esta forma sabemos los valores de cada elemento que compone el destino pero además la importancia que cada elemento tiene para el turista.

Desde su aparición, las RRSS han captado la atención de los investigadores que se han acercado a ellas desde muchas perspectivas: como canal de comunicación (Shilbury, Westerbeek, Quick, & Karg, 2014; Litvin, Goldsmith, & Pan, 2008; Placido, 2010), como fuente de información usada para tomar decisiones (Bieger & Laesser, 2004; Gursoy & McCleary, 2004; Xiang & Gretzel, 2010; Huertas, 2012; López & Sicilia, 2012; Kim & Tussyadiah, 2013) y también como fuente para conocer el comportamiento de compra y consumo de los clientes (Nezakati, Amidi, Jusoh, Moghadas, Aziz, & Sohrabinezhadtalemi, 2015; García, Carreras, Álvarez, & Royo, 2010; Roig, 2010) o el estudio de opiniones (Ruibal, 2010).

La mayoría de estos estudios implican desarrollar un análisis de contenido de lo expuesto en las RRSS. Las técnicas utilizadas para el análisis de contenido utilizadas hasta ahora son manuales (Tse & Zhang, 2013) o semi-manuales (Pan, MacLaurin, & Crotts, 2007; Jani & Hwang, 2011; Lim, Chung, & Weaver, 2012; Stepchenkova & Zhan, 2013; Oliveira & Panyik, 2015). Con estas técnicas los investigadores deben extraer los contenidos desde el sitio web objetivo página a página para almacenarlos en un archivo. La técnica tradicional de la extracción es un simple copiar y pegar. Además tras la extracción y el almacenamiento, los investigadores deben exportar el archivo donde se almacenan los contenidos a la herramienta de análisis de contenido. Esta manera de trabajar no es una técnica automática y además los

contenidos extraídos no son información en tiempo real dado que los contenidos de la página web se actualizan constantemente. Esto hace que resulten tediosos y de alcance limitado. No pudiéndose trabajar con gran cantidad de datos.

Sin embargo, podrían diseñarse otras opciones automáticas de extracción de datos que corregiría estas limitaciones y podría ser una metodología interesante para el uso de las RRSS como fuente de información sobre comportamientos de los mercados.

Esta situación nos hace descubrir un hueco de investigación que nosotros hemos establecido como el objetivo final de este trabajo: **obtención de una metodología para el análisis de la imagen de destinos a través de la extracción automática en las RRSS de los chinos parlantes.**

Para la contrastación de nuestro método hemos seleccionado el destino España. Las razones que lo hacen interesante son: que tiene una gama de productos muy variados entre los que se incluye el turismo cultural y de compras que son dos de las motivaciones más solicitadas por el turista chino-parlante (Baidewei, 2014) y que se trata de un destino líder y consolidado. España es el destino tercero por número de turistas recibidos y segundo destino en ingresos por turismo (OMT, 2014). Además esta situación viene repitiéndose desde hace más de una década (OMT, 2014) por lo que no cabe duda de que se trata de un liderazgo consolidado.

Queremos dejar claro que esta Tesis Doctoral no entra en el debate de las variables más idóneas para el estudio de la imagen de destinos. No pretendemos construir, enriquecer o ampliar los modelos existentes para el análisis de la imagen de destinos. Nuestros intereses son otros. En primer lugar, deseamos, basándonos en los modelos propuestos, diseñar un método para el estudio de la imagen de destino que extraiga los datos necesarios no de encuestas sino de la opinión que los turistas vuelcan en las RRSS. Además damos un paso más, diseñando una metodología que permita que dicha extracción se realice de forma automática y no manual o semimanual como proponen la mayoría de los trabajos hasta ahora.

Nuestra otra aportación está en el sujeto de estudio: el turista chino-parlante. Los escasos programas que permiten construir métodos de extracción automática de contenido están pensados para lenguas occidentales. Encontramos solamente un estudio que trabaja con el idioma chino (Tse & Zhang, 2013). Sin embargo en este estudio los investigadores

realizaron un análisis de contenido de forma manual. Esto se debe a la falta de metodologías sofisticadas para trabajar con los contenidos online en chino, lo cual hace necesario plantearnos crear un nuevo método de trabajo usando la técnica del procesamiento de lenguajes naturales para llevar a cabo correctamente esta investigación.

Para ello, acudimos a ayudas de los expertos sobre análisis de contenido chino. Tuvimos una estancia en el laboratorio del procesamiento de lenguaje natural chino en la Academia Sinica, el centro de investigación nacional de Taiwán. Este centro de investigación está considerado uno de los 10 mejores centros mundiales. Durante esta estancia, encontramos las herramientas adecuadas para nuestra investigación y nos iniciamos en su manejo y programación.

Ya de regreso y tras varias pruebas varios hitos, conseguimos crear un método de trabajo automático para el estudio de la imagen de destino de chino-parlantes.

Como en toda investigación, este trabajo no ha sido lineal y fácil. Varios han sido los hitos a solucionar, por ejemplo, el ensamblaje de varios programas que debían trabajar unidos y en armonía, o la búsqueda de estudios precedentes similares que nos sirven como referencia o justifican nuestro trabajo.

Todo nuestro trabajo se presenta estructurado en los siguientes capítulos: 1. Marco teórico, 2. Objetivos, 3. Modelo y propuesta para la extracción automática para el análisis de la imagen de destino, 4. Resultados, y por último, 5. Conclusiones, limitaciones y líneas futuras de investigación.

En el primer capítulo repasamos conceptos y teorías acerca de la imagen de destino y las RRSS. Revisando los estudios de los últimos años y los trabajos más relevantes relacionados con estos temas, resumiendo los conceptos, las clasificaciones y temas más tratados en este entorno, de allí, delimitamos el nicho del trabajo para desarrollar esta Tesis Doctoral.

A partir de los descubrimientos de la parte teórica, el segundo capítulo se dedica a concretar y justificar las tareas de esta investigación. Explicando la razón por la que estudiamos la imagen de destino de los chino-parlantes a través de las RRSS y proponemos una nueva metodología de trabajo.

El tercer capítulo constituye la parte más importante de la Tesis, ya que en él se presenta y consta nuestra propuesta metodológica. Este novedoso método es complejo pues se trata de un proceso encadenado con distintas etapas. Además, cada etapa se compone de varias tareas. Esto requiere, que para conseguir un resultado concluyente, necesitemos muchas pruebas. Para simplificar y ordenar la presentación, el método se recoge por secciones. Cada sección aquí está denominada con el nombre de la etapa, empezando por una introducción que aclara el motivo y la tarea por realizar, siguiendo la descripción con detalle de los pasos, y terminando con uno de los resultados obtenido en dicha etapa a manera de ejemplo.

Tras realizar una serie de pruebas con la metodología creada en el capítulo anterior, en el cuarto capítulo, demostramos el resultado obtenido con el implemento de la metodología completa. Aparte de demostrar la imagen de destino global poseen los chino-parlantes en las RRSS, revisamos los resultados de opiniones con el fin de evaluar la precisión de la metodología propuesta.

Finalmente, se presentan las conclusiones y futuro línea de investigación. En la conclusión explicamos la aportación de la Tesis a la investigación actual. Respecto a la futura línea de investigación, se ofrece la recomendación para mejorar el proceso de analizar el texto chino en RRSS. Al final, se añaden las bibliografías citadas en este trabajo para concluir la Tesis Doctoral.

# INTRODUCTION

## INTRODUCTION

In recent years, China has shown its influence on the world tourism market. According to a report published by the United Nations World Tourism Organization in 2014, China has become an important market, with increasingly growth in the number of tourists. Moreover, China leads growth in expenses from travelling abroad, reaching a record \$129,000 million US dollars. In addition, Chinese tourists continue to demonstrate the potential for consumption in the tourist sector, occupying the top place among the ten most important groups for outbound travel markets in the world during the first semester of 2013 (OMT, 2015).

Not surprisingly, tourism destinations have taken a serious interest in this market. This includes, for example, interest in understanding Chinese tourist habits in terms of consumption, and the opinions they have regarding particular destinations. Understanding the latter is the primary target in this study.

The study of images of tourism destinations is an appellant topic among researchers. Traditionally, surveys were utilized to ask tourists about certain features that apply to particular destinations. Nevertheless, the ubiquity of social networks has provided a new source of information for studying tourists' images of a destination.

The opinions that tourists share on social networks allow us to better understand how they feel about the destination. In addition, this method of collecting opinions is advantageous. For instance, the information is spontaneous; it is not directed by a survey. Thus, we can understand not only the values attached to each aspect of a tourist destination, but also the importance that each of these aspects has for the tourist.

However, most of the current content-analysis studies regarding consumers' opinions on social networks continue to proceed manually (Tse & Zhang, 2013) or semi-automatically (Bread et al., 2007; Jani & Hwang, 2011; Lim et al., 2012; Stepchenkova & Zhan, 2013; Oliveira & Panyik, 2015). Such methods are limited in scope because they are time-consuming and require the duplication of manual effort.

The absence of efficient techniques has motivated us to establish a final goal for this study: **obtaining a method to automatically extract comments from Chinese social-media content for analysing destination image.**

To verify this method, we selected Spain as the destination. Spain proved to be a destination with an enormous variety of tourism activities including shopping and cultural tourism that are two major travelling motivations of Chinese tourists (Baidewei, 2014). Besides, Spain is the third most visited international destination (UNWTO, 2014). For the above reasons, we can make sure that the chosen location has generated a diverse and vast amount of comments in order to obtain a significant result.

Also, we need to make clear that this thesis does not aim to search the most suitable variables for the study of image of destinations. Building, enhancing or expanding existing models for analyses of destination images are not our intention in this study. Our interest lies in: extracting the necessary data from the reviews, which tourists posted on social networks not from surveys. This process of data retrieving is based on the models proposed for the destination image study. In addition, it is a sequential, automatically processing, no manual or semi-manual as proposed by most of the studies so far.

As to the outline of this thesis, all contents are divided into the following chapters:

## CHAPTER 1. THEORETICAL FRAME

In the first chapter, concepts and theories about the tourism destination image and social networks are revised. After reviewing the recent studies and the most relevant works related to these themes, summarizing concepts, topics and methodologies applied in this field, we define tasks to develop this thesis.

## CHAPTER 2. OBJECTIVES

The second chapter is devoted to defining and justifying researching tasks and explains the reason of studying Chinese-speakers image about destination on social networks and propose a new methodology.

### CHAPTER 3. METHOD CONFIGURATION OF AUTOMATIC EXTRACTIONS FOR THE ANALYSIS OF THE DESTINATION IMAGES

The third chapter is the most important part of the thesis, since it presents and refines our proposed method. This proposed method is complex and contains a process with different stages chained together with several tasks. Each of stages needs extensive tests to achieve a conclusive result. To make the presentation easy to understand, this chapter is separated into sections. Each section is named of the stage, starting with an introduction that explains the purpose and the task at the beginning, following the description in detail of the steps, and ending with an example result obtained in each stage.

### CHAPTER 4. EMPIRICAL TESTING OF THE PROPOSED MODEL

After conducting a series of tests with the method developed in the earlier chapter, the fourth chapter shows the complete results obtained after processing the entire proposed method. Apart from the global destination image, hold by the Chinese-speakers in social networks, this chapter also provides examinations of resulting opinions to assess the accuracy of the proposed method.

### CHAPTER 5. CONCLUSIONS, LIMITATIONS AND FUTURE INVESTIGATION.

This chapter describes the contribution of the thesis to the current investigation. Regarding future research, the recommendation is provided to enhance the better result of Chinese text analyses from social networks.

# MARCO TEÓRICO

# 1. MARCO TEÓRICO

**Resumen:** En el presente capítulo ofrecemos una introducción respecto a la conceptualización y estado de la investigación sobre la imagen de destino turístico, redes sociales y mercado emisor chino. Recorriendo las tareas realizadas en las investigaciones de los últimos años, resumimos algunos descubrimientos propios que nos ayudan a plantear el objetivo de esta Tesis. Esto con dos fines, primero delimitar los conceptos que vamos a usar en esta investigación para que se pueda interpretar correctamente el trabajo. Segundo, conocer los temas más tratados de los últimos años para descubrir el hueco de investigación.

Antes de entrar en profundidad en nuestra investigación es necesario dejar claro y bien delimitado los conceptos con los que vamos a trabajar. Estos son: Imagen de los destinos y redes sociales, mercado. El fin es garantizar una adecuada lectura e interpretación del trabajo y de las conclusiones. Por otra parte, este capítulo también recoge el estado de la cuestión o estado del arte sobre la investigación en el tema que nos ocupa: la imagen de los destinos a través de las opiniones revertidas en las RRSS. Este análisis es vital para identificar qué ha preocupado, cómo y a quién dentro de la comunidad científica sobre el estudio de las RRSS y la imagen de los destinos. El diagnóstico del estado de la investigación nos permitirá detectar huecos, temas o aspectos no tratados o parcialmente resueltos, y nos ayudará a orientar nuestro trabajo y perfilar las metas de este.

## 1.1. Imagen de Destino Turístico

El destino turístico es uno de los factores más importantes para el marketing del turismo. Tal y como Wang citó, los elementos geográficos son uno de los componentes principales en la gestión y administración del turismo acompañado por los elementos de turistas y los elementos de industrias turísticas (2011, pág. 2). Dentro de los elementos geográficos, el destino turístico funciona como un factor que atrae turistas y ofrece un lugar

donde los visitantes pueden realizar actividades turísticas y tener experiencias. Por ello, el destino turístico forma parte de los componentes que sostiene el sistema de turismo.

Enriqueciendo la línea anterior, a través de varios estudios relacionados con el destino turístico (Martín Gutiérrez, 2005; OMT, 1998; Bull, 1994; Joppe, Martin, & Waalen, 2001; Murphy, Pritchard, & Smith, 2000; Cooper, Fletcher, Gilbert, & Wanhill, 1993; Gunn & Var, 2002; Hu & Ritchie, 1993; Font, 1997), averiguamos que un destino se puede tratar de un espacio, una ciudad, una región, o bien un país hacía donde los turistas se desplazan. En este lugar determinado, los visitantes pueden disfrutar de las instalaciones, encontrar servicios turísticos y cumplir sus objetivos de viaje. Por lo tanto, el destino turístico no cuenta solamente con un territorio geográfico sino también con los productos y servicios ofrecidos en este lugar concreto, es decir, el destino turístico es una integración de varios productos de turismo y ofrece la experiencia a sus visitantes.

Reflejo de ello, existe un consenso en la literatura en identificar como al destino como un macro producto turístico donde se incorporan la suma de los productos turísticos (oferta básica, de ocio, etc.) más los atractivos del lugar (clima, cultura, hospitalidad, gestión, etc.) creando sinergias que hacen que la propuesta del destino sean más que la suma de las ofertas particulares (Chen F.-W. , 2010; Wang Y. , 2011; Joppe, Martin, & Waalen, 2001; Martín Gutiérrez, 2005; Alcázar Martínez, 2002). Dado que el objetivo de esta Tesis no recae en el debate de que se entiende por destino turístico y que, además, hay un notable consenso sobre ello, tomaremos estas reflexiones como punto de referencia sobre que entendemos como destino turístico.

A partir de estas, en este apartado, pretendemos definir los elementos principales que afectan a la imagen de un destino concreto para saber cómo gestionarlo y promocionarlo.

Mariné-Roig manifestó que “la imagen turística es un concepto fundamental para entender el fenómeno turístico y ver cómo el turismo afecta y se relaciona con los destinos y la sociedad” (2010, pág. 62). En este sentido, Park y Njite (2010) acertaron afirmando que una mejor imagen representativa de un determinado destino turístico, supone una mayor probabilidad de que el turista visite dicho lugar. Con el motivo de mejorar la imagen del turismo, estudiamos la imagen origen de un grupo de turistas específicos y esperamos hallar una puesta en común y desarrollar un método sostenible para el turismo de un sitio concreto.

Para esto, es imprescindible conocer los componentes que afectan a la imagen turística y su influencia hacia otros factores solo así se puede convertir la imagen en una herramienta para trabajar la decisión y comportamiento de visitantes tal y como Baloglu y McCleary (1999) manifestaron: “Before image can be used to influence behavior, it is important to understand what influences image.”.

Por lo tanto, este estudio comienza por los conceptos generales de la IDT y la delimitación de los factores que la construyen para pasar luego a la revisión del estado de la cuestión en la actualidad.

### **1.1.1. Concepto de la imagen de destino turístico**

Desde mediados de los años 70, la imagen de destino turístico (IDT) se ha ido convirtiendo en uno de los conceptos más importantes en el marketing de destino cuya función sirve como un medio para diferenciar entre los destinos, y también, se ha destacado su importancia en el proceso de toma de decisiones de turistas (Shani & Wang, 2011; Hudson, Wang, & Gil, 2011; Park & Njite, 2010; Karim & Chi, 2010; Phillips, Wolfe, & Hodur, 2013; Wang H.-Y. , 2012; Correia Loureiro & Miranda González, 2008; Hernández-Lobato, Solis-Radilla, Moliner-Tena, & García, 2006; McCartney, Butler, & Bennett, 2009).

A lo largo de los años de investigación, el concepto de la imagen de destino turístico también logra tener varias definiciones a través de diferentes autores. Según Tasci, Gartner, & Cavusgil (2007), los autores cuya definición es citada con mayor frecuencia son: Hunt (1975), quien definió la imagen de destino como una impresión que un individuo o un grupo de personas tienen sobre un sitio donde él o ellos no residen; Reynold (1965), quien apuntó que el origen de la imagen de un destino se crea a través de la construcción mental desarrollada por unos visitantes potenciales basada en la selección de algunas impresiones desde una gran cantidad de imágenes, y que mediante este proceso las impresiones son elaboradas, ennoblecidas y ordenadas; y, en tercer lugar, Crompton (1979), quien manifestó que la imagen de destino es la suma de creencias, ideas e impresiones que posee una persona de un destino.

Las aportaciones al concepto de imagen se han prolongado hasta el día de hoy. Así, Mariné-Roig (2010) señaló que la imagen turística es la expresión de todo conocimiento

objetivo, impresiones, prejuicios, imaginaciones y pensamientos emotivos que un individuo o grupo pueden tener de un lugar concreto.

Como puede observarse, a pesar de que no existe solamente una definición en ellas observamos unos rasgos comunes: las impresiones sobre un destino que se procesan en la mente del turista y que tienen variadas fuentes de inspiración (ideas, vivencias, creencias, etc.).

Al margen de las definiciones algunos autores hacen matizaciones interesantes para tener una idea de que es y como se construye la IDT. En esta línea nos resulta interesante las aportaciones de Mohan (2010) cuando declara que la percepción de la imagen es simultáneamente subjetiva y comparativa según distintos receptores en diferentes lugares y Hunter (2010) que aclara que la imagen de destino puede referirse a la imagen percibida por los visitantes sobre un lugar o la imagen proyectada bien por las OGDs o bien por las empresas de turismo.

Lo que estos todos los estudios demuestran es que el concepto de imagen es de una gran complejidad, como Gallarza, Saura y García demuestran (2002) dado:

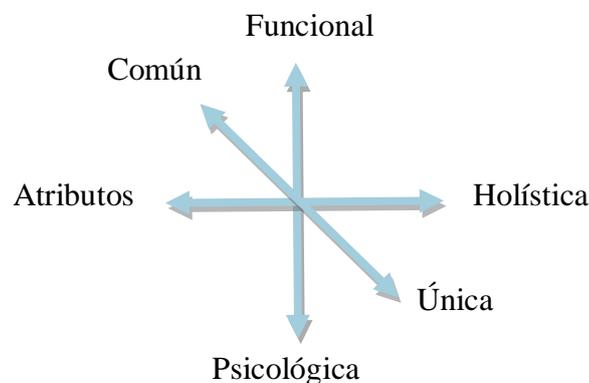
- la complicada naturaleza,  
Lo que permite más de una interpretación o cuya comprensión carece de un sentido único, como por ejemplo la definición y la conceptualización de la IDT que mencionamos anteriormente.
- la multiplicidad,  
La existencia de una multiplicidad de factores o variables que forman la identidad de la IDT tal y como lo que tratamos averiguar en el apartado siguiente.
- la relatividad,  
Es a la vez subjetivo (cambios de persona a persona) y comparativo (se refiere a las percepciones entre los diversos objetos). Una imagen siempre corresponde a algunas percepciones generales, y no todo el mundo tiene las mismas percepciones.
- y la dinámica  
La imagen no es estática sino que cambia, dependiendo fundamentalmente de dos variables: tiempo y espacio

### 1.1.1.1. Las Dimensiones de la Imagen de Destino

Hoy nadie pone en duda que la imagen es un concepto complejo y multidimensionales (Gallarza, Saura, & García, 2002), por lo que recopilamos los principales estudios que mencionan a las dimensiones de la imagen y los exponemos en este apartado con el objetivo de tener una visión completa de las dimensiones de la imagen.

Echtner y Ritchie (2003) revisaron varios estudios y clasificaron la imagen de destino según tres dimensiones esenciales tal y como se muestra en la figura 1.

Figura 1. Tres dimensiones de la IDT de Echtner y Ritchie



Fuente: Extraída de *The Meaning and Measurement of Destination Image*,  
 "Echtner y Ritchie, 2003, *The Journal of Tourism Studies*, Vol. 14, No.1

- Funcional vs. psicológica: las características funcionales, con aspecto tangible, están definidas como objetos que se pueden medir u observar directamente, como pueden ser el nivel de precios, cantidad de monumentos, etc. ; y son opuestas a los factores psicológicos, con aspecto intangible, que son ideas más subjetivas o sentimientos como son por ejemplo la simpatía y atmósfera que no pueden medirse evidentemente.
- Holística vs. atributos: las impresiones holísticas se refieren a las percepciones conjuntas y generales a cambio de los atributos que son imágenes particulares. Ejemplo de esto es la imagen holística de París, que podría ser romántica y la de atributo por su parte puede ser indiferente y grosera para algunos elementos del destino como por ejemplo hoteles.

- Común vs. única: las características comunes se refieren a componentes que pueden observarse generalmente en cualquier destino al contrario de las únicas, que existen solamente en un lugar determinado. Por ejemplo, la Gran Muralla en China, la Torre Eiffel en París, el Carnaval en Río, la Tomatina en Valencia, etc.

Echtner y Ritchie (2003) defienden que la clasificación de la IDT en estas dimensiones es muy dinámica e interesante. Sin embargo, ellos mismos manifiestan que existen ciertas imágenes que se superponen entre distintas dimensiones y no se pueden catalogar claramente; tal y como un atributo particular podría estar influido por una imagen holística, o la clasificación, por ejemplo, de la limpieza que puede catalogarse en la dimensión funcional o psicológica.

Para superar estas objeciones surge, poco más tarde, una clasificación de las dimensiones que abarca las propuestas de Echtner y Ritchie y se ha convertido en la alternativa más hablada por los autores (Beerli & Martín, 2004a; Hudson, Wang, & Gil, 2011; Ekinci & Hosany, 2006; Tasci, Gartner, & Cavusgil, 2007). Se trata del concepto multidimensional, formado por las imágenes cognitivas, afectivas más las imágenes conativas.

Concretamente las imágenes cognitivas son elementos físicos o características de elementos que los turistas perciben en un destino, es decir, los conocimientos turísticos que los visitantes poseen; y las imágenes afectivas son sentimientos de turistas hacia el destino (Millet, 2011; Hudson, Wang, & Gil, 2011).

El componente cognitivo de la imagen de lugar, según Martín y Rodríguez del Bosque (2008), tiene que ver con las creencias del individuo sobre un destino turístico. Se trata de un componente muy fuerte pues como Hernández-Lobato et al. (2006) declararon que el factor cognitivo es el primer elemento que influye en la satisfacción turística. Respecto a los componentes afectivos, estos están relacionados con el sentimiento que tienen los visitantes, como puede ser el sentimiento de comodidad, de seguridad, de caro o de natural, etc. Tras realizar un estudio sobre la imagen, la satisfacción de turista y la lealtad, Hernández-Lobato et al. descubrieron que la emoción tiene mayor influencia en la fidelidad de los turistas.

Además, existen bastante unanimidad en que el componente cognitivo está filtrado por las percepciones de los turistas (Baloglu and Brinberg 1997; Baloglu and cCleary 1999a,

1999b; Gartner 1993; Walmsley and Young 1998) es decir, la información que reciben de los destinos son interpretados a través de sus creencias y situaciones particulares.

Desde un punto de vista teórico, además, hay un acuerdo general de que el componente cognitivo es un antecedente del componente afectivo y que las respuestas de evaluación de los consumidores se derivan de su conocimiento del destino (Anand, Holbrook and Stephens 1988; Holbrook 1978; Russel and Pratt 1980; Stern and Krakover 1993).

Respecto a la dimensión conativa, esta dimensión se produce como fruto de las dos dimensiones mencionadas anteriormente; se trata de la reacción y el comportamiento que tienen los turistas cuando reciben la información relacionada con el destino y crean sus propias opiniones o sentimientos.

Tasci, Gartner y Cavusgil (2007) parten también de esta idea pero dan un paso más y proponen incluir estas tres dimensiones a las planteadas por Echtner y Ritchie (2003). Para estos autores las dimensiones propuestas por Echtner y Ritchie proporcionar una estructura más profunda que recoge como cada uno de los componentes (cognitivo-afectivo-conativo) se internalizan. Así para ellos, “la imagen de destino es un sistema de interacción entre pensamientos, opiniones, sentimientos, visualizaciones y comportamientos que tienen los visitantes respecto a un lugar determinado (pág. 200)”.

Figura 2. La interrelación de los componentes de la imagen de destino



Fuente: Extraída de Conceptualization and Operationalization of Destination Image  
*Journal of Hospitality & Tourism Research* 31:194 (Tasci, Gartner, & Cavusgil, 2007, pág. 200)

Como puede observarse la figura 2. En el núcleo del modelo interactivo propuesto hay un conocimiento cognitivo de los atributos comunes y únicos de destino y de la respuesta afectiva hacia esos atributos. Con la interacción entre el conocimiento de los atributos y sentimientos hacia ellos, sean estos únicos o comunes, se forma una imagen compuesta (holística o global) que es la que es utilizada por el turista para simplificar la tarea de toma de decisiones.

Este es un modelo dinámico e interactivo en la que cada elemento puede ser a la vez causa y efecto de un cambio en cualquier momento, y las dimensiones no se puede comprender aisladamente. Por lo tanto, deben ser estudiadas de manera integrada.

Esta propuesta, ampliamente aceptada, exige tener claro los atributos que componen dicha imagen. Atributos a través de los cuales se construye la imagen general u holística del destino. Una propuesta en este sentido es la que aporta Murphy, Pritchard y Smith (2000). Ellos investigaron los factores que influyen en la experiencia de los turistas y llegaron a la conclusión que eran la combinación de la calidad y el valor del destino concreto que perciben durante el viaje. A través de varias investigaciones, ellos clasificaron estos factores influyentes en dos grandes grupos como podemos ver a continuación:

- Por el entorno: Entorno natural, Factores político-legales, Factores tecnológicos, Factores económicos, Factores culturales, y Factores sociales
- Por los servicios: Compras, Recreo y atracción, Acomodación, Transporte, Restauración y Viaje

Observando la figura 3, podemos averiguar la relación entre tres niveles.

Figura 3. Factores que influyen la experiencia de turistas



Fuente: Extraída de The destination product and its impact on traveller perceptions, *Tourism Management* (Murphy, Pritchard, & Smith, 2000)

Para otra parte, Beerli y Martín (2004) tras un examen de las atracciones y atributos incluidos en las escalas con las que se estudia la imagen de destinos, clasifican e incorporan todos los factores que influyen en la evaluación de imágenes en las nueve dimensiones siguientes:

- Recursos naturales
- Infraestructura general
- Infraestructura turística
- Ocio y entretenimiento turístico
- Cultura, historia y arte
- Factores políticos y económicos
- Entorno natural
- Entorno social
- Ambiente del lugar

Cada dimensión está relacionada con un sub-grupo de atributos con el motivo de establecer una buena escala de medir la imagen (A.1. Las dimensiones de imagen de destino). Dependiendo del destino y el tema que un investigador estudia, las variables se modifican de acuerdo a sus necesidades. Adoptando las características de la imagen de destino, la

multiplicidad, la relatividad y la dinámica, esta propuesta consigue superar la complejidad natural del concepto de la percepción de destino e implantar un método sofisticado para la investigación de esta área.

Debemos terminar diciendo que la delimitación de los atributos es solo un paso. Para el análisis y evaluación de la imagen estos atributos deben ser valorados en el binomio: dimensión cognitiva y afectiva. Es decir, que sabemos acerca de ese atributo y que sentimos en referencia a dicho atributo.

#### **1.1.1.2. Factores que influyen en la formación de la imagen de destino en la mente del consumidor**

La formación de imagen turística se elabora a partir de una selección de impresiones sobre la que se desarrolla la construcción mental a través de un flujo de información de destino (Echtner & Ritchie, 2003).

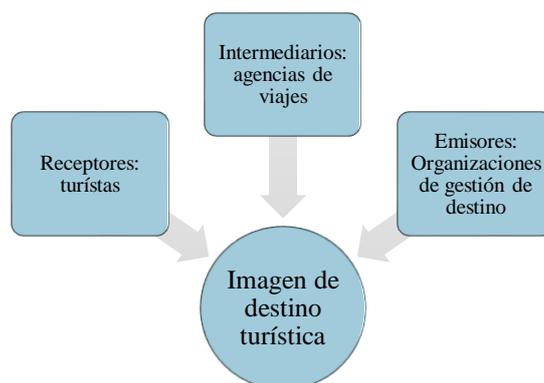
Partiendo de esta idea, el primer paso a estudiar es el tipo de imágenes o impresiones que entran en juego. Miossec (citado por McCartney et al., 2008) cataloga estas imágenes en tres tipos como observamos a continuación:

- Imágenes universales (universal images): origen de la influencia histórica
- Imágenes efímeras (ephemeral images): origen de la interpretación a través de otros medios, tal y como artes, músicas, cines, etc.
- Imágenes inducidas (induced images): origen de la acción comercial con el fin de influir en la decisión de las turistas

Por su parte, el flujo de información podría generarse desde diferentes fuentes y los emisores de esas fuentes podrían ser un grupo organizado, una persona individual, o una agencia de publicidad. Según Tasci y Gartner (2007), podemos clasificar el origen de la información en tres fuentes:

- Emisores: por parte del proveedor o de destino
- Intermediarios: pueden ser persona individual o autonomía
- Receptores: por parte de los demandantes o receptores de imagen

Figura 4. La formación de la imagen



Fuente: Extraída de *Destination image and its functional relationships*,  
*Journal of Travel Research*, 45:4 (Tasci & Gartner, 2007)

Además, la imagen de destino se puede transmitir a través de varios aspectos o herramientas como (Govers et al., 2007):

- Medios masivos de comunicación: publicidad, folletos, periódicos, revistas, televisión, películas, reportajes y documentales;
- Opiniones: familias, amigos, personas del sector turístico, los post-visitantes;
- Arte y cultura: pinturas, fotografías, música, novelas, etc.

Por otro lado, los turistas reaccionan de distintas maneras, interpretan los mensajes que reciben de diferentes emisores de varias formas y llegan a tener comportamientos particulares (Ekinci y Hosany, 2006). Tal y como señalan Stern & Krakover (1993), en la construcción de la IDT están implicados tanto la información obtenida de diferentes fuentes como las características del individuo. Estos autores proponen que el modelo que explique las construcciones de la IDT debe contemplar que las características de la información y del individuo tienen un efecto sobre el sistema de interrelaciones que rige los estímulos percibidos del entorno, produciendo una imagen compuesta. De acuerdo con esto, algunos autores (Baloglu & McCleary, 1999; Beerli & Martín, 2004a) proponen un modelo general donde los factores de formación de imagen se diferencian entre los factores de estímulo (fuentes de información, experiencia previa y distribución) y factores personales (psicológicos y sociales). De todas las variables personales, la percepción es la más citada (Woodside & Lysonsky, 1989; Um & Crompton, 1990; Stabler, 1995). De hecho, las propias definiciones de IDT hablan de impresiones de imágenes, es decir, de imágenes interpretadas a través de traductor y filtro que supone la percepción (Schiffman & Kanuk, 1997).

Pero también la experiencia previa está presente en el cómo se construye la imagen. En este sentido, Echtner y Ritchie (2003) explicaron la construcción mental de la IDT como un proceso que comienza por tener la imagen de viajar y dicha imagen se modifica después por la información adquirida antes de decidir el destino; la segunda modificación de la imagen ocurre a continuación de que el turista haya viajado al destino y haya participado en las actividades turísticas. Al igual que esta idea, Beerli y Martín (2004) indicaron que la imagen posee por un pre-visitante y un visitante es diferente. Un pre-visitante puede adquirir la información de segunda mano difundida por turistas que ya han estado en un destino o por las organizaciones de gestión de destinos, sin embargo, no importa si una información o un mensaje que va a llegar a los consumidores son de primera mano o no.

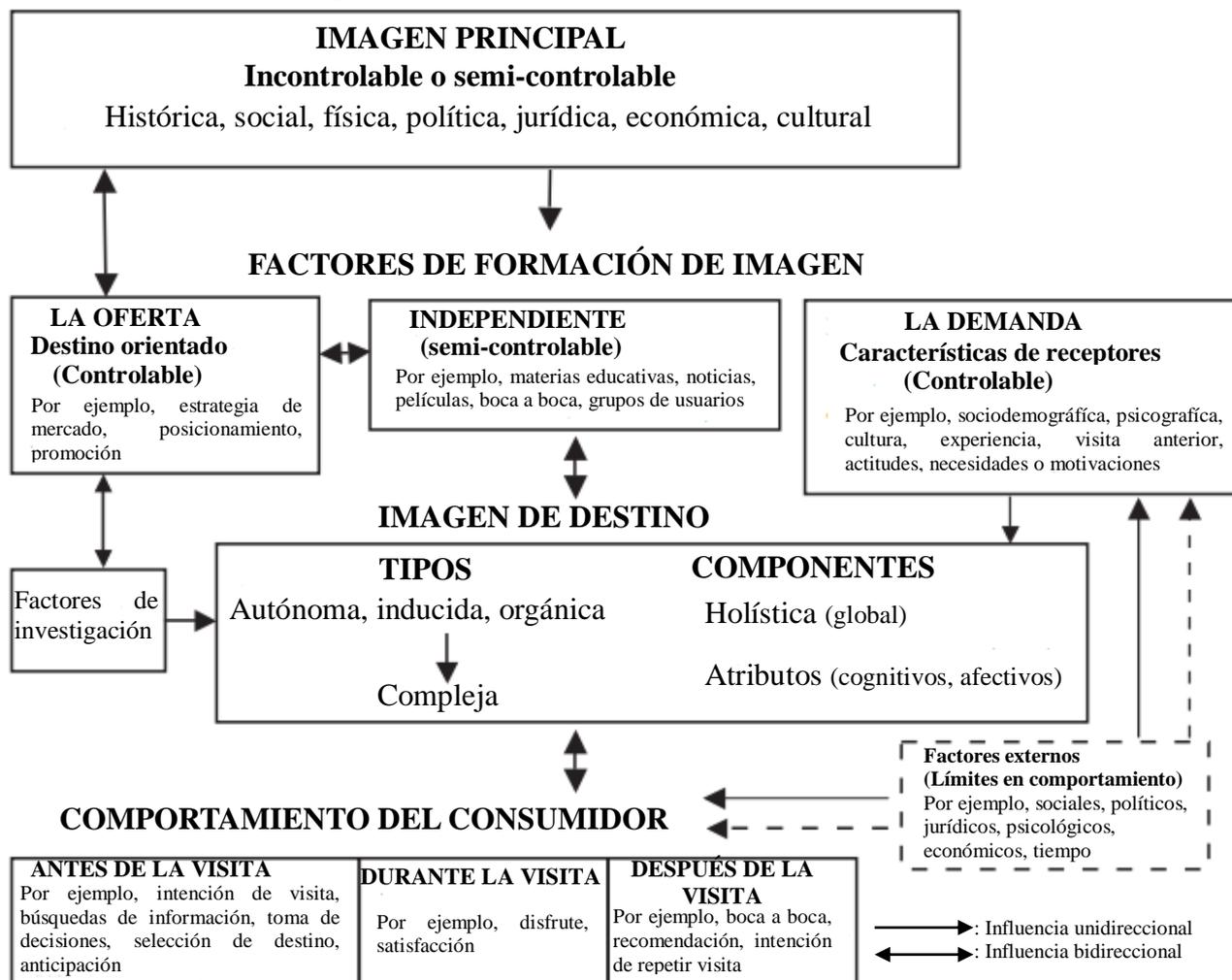
Por otra parte, la imagen influye en la decisión de los turistas a la hora de elegir un destino y también afecta al comportamiento y la satisfacción de los turistas. Como consecuencia, la imagen de destino se ve afectada por la experiencia de turistas después de haber viajado al destino (Boo & Busser, 2005). A continuación, resumimos el origen de la imagen de destino a través de estudios recientes.

Por último, en numerosos artículos (Richardson & Crompton, 1988; Pizam & Sussmann, 1995; Verlegh & Steenkamp, 1999; Reisinger & Turner, 2002; Prebensen, 2005; Karahanna, Evaristo, & Srite, 2005; Lee & Lee, 2009; Lee & Sparks, 2007; Tasci & Gartner, 2007) los investigadores mencionan que turistas de diversos países y perfiles o en diferentes momentos podrían resultar tener distintas imágenes del mismo destino. A este respecto nos resulta muy interesante la propuesta de Tasci y Gartner (2007). Estos autores revisaron las influencias socio-demográficas y de cultura en imágenes de destino y mencionaron que el origen cultural de turistas internacionales podría influir el resultado de la imagen de destino. En el estudio de Tasci y Gartner ordenaron tres tablas relacionando la interacción del perfil de turistas y la imagen de destino:

- la primera tabla donde se indican turistas de distintos perfiles y las influencias que surgen en la imagen de destino;
- la segunda tabla demuestra la diferencia de la imagen que poseen en turistas que ya han visitado el destino y turistas que aún no han visitado el lugar;
- y la última enseña los efectos que proporciona la imagen de destino en el comportamiento de los consumidores.

Finalmente, trabajan conjuntamente estas tablas y proponen un modelo global donde explicaron la relación entre la imagen de destino y la influencia de otros factores como se puede observar en la figura 5.

Figura 5. Modelo de imagen de destino y sus relaciones



Fuente: Extraída de *Destination Image and Its Functional Relationships*. 2007, *Journal of Travel Research* 45:413. (Tasci & Gartner, 2007, pág. 422)

### 1.1.2. Estado de la cuestión en la investigación sobre la imagen de destino

Una vez delimitado el concepto de estudio, la imagen de los destinos turísticos y abordado la forma en la que se construye y sus diferentes dimensiones es necesario profundizar en cómo se está estudiando este tema. Como apuntábamos en la introducción, el estudio de la imagen de destino es un tema ampliamente tratado y desde antiguamente por los investigadores, que además han cambiado el enfoque y propuestas metodológicas a lo largo de los años. Para nuestra investigación resulta interesante, acabar esta parte teórica con un repaso de los artículos publicados en la última década, para que, tengamos una imagen global

de cómo está en estos momentos la investigación sobre la imagen de destinos. Por ello, a continuación, presentamos los temas más relevantes, las metodologías aplicadas, las variables escogidas y los resultados obtenidos en los estudios, comparándolos y obteniendo las conclusiones pertinentes.

### **1.1.2.1. Temas más investigados**

Revisando los estudios realizados en torno de la imagen de destino en la última década, podemos clasificar dichos estudios en los siguientes temas:

- Revisiones de la conceptualización y las dimensiones de la IDT.
- Los procesos de la formación de imagen.
- Las medidas, las valoraciones y la evolución de la imagen de destino.
- Las interacciones e influencias de la imagen turística en otros componentes.

#### ***Revisiones de la conceptualización y la dimensión***

Un gran número de trabajos sobre la IDT se plantean propuestas de ordenar y clasificar las aportaciones realizadas hasta determinadas fechas. Dando lugar a numerosos estudios que se encuentran en este ámbito. Así encontramos a investigadores que se dedican a revisar, resumir y concluir lo que otros académicos descubren. Podemos citar entre ellos a Pike (2002) que hizo una revisión de las investigaciones realizadas en torno a este tema durante el año 1973 y el año 2000 y las clasificó por el tema y las regiones que investigan, las variables y las técnicas que usan; Echtner y Ritchie (1993; 2003) que revisaron numerosos artículos con la intención de diseñar una técnica apropiada para medir la imagen de destino o Tasci y Gartner (2007) que hicieron una revisión y propusieron un marco teórico de referencias para los futuros estudios. Al mismo tiempo, Tasci (2007) hizo también una revisión con Gartner y Cavusgil sobre los estudios desde el año 90; ellos resumieron las definiciones, las metodologías, los modos de coleccionar datos, número de factores y variables de las investigaciones. Stepchenkova y Mills (2010) estudiaron los trabajos de otros investigadores publicados entre los años 2000 y 2007 con el objetivo de identificar las tendencias actuales y emergentes en el ámbito de los estudios de la IDT. Por último, debemos destacar a Gallarza et al. (2002) que, tras examinar varios estudios, descubrieron la complejidad de la imagen y crearon un sistema ayudando a las organizaciones públicas y privadas a analizar y controlar las variables de la IDT.

Quizás el trabajo más interesante es el de Echtner y Ritchie (1993) debido a su gran aportación relacionada con la formación de la imagen a la teoría de la imagen de destino, siendo además el estudio más citado por otras investigaciones del mismo tema.

Aunque en este trabajo no podemos pasar por alto la recopilación de Stepchenkova y Mills (2010) por ser la más reciente en el tiempo, recoge la reflexión más actual sobre el tema de Internet y el análisis práctico computarizado.

### ***Los procesos de la formación de imagen***

La imagen de destino turístico es un esquema complejo, como ha quedado manifiesto en el punto anterior, y por ello es difícil de medirla. Por lo tanto, hay parte de investigadores que desarrollan estudios orientados a proponer, encontrar o mejorar un método para explicar la formación de la imagen. Por ejemplo, Beerli y Martín (2004) desarrollaron una plantilla que explica los factores válidos en la formación de imagen. Camprubí, Guia, y Comas (2008) propusieron una nueva manera de formalizar e inducir la imagen de turismo a través del análisis de la relación y la interacción entre la persona turística interna y externa. Wang (2008) investigó el proceso de la imagen proyectada por distintos emisores, tales como la organización pública o los proveedores locales. Polo-Peña et al. (2012) identificaron las dimensiones implicadas en la formación de la imagen de un destino rural. Li y Stepchenkova (2012a; 2012b) examinaron la imagen que poseen los turistas chinos e identificaron la percepción de la imagen recibida por los turistas chinos de los EEUU, examinando más tarde el proceso de formación de dicha imagen. Banyai (2011) demostró un proceso distinto de formación de imagen por los turistas y las guías del castillo de Bran.

En general puede decirse que los modelos propuestos sobre la formación de imagen se dividen en dos grupos, la imagen transmitida por los emisores (OGDs) y por los receptores (turistas), dado que estos dos grupos no muestran siempre la misma imagen respecto al mismo destino. Entre estos estudios, debemos destacar el de Beerli y Martín (2004). Su investigación nos ofrece una conceptualización sólida y clara, aportando una tabla que incluye los atributos que forma la imagen de destino y que están separados en nueve dimensiones combinando tanto los factores cognitivos como los afectivos.

### *Las medidas, las valoraciones y la evolución de la imagen de destino*

También existen autores que se dedican a medir, examinar, comprobar y evaluar la función de la imagen. Chen et al. (2012) buscaron la imagen más representativa en un destino concreto al igual que Chen, Chang y Cheng (2010) midieron la imagen del turismo rural; Jetter y Chen (2011) examinaron la imagen del destino preferido para convenciones. Jani y Hwang (2011), por su parte, analizaron la imagen a través de los contenidos extraídos desde Lonely Planet. Ramkissoon y Nunkoo (2011) examinan la imagen que poseen los propios residentes del destino y Ryan y Aicke (2010) examinaron diferentes imágenes que poseen residentes y turistas respetivamente. Tasci (2007) investigó las metodologías que se aplican al medir la imagen, descubriendo que el resultado de la imagen varía según el tipo de metodología aplicada por cada investigador. La investigación de Bell (2008) se dedicó a comprobar la imagen de Eco turismo en Nueva Zelanda y la oferta real que se ofrece in situ. McCartney et al. (2008) buscaron una metodología para desarrollar una imagen eficaz y óptima para atraer visitantes al destino y al año siguiente (2009) McCartney, Bulter y Bennett examinaron si la imagen proyectada en Macao influyó en la decisión de turistas de Beijing, descubriendo que los principales atributos de dicha imagen no se identifican con los motivos de visita. Djafarova y Anderson (2008) investigaron la función del lenguaje, concretamente el uso de la metáfora, el juego de palabras y la aliteración, en la publicidad donde se representa la imagen de un destino concreto. Yang et al. (2012) usaron un método para medir la imagen de destino. Tang et al. (2009) examinaron las páginas de dos destinos en inglés para comparar las imágenes representadas. Milman (2011) examinó la imagen que se transmite a través de postales. Li, Wang (2011), Law y Cheung (2010) examinaron la imagen a través de blogs. A diferencia de ellos, Wang (2012) además de examinar la imagen, buscó un modelo de proyectarla y que pueda influir en la decisión de turistas. Dolnicar y Grün (2013) investigaron la mejor escala para la medida de la imagen y afirmaron que la elección binaria entre "sí" y "no" es la mejor escala para averiguar la preferencia que tienen los turistas en seleccionar los atributos de destino.

En general, los estudios relacionados con esta línea tienen la intención de averiguar la imagen proyectada y a su vez, reflejan la relatividad de la imagen de destino. Es decir, aunque en los estudios se cambian los sujetos de investigación y las percepciones entre los diversos objetos, todos tienen el mismo objetivo: descubrir la imagen proyectada de cierto destino.

Entre estos, nos llama la atención el estudio de Dolnicar y Grün (2013). A diferencia de la mayoría de investigaciones donde se recomienda medir la imagen de destino con escalas de tipo *Likert*, ellos proponen usar la escala más simple, la elección binaria después de experimentar con distintas medidas.

### ***Las interacciones e influencias de la imagen turística en otros componentes***

Dentro de este grupo se recogen los trabajos que se acercan a analizar la relación de la imagen con variables externas. Así, Frochot (2003), Karim y Chi (2010) estudiaron la relación de la comida con la imagen de destino. Yao et al. (2013) examinaron la fiabilidad de la imagen proyectada y los comportamientos turísticos. Un estudio parecido elabora Feighey (2003). Hallmann y Breuer (2010) y Mohan (2010) examinaron la imagen recibida de las actividades deportivas con la imagen proyectada del mismo destino y Chen et al. (2012) contrastaron la imagen proyectada por la organización de turismo y por los blogueros. Boo y Busser (2005) examinaron las características de turistas y sus influencias en la percepción de la imagen de destino. Ekinci y Hosany (2006) adoptaron la teoría de otro investigador, la personalidad de marca de Aaker, y la aplicaron en la imagen de destino. Autores como Hernández-Lobato et al. (2006), Correia-Loureiro, Miranda-González (2008), Lee (2009) Park, Njite (2010), Phillips et al. (2013), Forga y Cánoves (2013) investigaron las interacciones entre la imagen de destino, las actitudes de turistas, sus motivaciones, sus satisfacciones y sus posibles comportamientos en el futuro. Tang et al. (2009) investigaron las diferentes imágenes que se muestran en las páginas web oficiales en distintos idiomas de un mismo destino. Jönsson y Devonish (2008) exploraron la relación entre la motivación de turistas y sus percepciones de imagen del sitio turístico. McCartney (2008) buscó el impacto producido por la diversidad de cultura en la formación de la imagen. Cornelissen (2005) investigó cómo la situación económica de un destino concreto perjudica a la imagen turística. Hudson et al. (2011) examinaron la influencia de la película "Motorcycle Diaries" en la percepción de la imagen de destino. Govers y Go (2004) analizaron la influencia de la imagen de destino que se proyecta a través de fotografías online. Govers et al. (2007) buscaron la relación de la promoción turística con la imagen. Phillips y Jang (2010) estudiaron la influencia que los visitantes anteriores pueden dar a un destino determinado. Ivanov et al. (2010) y Kneesel et al. (2010) trabajaron para encontrar la relación existente entre la marca de destino y su imagen. Elliot et al. (2011) exploraron una fuerte interacción entre el turismo, el producto y la imagen global de un sitio concreto. Cousin (2008), Lee y Lockshin (2012)

examinaron la relación de la imagen proyectada por la organización de turismo y la decisión de turistas. Existen además varios investigadores que examinaron la influencia de la percepción de imagen de destino de turistas a través de las fotografías como son MacKay y Couldwell (2004), Dorwart et al. (2006), Garrod (2009) y Hunter (2010).

Como podemos comprobar, esta línea de investigación ha sido muy prolífera y variada. Los estudios realizan comparaciones entre los atributos influyentes de la imagen, las maneras de proyectar la imagen, las actividades que se ofrecen en un destino, la imagen que posee a través de distintas fuentes, etc.

### **1.1.2.2. Metodologías, variables y escalas más usadas**

Después de revisar numerosos estudios, descubrimos que no existe tan sólo una forma de medir y analizar la imagen de destino ni una única escala común utilizada por los investigadores. Tal y como Gallarza et al. (2002) definieron, la imagen de destino es compleja, diversa, relativista y dinámica. Lo que lleva a una dificultad para el consenso en estas propuestas.

#### ***Metodologías***

Obviamente en el análisis de las metodologías empleadas debemos dejar aparte los trabajos de revisión bibliográfica. Estos artículos que revisan los estudios suelen realizar una conclusiones donde seleccionan los estudios más sobresalientes y diagnóstico de la materia hasta ese momento.

Observando el desarrollo de las investigaciones en torno a este tema, algunos estudios revelan la evolución en la metodología usada. En la revisión de Tasci et al. (2007), ellos analizan los métodos para estudiar la imagen, llegando a la conclusión de que los usos de la metodología cualitativa y su combinación con la cuantitativa aumentan después de la publicación de Echtner y Ritchie en 1991 y 1993 siguiendo las recomendaciones que propusieron en sus estudios. La revisión de Stepchenkova & Zhan (2013) demuestra que existe cierta cantidad de investigadores (Miller, 1993; Rezende-Parker, Morrison, & Ismail, 2003; Govers & Go, 2005; Ryan & Cave, 2005; Xiao & Mair, 2006; Choi, Lehto, & Morrison, 2007) que usaban técnicas no estructuradas tales como la entrevista, la pregunta abierta o los contenidos on-line para conseguir una investigación cualitativa, a diferencia de la investigación de Pike (2002), en los que la mayoría de estudios realizados durante los años

1973 y 2000 empleaban una técnica estructurada. Esto deja ver que también ha habido un paso de técnicas menos estructuradas a más estructuradas con el paso del tiempo.

En general, los estudios de la última década revelan que la combinación de la metodología cualitativa y cuantitativa representa la opción más usada para trabajar los datos. Por su parte, el cuestionario estructurado, el cuestionario abierto o bien la entrevista son las herramientas que se usan más a menudo en aquellos tipos de investigación que estudian los factores que componen, influyen o bien intervienen en la imagen de destino. Los artículos que revisan los estudios suelen realizar una *elicitation*, es decir, adoptan la información de un autor como la propia de forma fluida para concluir el estudio.

A la hora de analizar los datos, el análisis multivariable es comúnmente el método más adoptado, especialmente el análisis de regresión, el análisis de la varianza, como ANOVA, el análisis factorial o bien el análisis de ecuaciones estructurales. Sobre todo si la investigación busca la interacción de los factores que influyen en la construcción de la imagen.

No obstante, el método de análisis de contenido ha cobrado protagonismo en los estudios de la última década, usado para investigar la imagen a través de fotografías, postales, anuncios, webs o mediante blogs. Por esa tendencia, descubrimos las herramientas empleadas para investigaciones de este campo, análisis de contenido, como CATPAC, que es la más usada para analizar los textos. Este programa permite la realización de un estudio cualitativo. Además, existen otros sistemas informáticos empleados como Atlas. Ti, TextSTAT, WORDER, CAQDAS y Nvivo.

El software SPSS sigue representando la herramienta más usada para analizar e interpretar los datos cuantitativos, seguido de programas como AMOS y MATLAB.

La combinación de varias técnicas con distintas herramientas también pueden observarse a menudo en las investigaciones recientes. Por ejemplo, Liang et al. (2011) intentaron identificar y comparar la imagen de Macao y Hong-Kong a través de las webs oficiales de turismo en inglés utilizando el software WORDER para clasificar las categorías de datos, CATPACII para realizar el análisis clúster, MATLAB para analizar la matriz y realizar un análisis de contenido para desarrollar un análisis multidimensional. Al igual que muchos investigadores (Bell, 2008; McCartney et al., 2008; Li y Stepchenkova, 2011, 2012;

Mackay y Couldwell, 2004; Polo Peña et al., 2012; Dolnicar y Grün, 2013; Dorwart et al., 2006; Mohan, 2010; Jetter y Chen, 2011; Hallmann y Breuer, 2010) adoptan dos maneras de reunir los datos a través de un cuestionario estructurado para obtener datos cuantitativos y un cuestionario abierto o una entrevista para adquirir datos cualitativos.

No podemos concluir que las metodologías y herramientas usadas van en concordancia con el tema de estudio. Existen diversas formas de recopilar los datos y muchos métodos de analizar e interpretar la información en estudios de temáticas parecidos. Por ejemplo, Mohan (2010) y Hallmann y Breuer (2010) investigaron la imagen de destino y sus eventos deportivos, el primero utilizó el método de entrevista y encuesta, a diferencia de los posteriores que aplicaron solamente cuestionarios. También los dos estudios usaron distintos métodos para analizar los datos. Mohan aplicó un análisis de regresión para descubrir las correlaciones importantes entre los destinos y las audiencias, Hallmann y Breuer practicaron un análisis de correspondencias para catalogar las imágenes de destino y de los eventos deportivos, y además un análisis de regresión para examinar los valores significativos reflejados a través de diferentes indicadores. Podemos concluir que no existe una sola metodología apropiada para tratar cada tema. La metodología a aplicar en cada tipo de estudio depende de la decisión de los investigadores y sus objetivos de estudio.

### ***Variables***

En relación a las variables, las más usadas en los estudios anteriores son, por ejemplo, infraestructuras y elementos tangibles e intangibles de un destino y sirven para medir tanto la imagen cognitiva como la imagen afectiva. Los valores de estas variables se recogen basándose en las opiniones de turistas.. Es decir, los estudios previos suelen preguntar por los atributos que componen el destino y se completan con valoraciones generales de la visita o impresiones más globales del destino. A menudo para recoger la información no sólo se recurre a un cuestionario estructurado, sino también a una entrevista y cuestionario abierto. Se consideran como los datos para medir la imagen poseída o proyectada por los turistas o bien por las industrias del turismo (Stepchenkova y Mills, 2010). Tal y como habíamos mencionado, se considera generalmente que las IDTs están formadas por un conjunto de atributos que definen el destino (Beerli & Martín, 2004b). Si observamos los estudios de la última década, los atributos se encuentran catalogadas generalmente a través de tres características: común-estereotipada- única, funcional- psicológica y cognitiva-afectiva-

conativa. Evidentemente, según los atributos que los investigadores quieren medir, las variables se adaptan en función de los sujetos y objetos de la investigación, los motivos de estudios y los métodos de recopilar los datos.

Adoptando la dimensión de variables de Beerli y Martín (2004a), elaboramos una tabla donde se recogen las variables utilizadas por los estudios de los últimos diez años. Esta tabla no cuenta con los estudios en los que no se especifican sus variables, que se centran en algunas imágenes concretas, o bien que estudian varios destinos a través de diferentes variables (A. 2. Tabla de la revisión de las variables empleadas).

Contemplando dicha tabla, observamos que “el precio” es la que aparece con más frecuencia, seguida por variables como “la seguridad”, “los paisajes”, “los monumentos”, “los edificios”, “los museos” y “la hospitalidad local”. Pese a que los elementos determinados por Beerli y Martín son muy completos, siguen faltando algunas variables mencionadas en varios artículos como son la accesibilidad de recibir la educación en el destino (McCartney, 2008), los conocimientos que se pueden adquirir durante el viaje (Echtner y Ritchie, 2010) o bien la similitud o la relación del país origen con el destino (Chen, Chen, & Lee, 2010). También los atributos que se muestran en la lista de Beerli y Martín no son exactamente iguales ni incluyen todas variables que aparecen en otros estudios. Por ejemplo, Boo y Busser (2005) consideraron que el idioma como el dialecto es un valor único y representa el atractivo de un destino concreto a pesar de Beerli y Martín clasificaron la diferencia del idioma como un obstáculo para el turismo en la dimensión del medio social; en la dimensión del ambiente local, Beerli y Martín trabajaron con el ambiente orientado a la familia y no incluyeron otro tipo de ámbito tal y como es el ambiente orientado al adulto (McCartney, 2008), etc.

Tradicionalmente, la imagen de destino se mide a través de la escala de tipo Likert (MacKay y Couldwell, 2004). Sin embargo, últimamente, cada vez hay más investigadores que aplican distintas escalas en una misma investigación. Después de revisar varios estudios llegamos a la conclusión de que las escalas de tipo Likert de 1 a 5 y de 1 a 7 puntos siguen siendo las más usadas, seguidas de la escala bipolar. A pesar de que las escalas de tipo Likert podrían captar todo tipo de matiz hasta los mínimos detalles, autores como Ivanov, Illum y Liang (2010) manifestaron que dichas escalas no son intuitivas, sino que son escalas forzadas, es decir que obliga a los encuestados a elegir una de las opciones. Asimismo, en 2013, Dolnicar y Grün investigaron las medidas para el estudio de la imagen de destino y

descubrieron que el “forced-choice full binary”, ofreciendo la elección entre “sí” y “no”, actúa mejor en el momento de responder a los cuestionarios e interpretar los datos.

### **1.1.2.3. Conclusión de las investigaciones en torno a la imagen de destino en la última década**

Como mencionamos anteriormente, la imagen de destino es multidimensional y este aspecto puede dificultar su estudio. Los modos de adquirir datos y las técnicas analíticas varían en consonancia con la dimensión de la imagen que cada investigador quiere estudiar. Con un repaso a la literatura de la última década, llegamos a la conclusión de que la influencia y la relación de la imagen de destino sobre los visitantes son de los temas más estudiados, siendo este descubrimiento análogo a la observación de Pike (2002); la metodología más usada sigue siendo la que proporciona un el estudio de variables latentes o sintetiza información como es un análisis factorial que recoge los datos a través de cuestionarios o entrevistas, manteniendo las características que se encontraron en los artículos revisados por MacKay y Couldwell (2004) y Tasci et al. (2007).

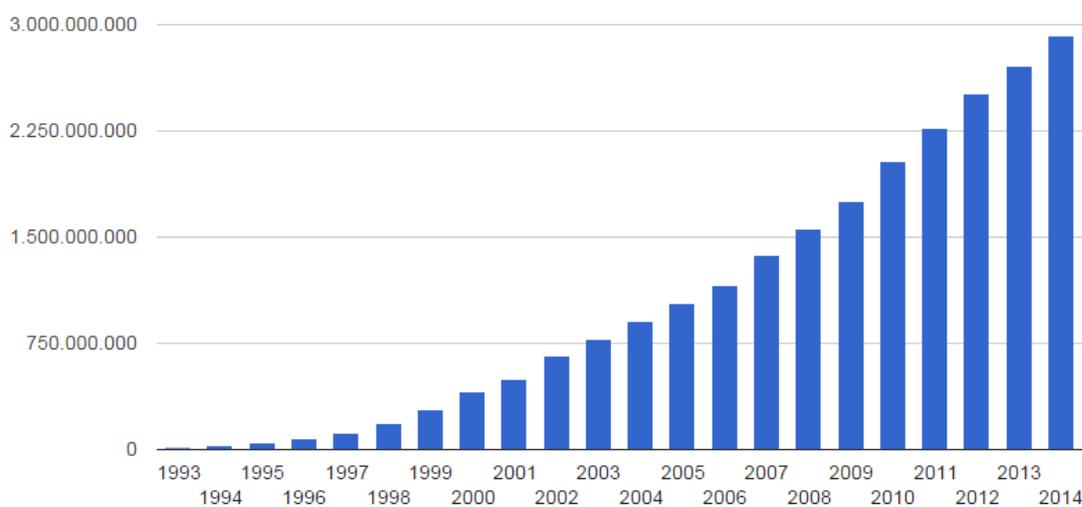
Observando los estudios de la última década, comprobamos que varias investigaciones (Frochot, 2003; Govers & Go, 2005; Ryan & Cave, 2005; Xiao & Mair, 2006; Choi, Lehto, & Morrison, 2007; Pan, MacLaurin, & Crotts, 2007; McCartney, 2008; Wang Y. , 2008; Kneesel, Baloglu, & Millar, 2010; Tasci A. D., 2007; Jani & Hwang, 2011; Lim, Chung, & Weaver, 2012; Tse & Zhang, 2013; Oliveira & Panyik, 2015, etc.) realizaron un análisis de contenido y este resultado expone lo contrario de lo que comentó Pike en su estudio (2002). Sin embargo, sólo parte de estas investigaciones estudian las opiniones y comentarios ofrecidos en las RRSS (Govers & Go, 2005; Choi, Lehto, & Morrison, 2007; Jani & Hwang, 2011; Tse & Zhang, 2013; Oliveira & Panyik, 2015; Li & Wang, 2011). Mientras que los otros estudios examinan los valores que son extraídos de cuestionarios estructurales preparados por los investigadores. Dichos valores pueden ser manipulados según el diseño de las preguntas; Ivanov et al. (2010) manifestaron que los cuestionarios pre-formalizados tuercen los datos y perjudican la investigación porque los encuestados están recordando e indicando las características y las singularidades de un destino por las repuestas pre-diseñadas. No obstante, las opiniones subconscientes, así como las conversaciones entre familias y amigos, y los textos en las RRSS son más fiables porque los sujetos ofrecen sus enjuiciamientos en un estado natural sin influencia exterior.

Tras una revisión bibliográfica de los estudios de los últimos años, nuestro tema de estudio, supone un hueco de investigación en la imagen de los destinos ya que por su novedad aún no ha sido muy atendido. Por otra parte, resulta muy recomendable la recopilación de datos a través del análisis del contenido on-line, ya que dicho método puede evitar la dificultad de encontrar los sujetos apropiados para estudiar, además del error en los datos debido al diseño del cuestionario.

## 1.2. Redes Sociales y el Turismo

Sin duda, el uso de Internet se ha generalizado en nuestra sociedad. Según ITU ICT Facts and Figures – The world in 2015 (Unión Internacional de Telecomunicación, 2015), el número de usuarios de Internet llega a 3,200 millones de personas en 2015. Con un crecimiento media interanual de tipo exponencial.

Figura 6. Crecimiento del número de usuarios de Internet



Fuente: Extraída de ITU ICT Facts and Figures – The world in 2015 (Unión Internacional de Telecomunicación, 2015)

Observamos que el número de usuarios de Internet llega a los primeros mil millones en el año 2005 y a las dos mil millones en el año 2010. Este número sigue creciendo cada año con un porcentaje notable. Respecto a las regiones de residencia de estos usuarios, Asia se sitúa en el primer puesto en comparación con otros continentes, seguida de América y Europa respectivamente. China, EEUU, India, Japón, Brasil, Rusia, Alemania, Nigeria, Reino Unido y Francia son los primeros 10 países que tienen más usuarios de Internet en 2014 (Internet Live Stats). No obstante, sitios webs como las RRSS han acompañado el crecimiento del uso de Internet.

En 2015 los usuarios de las RRSS llegan a contabilizar 2,078 millones (We Are Social Singapore, 2015). Dado que el aumento de los usuarios de las RRSS ayuda al crecimiento de contenidos en Internet, esto incrementa indirectamente el uso de Internet (Unión Internacional de Telecomunicaciones, 2014). El uso de las RRSS nos ha mostrado su potencia en el crecimiento del usuario y el tiempo que los usuarios pasan en este tipo de plataforma y siguen mostrando sus influencias en la vida cotidiana.

En este apartado, en primer lugar, estudiamos el origen de las RRSS y su desarrollo en Internet; en segundo lugar, distinguimos los tipos de RRSS y sus usos, y por último, estudiamos su impacto en el sector turístico.

### **1.2.1. Concepto de la red social**

Los estudios de la red social son muy complejos y diversos. Básicamente, las RRSS son redes de contacto, también llamadas "Networking", que facilita la relación entre personas y personas. Con un buen aprovechamiento del mecanismo de "Networking", logramos difundir la información a la gente de nuestro entorno sobre nuestra situación, necesidades, intereses, expectativas, etc. (Castañeda & Gutiérrez, 2010). John Barnes, el antropólogo que definió el concepto de Redes Sociales en 1954, acertó la red social como "un conjunto de puntos, algunos de los cuales están unidos por líneas" (citado por Martín, 2012).

Debido a esto, se representan siempre el gráfico del análisis de la red social compuesto por líneas y puntos los cuales representan la relación entre cada individuo.

A diferencia de las RRSS presenciales, las RRSS on-line sostienen la misma función pero sin límite de tiempo o de distancia gracias a las tecnologías desarrolladas hoy en día.

Para comenzar el estudio, en primer lugar, queremos concretar los términos según la descripción de Castañeda y Gutiérrez (2010) así como los conceptos de medios de comunicación social en red (Social Media), medios de seguimiento online (Lifestreaming) y las RRSS virtuales (Social Networks).

El término *Social Media* se refiere las herramientas telemáticas mediante las cuales los usuarios logran publicar los elementos concretos, también conocida hoy en día como Web 2.0. A diferencia de la Web 1.0, la Web 2.0 evalúa la forma de gestionar los contenidos de la página web. En una Web 2.0, los usuarios tienen no solamente el control de acceder a la información de su interés sino también el control de difundir el contenido respecto de ello

(Parra-López, Bulchand-Gidumalb, Gutiérrez-Tañoa, & Díaz-Armas, 2011; Alén González, Comesaña Rodríguez, & Fraiz, 2009; Guevarra Plaza, 2008; Cox, Burgess, Sellitto, & Bultjens, 2009). Los medios de comunicación social se componen de sitios virtuales donde los usuarios pueden publicar elementos de diferente naturaleza como imágenes, videos, audios, documentos, etc. simplemente a través de una cuenta de usuario y contraseña todo de forma gratuita. Los ejemplos más destacados de medio de comunicación online son Flickr para fotografías, Youtube para videos, Last.fm para música y Slideshare para presentaciones. Su OBJETIVO principal es el de compartir elementos.

*Lifestreaming*, una de las funciones incluidas en la Web 2.0, que facilita a los usuarios sincronizar las actividades en red producidas por la actualización de contenidos de un sitio web. A través de estas herramientas, los usuarios pueden poner al día las actividades de una página web, un blog, una red social o bien un perfil de un particular. Ejemplos de estas herramientas con esta característica son RSS y Plugging que son servicios de seguimiento o bien Twitter y Plurk que tienen características propias de medios de comunicación pero también con función de seguimiento. Su OBJETIVO principal es seguir informaciones.

*Social Networks* engloba a conjuntos de personas individuales o grupos pequeños dentro de una estructura formada por un conjunto de intereses parecidos y se conecta entre sí por distinta relación directa o bien indirecta. Mediante las herramientas telemáticas, los usuarios crean su propio perfil y forman parte de una red social, o bien crean su propia red de contacto. Ejemplos de este tipo de herramientas son Facebook, LinkedIn, Hi5, etc. Su OBJETIVO principal es establecer una relación con otros usuarios.

Sin embargo, destacamos la dificultad de catalogar bien todas herramientas dentro de los tres grupos que mencionamos previamente. Teniendo cuenta que con el desarrollo de la técnica de comunicación, existen muchos medios que contienen multifunciones que no solamente permiten a compartir elementos sino también permiten seguir informaciones y conectar con otros usuarios tal y como son foros y blogs. A través de blogs, los usuarios pueden ver los elementos publicados por el autor, dejar sus comentarios, mantener el dialogo con el autor y otros usuarios, suscribirse al RSS para el seguimiento de los blogs, etc. Al igual que los blogs, en los foros, los usuarios también pueden visualizar los elementos publicados por otros usuarios, dejar comentarios y ponerse en contacto con otros usuarios siempre siguiendo una serie de normativas del propio foro.

El concepto de la red social virtual es múltiple y se implica en varias áreas, según Rodríguez Ruibal (2010, págs. 3-4)

“las redes sociales son estructuras en las que puede estar representado un ser de manera individual o de forma colectiva, con una serie de líneas que se unen entre ellos. En muchos casos se pueden producir relaciones entre diferentes seres o grupos de seres de distinto tipo como intercambios financieros, comerciales, amistad, sexo, etc. y en donde existe una serie de interacción entre ellos mismos.”

En este estudio tratamos de analizar cómo los usuarios de las RRSS proyectan y reciben la imagen de destino. Por lo tanto, las RRSS que mencionamos en esta Tesis comprenden tanto a los medios de comunicación virtual como a las RRSS y sus seguimientos a la información práctica.

#### **1.2.1.1. La clasificación de las redes sociales virtual**

Tal y como mencionamos anteriormente, las RRSS son muy diversas y, por esto, existen muchas clasificaciones sobre ellas. Escogemos las aportaciones más citadas para aclarar los tipos de las RRSS.

Dependiendo de los usos principales de sus usuarios, Crespo y García (2010) clasifican las RRSS en tres grandes grupos como las redes de contacto (Facebook, LinkedIn, Tuenti, Hi5, etc.), los sitios de publicación social (SlideShare, Flickr, Youtube, etc.) y las herramientas de seguimiento (como por ejemplo, Twitter).

Martín A. V. (2012) clasifica las comunidades virtuales por las funciones de web en tres categorías: las redes generalistas, las redes profesionales y las redes especializadas.

- Las redes generalistas  
Ejemplos de las redes generalistas son Facebook, Twitter, Tuenti, etc. cuyas funciones son similar a las de foros de conversación de interacción. Estas plataformas facilitan la comunicación entre personas y el intercambio de opiniones o archivos. Los usuarios pueden crear su propio perfil y publicar su última noticia para que llegue a personas permitidas, al igual que las empresas pueden crear su perfil para publicar noticias sobre sus productos y sus ofertas, haciendo llegar la información al público

- Las redes profesionales  
Dichas redes tienen el fin de localizar cooperadores, compañeros de negocios o bien trabajadores, y sus ejemplos son Xing o LinkedIn, donde los usuarios pueden crear su perfil especificando el nivel de formación y las habilidades laborales.
- Las redes especializadas  
Por último, las redes especializadas, como son Flickr, Youtube o Last.fm, permiten a sus usuarios compartir determinados archivos como fotografías, música, películas o libros.

Según Burgueño (2009), investigador especializado en tecnología, privacidad y comercio electrónico, las RRSS se pueden clasificar en 3 ámbitos generales que mostramos a continuación.

- Por su público, objetivo y temática  
Respecto al tema y el objetivo de público, Burgueño divide las RRSSs con este aspecto en dos sectores que son redes sociales horizontales y verticales. Las primeras están dirigiendo a usuarios genéricos sin un fin definido como pueden ser usuarios de Facebook o Twitter. A diferencia de las redes horizontales, las segundas se basan en una función específica y un objetivo determinado como son Xing y LinkedIn, dirigidas a generar relaciones profesionales; Last.fm y GoodRead, por su parte, enfocadas a realizar o desarrollar las actividades de ocio.
- Por el sujeto principal de la relación  
En relación a esta clasificación, Burgueño añade tres subcategorías que son redes sociales humanas, de contenidos y de objetos. En la primera se encuentra el ejemplo más representativo en Tuenti. La función principal de este tipo de RRSS es establecer relaciones entre individuos. En la segunda observamos ejemplos como Scribd o Flickr que establecen relaciones con otros individuos a través de los contenidos publicados. En la última el autor no ofrece un ejemplo claro, simplemente explica que es un tipo de red social novedoso con función de unir marcas, automóviles y lugares entre personas.

- Por su localización geográfica

Burguño separa este tipo de red social en dos tipos: RRSS sedentarias y RRSS nómadas. Las primeras, como son por ejemplo Blogger o Plurk, funcionan mutuamente entre personas, los contenidos compartidos o los eventos creados. Como ejemplos del segundo tipo son Foursquare y Skout que se desarrollan basándose en la localización geográfica con el fin de encontrar y conocer usuarios cercanos.

Observando la línea de la clasificación, percibimos que no se encuentra del todo clara, y además observamos que una misma herramienta puede estar incluida en distintos grupos a la vez según el autor en concreto y su desigual forma de organizar las herramientas. Teniendo en cuenta de que los medios de RRSS se están desarrollando cada vez de forma más mestiza e híbrida, y que la tendencia de las RRSS será la de combinar multifunciones con multicanales al concepto de la futura Web 3.0, la clasificación de las herramientas resulta meramente orientativa (Crespo & García, 2010).

#### **1.2.1.2. Impacto de las redes sociales virtuales sobre la industria turística**

La web 2.0 también se refleja en términos tales como las RRSS, los medios sociales, el contenido generado por el usuario (UGC- *User Generated Content*), sitios de red social (SNS- *Social Networking Sites*) o medios de comunicación generados por el consumidor (CGM- *Consumer Generated Media*) (Cox, Burgess, Sellitto, & Buultjens, 2009; Lange-Faria & Elliot, 2012). Es de gran importancia, ya que propociona a los consumidores individuales acceso a la presentación, revisión y respuesta de los contenidos en línea.

Estas tecnologías de la información y comunicación han provocado un cambio tanto en las empresas como en la industria turística (Aldebert, Dang, & Longhi, 2011). En concreto, según los estudio de Albacete Sáez y Herrera Fernández (2012) las TICs han ofrecido diversas ventajas para las empresas:

- “Ha mejorado la eficiencia de las operaciones reduciendo así los costes operativos.
- Ha aumentado la posibilidad de mantener relaciones con el cliente mejorando así el servicio al cliente, la calidad del mismo y su satisfacción.
- Ha incidido positivamente en el desarrollo de diferentes redes.
- Ha incrementado y mejorado las relaciones con los proveedores.
- Ha facilitado el acceso de las empresas a mercados más globales.
- Ha desarrollado una mejor planificación del producto turístico al permitir su ajuste con horarios, precios, hacer reservas, obtener tickets y posibilitar pagos online (pág. 21).”

De un modo u otro estas ventajas han promovido el desarrollo tecnológico en el sector turístico. No obstante, junto con la evolución de las herramientas de TIC, los investigadores también mencionan algunos impactos concretos que han tenido las tecnologías sobre la industria turística.

Langheinrich y Karjoth (2010) señalan que las RRSS ofrecen: 1. Nuevas oportunidades de negocio, tanto para la adquisición de clientes como para la retención de los mismos (Gretzel, 2006), 2. Facilita la transferencia de conocimiento entre la empresa, y 3. Pueden influir positivamente en el clima de trabajo. Sin embargo, como aspecto no deseable, se puede decir que: 1. Las RRSS también pueden destruir rápidamente la imagen de una compañía que ha tardado años en construirla, 2. El uso de las RRSS podría dar lugar a una pérdida de productividad en el trabajo, y 3. Existe la posibilidad de recibir ataques procedentes de profesionales (Albacete Sáez & Herrera Fernández, 2012, págs. 20-21).

A través de la revisión de la literatura, observamos la contradicción entre el deseo del sector turístico de emplear las RRSS en el negocio y el miedo de recibir consecuencias negativas. Sin embargo, con el paso del tiempo, el uso de las RRSS por los consumidores se ha convertido en un fenómeno inevitable y representa un símbolo del cambio de etapa. Por lo tanto, afirmamos la necesidad de entender el comportamiento de los turistas en las RRSS para poder reaccionar rápido en cualquier situación inesperada.

### 1.2.2. Revisión conceptual de la red social virtual con el turismo

El uso de Internet se ha convertido en unas de las herramientas más aplicadas a la hora de realizar una búsqueda de la información turística (Lange-Faria & Elliot, 2012; Pan, MacLaurin, & Crotts, 2007; Rodríguez Ruibal, 2010; Cox, Burgess, Sellitto, & Buultjens, 2009). También se ha convertido en un medio importante de marketing para los profesionales de la industria. Esta herramienta proporciona una plataforma de comunicación para los turistas y también sus proveedores.

Tradicionalmente, los consumidores podían encontrar los productos que le interesan a través del Internet, sin embargo, los proveedores eran los que decidían qué información quieren mostrar a sus clientes; hoy en día, a través del desarrollo avanzado de la tecnología, la Web 2.0 en lugar de ser únicamente una plataforma de comercio para los vendedores, también es un lugar donde se permite a las turistas realizar una visita virtual a un destino a través de videos e imágenes que otros visitantes han publicado e intercambiado, así como la información sobre los servicios y productos de un destino concreto, además de las opiniones de otros consumidores sobre un destino turístico.

La intangibilidad y la imposibilidad de probar los servicios turísticos requieren de fuentes de información en las que el consumidor puede confiar. Para la búsqueda de esta información están disponibles diferentes fuentes, tales como los propios vendedores, expertos independientes y, cada vez más, otros clientes. El impacto de las revisiones de los clientes sobre el proceso de toma de decisiones de los turistas es muy significativo en el caso de la industria turística debido a la confianza que generan las informaciones (Gretzel, 2006; Langheinrich & Karjoth, 2010; Cox, Burgess, Sellitto, & Buultjens, 2009). La influencia interpersonal y la del boca-a-boca (Word of Mouth) se consideran como las fuentes de información más importantes con las que los consumidores cuentan a la hora de decidir sobre una compra (Litvin, Goldsmith, & Pan, 2008). Estas influencias son especialmente importantes en la industria de la hospitalidad y el turismo, cuyos productos son intangibles y difíciles de evaluar antes de su consumo. Cuando la función del boca-a-boca se convierte en digital, la naturalidad y gran escala del Internet inducen nuevas formas de capturar, analizar, interpretar y gestionar la influencia que un consumidor pueda tener sobre otros.

Sin embargo, existen estudios que demuestran una mayor fiabilidad de los turistas en los contenidos de las webs oficiales sobre la de las RRSS (Cox, Burgess, Sellitto, & Buultjens, 2009) Al contrario,, también existen estudios (Albacete Sáez & Herrera Fernández, 2012) que mencionan algunos casos de páginas fraudulentas que venden servicios que luego no son prestados, o que no son prestados tal y como se anuncian en estos sitios. Estos hechos provocan desconfianza en los consumidores, originando así la falta de actividad o consumo por parte de los mismos.

Los consumidores turísticos usan las RRSS para obtener información acerca de los viajes. La búsqueda de información implica tres fases: antes, durante y después de la realización del viaje (Albacete Sáez & Herrera Fernández, 2012). Antes de realizar un viaje, los usuarios se involucran en las RRSS para buscar información, consultar experiencias y comparar los servicios ofrecidos con la idea preconcebida que tenían. Durante el viaje, los usuarios compruban in situ la información recibida a través de las RRSS mientras actualizan la información. Después del viaje, los usuarios comparten su experiencia y dejan su comentario para otros usuarios.

Paralelamente las RRSS son una fuente de información para los gestores, empresas o destinos. Consideramos que la aparición de las RRSS es una buena oportunidad para conocer, acercarse y tener interacción con los turistas.

Finalmente, debemos reconocer que los destinos turísticos también se han visto muy beneficiados por las RRSS online. Tal y como Albacete Sáez & Herrera Fernández (2012) mencionan:

Las TICs a través de herramientas como la Web 2.0, y en concreto las redes sociales, han facilitado la creación de espacios virtuales en los que las diferentes actividades o experiencias de un lugar pueden ser agrupadas facilitando de este modo la interacción y búsqueda de información para los visitantes actuales y para los potenciales, es decir, las redes sociales pueden ser un instrumento interesante para promocionar los destinos turísticos. (pág. 32)

### 1.3. Estado de la cuestión en la investigación de redes sociales e imagen de destino

Tradicionalmente las OGDs (Organización de Gestión de Destino) transmitían y creaban la imagen positiva de su destino. Sin embargo, desde que el uso de las RRSS se hace cada vez más popular, el método de comunicación del destino se ve obligado a cambiar (Lim, Chung, & Weaver, 2012). Las RRSS ofrecen a los usuarios un lugar donde ellos pueden compartir sus experiencias y opiniones de viaje con otros usuarios (Xiang & Gretzel, 2010). De esta forma los canales no controlados por la empresa han tomado fuerza entre los medios de información y esta ya no es en solo un sentido (Stepchenkova & Zhan, 2013). Esto ha supuesto un incremento en la variabilidad de medios de comunicación y una diferenciación en la naturaleza y peso de estos.

Según la investigación de Google & IPSOS OTX Media CT (2011), el 45% de los viajeros planifican su viaje de acuerdo a comentarios y experiencias de otros. Por lo tanto los comentarios online resultan cada vez más importantes para los turistas a la hora de planificar su viaje.

A su vez, las RRSS son una fuente importante de información para las OGDs. Las opiniones que los turistas dan sobre los destinos nos permiten conocer qué imagen tienen de nosotros. Además esta información es espontánea, no dirigida por una encuesta. Así las opiniones encontradas ponen de manifiesto lo que realmente son aspectos que les preocupan a los turistas. Es decir, nos permite ver además de su valoración su peso. Por todo esto nos resulta interesante profundizar en el conocimiento de las opiniones de los turistas en las RRSS.

Los estudios sobre RRSS y destino son escasos principalmente por lo novedoso de esta línea de investigación. Uno de los primeros trabajos sobre el tema es el de Pan et al. (2007) y se trataba de averiguar la experiencia de turistas en un destino a través de blogs turísticos.

En este tipo de estudios, las fuentes de los contenidos pueden tener origen en la web oficial de OGD (Stepchenkova & Zhan, 2013), la web turística como Lonely Planet (Jani & Hwang, 2011), las RRSS como Youtube, Flickr, blogs, etc. (Pan et al., (2007), Lim et al., (2012), Stepchenkova & Zhan (2013), Tse & Zhang (2013))

La revisión bibliografía sobre la imagen online en los últimos años, pone de relieve que los estudios están aplicados a dos tipos de contenidos que son:

- el contenido textual (opiniones y comentarios) como el citado trabajo de Pan et al. (2007), o los de Jani & Hwang (2011), Lim et al. (2012), Tse & Zhang (2013) y Oliveira & Panyik (2015);
- y el contenido visual (imágenes y videos) como la investigación de Stepchenkova & Zhan (2013).

Respecto al análisis de contenido textual, las metodologías que se suelen aplicar en las investigaciones es la metodología cualitativa. En dicha metodología se suele realizar un análisis semántico como podemos ver en el trabajo de Pan et al. (2007). El análisis de contenido se basa en sus atributos y suele buscar la frecuencia, la coocurrencia, la agrupación (clustering) o temas relativos entre los atributos (Pan, MacLaurin, & Crotts, 2007; Jani & Hwang, 2011; Lim, Chung, & Weaver, 2012; Stepchenkova & Mills, 2010; Tse & Zhang, 2013; Oliveira & Panyik, 2015).

Las herramientas empleadas en la mayoría de las investigaciones son CATPAC II, Nvivo, TextSAT o ATLAS. ti., que son herramientas pre-diseñadas para el análisis de contenido off-line. Es decir, los usuarios tienen que extraer en primer lugar los textos manualmente desde páginas web y después insertarlos en dichos programas. Esto supone un trabajo muy tedioso y el tiempo en la obtención del resultado aumenta considerablemente. Además, los usuarios tienen que crear una serie de palabras clave manualmente para que los sistemas analicen y busquen dentro de los textos insertados. Por estos motivos, los consideramos como herramientas semi-manuales. Estos métodos, además, son más proclives al sesgo ya que para trabajar necesitan que el investigador introduzca las palabras de búsqueda según su criterio.

#### 1.4. El mercado chino-parlante y el uso de las RRSS

Contemplando los datos de la OMT (OMT, 2015), no es difícil encontrar países chino-parlantes como China, Singapur, Hong Kong, Taiwán o Malaysia en los primeros puestos del ranking de gasto por turismo internacional. Además, según *Digital, Social & Mobile in 2015* (We Are Social Singapore, 2015), podemos observar también que en Asia del Este es donde existen más cuentas de RRSS activas, ocupando el primer lugar de ranking, seguido por América del Norte y Sudeste de Asia. En el mismo informe se demuestra que el 64% de la población de Hong-Kong tiene cuentas de RRSS activas, el 55% en Malaysia y el 66% en Singapur. Esto nos confirma la importancia y la posibilidad de estudiar la imagen de destino de esta región a través de las RRSS.

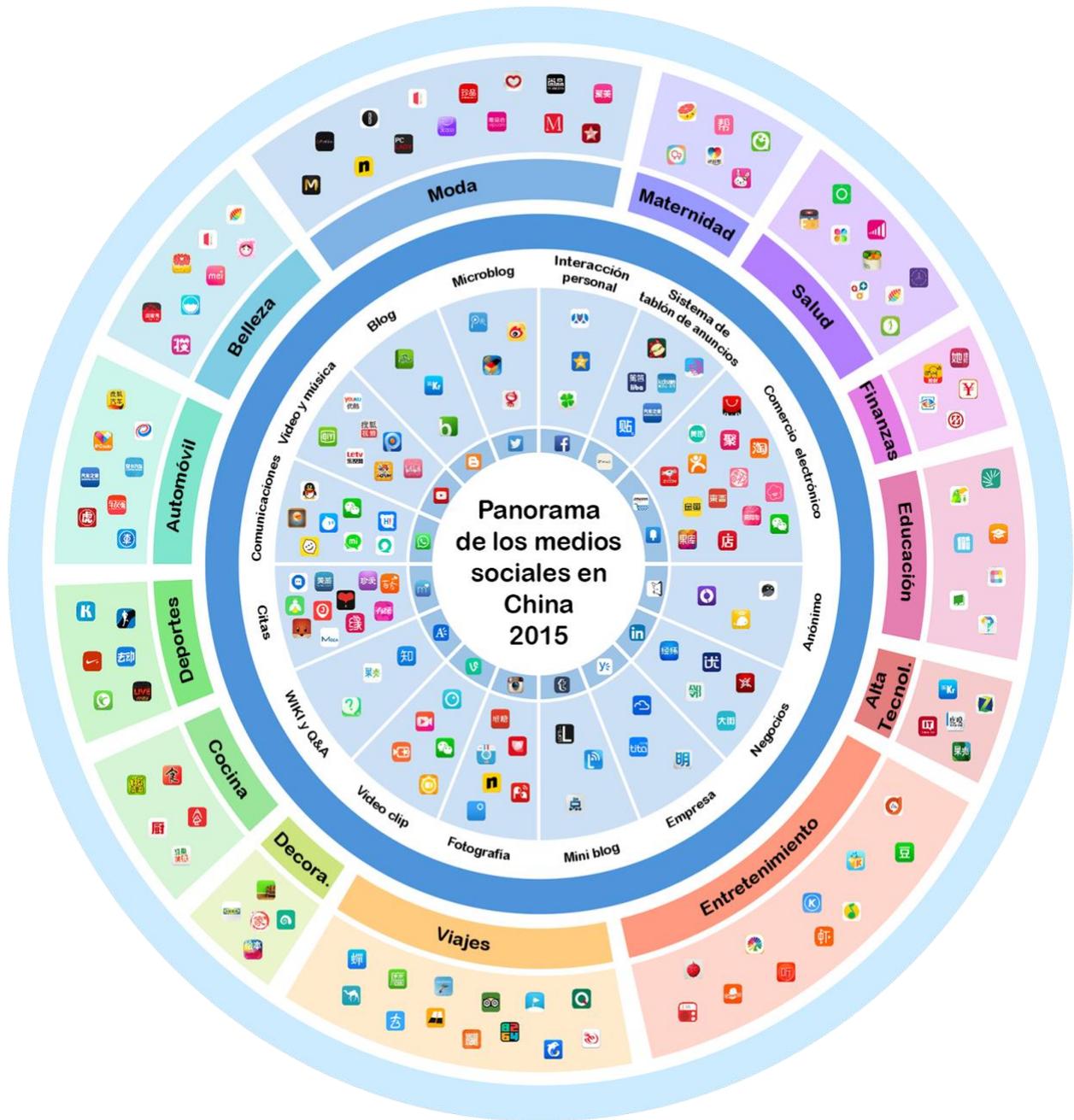
Destacamos la misma situación desde un informe estadístico nacional de la República Popular China (el mercado chino-parlante más potente según el dato de OMT en 2014) publicado en enero de 2015 por el Centro de Información de Red del Internet de China (CNNIC, 2015). Según este informe, el número de usuarios de Internet alcanzó 649 millones de personas a finales de diciembre de 2014, de los cuales 257 millones son usuarios habituales de RRSS en el primer semestre de 2014. Se trata de casi la mitad de los usuarios de Internet los que utilizan las RRSS en su vida cotidiana.

Y además en el mismo informe nos muestra que los medios sociales no se fortalecen solamente en la función de la comunicación, sino que también aumenta en su función de compartir la información. El 54.5% de los usuarios demuestran confianza en la información y mensajes que circulan por los medios sociales. La investigación reveló que el 60% de los usuarios chinos ofrecen una actitud positiva acerca de compartir información en Internet. En general, el 53.1% de usuarios sienten la necesidad de usar las RRSS, de los cuales este impulso aumenta según el nivel de educación (44.9% de primaria o estudios inferiores, 49.5% de secundaria, 54.2% de bachiller o grado, 60.2% de grado superior y 63.9% en los estudios universitarios o superiores). Este uso extenso y con estos fines hace que las RRSS sean para el caso de los chino-parlantes una fuente interesante para estudiar la opinión que de un destino tienen estos turistas o potenciales turistas.

Sin embargo, debido a la censura de determinados sitios webs en la República Popular China, los usuarios de dicho país emplean otras herramientas alternativas. Esto sin duda aumenta la dificultad de investigar los comportamientos de los usuarios de esta región, dado

que el gobierno de dicha región bloquea el acceso de algunas RRSS, como por ejemplo Facebook o Twitter, que son precisamente las más usadas al nivel mundial según el ranking de Alexa.com. A continuación, en la figura 7 se muestra la equivalencia de medios sociales empleados en China organizada por CIC (2015).

Figura 7. Panorama de los medios sociales en China



Fuente: Extraída desde [http://www.ciccorporate.com/index.php?option=com\\_content&view=article&id=1296:the-state-of-chinese-social-media-in-2015-what-you-need-to-know-six-takeaways-about-wechat-weibo-and-up-and-coming-platforms&catid=99:archives-2014&Itemid=208&lang=en](http://www.ciccorporate.com/index.php?option=com_content&view=article&id=1296:the-state-of-chinese-social-media-in-2015-what-you-need-to-know-six-takeaways-about-wechat-weibo-and-up-and-coming-platforms&catid=99:archives-2014&Itemid=208&lang=en)

En el primer anillo interior se muestran las herramientas censuradas en China. Junto a él, en un segundo anillo se encuentran las herramientas que las sustituyen en dicho país. A continuación en el anillo de fondo blanco se encuentra la clasificación de las herramientas de los dos anillos anteriores según su función principal. Por ejemplo, Renren (人人网) y Kaixin (开心网) son las dos herramientas más usadas en China que sustituyen a Facebook, siendo su función principal la de ofrecer un servicio de interacción personal (Servicio de Red Social). En el anillo exterior se encuentran las herramientas segmentadas por temas. En el caso de viaje, encontramos algunas herramientas como TripAdvisor (到到网), Qyer (穷游网), Qunar (去哪儿网), Ctrip (携程旅行网), eLong (艺龙旅行网), 8264(户外资料网), Alitrip (阿里旅行), etc. que son las herramientas de uso más frecuente en dicho país.

La creciente importancia de las RRSS como fuente de información de las opiniones turísticas y el auge del mercado chino son las dos razones que justifican este trabajo. En base a esto, surge un hueco de investigación que constituye el objetivo de este trabajo que es la de estudiar la imagen que poseen los chino-parlantes a través de las RRSS que usan.

# OBJETIVOS

## 2. OBJETIVOS

**Resumen:** En este capítulo presentamos los objetivos generales y específicos, esbozando las líneas de trabajo destacando las áreas de conocimiento requeridas para poder alcanzar los objetivos. Empezamos justificando el interés de su estudio y al final concretamos en qué consiste nuestra metodología del trabajo.

La literatura analizada en el capítulo anterior sobre la investigación de la imagen de destino y el uso de las RRSS en el turismo nos pone de manifiesto interesantes huecos de estudio que son las que han inspirado los objetivos de este trabajo.

A saber:

1. El creciente uso de las RRSS por el parte de los turistas tanto como fuente de información como para la creación de contenido.
2. La siempre necesidad de profundizar en el estudio de la imagen de destino.
3. La importancia que en el estudio de la imagen de destino tienen los sentimientos y no sólo en la valoración de atributo.
4. La falta de una metodología automática que permita trabajar con grandes cantidades de información extraída desde las RRSS.
5. La importancia de los mercados chino-parlantes.
6. La falta de una metodología para trabajar correctamente con la lengua china.

Dos son los objetivos generales que marcamos para este trabajo, uno de carácter metodológico y otro de carácter empírico. Como en todo trabajo también en el nuestro la persecución de estos objetivos generales implica plantearse objetivos más específicos o concretos que guían la Tesis hasta los objetivos generales.

Pasamos en numerar estos objetivos y tratar con detalles las razones que avalan el estudio.

## 2.1. Objetivos generales y específicos

El presente trabajo se plantea dos objetivos paralelos.

*El primero de ellos diseñar un método que, basado en la extracción automática, permita el análisis de la imagen de un destino a través de la información volcada en las RRSS. Más concretamente, el método a diseñar está pensado en los turistas chino-parlantes.*

Nuestro reto, por tanto, se inspira en varios huecos dentro del campo de la investigación: la oportunidad que brinda las RRSS como fuente de información de opiniones; la falta de métodos automáticos que puedan procesar gran cantidad de información. Lo que hace que los resultados sean representativos; pero, además nuestro trabajo en lugar de analizar los mercados occidentales tradicionales y más estudiados, damos un giro y buscamos la imagen en los mercados orientales, precisamente en los países chino-parlantes (China, Taiwán, Singapur, Malaysia, etc.).

La literatura pone de manifiesto que la imagen se compone de tres dimensiones: conativa, afectiva y cognitiva. Partiendo de esto, resulta clave para el estudio de la imagen no solo conocer la valoración de los atributos sino también los sentimientos unidos a estos.

Las RRSS nos ofrecen, en este caso, una ocasión única con muchas ventajas sobre los cuestionarios y entrevistas, puesto que son contenidos generados por usuarios en su mayoría anónimos y no están dirigidos por series de preguntas ni están supervisados a la hora de crearlos. Es por ello que esta fuente de información tiene más credibilidad que otras.

Sin embargo, para poder obtener un resultado representativo, los trabajos deben realizarse con un gran número de aportaciones y los métodos hasta ahora, al ser manual o semi-manuales, presentan una limitación para solucionar este reto.

Por otra parte, el potencial de crecimiento como emisores de esta zona geográfica y la ausencia de estudios específicos lo justifican como un tema interesante de estudio.

A pesar de las relevancias de esta información, las grandes diferencias entre el idioma chino y los idiomas occidentales provocan que las metodologías de trabajo hasta la fecha empleadas no resulten eficientes. Observando los estudios realizados en los últimos años, destacamos que la mayoría de ellos trabajan con textos en inglés. Tan solo hemos encontrado una investigación (Tse & Zhang, 2013) en la que se estudian textos en idioma chino y además

se realiza el análisis de forma manual. La razón principal de esta escasez podría ser la dificultad de interpretar el idioma. Esta limitación, proviene de la propia complejidad del lenguaje chino. También debido a que la mayoría de los programas actuales no están preparados para realizar un análisis semántico en texto chino. Es decir, los usuarios pese a poder insertar texto chino, no pueden obtener una lista con las palabras más frecuentes. La razón de esto es que dichos programas aíslan los términos a través de los espacios existentes entre unos y otros, y en un texto chino las palabras no se encuentran separadas mediante espacios.

Teniendo en cuenta el volumen de turistas, la importancia que como emisor está teniendo China y sus particularidades debido a las grandes diferencias entre la lengua china y lengua occidental, observamos la necesidad de crear un método de realizar este tipo de investigación. En este estudio, queremos crear una metodología que puede interpretar los textos en chino, analizar gran cantidad de información en un instante y además que sea automática.

De este objetivo general pueden extraerse **cinco objetivos específicos** que coinciden con los retos que deben abordarse para extraer la imagen del destino. Estos son:

1. Determinación de las RRSS relevantes como sujetos de estudio.
2. Extracción y organización de los comentarios depurados.
3. Técnica para analizar sintácticamente los comentarios en chino.
4. Diseño de una estructura de la imagen de destino para determinar los atributos influyentes en los comentarios.
5. Detección de las opiniones procedentes de los sujetos de estudio.

Como para el desarrollo de este método y para su contratación empírica necesitamos concretar un destino, se ha seleccionado España. La selección de España como destino de prueba no es casual. Como se comentará más adelante se trata de un destino líder y muy variado en su gama de productos como ya mencionamos en el apartado anterior.

Esto da pie a nuestro **segundo objetivo general**. *La contrastación empírica nos va a permitir conocer la imagen de España entre los turistas chino-parlantes.*

Este objetivo general, a su vez, implica plantearse:

1. Los atributos que más intervienen en la construcción de la imagen de España entre los chino-parlantes.
2. La identificación de los atributos mejor valorados.
3. La identificación de los atributos con valoraciones negativas.
4. A su vez, nos permite identificar si, como expone la teoría, la imagen se construye con un modelo en el que las dimensiones están interrelacionadas entre ellas, no pudiéndose interpretar cada dimensión de forma independiente.

Todos ellos suman los **cuatro objetivos específicos** del segundo objetivo general.

**MODELO Y PROPUESTA DE LA  
EXTRACCIÓN AUTOMÁTICA  
PARA EL ANÁLISIS DE LA  
IMAGEN DE DESTINO**

### 3. MODELO Y PROPUESTA DE LA EXTRACCIÓN AUTOMÁTICA PARA EL ANÁLISIS DE LA IMAGEN DE DESTINO

**Resumen:** En este capítulo presentamos el modelo propuesto para realizar las tareas que componen la Tesis, describiendo como se ha construido dicho modelo. Para ello, explicamos el motivo del diseño de cada fase y los recursos que adoptamos para cumplir el objetivo específico. Además de la presentación de cada herramienta, demostramos el uso y el resultado que podemos adquirir tras su implementación.

La metodología de estudio que se pretende llevar a cabo es innovadora. A diferencia de otros estudios que usan cuestionarios personales para adquirir la imagen general de turistas sobre un lugar en concreto, estudiamos la conversación de los viajeros a través de las RRSS. Esta fuente nos permite conseguir una imagen más real sin estar manipulada por otros factores. El individuo no se siente observado y, por lo tanto, se encuentra más predispuesto a dar su opinión de forma más sincera.

Y, por otro lado, como puede verse, la Tesis que mostramos es claramente interdisciplinar pues para llevar a cabo, se requieren varios ámbitos de conocimiento tal y como son los sistemas informáticos, el Marketing, el turismo y por último el conocimiento de la lengua china y la cultura china. En la figura 8 se puede observar el esquema de las ramas de estudio.

Figura 8. Esquema de las ramas de estudio



Fuente: Elaboración propia

### 3.1. Planeamiento del trabajo

Partiendo la idea de que esta Tesis supone un trabajo interdisciplina, profundizamos en estudios específicos y publicaciones científicas relacionadas con estos temas para elaborar el estado de la cuestión sobre imagen de destino y uso de redes sociales como fuente de información de la demanda y hábitos de consumo turísticos. Así, como estudios en estos ámbitos en los que el turista chino sea el protagonista. El estudio de estas publicaciones nos está permitiendo recopilar las variables claves, su peso y relación entre ellas y dibujar un modelo lógico de cómo se construye la imagen de los destinos en la mente de los potenciales turistas. Este será nuestro punto de arranque. Asimismo, estamos recopilando las metodologías más usadas para el tratamiento del lenguaje natural y sus puntos fuertes y débiles. A su vez la extracción de datos de las RRSS implicaba conocer su uso y las estructuras internas de trabajo de estas plataformas.

A través de la revisión bibliografía, observamos que en muchos casos los investigadores creen que para conocer las opiniones de turistas basta con extraer simplemente los adjetivos o bien adverbios de una frase. Sin embargo, según Turney (2002, pág. 418) “aunque un adjetivo aislado puede indicar subjetividad, puede haber contexto insuficiente para determinar la orientación semántica.” A continuación vemos un ejemplo en español extraído desde el foro de Los Viajeros (Manuzarc, 2014).

*“Creo que ese apartamento no vale el dinero que pagamos por él, hemos visto en otro sitio apartamentos mucho más barato que estos y mejores condiciones.”*

Aparte del error gramatical en este comentario, si nos fijamos solamente en los adjetivos y adverbios, encontramos las expresiones “más barato” y “mejores condiciones”. Literalmente, estas expresiones son positivas, en cambio, según el contexto, observamos que el usuario está criticando la calidad del alojamiento. Por lo tanto, si empleamos este método podemos llegar a obtener un resultado bastante equivocado.

Otros investigadores aplican ciertas herramientas prediseñadas, tal y como CATPAC II, Nvivo, TextSAT o ATLAS.ti., etc., para analizar los contenidos. Todas estas propuestas trabajan de forma semi-manual. Esto supone un trabajo muy tedioso y proclive al sesgo, ya que los usuarios tienen que extraer en primer lugar los textos manualmente desde las páginas web y a continuación insertarlos en dichos programas. Posteriormente, deben crear una serie de palabras clave manualmente según su criterio para que los sistemas la busquen y localicen dentro de los textos insertados. Este método de trabajar semi-manualmente es muy subjetivo y de alcance limitado, debido a que la salida de datos demuestra nada más la información requerida por los usuarios y por lo tedioso impide tratar gran cantidad de datos.

También destacamos que la mayoría de los estudios en los últimos años trabajan con textos en inglés. Tan sólo hemos encontrado una investigación (Tse & Zhang, 2013) en la que se estudian textos en idioma chino y además se realiza el análisis de forma manual. La razón principal de esta escasez podría ser la dificultad de interpretar el idioma. Esta limitación proviene de la propia complejidad del lenguaje chino y también debido a que la mayoría de los programas actuales no están preparados para realizar un análisis semántico en texto chino. Es decir, los usuarios pese a poder insertar texto chino, no pueden obtener una lista con las palabras más frecuentes. Esto se debe a que dichos programas aíslan los términos a través de los espacios existentes entre unos y otros, y en un texto chino las palabras no se encuentran separadas mediante espacios.

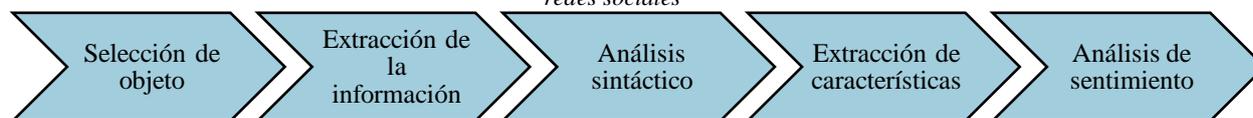
Una vez tengamos los conocimientos suficientes acerca de la teórica, metodologías de trabajo mencionadas en los estudios relativos al tema y funcionamiento de las plataformas nos centramos en la prueba y selección de las herramientas. Al final, decidimos hacer el uso del lenguaje de programación Python, ya que resulta más sencillo y flexible de programar y permite aplicarse en textos chinos y descartar la información inútil como los anuncios.

Para la consecución de los objetivos marcados no nos valen las técnicas usadas hasta ahora. Debemos diseñar una metodología más compleja que:

- Interprete bien el idioma chino.
- Realice un análisis semántico de los textos en las RRSS.
- Sirva para analizar el sentimiento y la opinión en un lenguaje natural.
- Sea rápida y permita analizar una gran cantidad de información.

Es lógico pensar que el método propuesto debe contener varias etapas puesto que son varias las tareas que deben de realizarse de secuencial para obtener la imagen del destino. En concreto las tareas serían como las que muestran en la figura 9.

Figura 9. La propuesta de la metodología para el análisis de la imagen de destino a través de la extracción automática en redes sociales



En las siguientes secciones se presenta nuestra propuesta paso a paso, tratando de definir el motivo y la tarea de cada proceso.

### 3.2. Selección de objeto de ensayo

#### 3.2.1. Objetivo de la etapa

Nuestro interés se centra en analizar los comentarios de viajeros chinos parlantes en foros turísticos on-line comúnmente usados por ellos, tanto por los usuarios de chino tradicional (usado en Taiwán, Hong-Kong y el Asia-Pacífico) como por los de chino simplificado (usado en la China continental). Como evidentemente no pueden analizarse todos los foros utilizados por los turistas, la clave se encuentra en elegir bien los foros de estudio. Para su elección se tuvo en cuenta que deberían ser los más usados y que no sean comerciales porque así garantizamos que esto refleja una opinión neutra.

Con esto se aborda **el primer objetivo específico del primer objetivo general.**

#### 3.2.2. Herramienta aplicada

Para conseguir las metas anteriores, debemos usar una herramienta web analítica. Dicha herramienta nos permite entender la información básica de un sitio web, tanto el perfil de sus usuarios como el comportamiento de estos en dicha web. Existen muchas herramientas

que se dedican a analizar los datos de sitios web, algunas de ellas están diseñadas para el uso de la gestión de página web y otras para el estudio de mercado. Para nuestro estudio, necesitamos una herramienta que nos permita conocer el ranking de la web, el tráfico que genera, el perfil de usuarios y que sea un medidor internacional. Con estas premisas usamos la web de Alexa.com para realizar esta tarea. La razón por la que elegimos esta herramienta no es sólo porque su función se adapta a nuestra necesidad (Muñoz Vera & Elósegui, 2011, pág. 166) sino también por su popularidad (Betancort García, Ocón Carreras, Galán Moreno, & Rubio Royo, 2008; Míguez, Roig, & Huertas, 2014; Placido, Interactividad usuario-usuario y redes sociales online en el sector turístico. Análisis de las páginas web turísticas oficiales de las administraciones andaluzas., 2010).

### 3.2.3. Resultado de esta etapa

Tras seleccionar los criterios de búsqueda en Alexa, el resultado nos ofreció la web Backpackers.com.tw como primera dentro del ranking de web en chino tradicional, disponiendo de un foro para intercambiar información turística de los destinos internacionales. A continuación podemos ver la categoría y sub-categoría donde encontramos esta información:

*Top Sites in: All Categories > World > Chinese Traditional TW > 區域 > 亞洲 > 台灣 > 旅遊 > 旅行資訊*

Significa: Mejores sitios en: Todas categorías> Mundo> Chino Tradicional Taiwán> Turismo> Información de turismo.

Del mismo modo encontramos nuestro objetivo en chino simplificado, Qyer.com ocupando el quinto lugar del ranking. Elegimos esta web porque dispone de un foro no comercial que las otras no poseen. En la siguiente línea, podemos encontrar el ranking de webs:

*Top Sites in: All Categories > World > Chinese Simplified CN > 休闲 > 旅游 > 旅行服务.*

Es decir, Mejores sitios en: Mundo> Chino Simplificado China> Ocio> Turismo> Servicios turísticos.

La web Alexa.com utiliza una rutina de búsqueda distinta según los países, por ello la rutina de chino tradicional es distinta a la empleada para el chino simplificado.

### 3.3.Extracción de la información

#### 3.3.1. Objetivo de la etapa

Para comenzar la extracción de datos, seleccionamos las secciones sobre España a partir de foros con los que trabajar para realizar la investigación, que son (<http://www.backpackers.com.tw/forum/forumdisplay.php?f=117>) y (<http://bbs.qyer.com/forum-18-1.html>).

Observamos que en cada uno de estos foros se emplea distinto tipo de escritura china (caracteres tradicionales o simplificados), y además cada uno de ellos dispone de una estructura y diseño web distinto como podemos ver en la figura 10 las capturas de pantalla de *Backpackers* y *Qyer*, respectivamente.

Figura 10. Capturas de pantalla de la web de Backpackers y Qyer



Además, los contenidos en ambas webs se actualizan en tiempo real al mismo ritmo que los usuarios publican nuevos comentarios. Estas son las dificultades a las que nos enfrentamos en este paso.

Con el fin de capturar la información adecuada para nuestra investigación, debemos crear un módulo apropiado cuyo objetivo consiste en adaptarse al foro en concreto. De forma

preliminar, descargamos por completo los contenidos de estas dos páginas para definir el rango de trabajo y así facilitar el trabajo en un momento concreto ya que el contenido de las RRSS está continuamente cambiando.

Con esto se aborda **el segundo objetivo específico del primer objetivo general**.

### **3.3.2. Herramienta aplicada y procesos**

Para poder extraer los contenidos sin publicidad, inspeccionamos los elementos del código HTML de ambas páginas para poder acceder a la información que nos interesa analizar. Una vez visualizados y seleccionados los contenidos concretos de los dos foros, comenzamos la extracción de los datos. Para ello, trabajamos con el lenguaje de programación Python, ya que resulta más sencillo y flexible de programar y permite su aplicación en textos chinos, descartando la información no útil para nosotros como son los anuncios. Los módulos principales que usamos en este paso son *BeautifulSoup* y *pymongo*. El uso de *Beautifulsoup* es para facilitar la exploración del contenido de una web. En cuanto a *pymongo*, lo usamos para organizar y guardar los datos que extraíamos para un uso posterior.

#### **3.3.2.1.El caso de Backpackers**

Cuando entramos en el foro de Backpackers, podemos observar que en la página se encuentran numerosos anuncios. Observamos que la información útil para este estudio se sitúa dentro de la parte central-izquierda de la página web en los contenidos principales (<body >). Sin embargo, para extraer estos contenidos, tenemos que seguir varios hipervínculos, teniendo en cuenta además que dichos contenidos están mezclados con publicidad. Por suerte, observamos que este foro dispone de una sección de archivo donde los usuarios interesados pueden consultar los comentarios. Su dominio es *backpackers.com.tw/forum/archive/index.php/f-117.html*. Al entrar en la página, observamos que el contenido se encuentra bien organizado por temas. Al inspeccionar el contenido, encontramos lo siguiente (figura 11):

Figura 11. El código HTML de la web Backpackers

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" dir="ltr" lang="zh-TW">
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
    <meta name="keywords" content>
    <meta name="description" content="西班牙旅遊討論區">
    <title>西班牙 #1</title>
    <link rel="stylesheet" type="text/css" href="http://c.bbkgz.net/forum/archive/
archive.css">
    <script async src="http://www.google-analytics.com/analytics.js"></script>
    <script type="text/javascript">...</script>
    <style type="text/css">
      img {max-width: 100%;}
    </style>
  </head>
  <body>
    <div class="pagebody">
      <div id="navbar">...</div>
      <hr>
      <div class="pda">...</div>
      <h1 style="font-size:20px">...</h1>
      <hr>
      <div id="pagenumbers">...</div>
      <hr>
      <div id="content">...</div>
      <div style="z-index:100;position:fixed;font-size: 16px;font-weight: bold;
bottom: 0px;left: 25px;margin-right: 25px;text-align:center;padding: 15px
15px;border:1px dashed;background: #B0C4DE;">...</div>
      <script type="text/javascript">...</script>
      <div id="copyright"></div>
    </div>
  </body>
</html>

```

En primer lugar, observamos que la web se encuentra escrita en chino tradicional (lang="zh-TW"), el contenido de la página es código HTML (content="text/html") y su codificación de caracteres es UTF-8 (charset=UTF-8"). Después observamos que los contenidos requeridos se encuentran en la sección <div id="content">. Entrando en el primer enlace de la lista (<a href="http://www.backpackers.com.tw/forum/archive/index.php/t-204.html">巴塞隆納市區交通</a>), descubrimos los comentarios bien ordenados por fecha de publicación, sin embargo, tan sólo aparece el nombre del autor de cada comentario en cuestión. Para conseguir más datos del autor, inspeccionamos los elementos del código HTML y encontramos un enlace donde aparece la información completa de la persona que hizo el comentario como por ejemplo, su nombre de usuario, la cantidad de texto que ha publicado, fotos de viaje, agradecimientos recibidos, la fecha de registro de su cuenta, su sexo y experiencia en el foro contabilizada a través del número de estrellas. También encontramos las veces que otros usuarios han leído dicho comentario y sus repuestas.

Después de localizar el objetivo, el paso preliminar es descargar directamente toda la información necesaria desde <http://www.backpackers.com.tw/forum/archive/index.php/f-117.html> . Este paso es imprescindible, ya que tal y como mencionamos previamente, la información en las RRSS cambia constantemente. Por ello, creamos las primeras funciones (ver Anexo bpkptest.py) con el motivo de descargar el contenido on-line completo para poder trabajar off-line continuamente. A continuación, creamos varias funciones para extraer, clasificar y almacenar los contenidos.

### 3.3.2.2.El caso de Qyer

El proceso de extracción de datos del foro Qyer es más complejo que el de Backpackers debido a la falta de una sección de comentarios archivados, la gran cantidad de información que se actualiza en cuestión de minutos y además el gran número de fotos insertadas en los comentarios en comparación con el foro anterior.

Al inspeccionar la estructura de este foro, encontramos que España como destino turístico se encuentra agrupada junto a Portugal y Andorra, debido a que comúnmente, los turistas procedentes de China suelen visitar estos tres destinos en conjunto. Junto a este grupo existen otras agrupaciones de países. En el caso de que si un usuario viaja a más que un destino, sus comentarios relacionados con varios destinos pueden ser colocados en cualquiera de los destinos en los que ha estado. Si un usuario se equivoca de sección y coloca su texto en un sitio inadecuado, los administradores recolocarán dicho texto.

Una vez que nos situamos en la sección de España, comenzamos a estudiar su estructura HTML. Como podemos ver en la pantalla capturada en la figura 12, esta web tiene el contenido parecido a la de *Backpackers* (`<meta http-equiv="Content-Type" content="text/html; charset=utf-8">`), con la diferencia de que el idioma usado es chino simplificado (`lang="zh-CN"`). En los contenidos principales, descubrimos la lista de textos (`<ul class="bbs_threadlist">`) donde se alojan los comentarios.

Figura 12. El código HTML de la web Qyer

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" lang="zh-CN">
  <head>...</head>
  <body id="forumdisplay" onkeydown="if(event.keyCode==27) return false;">
    <link href="http://v3.jiathis.com/code/css/jiathis_share.css" rel=
      "stylesheet" type="text/css">
    <iframe frameborder="0" style="position: absolute; display: none; opacity:
      0;">...</iframe>
    <div class="jiathis_style" style="position: absolute; z-index: 1000000000;
      display: none; top: 50%; left: 50%; overflow: auto;"></div>
    <div class="jiathis_style" style="position: absolute; z-index: 1000000000;
      display: none; overflow: auto;"></div>
    <iframe frameborder="0" src="http://v3.jiathis.com/code/
      jiathis_utility.html" style="display: none;">...</iframe>
    <div id="append_parent"></div>
    <div id="ajaxwaitid"></div>
    <div id="header">...</div>
    <div class="qyer_head_crumbg">...</div>
    <div class="bbs_listtop clearfix">...</div>
    <div class="bbs_sldline2"></div>
    <div class="lay_wrap">
      <div class="lay_main">
        <div class="bbs_listcbbd">...</div>
        <script>...</script>
        <dl class="bbs_subforums clearfix">...</dl>
        <div class="bbs_sldline2"></div>
        <!--列表导航-->
        <div class="bbs_maintit">...</div>
        <div id="bbs_list_div">
          <!--列表nav分类-->
          <ul class="bbs_tabnav">...</ul>
          <!--列表信息-->
          <div class="bbs_listinfos">...</div>
          <!--纯文列表-->
          <div id="threadlist" style="position: relative;">
            <form method="post" name="moderate" id="moderate" action=
              "topicadmin.php?action=moderate&fid=18&infloat=yes&nopost=yes">
              <input type="hidden" name="formhash" value="2f7674d4">
              <input type="hidden" name="listextra" value="page%3D1">
            <ul class="bbs_threadlist">...</ul>
          </div>
        </div>
      </div>
    </div>
  </body>
</html>

```

Desplegando la estructura, observamos todos los títulos de los comentarios, además de sus enlaces de la forma `<a href="thread-cifras-página.html" class="tit" title="título de post"...</a>`, a través de los cuales podemos acceder al texto completo del tema en cuestión.

Al igual que la forma de tratar la web de *Backpackers*, con las funciones que construimos con Python, descargamos completamente el contenido del siguiente enlace: <http://bbs.qyer.com/forum-18-1.html> y aplicamos otras funciones para extraer, organizar y almacenar dichos contenidos (ver Apéndice Qytest.py).

### 3.3.3. Resultado de esta etapa

En total existen 16,371 comentarios relacionados con España en el foro de *Backpackers* a fecha fin de agosto 2014. Los datos extraídos se han almacenado organizándolos en una base de datos de formato BSON a través de los siguientes campos:

- username - Nombre de usuario
- source - Fuente del comentario,
- gender - Sexo,
- title - Título del comentario,
- content - Contenido del comentario,
- thanks - Número de agradecimientos para este comentario/ Cantidad de post que el autor ha publicado en este foro,
- date - Fecha de publicación - date,
- \_id - Identificador que ha creado el sistema para cada comentario.

En el caso de *Qyer*, en total, hemos capturado 32,771 comentarios relacionados con España igualmente a fecha fin de agosto 2014. Extraemos los contenidos de dichos comentarios y los organizamos según la información establecida en la web:

- username - Nombre de usuario,
- content - Contenido del comentario,
- title - Título del comentario,
- experience - Experiencia del usuario,
- source - Fuente del comentario,
- date - Fecha de publicación el post,
- \_id - Identificador creado para cada comentario.

A continuación podemos ver un ejemplo de *Backpackers* capturado directamente desde la interfaz de Python. Este ejemplo se muestra el formato de texto chino codificado a Unicode.

- { 'username': 'Nevah',
- 'source': 'all\_bkcont/2814.html',
- 'gender': 'unknown',
- 'title': '\u56de\u8986: \u63a8\u85a6\u99ac\u5fb7\u91cc\u4f4f\u5bbfHotel  
Silken Puerta Castilla Madrid',
- 'content': [ ' <!-- -->',  
 '\u771f\u5de7\u5de7\u624d\u525b\u525b\u8ddf\u670b\u53cb\u63a8\u85a6  
 Silken Puerta Castilla\u5c31\u6709\u4eba po \u7167\u7247\u4e86\u5c01',  
 'Silken Puerta Castilla  
 \u771f\u662f\u4e00\u9593\u99ac\u5fb7\u91cc\u5f88\u4e0d\u932f\u7684\u98ef\u5e97\u5c01  
 \u4f4d\u7f6e\u5f88\u5b89\u5168\u5c01\u4ea4\u901a\u4e5f\u65b9\u4fbf\u5c01\u800c\u4e14\u5c01\u6211\u4ee5\u53ca\u6211\u7684\u670b\u53cb\u90fd\u4f4f\u904e\u5c01\u4e00\u665a\u624d 80- 100 \u6b50\u5143\u800c\u5df2\u5c01\u662f  
 4 \u661f\u7684\u5c01\u4e00\u4eba\u4e5f\u624d 40-50  
 \u6b50\u5143\u800c\u5df2\u5c01\u9084\u6709 wifi \u5462\u5c01\u6709\u4e9b  
 hostel  
 \u6c11\u5bbf\u4e5f\u4fbf\u5b9c\u4e0d\u4e86\u591a\u5c11\u5c01\u6d74\u5ba4\u7684\u885b\u751f\u6c34\u5e73\u6839\u672c\u6bd4\u4e0d\u4e0a 4  
 \u661f\u98ef\u5e97\u5c01'],
- 'thanks': [736, 472],
- 'date': datetime.datetime(2011, 12, 11, 22, 9),
- 'id': ObjectId('5406a2563480ad1b9b8289f4') }

Esta ficha nos muestra que:

1. Nevah es el autor de este comentario,
2. La página original está guardada en nuestro equipo con la ruta *all\_bkcont/2814.html*,
3. No aparece el género,

4. El título de este comentario es <sup>1</sup>回覆: 推薦馬德里住宿 *Hotel Silken Puerta Castilla Madrid*,
5. Su comentario es <sup>2</sup>
6. “真巧，才剛剛跟朋友推薦 *Silken Puerta Castilla*，就有人 po 照片了!
7. *Silken Puerta Castilla* 真是一間馬德里很不錯的飯店，位置很安全，交通也方便，而且，我以及我的朋友都住過，一晚才 80- 100 歐元而已，是 4 星的，一人也才 40-50 歐元而已，還有 wifi 呢! 有些 *hostel* 民宿也便宜不了多少，浴室的衛生水平根本比不上 4 星飯店。”
8. En total 736 usuarios agradecen el contenido de este comentario y Nevah tiene 472 comentarios publicados en este foro,
9. La hora y fecha de publicación este comentario es a las 22:09 del día 11 de diciembre de 2011, hora y fecha local de Taiwán,
10. El número de identidad que asignamos a esta ficha es *5406a2563480ad1b9b8289f4*.

De esta forma, obtenemos no solamente los comentarios, sino también los aspectos demográficos de los usuarios de foro. Conociendo la opinión y el perfil de los autores, podremos obtener más adelante conclusiones acerca del comportamiento de los mismos.

### 3.4.Realización del análisis sintáctico

#### 3.4.1. Objetivo de la etapa

Entre las características principales de la lengua china se encuentran la carencia de alfabeto y la inexistencia de espacios entre los términos. Sin embargo, la unidad básica para analizar frases en chino es la palabra (Chen & Liu, Word identification for Mandarin Chinese sentences, 1992). Dicho esto, el objetivo principal de esta etapa es identificar las palabras dentro del conjunto de caracteres que compone una frase para facilitar el trabajo del siguiente

---

<sup>1</sup> Su traducción en español es, Responder: recomienda alojamientos en Madrid, Hotel Silken Puerta Castilla de Madrid

<sup>2</sup> Su traducción en español es, "¡Qué casualidad, justo recomendando Silken Puerta Castilla a un amigo, enseguida alguien ha puesto las fotos! Silken Puerta Castilla de Madrid es realmente un hotel muy agradable, la ubicación es muy segura y fácil de acceder. Mi amigo/a (mis amigos) y yo hemos estado allí. Por noche cuesta solamente entre 80 € y 100 €. Es de 4 estrellas. Por persona sólo cuesta entre 40 € y 50 €. ¡Además hay wifi también! Algunos hostales o Bed and Breakfast tampoco salen mucho más baratos. Además la higiene del baño no es nada comparable con la de un hotel de 4 estrellas."

paso. Antes de comenzar este trabajo, debemos tener en cuenta que el idioma chino es un lenguaje muy difícil de interpretar y esto sin duda complica el trabajo. Además notamos la ausencia de signos ortográficos y errores o confusiones en la gramática en el lenguaje natural. Para ello, tenemos que acudir a herramientas específicas para la extracción de palabras de textos chinos (parsing tool).

Por otra parte es importante indicar que existen pocas herramientas para realizar el análisis sintáctico en chino con un gran corpus establecido y dentro de estas tan sólo se encuentran disponible dos de ellas de libre uso. Estas son el sistema de Conocimiento Chino y Procesamiento de la Información (CKIP, en sus siglas en inglés) desarrollada por la Academia Sinica en Taiwán y Stanford Chinese Parser por la Universidad de Stanford en Estados Unidos.

Con esto se aborda **el tercer objetivo específico del primer objetivo general**.

### **3.4.2. Selección de la herramienta**

Tal y como mencionamos anteriormente, CKIP y Stanford Parser son los dos programas más adecuados para nuestro estudio. A continuación hacemos un resumen de la definición y características elementales de cada uno, a fin de elegir entre estos la herramienta más adecuada para complementar la tarea de este paso.

#### **3.4.2.1. CKIP (Chinese Knowledge and Information Processing)**

El sistema de Conocimiento Chino y Procesamiento de la Información creado por el Instituto de la Ciencia Informática de la Academia Sinica, equivalente al CSIC en Taiwán, se dedica a identificar y clasificar palabras chinas según la oración dentro de un contexto. El desarrollo de este sistema comienza en 1992 y hoy en día se encuentra en proceso de mejora de su precisión. La entrada de datos para este sistema está limitada a textos chinos de escritura tradicional, ya que este tipo de escritura es el usado comúnmente en Taiwán.

#### **3.4.2.2. Stanford Parser**

Análogamente, Stanford Parser es un analizador sintáctico creado a finales de 2002 por la Universidad de Stanford en EEUU. Dicho programa trabaja con la estructura gramatical de las oraciones y en un principio se creó para aplicarlo en textos en inglés. Hoy en día, este programa ya puede analizar sintácticamente textos en idiomas como inglés,

italiano, búlgaro, portugués y se combina con otras herramientas para poder analizar oraciones en otros idiomas como son chino, árabe o alemán.

### 3.4.2.3. Comparación de las dos herramientas

Como hemos comentado, CKIP y Stanford Parser disponen de la función de analizar textos chinos ya que ambos programas nos permiten descomponer la frase en palabras y clasificarlas, aunque lo hacen de manera distinta.

A continuación podemos ver un ejemplo de cómo se ha trabajado el análisis de palabras para evaluar las dos herramientas.

Partamos de la siguiente frase seleccionada:

今年七月去歐洲自助時因為網路上找不到西班牙好的旅館  
決定當地去再找

Cuya traducción al español es:

“EN JULIO DE ESTE AÑO CUANDO FUI A VIAJAR DE MOCHILERO POR EUROPA, COMO EN INTERNET NO HABÍA (O HABÍAMOS) ENCONTRADO UN BUEN HOTEL, DECIDÍ (O DECIDIMOS) BUSCAR EL HOTEL CUANDO LLEGASE (O LLEGÁRAMOS) AL DESTINO.”

A continuación observamos a modo de comparación, un ejemplo procesado por CKIP y Stanford Parser, y comparamos los resultados obtenidos por cada uno de ellos.

Escogemos la frase anterior original escrita en chino tradicional:

“今年七月去歐洲自助時因為網路上找不到西班牙好的旅館  
決定當地去再找”

Insertamos los datos en el software CKIP, resultando como salida la siguiente:

今年(ND) 七月(ND) 去(VCL) 歐洲(NC) 自助(VA) 時(NG) 因為(CBB) 網路(NA) 上(NCD)  
找(VC) 不(D) 到(VCL) 西班牙(NC) 好(VH) 的(DE) 旅館(NC) 決定(VE) 當地(NC) 去  
(D) 再(D) 找(VC)

Observando la salida del texto una vez procesado, comprobamos cómo el programa ha segmentado las frases en palabras, añadiéndole además entre paréntesis el tipo de palabra (POS-tagging) a continuación de cada término. Así se indica si la palabra es de tiempo con Nd, palabra de lugar con Nc, predicado verbal de acción intransitivo con Va, posposición con Ng, conjunción subordinantes con Cbb, nombre normal con Na, palabra en explicación de posición con Ncd, los caracteres como <的. 之. 得. 地> con DE, adverbio con D (CKIP, 1993), etc (A.3. Etiqueta usado por CKIP)

Usando Stanford Chinese Parser a modo de comparación, obtenemos como resultado:

今年/NT 七月/NT 去/VV 歐洲/NN 自助/NN 時因/NN 為網/NN 路上/NN 找/VV 不/AD 到/P 西班牙/NNR 好的/NNR 旅館/NN 決定/NN 當地/NN 去/VV 再找/NN

Al igual que CKIP, Stanford Chinese Parser también puede analizar sintácticamente y segmentar las palabras desde la frase generando etiquetas (POS-tagging) para indicar el tipo de palabra que sigue a cada una, ya sea sustantivo de tiempo con NT, otros verbos con VV, sustantivo normal con NN, adverbio con AD, preposición con P, nombre propio con NR (Xia, 2000), etc (A.4. Stanford Parser Post Tag).

Si nos fijamos en la parte subrayada y en negrita de ambos resultados, descubrimos dos errores de interpretación en las salidas generados por el sistema de Stanford a la hora de segmentar las palabras como podemos ver a continuación:

- CKIP: 自助(VA) 時(Ng) vs. Stanford Chinese Parser: 自助/NN 時因/NN
- CKIP: 網路(Na) vs. Stanford Chinese Parser: 路上/NN

En el primer caso, observamos que el resultado obtenido por CKIP ha clasificado la palabra “時 (Ng -posposición)”, que significa “momento” como un modificador para el verbo “自助 (VA- verbo intransitivo)” que significa literalmente “viajar en mochilero”. El significado global sería “cuando fui a viajar de mochilero”. Sin embargo, a través del programa Stanford Chinese Parser, debido a una segmentación errónea de la oración, la palabra “自助” no ha sido detectada como un verbo sino como un sustantivo que significa “autoservicio”, con lo cual deja de tener sentido global la frase.

Este no es el único fallo. Podemos comprobar que Stanford Chinese Parser también comete un error a la hora de segmentar otra palabra como: 路上/NN- sustantivo, que significa “en la calle” en lugar de 網路 (Na - sustantivo), que significa “Internet”.

### 3.4.3. Resultado de esta etapa

Concluyendo, si adoptamos el programa Stanford Chinese Parser, perdemos el “momento” de realizar la acción “viajar de mochilero” y el lugar donde se crea la causa “en Internet”, originando de esta forma la inclusión de elementos erróneos en la base de datos. Estos errores no son aislados sino que comprobamos que se repiten con frecuencia en otros textos analizados. Por lo que nos inclinamos por CKIP como herramienta para el análisis sintáctico.

Después de decidir el programa, hacemos un módulo dedicado a importar los textos de nuestra base de datos, convertir las escrituras de chino simplificado a tradicional, conectar con la base de datos de CKIP, y devolver los resultados de nuevo a nuestra base de datos. A continuación, observamos el mismo ejemplo que usamos en la sección anterior con su resultado después de un análisis sintáctico en la tabla 1.

Tabla 1. Ejemplo de CKIP

Antes	Después
<!-- --> 真巧, 才剛剛跟朋友推薦 Silken Puerta Castilla, 就 有人po照片了! Silken Puerta Castilla 真是一間馬德里很不錯的飯店, 位置很安全, 交通也方便, 而且, 我以及我的朋友都住過, 一晚才 80-100 歐元而已, 是 4 星的, 一人也才 40-50 歐元而已, 還有 wifi 呢! 有些 hostel 民宿也便宜不了多少, 浴室的衛生水平根本比不上 4 星飯店。	<(PARENTHESISCATEGORY) !(EXCLAMATIONCATEGORY) -- (DASHCATEGORY) --(DASHCATEGORY) >(PARENTHESISCATEGORY) 真(D) 巧(VH) , (COMMATEGORY) 才(Da) 剛剛(D) 跟(P) 朋友(Na) 推薦(VC) Silken(FW) Puerta(FW) Castilla(FW) , (COMMATEGORY) 就(D) 有(V_2) 人(Na) po(FW) 照片(Na) 了(T) ! (EXCLAMATIONCATEGORY) Silken(FW) Puerta(FW) Castilla(FW) 真是(D) 一(D) 間(Ncd) 馬德里(Nc) 很(Dfa) 不錯(VH) 的(DE) 飯店(Nc) , (COMMATEGORY) 位置(Na) 很(Dfa) 安全(VH) , (COMMATEGORY) 交通(Na) 也(D) 方便(VH) , (COMMATEGORY) 而且(Cbb) , (COMMATEGORY) 我(Nh) 以及(Caa) 我(Nh) 的(DE) 朋友(Na) 都(D) 住(VCL) 過(Di) , (COMMATEGORY) 一(D) 晚(VH) 才(Da) 80(Neu) - (DASHCATEGORY) 100(Neu) 歐(Nc) 元(Nd) 而已(T) , (COMMATEGORY) 是(SHI) 4 星(Nb) 的(DE) , (COMMATEGORY) 一(D) 人(Na) 也(D) 才(Da) 40(Neu) - (DASHCATEGORY) 50(Neu) 歐(Nc) 元(Nd) 而已(T) , (COMMATEGORY) 還(D) 有(V_2) wifi(FW) 呢(T) ! (EXCLAMATIONCATEGORY) 有(V_2) 些(Nf) hostel(FW) 民宿(Nc) 也(D) 便宜(VH) 不(D) 了(Di) 多少(Neqa) , (COMMATEGORY) 浴室(Nc) 的(DE) 衛生(Na) 水平(Na) 根本(D) 比不上(VJ) 4 星(Nb) 飯店(Nc) 。(PERIODCATEGORY)

Después de esta etapa, en la base de datos creado en el caso de *Backpackers* se añade un campo denominado *parsed* para almacenar los textos analizados sintácticamente. Respecto a la base de datos de *Qyer*, añadimos los campos *cht\_contents* y *parsed* en los que se incluyen los textos traducidos a la escritura tradicional y textos analizados sintácticamente, respectivamente.

### **3.5.Extracción de características**

#### **3.5.1. Objetivo y tareas de la etapa**

A través del paso previo, conseguimos separar las palabras incluidas en los textos. En esta etapa, el primer trabajo a realizar es decidir cuáles de ellas son las palabras claves. Se trata de averiguar los atributos influyentes para la percepción de las imágenes, como por ejemplo, el transporte, el monumento o el paisaje que puede ser los atributos que componen la imagen de destino.

En cuanto tengamos la lista de palabras claves, la siguiente tarea es conseguir la información sobre con qué relaciona la gente cuando menciona estos atributos. En otras palabras, cuáles son los elementos o las características de cierto atributo, como por ejemplo, la gente suele hablar del metro cuando menciona el transporte. Entonces, el metro es una característica del transporte.

El objetivo principal es construir una estructura representativa para la imagen percibida por los turistas chino parlantes con las distintas características de un cierto atributo. Con esto se aborda **el cuarto objetivo específico del primer objetivo general**.

#### **3.5.2. Herramienta aplicada**

Se trata a priori de determinar los atributos esenciales para poder construir la estructura básica de la imagen. Para ello, integramos varias herramientas que engloban áreas de turismo y del procesamiento del lenguaje natural con el fin de establecer un mapa para el uso de la extracción de la imagen. A continuación detallamos la información de cada herramienta y las tareas que ejercemos paso a paso con cada herramienta.

### **3.5.2.1.Las dimensiones de la imagen**

Para construir los pilares de la estructura, nos guiaremos por el estudio de Beerli y Martín (2004a) sobre los factores y la dimensión de la imagen. Ellos propusieron una tabla como instrumento de medición de la imagen, revisando y comparando varios estudios de la imagen de destino turístico (A.1. Las dimensiones de imagen de destino). Esta tabla contiene nueve dimensiones con los factores detallados que afectan a la percepción de la imagen. Las nuevas dimensiones son: los recursos naturales, la infraestructura general, la infraestructura turística, ocio y entretenimiento turístico, cultura, historia y arte, los factores políticos y económicos, el entorno natural, el entorno social, el ambiente del lugar.

Cada dimensión está relacionada con un sub-grupo de atributos. Dependiendo del destino y el tema que un investigador estudia, las variables se modifican a su propósito. Adoptando las características de la imagen de destino, la multiplicidad, la relatividad y la dinámica, esta dimensión consigue superar la complicidad natural del concepto de la percepción de destino e implantar un método sofisticado para la investigación de esta área. Adaptamos esta clasificación por los factores detallados y también por el uso frecuente de esta tabla en otros estudios (Mohan, 2010).

Para nuestra investigación, escogemos las primeras ocho dimensiones ya que la última dimensión se refiere a los sentimientos (factores afectivos) que es la tarea del siguiente paso: análisis de sentimientos. Adaptamos su estructura hasta el primer sub-grupo y aplicamos un pequeño ajuste para definir los atributos. A continuación podemos ver un ejemplo de este proceso en la tabla 2.

Tabla 2. Ejemplo de la adaptación a la teoría de Martín y Beerli

El origen	Nuestra adaptación
Los recursos naturales	Los recursos naturales
Clima	Clima
Temperatura	Playas
Las precipitaciones	Riqueza de las zonas rurales
Humedad	La variedad y especialidad de flora
Horas de sol	y fauna
Playas	
Calidad del agua del mar	
Arena o playas rocosas	
Longitud de las playas	
El hacinamiento de las playas	
Riqueza de las zonas rurales	
Reservas naturales protegidas	
Lagos, montañas, desiertos, etc.	
La variedad y especialidad de flora	
y fauna "	

La razón por la que escogemos los dos niveles más altos de la tabla es para no restringir las características de los atributos y evitar los prejuicios, como por ejemplo, podemos pensar que un turista cuando habla de un hotel siempre menciona la categoría del hotel. Sin embargo, esto no siempre es cierto. Queremos construir un mapa donde los atributos y sus características nos puedan aportar mayor información posible sobre lo que piensan los usuarios.

Aparte de esto, debemos tener en cuenta que los atributos determinados y sus características en el estudio de Beerli y Martín suelen ser aplicados a cuestionarios y los encuestadores pueden interpretar las preguntas a su manera. Sin embargo, en nuestro caso, tenemos que indicarle claramente al programa las palabras claves correspondientes a los atributos para ejecutar la extracción de información. Es por ello que se hace necesario definir las palabras exactas. Para ello, elegimos el tesoro de la OMT (2001).

### 3.5.2.2.El tesoro turístico de OMT

El tesoro de turismo y ocio de la Organización Mundial del Turismo (OMT, Tesoro de Turismo y Ocio, 2001) tiene un extenso corpus que consta de 8,185 términos turísticos (suma de los descriptores y no-descriptores en francés, inglés y español) y cubre 20 campos semánticos jerarquizados en un máximo de 5 niveles que representan la actividad turística bien definidos a través de términos específicos del vocabulario turístico. Esto nos facilita obtener definiciones y sinónimos de los términos turísticos.

Los 20 campos que representan las actividades turísticas son: Actividades deportivas, Legislación turística, Ecología del turismo, Economía del turismo, Instalaciones turísticas, Movimientos turísticos, Formación y empleo, Alojamiento, Ocio, Acontecimientos turísticos, Patrimonio turístico, Política turística, Servicios turísticos, Profesionales de turismo, Promoción turística, Ciencia e información, Sociología del ocio, Turismo sectorial, Transporte, Países y agrupaciones de países. Observando los sub-grupos de estos campos, descubrimos que:

1. Existen muchas terminologías profesionales.
2. Muchos términos no están relacionados con la imagen de destino.
3. Existen muchas palabras compuestas que dificultan su traducción al chino.

Las dos primeras observaciones constituyen razones suficientes por lo que no podemos adoptar directamente la estructura de OMT y acudimos al uso de la dimensión de Beerli y Martín. La última es la causa por lo que tenemos que acudir a un recurso léxico especificado en chino.

A pesar de esto, el tesoro de la OMT nos ayuda a definir bien los atributos aparecidos en la dimensión de Beerli y Martín. A continuación observamos un ejemplo de este proceso siguiendo la tabla 3.

*Tabla 3. Ejemplo de la adaptación del tesoro de OMT*

Nuestra adaptación	Con el tesoro de OMT
Los recursos naturales	Los recursos naturales
Clima	Clima
Playas	Playas
Riqueza de las zonas rurales	Zonas rurales
La variedad y exclusividad de flora y fauna	Reserva natural
	Lago
	Montaña
	Desierto
	Flora
	Fauna

Podemos observar que palabras como “reserva natural”, “lago”, “montaña” y “desierto” han sido recuperadas después de revisar el tesoro de OMT. La razón por la que recuperamos estas palabras es porque estas palabras no se encuentran agrupadas dentro del conjunto de “zonas rurales” según el criterio de la OMT.

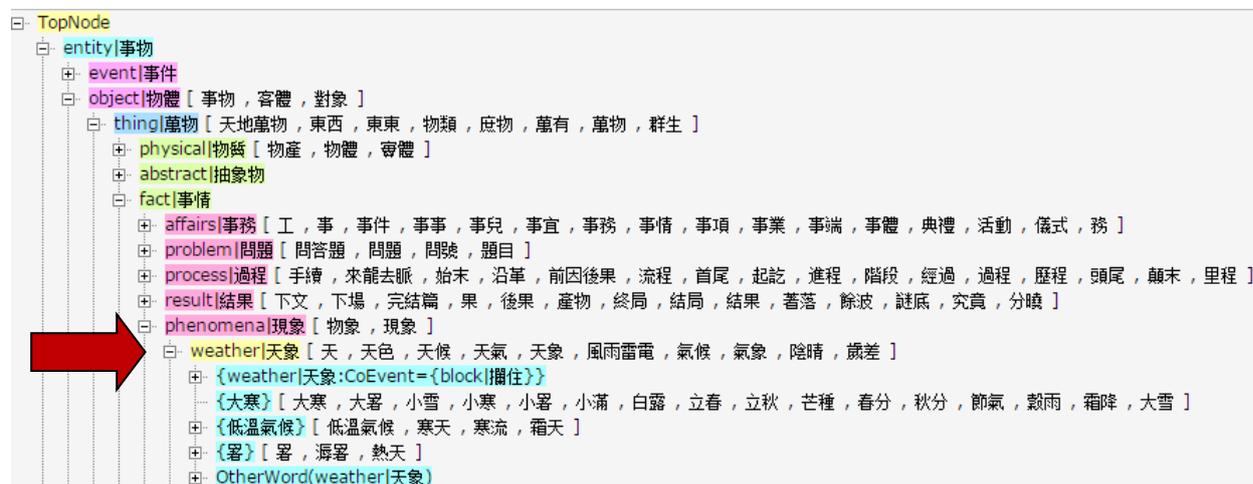
### 3.5.2.3.E-HowNet (Extended-HowNet)

Anteriormente mencionamos la dificultad de traducir los atributos que contienen palabras compuestas en chino. Asimismo, cómo obtener una traducción objetiva también es un tema importante en este paso. Para solucionar estos problemas, acudimos a la ayuda de Extended- HowNet (E-HowNet).

Basada en el módulo de HowNet (Dong & Dong, 1993), la creación del sistema E-HowNet se utiliza para la comprensión del lenguaje natural. Su desarrollo comienza en el año 2003 (Chen K.-J. , 2013). Se trata de un sistema que incluye una base de datos de léxicos chinos e ingleses con los conceptos de indexación basados en los conocimientos comunes on-line.

Este sistema no sólo nos muestra el término chino al que corresponde a nuestra búsqueda, también nos ofrece sus sinónimos y las palabras subordinadas. A continuación observamos un ejemplo del resultado cuando realizamos la búsqueda “weather” que significa clima en inglés (figura 13).

Figura 13. Ejemplo de E-HoeNet



Observamos que el sistema nos responde con la traducción exacta de “weather” en chino (天气). Aparte de esto también nos ha dado una lista de palabras que son los sinónimos de 天气, con lo cual se garantiza la objetividad de la traducción y la veracidad de los sinónimos. Seguimos con el ejemplo para mostrar el resultado de esta etapa en la tabla 4.

Tabla 4. Ejemplo de la adaptación de E-HowNet

Dimensión	Atributo	Inglés	Términos en chino
Recursos naturales	Clima	Weather	天, 天色, 天候, 天氣, 天象, 風雨雷電, 氣候, 氣象, 陰晴, 歲差
	Playas	Beach	海灘, 灘, 海邊
	Zonas rurales	Field	山野, 郊原, 郊野, 原野, 荒, 荒土, 荒地, 荒郊, 荒郊野外, 荒原, 荒漠, 草莽, 草澤, 野外, 野地, 曠野, 草野
	Reserva natural	Natural thing/ nature preserve	大自然, 天物, 天然物, 自然, 自然物, 造化/ 保護區
	Lago	Lakes	水潭, 淵, 湖, 湖泊, 潭, 灣潭
	Montaña	Mountains	山, 山地, 山岡, 山岳, 山崗, 山陵, 岑, 岡陵, 青山, 崦, 碧山, 翠微, 嶠, 嶺, 疊嶂, 陵
	Desierto	Desert	沙漠, 漠地
	Flora	Plant	一草一木, 花木, 花樹, 草木, 植物, 植物體, 植株
	Fauna	Animal	獸, 飛禽走獸, 動物, 野獸, 鳥獸, 禽獸, 鱗毛

Sin embargo, en el proceso de la búsqueda de sinónimos, aparecen términos que tienen más de un tipo de asociación semántica, es decir más de un significado. Ejemplo de esto lo tenemos con la palabra “經濟” que puede referirse a la estructura o régimen económicos de un sistema. A continuación, podemos ver la lista de sinónimos de dicha palabra, en este caso formada por un término.

[WORD('商.NA.1')]

En cambio, también podría referirse al significado de ahorrativo y sencillo. En este caso la lista de sinónimos sería la siguiente:

[WORD('秤薪而爨.VH.1'), WORD('從簡.VH.1'), WORD('節儉.VH,NV.1'), WORD('精打細算.VH.1'), WORD('儉.VH.1'), WORD('儉省.VH.1'), WORD('儉約.VH.1'), WORD('儉樸.VH.1'), WORD('質樸.VH.1'), WORD('樸拙.VH.1'), WORD('樸實.VH,NV.1'), WORD('克儉.VH.1'), WORD('因陋就簡.VB.1'), WORD('自奉甚儉.VH.1'), WORD('省儉.VH.1'), WORD('食無求飽.VH.1')]

En nuestro trabajo, buscamos el significado de la estructura económica de un sistema, es decir la primera lista. Por lo tanto, para casos como este, añadimos una función de aviso al módulo de extracción de sinónimos. Cuando el sistema detecta una palabra que tiene más de un tipo de asociación semántica nos avisará. A continuación describimos un ejemplo. En primer lugar buscamos las palabras pertenecen a la dimensión de factores políticos y económicos (FPE)

fpe\_dic = getEhowSynonymWord(FPE)

Una vez ejecutada la función de buscar los sinónimos, el sistema conectará con la aplicación E-HowNet y empieza a buscar los sinónimos de las palabras en la lista. Tras localizar los sinónimos, aparece el siguiente mensaje en la pantalla.

APPENDING VALUE ->

```
[U^U5E02\U50F9', U^U552E\U50F9', U^U6DE8\U503C', U^U8CA8\U50F9', U^U55AE\U50F9',
U^U724C\U50F9', U^U50F9\U76EE', U^U50F9\U4F4D', U^U50F9\U6B3E', U^U50F9\U78BC',
U^U50F9\U78BC\U5152', U^U6210\U4EA4\U503C', U^U6A19\U50F9', U^U8CB4\U8CE4', U^U50F9',
U^U884C\U60C5', U^U6279\U767C\U50F9', U^U5E95\U50F9', U^U7269\U50F9', U^U8A02\U50F9',
U^U503C', U^U8D77\U50F9'] TO KEY PRECIO
```

En el caso de que la palabra insertada no tiene otro sinónimo, el sistema recoge la palabra en sí-misma.

APPENDING VALUE ->

```
[U^U653F\U6CBB'] TO KEY POLÍTICO
```

En el caso de que el término insertado tenga más de un significado, el sistema nos avisará y saltará esa palabra.

\*\*\*\*\*ERROR: THE WORD 經濟 HAS MORE THAN ONE SEMANTIC TYPE \*\*\*\*\*

APPENDING VALUE ->

```
[] TO KEY ECONOMÍA
```

Por otra parte, creamos una función que muestra distintas listas de sinónimos en función de su significado. A continuación elige la lista correcta para adjuntarla a nuestra lista de palabras claves como lo que podemos ver a continuación:

```
IN [46]: FPE_DIC[U'ECONOMÍA'].APPEND(ADDVALUEWITHERROR(U'經濟'))
```

0 =====>

```
[WORD('秤薪而爨.VH.1'), WORD('從簡.VH.1'), WORD('節儉.VH,Nv.1'), WORD('精打細算.VH.1'), WORD('
儉.VH.1'), WORD('儉省.VH.1'), WORD('儉約.VH.1'), WORD('儉樸.VH.1'), WORD('質樸.VH.1'), WORD('樸
拙.VH.1'), WORD('樸實.VH,Nv.1'), WORD('克儉.VH.1'), WORD('因陋就簡.VB.1'), WORD('自奉甚
儉.VH.1'), WORD('省儉.VH.1'), WORD('食無求飽.VH.1')]
```

1 =====>

```
[WORD('商.NA.1')]
```

```
<< SELECT THE NUMBER OF TOKEN LIST >>
```

1

```
ADD 商'S WORDSET IN THE LIST
```

### 3.5.2.4.PMI

En términos generales, en lingüística, para clasificar palabras no debemos sólo basarnos en sus significados sino también se deben examinar sus coocurrencias con otras palabras (Church & Hanks, 1990). Por lo tanto, en este paso, seleccionamos la fórmula de PMI (*Point-wise Mutual Information* significa información mutua puntual) para conocer la relación de los atributos con otras palabras.

PMI es un tipo de medida empleada para asociar dos variables y se aplica frecuentemente en los estudios lingüísticos para encontrar la relación de unos atributos con sus características (Mata, 2011; Popescu & Etzioni, 2007). Los atributos anteriores se refieren a las palabras y sus características se refieren a características lingüísticas, también se llaman *feature* en el término lingüístico. Escalabilidad y simplicidad son algunas de las propiedades de PMI (Recchia & Jones, 2009). Se suele utilizar cuando se compara la probabilidad de encontrar una palabra  $x$  y una palabra  $y$  juntas (*the joint probability*) con las probabilidades de encontrar una palabra  $x$  y una palabra  $y$  de forma independiente (*chance*) (Church & Hanks, 1990; Turney, Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, 2002).

La forma de calcular este algoritmo es dividiendo el logaritmo decimal de la probabilidad existente en que las palabra  $x$  e  $y$  concurran entre la probabilidad de que las palabra  $x$  e  $y$  aparezcan de forma independiente.

Es decir, sean  $x$  e  $y$  dos términos distintos. Se denomina  $p(x)$  a la probabilidad discreta de que la palabra  $x$  aparezca. Esta posibilidad es el cociente entre el número de palabras  $x$  y número de comentarios (*posts*), calculándose como sigue:

$$p(x) = \frac{\text{número de } x}{\text{número de posts}}$$

De forma similar,  $p(y)$  es el cociente entre el número de palabras  $y$  y número de comentarios, que se calcula de la misma manera:

$$p(y) = \frac{\text{número de } y}{\text{número de posts}}$$

Se define la probabilidad conjunta ( $p(x, y)$ ) como la probabilidad existente en que la palabra  $x$  e  $y$  concurren en la misma ocasión, es decir, en el mismo comentario.

$$p(x, y) = \frac{\text{veces que concurren } x \text{ e } y}{\text{número total de posts}}$$

De esta manera, podemos calcular la información mutua puntual de la palabra  $x$  e  $y$  como se muestra a continuación:

$$PMI(x, y) = \log \frac{p(x, y)}{p(x) * p(y)} = \log \frac{p(x|y)}{p(x)} = \log \frac{p(y|x)}{p(y)}$$

En este paso, utilizamos la fórmula de PMI para conocer la relación entre las palabras y sus *features* existentes dentro de los comentarios. Cuando el valor de PMI de palabras  $x$  e  $y$  es más alto que otros, significa que tiene más relación comparado con otros. El fin de este proceso es el de conocer cuáles son las *features*, es decir, los términos más usados cuando los usuarios citan un elemento en concreto y también cuánta información podemos obtener desde un elemento en particular que nos ayuda a conseguir el mejor rendimiento posible. Como paso previo a la construcción de las características turísticas más relevantes, se genera una lista de términos aparejados. Esta lista de términos aparejados puede presentar en dos problemas:

1. En algunos caso los números de comentario son muy numerosos y/o algunos de las parejas de palabras no tienen significados, por lo tanto, es necesario aplicarse una condición especial que consiste en la limitación de la coocurrencia mínima de comentarios por pareja de palabras para mejorar el resultado. La literatura que utiliza PMI no aplica ningún criterio para reducir el número de comentarios. Para decidir el número de coocurrencia con el que quedarnos podemos hacer simulaciones para ver resultados de diferentes coocurrencias y elegir la opción que menos ruido presente.
2. El conjunto de parejas de palabras obtenidos resulta muy voluminoso, por lo que para ser operativo, pasamos a reducir el número de parejas con las que vamos a trabajar. La literatura que utiliza PMI no aplica ningún criterio para reducir el número de parejas. Para decidir el número de parejas de palabras recurrimos a técnicas cualitativas. En concreto recomendamos un estudio de Delphi con expertos en turismo y comportamiento de los turistas en redes.

A continuación se muestra un ejemplo con la palabra “氣候” que significa clima.

地中海型 氣候 HAVE PMI VALUE---> 5.3616344389

大雪 氣候 HAVE PMI VALUE---> 4.95616933079

氣候 高山 HAVE PMI VALUE---> 4.71104687275

氣候 鄉下 HAVE PMI VALUE---> 4.4125538842

氣候 鄉村 HAVE PMI VALUE---> 4.33998319136

La primera pareja significa clima mediterráneo – clima y obtiene un valor de PMI 5,3616344389 sin tener en cuenta el orden de la palabra, es decir que no contamos qué palabra aparece delante de un comentario. Las siguientes parejas de palabras son *fuerte nevada – clima*, *clima – montaña alta*, *clima – el campo*, siendo la última *clima – pueblo*.

### 3.5.3. Resultado de esta etapa

Reuniendo los resultados de los pasos anteriores, conseguimos la última versión de la estructura de imagen de destino que mostramos a continuación. En la figura 14, la imagen de destino se encuentra en el centro de la figura denominada IDT y de color verde, las 8 dimensiones de la imagen se encuentran alrededor de la IDT en color rojo y nombradas con las letras iniciales de sus nombres respectivamente. Estas son las infraestructuras generales (IG), los recursos naturales (RN), las infraestructuras turísticas (IT), ocio y entrenamiento turístico (OET), las culturas, historias y artes (CHA), los factores políticos y económicos (FPE), el entorno natural (EN) y el entorno social (ES). A continuación, partiendo de cada dimensión observamos en color azul los atributos. El número total de los atributos finales es 51. Por último, de cada uno de ellos parten las palabras claves en chino relacionadas con dichos atributos en negro.



En total existen 384 palabras claves en chino (A.5. Glosario de términos para la medición de la IDT). Cada palabra clave situada en la parte exterior del círculo ha sido introducida en el sistema y calculada con el algoritmo de PMI para localizar sus características lingüísticas. Finalmente, la palabra clave se empareja con sus 5 características lingüísticas (*features*) más relevantes para seguir el proceso de extracción de los comentarios que se encuentran relacionados con nuestros atributos. Como resultado de lo anterior, todos los comentarios que contienen las parejas de palabras han sido extraídos, separados y nombrados con el nombre de pareja, asignándole el orden de la extracción todo en formato de tipo de archivo de texto. Ejemplo de ello es el archivo nombrado como, 床單\_民宿 1.txt, donde “床單\_民宿” es la pareja de palabras, 1 es el orden de la extracción y .txt es el formato de fichero con el que archivamos este comentario. Aparte de archivar todos los comentarios que contienen las parejas de palabras, ordenamos estos archivos y los clasificamos según los atributos que poseen cada pareja de palabras. Es decir, en el caso del archivo 床單\_民宿 1.txt, la pareja “床單\_民宿” significa *ropa de cama\_albergue* y albergue, perteneciendo al atributo alojamiento, siendo por lo tanto, ropa de cama una de las características lingüísticas de albergue. Por lo tanto, guardamos el archivo 床單\_民宿 1.txt en una carpeta denominada *alojamiento*.

### 3.6. Análisis del sentimiento

#### 3.6.1. Objetivo de la etapa

Con motivo de añadir los sentimientos extraídos a partir de textos e integrar el resultado en la estructura prediseñada, tenemos que realizar una extracción de opiniones sobre las palabras claves que ya extrajimos en el paso anterior.

Para la extracción de la opinión, tenemos que identificar aquellas palabras que reflejen opiniones. Para ello, podemos adoptar algún corpus de opinión existente. No obstante, debemos tener en cuenta que para averiguar la opinión que se muestra en los comentarios, no podemos centrarnos solamente en palabras, sino que también debemos observar la construcción y la composición de las frases, así como el texto completo (Ku & Chen, 2010). Esto dificulta la interpretación ya que una frase dependiendo del modo de combinar vocabularios o la forma de estructurarse puede presentar diferentes significados. Para una correcta interpretación de las frases, tenemos que contemplar también el contexto en el que se escriben. Por lo tanto, se considera que un análisis de opinión dentro de las frases es más

ambiguo que dentro de un contexto, ya que en un contexto podemos encontrar varias palabras de opinión que nos señalan el interés del autor (Ku & Chen, 2010). Se trata de un trabajo muy complicado y requiere mucho tiempo y recursos.

Afortunadamente, existe un programa especialmente diseñado para esta función que trabaja de manera correcta con el lenguaje chino: CopeOpi (Chinese opinion extraction system for opinionated information). Este sistema logra recoger de forma correcta la valoración de sentimientos positivos o negativos que expresan los usuarios del foro en más de un 91% de los casos para el chino tradicional, alcanzando 0.68 puntos de valor-F<sub>1</sub>. Respecto al chino simplificado, consigue interpretar correctamente el 96% de los casos, obteniendo un valor-F<sub>1</sub> de 0.85. (Ku et al., 2008). Con esto se aborda **el último objetivo específico del primer objetivo general**.

### 3.6.2. Herramienta aplicada

CopeOpi es un sistema basado en la web y que se desarrolla a partir de documentos de noticias. Este sistema es capaz de extraer palabras, frases o documentos de carácter sentimental y determinar la polaridad de opinión (positiva o negativa) y la fuerza de estas opiniones. CopeOpi valora las relaciones entre los objetos basándose en el seguimiento de opiniones y la colocación de los términos dentro del texto que se está analizando. Este sistema se basa en tres factores principales que son: palabras sentimentales (sentiment words), parámetros de opinión (opinion operator) y palabras negativas (negation operator) (Ku et al., 2008).

Con este sistema trabajamos de la forma siguiente. Escogemos una pareja de palabras y extraemos todos los comentarios donde aparezca esta pareja de palabras y con ellos formamos un grupo de comentarios. Repetimos esto con cada una de las parejas de palabras seleccionadas anteriormente. Los grupos contienen los textos sin modificar su estructura. Por último, cada grupo se exporta a un archivo para poder calcular el valor de los sentimientos según el contexto.

CopeOpi genera un número de valores para cada comentario analizado por cada frase como podemos ver a continuación:

今年/0 七月/0 去/0.0 歐洲/0 自助/0.4579737 時/0 因為/0.0 網路/0.0 上/0 找/0.0 不/0.0 到/0.0 西班牙  
/0 好/0.0 的/0.0 旅館/0 決定/-0.47500015 當地/0 去/0.0 再/0.0 找/0.0

今年七月去歐洲自助時因為網路上找不到西班牙好的旅館決定當地去再找

\*\*\*SCORE=-1.0887929

有/0.0 人/0.0 推薦/0.019754000000000032 一/0 家/0 在/0 小間/0 的/0.0 INGEL'S/0 百貨/0.0 對面/0 巷  
子/0.0 的/0.0 VICTORIA/0 HOSTAL/0 有/0.0 進去/0.0 看看/0.0 還/0.0 OK/0 但/0.0 已經  
/0.029755600000000017 額滿/0.0

有人推薦一家在小間的INGEL'S百貨對面巷子的VICTORIAHOSTAL有進去看看還OK但已經額滿

\*\*\*SCORE=-0.3102905999999999

後來/0 在/0 附近/0 找/0.0 了/0.0 一/0 個/0 下午/0 發現/0.111519899999999999 一/0 家/0 位於/0.0 小  
間/0 INGEL'S/0 旁邊/0 巷子/0.0 超/0.0 不/0.0 顯眼/0.0 的/0.0 旅館/0 沒有/-0.0921865 招牌/0.0 是/0 因  
為/0.0 他/0 陽台/0.0 掛/0.0 著/0.0 多/0 國/0 國旗/0.0 才/0.0 注意到/0.0

後來在附近找了一個下午發現一家位於小間 INGEL'S 旁邊巷子超不顯眼的旅館沒有招牌是因為他陽台掛著多國國旗才注意到

\*\*\*SCORE=-0.5248398

小小/0.0 間/0 類似/0.0 民宿/0 的/0.0 感覺/0.5281246

小小間類似民宿的感覺

\*\*\*SCORE=-0.18378259999999999

衛浴/0.0 共用/0.0 房/0 內/0 有/0.0 洗臉盆/0.0 單人房/0.0 2 5/0 雙人房/0.0 4 0/0 歐/0 元/0

衛浴共用房內有洗臉盆單人房2 5 雙人房4 0 歐元

\*\*\*SCORE=0.0

Observamos que la salida del sistema coloca un número detrás de cada palabra. Las palabras sentimentales se encuentran valoradas con un número que representa la fuerza de opinión, es decir, una opinión positiva con número positivo o una opinión negativa con el número negativo que se expresa con el signo “-” delante. Al final de cada frase, obtenemos la puntuación total de la opinión basada en la oración.

El sistema también nos permite ver una valoración global del texto a través de un archivo CSV (comma-separated values) como podemos ver a continuación y relacionado con el ejemplo anterior:

0001,-2.2380587,Negative

La primera cifra representa el número ordenado de textos insertados seguida de la valoración global del contexto. El último dato es la opinión global del comentario, como podemos ver *Negative* en nuestro ejemplo. Esta valoración recoge los valores de cada palabra de opinión corregida gracias al contexto y según la colocación de cada palabra. La longitud del texto también afecta a su valoración. Un texto extenso tiende a amortiguar la valoración, ya sea positiva o negativa, sin embargo, un texto corto acentúa la valoración exagerando la positividad o negatividad del sentimiento.

### 3.6.3. Resultado de esta etapa

Insertando los comentarios que extrajimos en la etapa anterior, podemos obtener la opinión global para las parejas de palabras de un atributo concreto. A continuación observamos una parte del resultado de alojamiento a partir de los comentarios del foro Backpackers.

Tabla 5. Ejemplo del resultado de CopeOpi por el atributo

Pair	Positive	Neutral	Negative
(bkall_contentalojamiento - 旅館_橋樑)	3	0	0
(bkall_contentalojamiento - 旅館_農場)	6	0	0
(bkall_contentalojamiento - 旅社_街)	6	0	3
(bkall_contentalojamiento - 客棧_賓館)	3	0	0
(bkall_contentalojamiento - 店_酒店)	74	6	48

En la primera columna se encuentran las parejas de palabras que son 旅館\_橋樑 (hotel\_puente), 旅館\_農場 (hotel\_granja), 旅社\_街 (hotel\_calle), 客棧\_賓館 (posada\_albergue), 店\_酒店 (tienda\_hotel), etc. La segunda, tercera y cuarta columna indican el número de comentarios donde se incluyen cada pareja de palabras con valoración positiva, neutral y negativa respectivamente. Esta información nos ayuda a conocer cuáles son las palabras relacionadas con el atributo y con más presencia en los comentarios. Observando como, por ejemplo, la pareja de palabras, “店\_酒店” relacionada con el alojamiento, parece en mayor número de comentarios, ya sean positivos, neutrales o negativos.

CopeOpi también nos proporciona las puntuaciones de los comentarios individuales dentro de un atributo concreto. A continuación en la tabla 6 podemos observar un ejemplo con la pareja de palabras “旅社\_街” que significa en español hotel\_calle.

Tabla 6. Ejemplo del resultado de CopeOpi por las parejas de palabras

旅社_街6.txt	-3,416349817	Negative
旅社_街4.txt	2,058084	Positive
旅社_街7.txt	-3,112098617	Negative
旅社_街3.txt	7,695136792	Positive
旅社_街2.txt	16,50536367	Positive
旅社_街5.txt	-3,849431367	Negative
旅社_街9.txt	0,116040833	Positive
旅社_街8.txt	21,7656971	Positive
旅社_街1.txt	21,24169608	Positive

Observando el resultado que aparece en la tabla superior, observamos que el comentario archivado con el nombre *旅社\_街5.txt* obtiene la peor puntuación de todos ellos. Si queremos estudiar la razón por la que tiene una puntuación tan negativa, CopeOpi también nos ofrece el resultado individual por comentario como podemos ver en el ejemplo a continuación.

</0.0!/0.0--/0.0--/0.0>/0.0

<!---->

\*\*\*SCORE=0.0

我/0.0 只/0.0 回答/0.0 巴塞隆納/0.0 住宿/0.0 部份/0.0 。/0.0

我只回答巴塞隆納住宿部份。

\*\*\*SCORE=-0.6167512

巴塞隆納/0.0 分成/0.0 10/0.0 個/0.0 區/0.0 ，/0.0 基本/0.05 上/0.0 ，/0.0 是/0.0 北邊/0.0 的/0.0 區  
/0.0 比較/0.0 好/0.0 ，/0.0 大/0.0 至/0.0 來/0.0 說/0.0 ，/0.0 LES/0.0 CORTS/0.0 、/0.0 SARRI/0.0 到/0.0 -  
/0.0 SANT/0.0 GERVASI/0.0 、/0.0 L'EIXAMPLE/0.0 都/0.0 是/0.0 很/0.0 安全/0.6465813 的/0.0 區/0.0 ，  
/0.0

巴塞隆納分成 10 個區，基本上，是北邊的區比較好，大至來說，LESCORTS、SARRI-SANTGERVASI、L'EIXAMPLE 都是很安全的區，

\*\*\*SCORE=-0.2625

CIUTAT/0.0 VELLA/0.0 這/0.0 一/0.0 區/0.0 太/0.0 雜/0.0 太/0.0 亂/0.0 ，/0.0 雖然/0.0 移民/0.0 不/0.0  
都/0.0 一定/0.27317919999999996 是/0.0 小偷/0.7059536 ，/0.0 但是/0.0 ，/0.0 良/0.0 莠/0.0 不/0.0  
其/0.0 ，/0.0 不/0.0 建議/0.0 議/0.0 遊客/0.0 住/0.0 在/0.0 這裡/0.0 。/0.0

CIUTATVELLA 這一區太雜太亂，雖然移民不都一定是小偷，但是，良莠不其，不建議議遊客住在這裡。

\*\*\*SCORE=-0.9791327999999999

原文/0.0 載/0.0 於/0.0 :/0.0 背包/0.0 客棧/0.0 自助/0.4579737 旅行/0.0 論壇/0.0

原文載於: 背包客棧自助旅行論壇

\*\*\*SCORE=0.1079737

</0.0!/0.0--/0.0--/0.0>/0.0

<!---->

\*\*\*SCORE=0.0

很/0.0 可惜/0.18758610000000006 的/0.0 ， /0.0 CIUTAT/0.0 VELLA/0.0 是/0.0 許多/0.4874240000000001 遊客/0.0 在/0.0 不/0.0 知情/0.0 的/0.0 情況/0.0 下/0.0 選擇/0.0210186 的/0.0 住宿/0.0 地點/0.0 ， /0.0 因為/0.0 ， /0.0 CIUTAT/0.0 VELLA/0.0 屬於/0.262271 舊/0.0 城區/0.0 ， /0.0 最/0.0 熱鬧/0.19365780000000002 （/0.0 小偷/0.7059536 也/0.0 最多/0.38002240000000004 ）/0.0 的/0.0 RAMBLA/0.0 就/0.0 在/0.0 CIUTAT/0.0 VELLA/0.0 這/0.0 一/0.0 區/0.0 ， /0.0 而/0.0 旅館/0.0 、 /0.0 民宿/0.0 或/0.0 青年/0.0 旅社/0.0 最愛/0.5801090999999999 用/0.0 DOWNTOWN/0.0 或是/0.0 IN/0.0 THE/0.0 HEART/0.0 OF/0.0 BARCELONA/0.0 來/0.0 形容/0.0 此/0.0 區/0.0 ， /0.0 因此/0.0 ， /0.0 遊客/0.0 在/0.0 不/0.0 知情/0.0 的/0.0 情形/0.0 下/0.0 住到/0.0 最/0.0 亂/0.0 、 /0.0 最/0.0 不/0.0 安全/0.6465813 的/0.0 地方/0.0 來/0.0 。 /0.0

很可惜的，CIUTATVELLA 是許多遊客在不知情的情況下選擇的住宿地點，因為，CIUTATVELLA 屬於舊城區，最熱鬧（小偷也最多）的RAMBLA 就在 CIUTATVELLA 這一區，而旅館、民宿或青年旅社最愛用 DOWNTOWN 或是 INTHEHEARTOFBARCELONA 來形容此區，因此，遊客在不知情的情形下住到最亂、最不安全的地方來。

\*\*\*SCORE=-1.726588

其實/0.38113745000000004 ， /0.0 巴塞隆納/0.0 的/0.0 景點/0.0 沒有/0.0921865 通通/0.0 集中/0.011056000000000045 在一起/0.08611013333333332 ， /0.0 無論/0.31795875 住/0.0 在/0.0 哪裡/0.0 ， /0.0 只要/0.14611355000000004 有/0.0 地鐵/0.0 就/0.0 很/0.0 方便/0.17611159999999998 ， /0.0 根本/0.022774949999999985 不/0.0 需要/0.0 住/0.0 在/0.0 CIUTAT/0.0 VELLA/0.0 這/0.0 一/0.0 區/0.0 。 /0.0

其實，巴塞隆納的景點沒有通通集中在一起，無論住在哪裡，只要有地鐵就很方便，根本不需要住在 CIUTATVELLA 這一區。

\*\*\*SCORE=0.12518303333333333

千萬/0.028981400000000004 記得/0.2896336 ， /0.0 CIUTAT/0.0 VELLA/0.0 包括/0.05 EL/0.0 RAVAL/0.0 、 /0.0 EL/0.0 G/0.0 ò/0.0 TIC/0.0 、 /0.0 LA/0.0 BARCELONETA/0.0 、 /0.0 SANT/0.0 PERE/0.0 、 /0.0 SANTA/0.0 CATERINA/0.0 I/0.0 LA/0.0 RIBERA/0.0 這/0.0 五/0.0 區/0.0 ， /0.0 都/0.0 是/0.0 「/0.0 紅色/0.0 警戒區/0.0 」/0.0 ， /0.0 街道/0.0 很/0.0 小/0.0 ， /0.0 彎彎曲曲/0.0 的/0.0 ， /0.0 小偷/0.7059536 下手/0.44084225 後/0.0 逃到/0.0 巷子/0.0 裡/0.0 就/0.0 可以/0.15517970000000003 脫身/0.0 ， /0.0 在/0.0 天黑/0.0 後/0.0 要/0.0 盡量/0.13696245 避免/0.05408154999999998 CIUTAT/0.0 VELLA/0.0 這/0.0 一/0.0 區/0.0 。 /0.0

千萬記得，CIUTATVELLA 包括 ELRAVAL、ELGÒTIC、LABARCELONETA、SANTPERE、SANTACATERINAILARIBERA 這五區，都是「紅色警戒區」，街道很小，彎彎曲曲的，小偷下手後逃到巷子裡就可以脫身，在天黑後要盡量避免 CIUTATVELLA 這一區。

\*\*\*SCORE=-0.6038016999999998

CIUTAT VELLA 這一區的地鐵站包括：LICEU、DRASSANES、BARCELONAETA、JAUME I 這幾個，另外，URQUINAONA 地鐵站以南、PARAL

CIUTATVELLA 這一區的地鐵站包括：LICEU、DRASSANES、BARCELONAETA、JAUME I 這幾個，另外，URQUINAONA 地鐵站以南、PARAL

\*\*\*SCORE=0.05

LEL 地鐵站以西、SANTANTONI 地鐵站以西也屬於 CIUTATVELLA 這一區，所以，建議大家盡量避免住在以上那幾個地鐵附近。

LEL 地鐵站以西、SANTANTONI 地鐵站以西也屬於 CIUTATVELLA 這一區，所以，建議大家盡量避免住在以上那幾個地鐵附近。

\*\*\*SCORE=0.4362292

首先要進去巴塞隆納市政府的官方網頁：

\*\*\*SCORE=-0.2309721

[\\_EN.ASPX#X%3D28364%...%26W%3D1860%26H%3D1178](#)

\*\*\*SCORE=0.0

然後在左欄填上住宿旅館的地址，例如 MALLORCA 300

然後在左欄填上住宿旅館的地址，例如 MALLORCA300

\*\*\*SCORE=-0.30425120000000005

再/0.0 利用/0.49929475 左/0.0 上方/0.0 的/0.0 「/0.0 -/0.0 」/0.0 縮小/0.6559535999999999 鍵/0.0 ，/0.0 就/0.0 可以/0.15517970000000003 查出/0.0 是/0.0 在/0.0 哪/0.0 一/0.0 區/0.0 了/0.0 ！/0.0 再/0.0 利用/0.0 左/0.0 上/0.0 方/0.0 的/0.0 「/0.0 -/0.0 」/0.0 縮小/0.6559535999999999 鍵/0.0 ，/0.0 就/0.0 可以/0.15517970000000003 查出/0.0 是/0.0 在/0.0 哪/0.0 一/0.0 區/0.0 了/0.0 ！/0.0

\*\*\*SCORE=0.15517970000000003

Traduciendo y resumiendo este comentario, observamos que el autor indica las zonas norte Les Corts, Sarrià-Sant Gervasi y L'Eixample como zonas más seguras que Ciutat Vella porque en dicha zona se dan muchos casos de robos siendo también la zona más concurrida por turistas. Resaltando que muchos hoteles publicitan su alojamiento por su localización central en la calle Rambla, el autor recomienda firmemente de que no se alojen en aquella zona por temas de robos sobre todo en los callejones de las zonas El Raval, El Gòtic, La

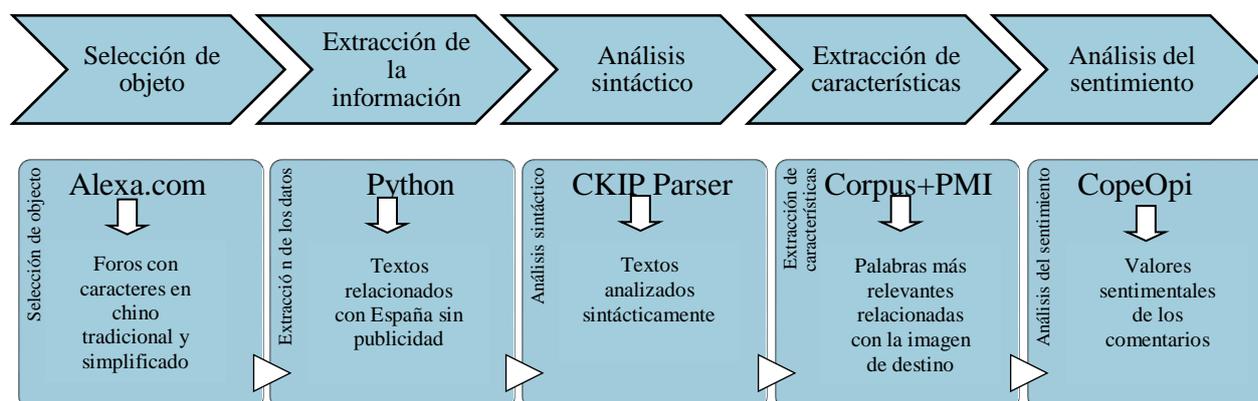
Barceloneta, Sant Pere y Santa Caterinaila Ribera donde los ladrones pueden robar y salir huyendo fácilmente. También cita otros sitios peligrosos en el centro de Barcelona.

Tras leer este comentario, comprendemos que la falta de seguridad supone la razón principal por la que este comentario resulta negativo y provoca por tanto la puntuación negativa de la pareja *hotel\_calle*. A través de este método, podemos confirmar si los resultados de las valoraciones son correctos.

### 3.7.Resultado obtenido con el modelo de extracción propuesto

Esta metodología no representa simplemente la propuesta de una técnica sino de una rutina de trabajo a través de distintas etapas. Es decir, de un método o forma de trabajar. Partiendo de un flujo de trabajo, observamos un resumen del resultado obtenido en cada paso según se indica en la figura 15:

Figura 15. Flujo del trabajo



La parte superior del gráfico nos muestra el motivo de cada paso. En la zona intermedia podemos ver el nombre de la herramienta que se aplica en cada paso y a continuación, en la parte inferior sus resultados.

Nuestra propuesta, se trata de un flujo de trabajo novedoso en varios aspectos. Primero, la forma de estructurar el mapa de la imagen se basa en los comentarios de los foros. Segundo, se propone una metodología automática lo que permite respuestas ilimitadas y menos posibilidades de sesgo. Esto cumple nuestro objetivo de obtener respuestas directas de los turistas y de conocer qué es lo que realmente les importa cuando viajan en España con representatividad en los resultados.

Cada una de estas fases está diseñada para abordar un objetivo concreto que supone un escalón para alcanzar el objetivo final. Debemos mencionar que el diseño de las actividades de cada fase no ha sido fácil y ha topado con hitos o problemas que tuvieron que abordarse para obtener finalmente una propuesta metodológica congruente y consistente.

Este método de trabajo nos ahorra mucho tiempo en la visualización de cada comentario y la clasificación de palabras pero a la vez, es flexible y, si queremos conocer más detalle, tan solo necesitamos contemplar los comentarios que nos interesan (que sea con opinión positiva o negativa), en lugar de todos los comentarios relacionados con estas parejas de palabras. Incluso, podemos estudiar la nota individual de cada comentario para conocer los puntos fuertes y débiles de nuestros atributos. Así, los investigadores con el modelo propuesto sólo necesitan estudiar el resultado del último paso para obtener una imagen completa, en nuestro caso sobre España.

# **CONTRASTACIÓN EMPÍRICA DEL MODELO PROPUESTO**

## 4. CONTRASTACIÓN EMPÍRICA DEL MODELO PROPUESTO

3. **Resumen:** En este capítulo resumimos los resultados obtenidos tras implementar el modelo prediseñado. A partir de los resultados, presentamos la imagen de destino principal que poseen los turistas chino-parlantes sobre España, las diferentes imágenes proyectadas por los usuarios de distintas redes sociales. También estudiamos los componentes principales y los comentarios positivos o bien negativos de cada atributo para conocer los factores principales de los mismos y las causas que provocan estas opiniones. Al final de este capítulo, ofrecemos una evaluación del modelo propuesto para medir la efectividad de este método del trabajo y concluimos la imagen general que poseen estos turistas sobre España.

Después de diseñar el modelo propuesto para la extracción automática de la imagen del destino en redes sociales chino-parlantes, necesitamos contrastar este modelo sobre un caso real. Esta contrastación nos permite ver la funcionalidad del modelo propuesto. Para ello, hemos seleccionado el destino España. Las razones que lo hacen interesante son: que tiene una gama de productos muy variados entre los que se incluye el turismo cultural y de compras que son dos de las motivaciones más solicitadas por el turista chino-parlante (Baidewei, 2014) y que se trata de un destino líder y consolidado.

Con ello, contribuimos a los dos objetivos generales planteados. Uno metodológico pues comprobamos que el método diseñado consigue el objetivo de utilidad como herramienta para el estudio de la imagen en redes sociales chino-parlantes. El otro, de carácter empírico, el propio análisis de la imagen de España en las RRSS chino-parlantes.

#### 4.1.Planteamiento del trabajo

Como hemos comentado antes, el modelo propuesto es aplicable para cualquier destino pero siempre necesitamos ajustar el modelo para adaptarlo al caso concreto. Necesitamos repasar los atributos de la imagen de destino por si hablamos de un destino muy específico y, sobre todo, fijarnos en la calidad de las parejas de palabras obtenidas y la cantidad de coocurrencia asociada a cada parejas de palabras. España como el destino del estudio no se trata de un destino muy específico por lo que no modificamos los atributos con lo que trabajamos. Pero si debemos trabajar en la calidad de las parejas de palabras obtenidas y la cantidad de coocurrencia.

Como en el caso de España los números de comentario son muy numerosos y/o algunos de las parejas de palabras no tienen significados, necesitamos limitar la frecuencia de coocurrencia de parejas de palabras. Por lo tanto, a los resultados obtenidos se aplica una condición especial que consiste en la limitación de la coocurrencia mínima. De este modo, cuando el número de la coocurrencia es uno significa que todas las parejas de palabras deberían aparecer como mínimo en un comentario. Los resultados obtenidos en la Tesis es el fruto de aplicar 4 coocurrencias, es decir, cada pareja de palabras debería aparecer como mínimo en 4 comentarios si no omitimos el resultado de dicha parejas que no cumple la condición. La razón por la que decidimos 4 coocurrencias no es infundada sino es el resultado de otro experimento. A continuación explicamos cómo determinamos el número de coocurrencia.

En el paso de calcular las características más relevantes de los atributos, hacemos un cálculo de PMI estableciendo con límites de coocurrencia entre una y cinco. Visualizando los resultados obtenidos teniendo en cuenta los anteriores límites, descubrimos que en el caso de una coocurrencia aparecen parejas de palabras que contienen números insignificantes, como por ejemplo “19.3\_節, 89.2\_信息, 15.9\_港口” que significan “19.3\_festival, 89.2\_mensaje y 15.9\_puerto” respectivamente. Observamos que estos números no se parecen a ningún número de teléfono, ni fechas tratándose por lo tanto de ruido. Sin embargo, conforme vamos aumentando el número de coocurrencias hasta cinco, observamos que va disminuyendo paulatinamente la aparición de los ruidos que mencionamos anteriormente. No obstante, descubrimos un nuevo tipo de ruido, como por ejemplo parejas de palabras que contienen pronombres personales. Estas palabras son “你, 我, 他, 你們, 我們, 他們” que significan “tú,

yo, él o ella, vosotros o vosotras, nosotros o nosotras, y ellos o ellas” respectivamente. Esto es consecuencia de aumentar demasiado la coocurrencia. Teniendo en cuenta que los pronombres personales representan las palabras que aparecen con más frecuencia en todo tipo de textos, los investigadores consideran a estos como ruido siendo no admitidos como variables (Turney, 2002).

Teniendo en cuenta los ensayos anteriores, y habiendo experimentado con un número de coocurrencias entre una y cinco, finalmente nos quedamos con cuatro coocurrencias dado que su resultado es lo más aceptable. Tras aplicar esta condición con la metodología creada, obtuvimos los resultados individuales para cada atributo y un resultado conjunto de la dimensión (A.6. Parejas de palabras obtenidas con sus respectivos valores de PMI).

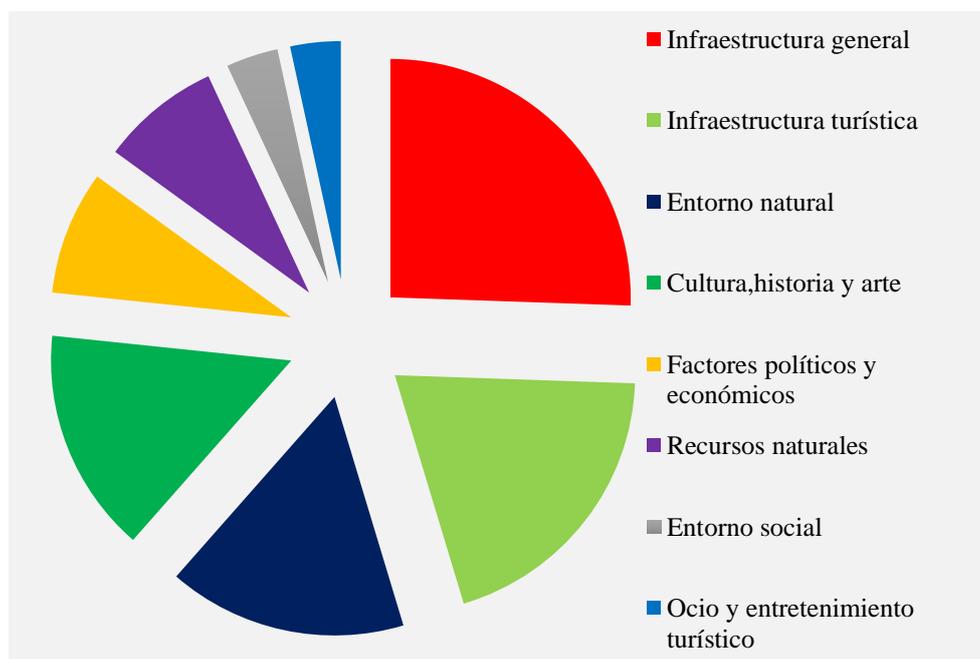
En el caso de España también el conjunto de parejas de palabras obtenidos resulta muy voluminoso, por lo que para ser operativo, pasamos a reducir el número de parejas con las que vamos a trabajar. La literatura que utiliza PMI no aplica ningún criterio para reducir el número de parejas. Para decidir el número de parejas de palabras recurrimos a técnicas cualitativas. En concreto usamos un estudio de Delphi con expertos en turismo y comportamiento de los turistas en redes. Este estudio Delphi se desarrollo con 10 expertos a los que se les explicó el trabajo y los objetivos. Después se les invitó con ejemplos de los resultados de parejas a que determinaran el número de parejas de palabras que recogía de forma oportuna el comportamiento turístico y que nivel implicaba la aparición de ruido. Fueron necesarias dos rondas de consulta para llegar a un acuerdo por unanimidad.

Finalmente el criterio consensuado fue quedarnos con las 5 parejas que obtuvieran el valor de PMI más alto. En los siguientes párrafos, interpretamos, en primer lugar, la imagen conjunta que poseen los chino-parlantes. Después, revisamos los resultados independientes de cada foro para verificar si existe distinta imagen entre dos mercados emisores de chino-parlantes. A partir de las revisiones, al final de este capítulo, ofrecemos una evaluación de la metodología diseñada.

## 4.2.Resultado del trabajo empírico

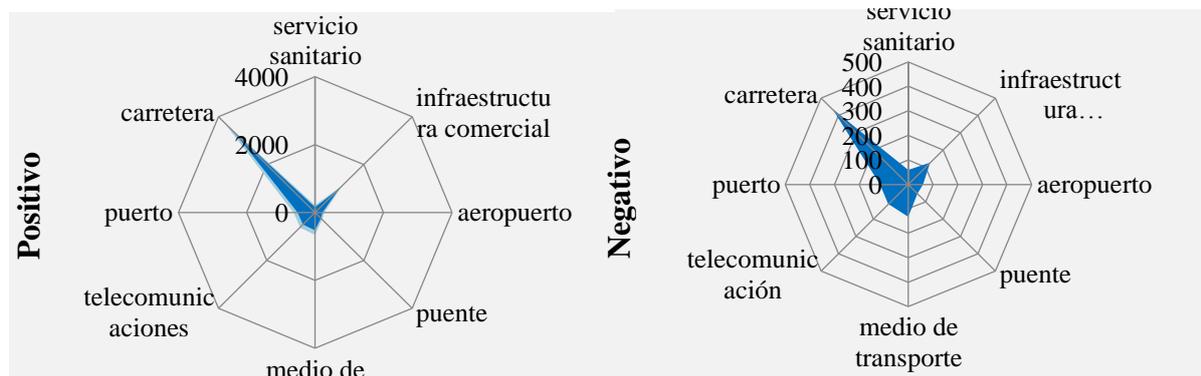
Tras finalizar de ejecutar la metodología prediseñada, descubrimos las características lingüísticas (*features*) más relevantes de cada palabras claves y las emparejamos. Introduciendo en el sistema dichas parejas de palabras, extraemos los comentarios relacionados con los atributos que buscamos. La siguiente figura representa gráficamente el porcentaje de comentarios donde aparecen los términos empleados en cada dimensión. Podemos ver por ejemplo que la dimensión compuesta por mayor número de comentarios es la infraestructura general. Siguiendo el gráfico en el sentido de las agujas del reloj, podemos observar el resto de dimensiones ordenadas de mayor a menor porcentaje de comentarios. Esto resultados nos reflejan la importancia que los turistas chino-parlantes le dan a cada una de las dimensiones que constituye la imagen de destino. Con esto abordamos **el primer objetivo específico del segundo objetivo general**.

Figura 16. Peso de cada dimensión



Centrándonos en la dimensión de la infraestructura general, podemos observar comparativamente a través de los gráficos expuestos a continuación la cantidad de comentarios donde aparecen las parejas de palabras correspondientes a los atributos concretos.

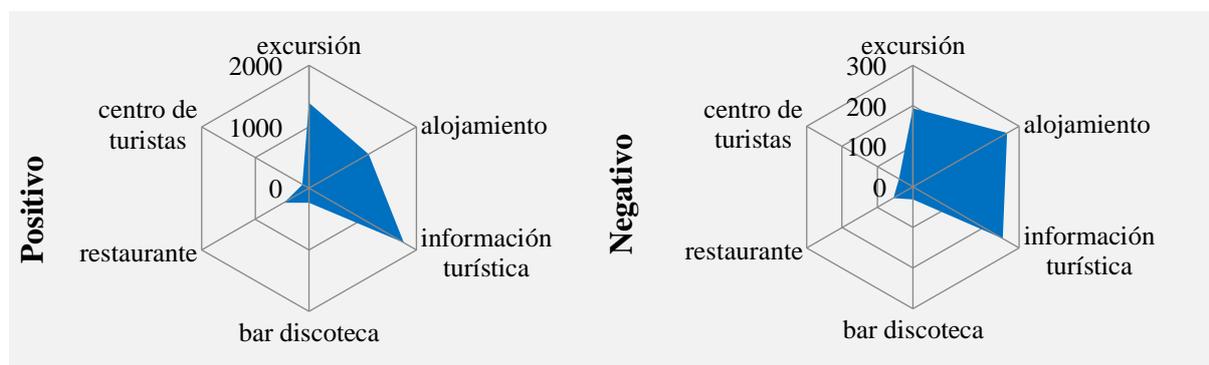
Figura 17. Cantidad de comentarios relacionados con la dimensión infraestructura general



En el gráfico de la izquierda se muestran los comentarios positivos, mientras que en el de la derecha están los comentarios negativos obtenidos. En el gráfico de la infraestructura general, contemplamos que las parejas de palabras relacionadas con el atributo *carretera* son las que aparecen con más frecuencia en los comentarios, también son las que reciben más comentarios negativos. Sin embargo, el atributo *infraestructura comercial* es el segundo que recibe más comentarios pero se encuentra en tercer lugar como atributo que obtiene más comentarios negativos asociados a él.

Centrémonos ahora en la segunda dimensión que recibe más comentarios relativos, la dimensión *infraestructura turística*. Observamos que *información turística* es el atributo asociado con más comentarios, seguido de *excursión* y *alojamiento*. Sin embargo, *alojamiento* es el atributo que se encuentra asociado con más comentarios negativos, seguido de *información turística* y *excursión*.

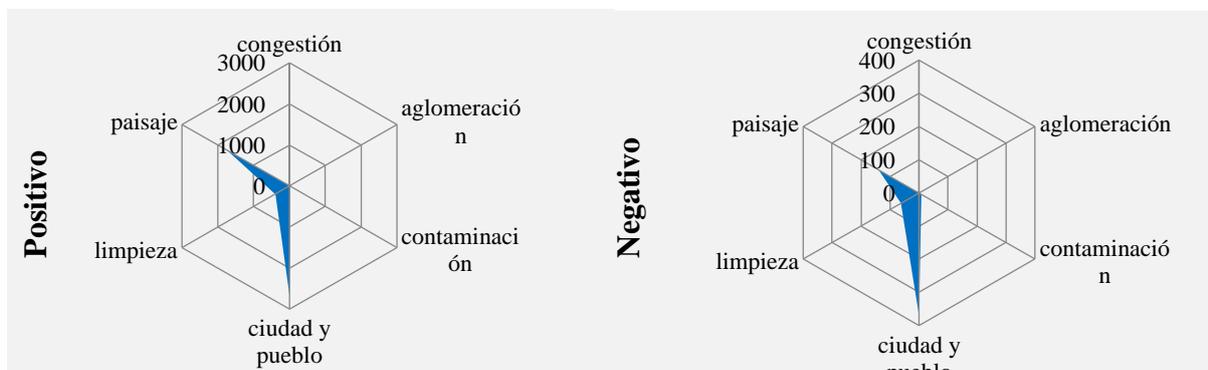
Figura 18. Cantidad de comentarios relacionados con la dimensión infraestructura turística



Respecto a la dimensión del entorno natural, la tercera dimensión con más comentarios asociados, observamos que no existe ningún comentario relacionado con el atributo *congestión* ni *aglomeración*. Las parejas de palabras relacionadas con *ciudad* y

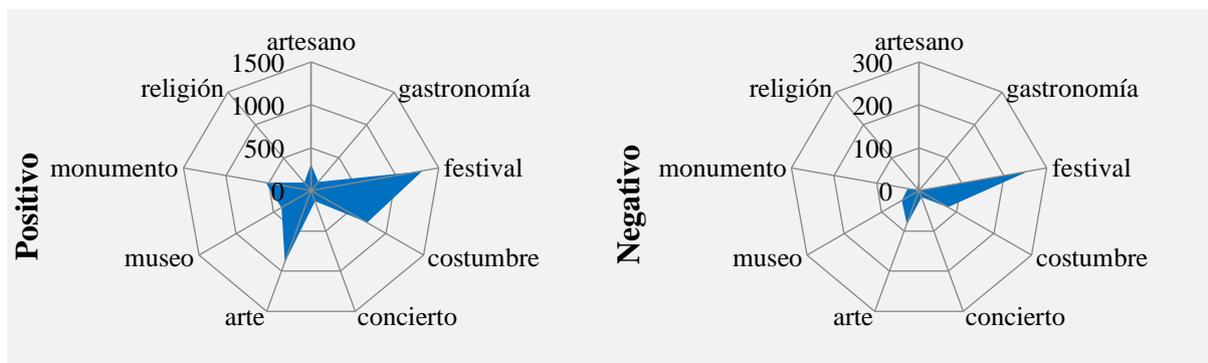
*pueblo* son las que aparecen con más frecuencia en los comentarios positivos. También son los atributos que están más asociados con los comentarios negativos como podemos observar en el siguiente gráfico.

Figura 19. Cantidad de comentarios relacionados con la dimensión entorno natural



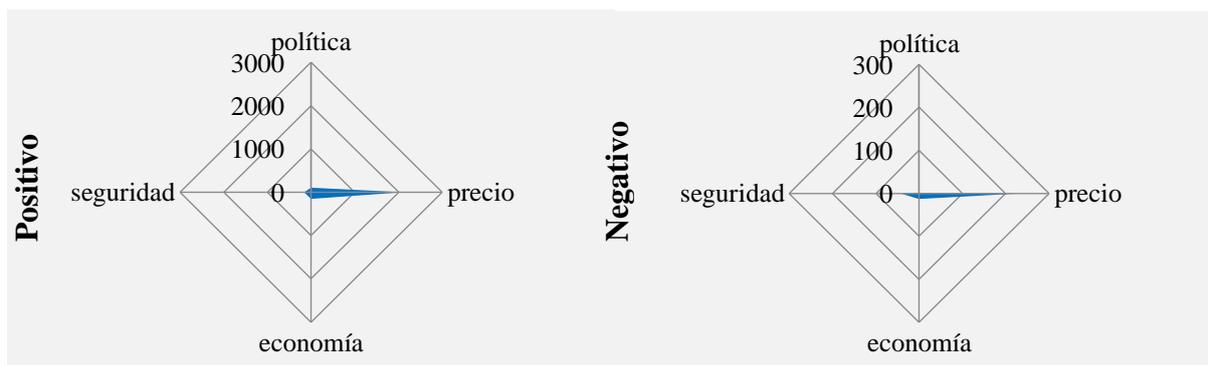
En cuanto a la dimensión *cultura, historia y arte*, las parejas de palabras relacionadas con los atributos *festival, arte y costumbre* son las que aparecen en mayor cantidad de comentarios y también en la mayor parte de comentarios negativos.

Figura 20. Cantidad de comentarios relacionados con la dimensión cultura, historia y arte



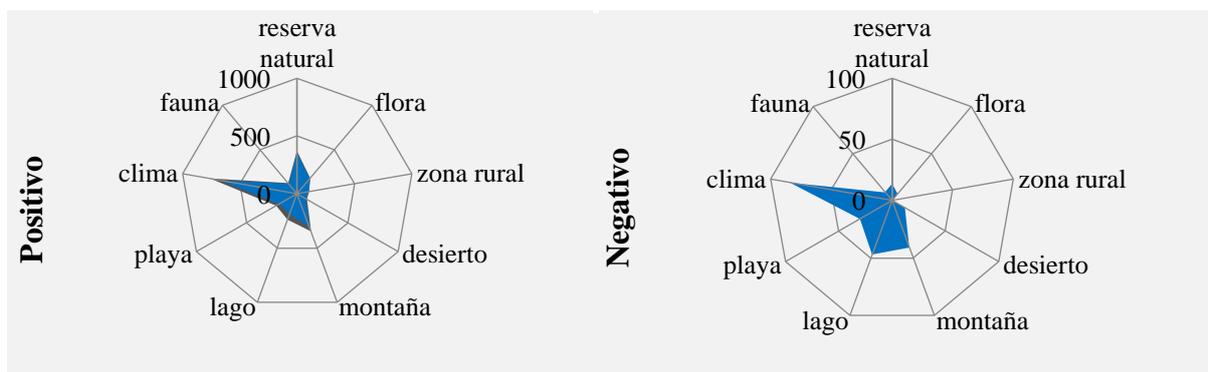
En relación a la dimensión *factor político y económico*, observamos que los usuarios mencionan más los temas de precios que otros atributos en esta dimensión. También este atributo genera más comentarios negativos asociados a sus parejas de palabras que las de otros atributos.

Figura 21. Cantidad de comentarios relacionados con la dimensión factor político y económico



Observando la dimensión *recurso natural*, comprobamos que el atributo *clima* es el más hablado dentro de los comentarios, seguido de *montaña* y *reserva natural*. *Zona rural* es el único atributo que no aparece en ningún comentario negativo. Por el contrario, *clima*, *lago* y *montaña* son los que están recogidos mayormente en los comentarios negativos como podemos contemplar en el gráfico a continuación.

Figura 22. Cantidad de comentarios relacionados con la dimensión recurso natural



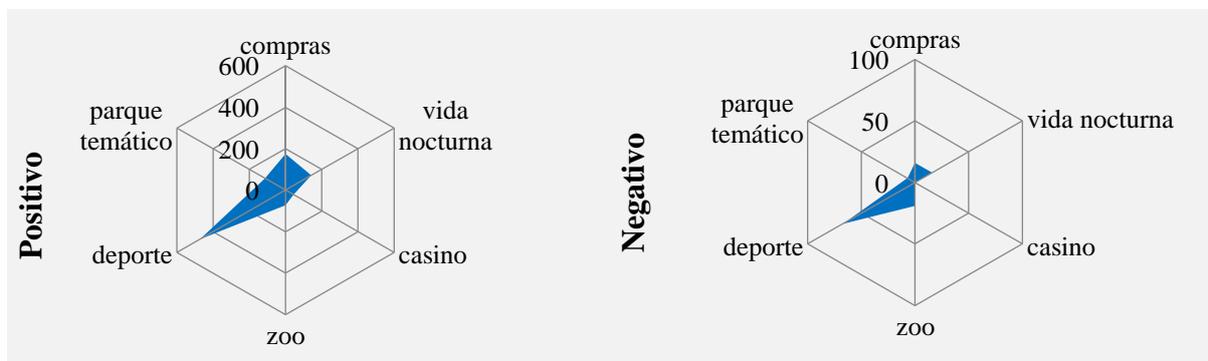
Observando el resultado de la dimensión *entorno social*, observamos que las parejas de palabras que incluyen *comunicación* aparecen en más comentarios y también se encuentran más presentes en las opiniones negativas.

Figura 23. Cantidad de comentarios relacionados con la dimensión entorno social



Por último, la dimensión *ocio y entretenimiento turístico* es la que menos presencia tiene en los comentarios de estos dos foros. Las parejas de palabras de deporte son las que aparecen en mayor cantidad de comentarios y también son las que aparecen con más frecuencia en los comentarios negativos. El atributo casino es el único atributo que no contiene las parejas de palabras asociadas con comentarios negativos, tal y como podemos observar en el siguiente gráfico.

Figura 24. Cantidad de comentarios relacionados con la dimensión ocio y entretenimiento turístico



Estudiando la percepción de la imagen a través de atributos entre los usuarios de dos foros distintos como se muestra en la figura 25, observamos que *ciudad pueblo* y *carretera* son los dos atributos que les llama más la atención. Con este resultado, podemos concluir que los elementos relacionados con *carretera* son lo más hablados entre ellos y que la visita a ciudades o pueblos supone una actividad esencial cuando los turistas chino-parlante vienen a España.

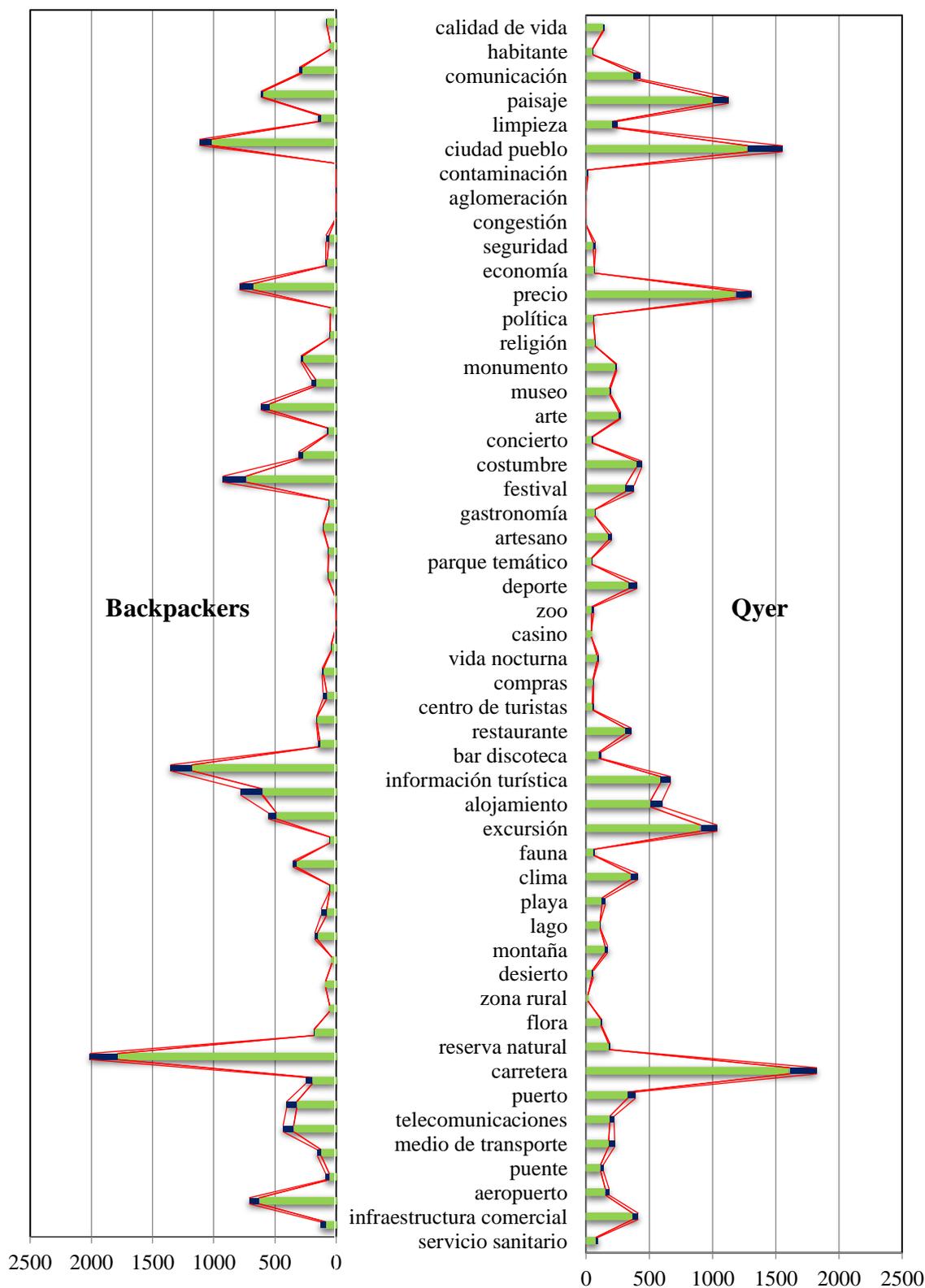
Tras este análisis, descubrimos que los atributos mejor valorados son *carretera*, *ciudad* y *pueblo*, *precio*, *información turística* y *paisaje* a partir de los resultados del análisis de sentimiento de los comentarios. Los que resultan peor valorados son *carretera*, *ciudad* y

pueblo, alojamiento, festival y información turística. Resulta interesante comprobar que existen ciertos atributos que según qué usuario tiene valoración positiva o bien negativa. Con este resultado abordamos **el segundo y tercer objetivo específico del segundo objetivo**.

También observamos las diferencias entre los usuarios de estos dos grupos. Por ejemplo, a los usuarios de *Backpackers* les interesa más los temas relacionados con el arte y los festivales. Además, buscan y comparten más comentarios sobre información turística y alojamiento. Llegamos a la conclusión de que este grupo de usuarios depende un poco más de los temas de *telecomunicaciones* y *medio de transporte* que el otro grupo. Sin embargo, a los usuarios de *Qyer* les interesan y, por lo tanto, comparten más temas sobre paisaje, precio, deporte, restaurante y excursión.

En cuanto a los comentarios negativos que publicados por los usuarios de los dos grupos, observamos que en la mayoría de casos la cantidad de comentarios negativos se encuentra en proporción directa con la cantidad de comentarios que recibe cada atributo. Es decir, a mayor número de comentarios, mayor número de opiniones negativas. A continuación, podemos ver la proporción de comentarios negativos situada en el extremo de cada barra y rodeada con una línea roja.

Figura 25. Comparativa de la imagen de destino a través de los dos foros



#### 4.2.1. Visualización de los comentarios en Backpackers

Observando la tabla 7 de resultados de Backpackers, comprobamos que los temas relacionados con las infraestructuras generales, infraestructuras turísticas y con la cultura, historia y arte son los más comentados en este foro. Las palabras relacionadas con las infraestructuras generales suman el 26% de los comentarios totales, el 19% de comentarios contienen palabras relacionadas con las infraestructuras turísticas y el 16% de comentarios están vinculados a temas de cultura, historia y arte. No obstante, todos estos temas son los que más generan opiniones positivas y también negativas.

Los atributos que reciben imágenes 100% positivas son *flora*, *zona\_rural*, *desierto*, *político* y *habitante*. Por otro lado, los atributos como *lago*, *aeropuerto*, *servicio\_sanitario* y *seguridad* son los que tienen peor percepción. Cada uno de estos atributos recibe más de un 30% de comentarios negativos. A continuación, exploramos una por una la dimensión de cada imagen para descubrir lo que comentan los usuarios de este foro. En los párrafos que se muestran a continuación, detallamos cada dimensión acompañándola de sus atributos.

Tabla 7. Resultado de Backpackers

Dimensión	Atributos	bkall_content		
		Positivo	Neutral	Negativo
Infraestructura_general	servicio_sanitario	86	0	43
	infraestructura_comercial	626	5	80
	aeropuerto	55	1	29
	puente	123	0	31
	medio_de_transporte	347	1	86
	telecomunicaciones	320	3	81
	puerto	192	3	53
	carretera	1785	0	227
Recursos_naturales	reserva_natural	179	0	4
	flora	53	0	0
	zona_rural	86	0	0
	desierto	35	0	0
	montaña	154	0	21
	lago	78	0	45
	playa	51	0	3
	clima	323	1	29
	fauna	51	0	2
Infraestructura_turística	excursión	487	1	65
	alojamiento	598	10	174
	información_turística	1177	1	175
	bar_discoteca	133	0	15
	restaurante	157	0	4
	centro de turistas	75	2	30
Ocio_y_entretimiento_turístico	compras	103	0	12
	vida_nocturna	38	0	4
	casino	0	0	0
	zoo	8	0	3
	deporte	66	0	2
	parque temático	62	0	3
Cultura_historia_y_arte	artesano	102	0	5
	gastronomía	56	0	1
	festival	732	4	188
	costumbre	275	0	35
	concierto	67	0	7
	arte	544	1	71
	museo	165	0	37
	monumento	270	0	18
	religión	49	0	2
Factores_políticos_y_económicos	política	46	0	0
	precio	677	4	110
	economía	73	0	12
	seguridad	59	0	25
Entorno_natural	congestión	0	0	0
	aglomeración	0	0	0
	contaminación	0	0	0
	ciudad_pueblo	1019	1	96
	limpieza	125	0	24
	paisaje	596	0	17
Entorno_social	comunicación	278	0	21
	habitante	47	0	0
	calidad de vida	76	0	2

Visualizando el resultado de las parejas de palabras compuestas por las palabras claves y sus características lingüísticas de cada atributo, revisamos los comentarios de aquellas parejas que reciben más comentarios positivos o negativos buscamos las razones por las que reciben opiniones positivas y negativas.

#### 4.2.1.1. La dimensión de infraestructura general

##### *Carretera*

Respecto a esta dimensión, observamos que el atributo carretera es el más comentado. Este atributo está representado por 132 parejas de palabras. Las más relevantes, es decir, las que aparecen con más frecuencia en los comentarios, son “天\_途” que significa *día o tiempo\_camino*, “旅行\_途” que significa *viaje\_camino*, “客棧\_途” que significa *posada\_camino*, “原文\_途” que significa *texto original\_camino*, “論壇\_途” que significa *foro\_camino* y “行程\_途” que significa *itinerario\_camino*. Cada una de estas parejas supone más del 5% de los comentarios que contienen dicho atributo “carretera”. Las 3 parejas de palabras con más opiniones negativas son la palabra *lago* acompañada de *ruta* (湖泊\_路線), *calle* junto con *puesto callejero* (路\_路邊攤), y *boca del callejón* que aparece junto con *callejón* (巷口\_巷子), representando un 50%, 40% y 40% de opiniones negativas respectivamente.

Visualizando los contenidos en los que aparecen sentimientos positivos, estos se deben generalmente a la buena organización de viajes tanto para ahorrar tiempo como dinero. Estos comentarios se encuentran relacionados con varios temas, como los de alojamiento, de transporte, de restaurantes y los de las ciudades, etc.

Examinando los textos originales en los que aparecen opiniones negativas, observamos, por ejemplo; que en el caso de *lago* y *ruta* la percepción resulta negativa porque, por ejemplo, en uno de los comentarios aparece un autor que menciona el lago en el Parque de Ciudadela y recomienda la ruta de metro para llegar allí. Sin embargo, también cuenta que su ruta de visita transcurre por la playa y si no tienen precaución al hacer fotos, aparecerán chicas en topless al fondo del paisaje (véase anexo 湖泊\_路線 2.txt). Otro caso es cuando el autor viaja con un grupo de amigos en coche y dirigiéndose hacia un lago situado en zona rural, se pierden en el campo por la mala indicación del GPS (véase anexo 湖泊\_路線 1.txt).

En el caso de *calle\_puesto callejero* aparecen opiniones negativas principalmente al compartir un usuario su experiencia de estar a punto de ser robada en la calle entre los puestos callejeros a pesar de que los policías locales se encontraban cerca. Respecto a la pareja “boca del callejón y callejón”, el resultado es negativo debido a la ubicación del alojamiento en un callejón y percibiendo que la calle no era muy segura. A continuación capturamos un párrafo incluido en el fichero 巷口\_巷子 4.txt y lo traducimos al español.

- El texto original con las puntuaciones es:

在/0.0 選擇/0.0210186 公寓/0.0 前/0.0 , /0.0 網路/0.0 評比/0.0 有/0.0 人/0.0 說/0.0 這裡/0.0 頗/0.0 讓  
/0.0 人/0.0 擔心/-0.2902095 的/0.0 是/0.0 , /0.0 下樓/0.0 出門/0.0 到/0.0 蘭布拉/0.0 大街/0.0 需要  
/0.0 走/0.0 一段/0.0 約/0.0 200/0.0 公尺/0.0 遠/0.0 , /0.0 看似/0.05 髒亂/0.0 沒有/-0.0921865 安全性  
/0.0 的/0.0 窄巷/0.0 ; /0.0 當時/0.0 只有/-0.2578135 一/0.0 個/0.0 直覺/0.5240750000000001 , /0.0  
或許/0.0 這/0.0 個/0.0 旅人/0.0 沒有/-0.0921865 去/0.0 過/0.0 印度/0.0 或/0.0 埃及/0.0 比較/0.0 古老  
/-0.249942600000000004 的/0.0 國家/0.0 , /0.0 猜想/-0.076111599999999999 這/0.0 種/0.0 感受/0.0 會  
/0.0 因/0.0 旅人/0.0 的/0.0 旅行/0.0 地區/0.0 和/0.0 經驗/0.070244399999999998 而/0.0 來/0.0 , /0.0  
所以/0.0 還是/0.1 孤注一擲/0.0 。 /0.0

\*\*\*SCORE=-0.3226551500000002

住進/0.0 這/0.0 個/0.0 公寓/0.0 後/0.0 , /0.0 覺得/0.5868026 這/0.0 段/0.0 狹小/-0.4809536 讓/0.0 人  
/0.0 害怕/-0.7250000000000001 的/0.0 巷子/0.0 本/0.0 就是/0.0 當地/0.0 典型/0.0 的/0.0 巷弄/0.0 ,  
/0.0 打烊/0.0 的/0.0 商店/0.0 拉下/-0.65 的/0.0 鐵門/0.0 上/0.0 總是/0.05 塗滿/0.0 髒亂/0.0 的/0.0 塗  
鴉/0.0 , /0.0 行人/0.0 和/0.0 商店/0.0 不/0.0 多/0.0 , /0.0 且/0.0 巷口/0.0 還是/0.1 有/0.0 幾/0.0 輛  
/0.0 警車/0.0 武裝/-0.4 站崗/0.0 , /0.0 若/0.0 一/0.0 個/0.0 女子/0.0 獨行/0.0 , /0.0 還/0.0 真/0.0 有點  
/0.0 恐懼/-0.7 , /0.0 不過/-0.5030712 後來/0.0 自己/0.0 亦/0.0 習慣/0.0830536 了/0.0 , /0.0 也/0.0 可  
以/0.155179700000000003 於此/0.0 獨自/0.0 穿梭/0.05 遊走/0.0 。 /0.0

\*\*\*SCORE=-4.5482178

當然 /0.337995750000000004 , /0.0 聽 /0.0 過 /0.0 一些 /0.0 假裝 /-0.675 警察 /0.0 檢查  
/0.0289814000000000004 證件/0.0 趁機/0.0 勒索/-0.325 和/0.0 偷竊/-0.55 、 /0.0 搶劫/-0.2 在/0.0 巴塞  
隆納/0.0 還是/0.1 時刻/0.0 發生/0.117464399999999997 著/0.0 , /0.0 行走路/0.0 上/0.0 不得不/-  
0.2142008000000000005 小心/-0.0380186000000000013 為/0.0 是/0.0 。 /0.0

\*\*\*SCORE=-1.4780584499999998

- y su traducción literal al español:

Antes de seleccionar este apartamento, veo algunos comentarios por Internet que dicen que la zona es bastante preocupante y dicen que tras bajar del apartamento se deben caminar unos 200 metros de callejones sucios e inseguros para poder llegar a Les Rambles. Yo no le presté mucha atención y creía que tal vez estos viajeros no han estado en países relativamente antiguos como la India o Egipto, por lo que supongo que esta sensación fue percibida debido a su experiencia de viaje en otra región. Por lo tanto, aún así, insisto en alojarme en aquel apartamento.

\*\*\*Puntuación = -0.3226551500000002

Después de alojarme en este apartamento, pienso que estos estrechos callejones dan sensación de miedo pero son típicos de la zona. Las puertas de las tiendas siempre se ven pintadas con grafiti cuando éstas cierran. No se ven muchos peatones ni comercios, y además en la boca del callejón hay un par de policías armados montando guardia. Realmente da un poco de miedo para una mujer sola. Pero después de pasar un tiempo, estoy acostumbrada y puedo caminar por este lugar.

\*\*\* Puntuación = -4.5482178

Por supuesto, también he escuchado algunos casos sobre gente que finge ser policía, pide tu documento y aprovecha para chantajear y robar. Casos de atraco ocurren continuamente en Barcelona. Es mejor tener cuidado cuando se camina por la calle.

\*\*\* Puntuación = -1.4780584499999998

### ***Puerto***

En total existen 25 pares de palabras relacionadas con el atributo puerto. Dentro de ellas, la pareja de palabras *puerto\_puerto* (港\_港口) constituye el 31% del total de los comentarios que incluye el atributo puerto como uno de sus elementos. En el 6% de los comentarios aparece la pareja *puerto\_barco* (港口\_船), en el 5.6% *puerto\_hospital* (港\_醫院), en el 5% *juego olimpico\_puerto* (奧運\_港) y en el 5% *playa\_puerto* (海邊\_港口).

Generalmente, los turistas de este foro obtienen sentimientos positivos sobre este atributo debido a los bonitos paisajes, el buen clima o la simpatía de la población local.

La pareja de palabras que recibe peor opinión es *puerto\_barco* (港口\_船) con un 40% de opiniones negativas. Revisando los textos en los que se incluyen estas dos palabras, entendemos que la mayoría de las opiniones negativas surgen por el alto precio del billete del barco y piensan que no merece la pena cogerlo.

### **Telecomunicaciones**

En total existen 47 parejas con este atributo, siendo las parejas de palabras *móvil\_teléfono* (手機\_電話), *internet\_telecomunicaciones* (網路\_電信), *móvil\_telecomunicaciones* (手機\_電信) y *tarjeta\_telecomunicaciones* (卡\_電信) las parejas de palabras más relevantes, ocupando cada una de ellas el 9%, el 8%, el 6%, el 5% y el 5% del total de comentarios respectivamente del atributo Telecomunicaciones.

Los turistas poseen buena imagen de este atributo principalmente por las tarifas económicas de las llamadas internacionales ofrecidas por varios operadores tanto locales como tarjetas específicas para llamadas internacionales de ciertos países, como por ejemplo, la tarjeta Jazz panda.

Relacionadas con este atributo, las parejas que obtienen peor puntuación son *móvil\_comisaría* (手機\_警局), con un 60% de opiniones negativas seguida de *móvil\_vista* (手機\_視線), que recibe un 50% de opiniones negativas. Revisando los textos que contienen estas parejas de palabras, observamos que en la mayoría de los comentarios aparecen las palabras *móvil* y *comisaría* mayormente por el tema de robos, al igual que los relacionados con *móvil* y *vista*.

### **Medio de transporte**

Este atributo ofrece 21 pares de palabras. Las parejas más relevantes de este atributo son *maleta\_transportar* (行李\_運) presente en el 44% de los comentarios que incluyen este atributo. A continuación, el 7% por la pareja *cielo o día\_transporte* (天\_運輸), 6% por la pareja *viaje\_transporte* (旅行\_運輸), 6% por *tráfico\_transporte* (交通\_運輸), 5% por *herramienta\_transporte* (工具\_運輸) y 5% por *posada\_transporte* (客棧\_運輸).

Las imágenes positivas recibidas por este atributo se deben a las experiencias y la información intercambiada entre los usuarios sobre las limitaciones de medidas y peso de los equipajes de cada compañía, el tiempo de la espera de la facturación, cómo conseguir una tarifa más económica a la hora de facturar el equipaje, etc.

Por otro lado, las parejas que reciben más opiniones negativas son *tráfico\_folk* (交通\_民風) en un 67% de los comentarios donde aparece estas palabras y *tráfico\_precio* (交通\_行情) en el 50%. En el caso de *tráfico\_folk*, observamos que las puntuaciones resultan negativas debido al contexto hablado del cuidado con el robo en los medios de transporte. Sin embargo los párrafos donde aparece la palabra folk corresponden a comentarios positivos que mencionan el honesto de los habitantes de pueblos. Respecto a la pareja *tráfico\_precio*, los comentarios en los que se incluyen esta pareja de palabras ofrecen información sobre los precios de transporte. Las opiniones negativas surgen debido a que los autores piensan que es poco accesible algún tipo de transporte como es el caso de un velero turístico.

### **Puente**

Este atributo se encuentra representado por 19 parejas de palabras. Las pares de palabras que aparecen con más frecuencia son *puente\_acueducto* (橋\_水道) en el 42% de comentarios relacionados con puente, *puente\_puente* (橋\_橋樑) en el 6.5%, *puente\_oro* (橋\_黃金) en el 6.5%, *puente\_río* (橋\_河流) en el 5.2%, *puente\_María* (橋\_瑪麗亞) en el 5%, y *puente\_riberaño del río* (橋\_河岸) en el 5%.

Visualizando los comentarios positivos relacionados con este atributo, no es difícil encontrar las relaciones de este atributo con otros como son monumento, museo, clima y paisaje. La combinación de este atributo con otros proyecta una buena imagen a los turistas de este foro.

Respecto a los pares de palabras con más opiniones negativas, encontramos la pareja *puente\_acueducto* (橋\_水道), conteniendo un 38% de comentarios negativos. Sin embargo, detectamos un error de análisis sintáctico, es decir, el problema de *parser*. En este caso, visualizamos los comentarios originales y descubrimos que 橋\_水道 es una única palabra y significa puente de acueducto. Por lo tanto, la puntuación negativa surge en muchos casos debido a otros elementos en el mismo comentario.

### ***Aeropuerto***

El atributo aeropuerto está representado por 11 pares de palabras. Entre ellos, parejas como *aeropuerto\_terminal* (機場\_航站), *compañía\_terminal* (公司\_航站), *terminal\_oficina* (航站\_處), *aviación\_terminal* (航空\_航站), *terminal\_conexión* (航站\_轉機), *terminal\_información* (航站\_資料) y *sección\_terminal* (段\_航站) son las que coocurren con más frecuencia. Los porcentajes de la ocurrencia son el 29%, 11%, 11%, 11%, 8%, 8% y 6% respectivamente.

Los sentimientos positivos que provocan este atributo se deben a los intercambios de experiencias y conocimientos relacionados con la infraestructura y servicios dentro del aeropuerto que son la comunicación entre transportes, la localización de las compañías aéreas, los procesos de devolución de IVA, etc.

Es estas parejas de palabras, las que reciben mayor porcentaje de comentarios negativos son *aviación\_terminal* con un 44%, *aeropuerto\_terminal* con un 44% y *terminal\_conexión* con el 43%. La razón por la que la pareja *aviación\_terminal* recibe una puntuación negativa es debida a que los turistas no localizan fácilmente la terminal de llegada o salida de su avión. En relación a *aeropuerto\_terminal*, las opiniones negativas surgen por no saber en qué lugar de la terminal se pueden encontrar los medios de transporte disponibles para salir del aeropuerto.

### ***Infraestructura comercial***

El atributo infraestructura ofrece 66 parejas de palabras, siendo los pares de palabras que aparecen en más comentarios las siguientes: *establecimiento* o *sitio\_vino* (莊\_酒) en el 9% de comentarios, *paisaje\_establecimiento* o *sitio* (景\_莊) 7%, *castillo\_establecimiento* o *sitio* (城\_莊) 7%, *distinto\_puesto* (另\_攤) 5%, *montaña\_establecimiento* o *sitio* (山\_莊) 5% y *trozo\_establecimiento* o *sitio* (片\_莊) con el 5%. Sin embargo, la palabra china “莊” puede tener muchos significados. Dependiendo de los caracteres que la acompañan y el contexto, “莊” puede referirse a un lugar, una granja, un pueblo, etc.

Los sentimientos positivos relacionados con este atributo tienen su origen en las buenas experiencias que varios turistas tienen sobre los establecimientos.

Las parejas que tienen más porcentaje de opiniones negativas (40%) son *unidad\_venta* (個\_門市) causada por la poca cantidad de ventas, *calle\_puesto callejero* (路\_路邊攤) por temas de seguridad y engaños, y *persona\_pequeño puesto* (人\_小攤) por la antipatía mostrada por el dependiente. Las parejas *yo\_puesto callejero* (我\_路邊攤) y *tú\_puesto callejero* (你\_路邊攤) también suman el 40% de comentarios negativos pero no deben contabilizarse. Tal y como mencionamos con anterioridad, los pronombres personales son las palabras más frecuentes en cualquier tipo de texto.

### ***Servicio sanitario***

Con este atributo tenemos 10 pares de palabras relacionadas con él. Las parejas de palabras *St. Pablo\_hospital* (聖保羅\_醫院), *monumento\_hospital* (紀念塔\_醫院) y *mansión\_hospital* (府\_醫院) son las más relevantes, apareciendo en el 32%, 10% y 10% de los comentarios respectivamente.

Los contenidos que expresan un sentimiento positivo son debidos a una buena organización del itinerario entre un hospital y otros puntos de visita. Concretamente este caso se refiere al hospital Santa Pau.

Descubrimos que la pareja *St. Pablo\_hospital* (聖保祿\_醫院) arroja un 50 % de opiniones negativas debido al horario de apertura ya que a veces se encuentra abierto y otras no. También los usuarios mencionan otros puntos de visita en el mismo comentario por lo que afecta al resultado de esta pareja. Además, estas dos palabras mencionan el mismo sitio que es el hospital Santa Pau. Al igual que el caso anterior, la pareja de palabras *mansión\_hospital* también se refiere al mismo sitio que es Palacio Güell, mostrando un 50% de opiniones negativas debido al hecho de encontrarse en reparaciones.

#### **4.2.1.2.La dimensión recursos naturales**

##### ***Reserva natural***

Observando los 20 pares de palabras encontrados con este atributo, las parejas *naturaleza\_natural* (大自然\_自然), *natural\_inspiración* (自然\_靈感), *curva\_natural* (曲線\_自然), *Dios\_natural* (上帝\_自然), *planta\_natural* (植物\_自然) y *humano\_natural* (人類\_自

然) son las más presentes entre todas, ocupando el 21%, 8%, 7%, 7%, 6% y 6% respectivamente del total de comentarios con este atributo.

La mayoría de los sentimientos positivos relacionados con este atributo se debe al diseño natural de la Sagrada Familia. También existen comentarios cuyo objetivo es animar a la gente a visitar sitios rurales en España y conocer espectáculos de la naturaleza, como por ejemplo la cueva de Nerja.

Generalmente, las puntuaciones para los pares de palabras de este atributo son positivas. Si nos paramos a analizar algunas parejas de palabras que contienen opiniones negativas, podemos comprobar que los comentarios en los que se incluyen las palabras 大自然\_天色 que significa *naturaleza\_color del cielo* u *hora del día* tienen un 13% de comentarios negativos debido al mal tiempo que deja el cielo oscuro en la zona rural.

### **Zona rural**

Este atributo está representado por 21 pares de palabras. Las parejas de palabras más llamativas son *movimiento\_zona rural* (動態\_原野) que aparece en el 8% de comentarios, *zona rural\_racing* (原野\_賽車) en el 8%, *valor\_zona rural* (值\_原野) en el 7%, *cielo o día\_zona rural* (天\_野外) en el 6%, *zona rural\_oeste* (原野\_西邊) en el 6%, *zona rural\_terreno* (原野\_地勢) en el 6%, *zona rural\_montaña alta* (原野\_高山) en el 6% y *zona rural\_lejano* (原野\_遠方) en el 6%.

Los sentimientos positivos respecto a este atributo están relacionados con los pueblos. Los usuarios admiran el paisaje de la zona rural y se sienten atraídos por la tranquilidad y el ambiente relajado de estos sitios. Ninguna de las parejas de palabras refleja opiniones negativas.

### **Flora**

Observando 10 parejas de palabras con este atributo, existen 4 de ellas más llamativas que otras que son *naturaleza\_planta* (大自然\_植物) constituyendo el 17% del total de comentarios del atributo flora, *planta\_ángulo* (植物\_角度) el 13%, *frente\_planta* (前方\_植物) el 13% y *planta\_ciudad* (植物\_都市) el 11%.

Los sentimientos positivos de los turistas con respecto a este atributo se deben a la gran cantidad de plantas que se pueden ver en sitios de visita como jardines, casas rurales, o zona rural. También se debe al diseño natural de figuras simulando flora dentro de la Sagrada Familia.

Respecto a los pares con más opiniones negativas, no observamos ningún comentario negativo relacionado con este atributo.

### ***Desierto***

Este atributo ofrece 10 pares de palabras. Entre ellas, las parejas *viaje\_desierto* (旅遊\_沙漠), *desierto\_la vida* (沙漠\_生活), *posada\_desierto* (客棧\_沙漠), *viaje\_desierto* (旅行\_沙漠) e *iglesia\_desierto* (教堂\_沙漠) son las más relevantes, ocupando cada una de ellas el 11% del total de los comentarios relacionados con el atributo desierto.

Los turistas poseen un buen sentimiento sobre este atributo, sin embargo, descubrimos que el uso de la palabra desierto que aparece en los comentarios no se refiere al desierto en España, sino que se refiere a la cultura árabe. En el caso del atributo desierto, no percibimos ninguna opinión negativa acerca de este atributo.

### ***Montaña***

El atributo montaña cuenta con 20 pares de palabras. De entre todas, las parejas *montaña\_palabras empiezan con la pronunciación “mon”* (山\_蒙), *montaña\_montaña alta* (山\_高山) y *montaña\_campo* (山\_鄉間) son las que aparecen con más frecuencia, ocupando cada una de estas parejas el 23%, el 13% y el 6% en los comentarios totales relacionados con este atributo.

Los sentimientos positivos relacionados con este atributo se deben a las actividades que se puede realizar en las montañas como senderismo, vista de panorámicas y el intercambio de información para acceder a la montaña.

Por otro lado, la pareja de palabras que recibe más opiniones negativas es *montaña\_estación de teleférico* (山\_纜車站), suponiendo un 43% de comentarios negativos. Esto surge, por ejemplo, un comentario que menciona la dificultad de encontrar la entrada del teleférico de Montjuic en Barcelona.

## Lago

En total existen 20 parejas de palabras relacionadas con el atributo lago. Entre estas parejas se encuentran *lago\_filo* (湖\_邊), *montaña\_lago* (山\_湖), *pueblo\_lago* (小鎮\_湖), *museo\_lago* (博物館\_湖) que son las más relevantes en los comentarios relacionados con lago. Cada una de ellas ocupa respectivamente el 15%, el 13%, el 8% y el 7% del total de comentarios que incluye el atributo lago como uno de sus elementos.

Las buenas imágenes percibidas por los turistas relacionadas con el atributo lago la mayoría son debido al paisaje, los monumentos o los sitios turísticos que se encuentran junto al lago.

Respecto a las parejas que obtienen peores comentarios, encontramos las parejas *lago\_compras* (湖\_購物) con el 67% de opiniones negativas, *lago\_restaurante* (湖泊\_餐廳) el 67%, *viaje\_lago* (旅行\_湖泊) el 50%, *posada\_lago* (客棧\_湖泊) el 50% y *lago\_ruta* (湖泊\_路線) el 50%. En el caso de *lago\_compras*, no percibimos ninguna opinión negativa directa vinculada con esta pareja. No obstante, la palabra *compras* está relacionada con el tema del robo. Por otra parte, debido al error ortográfico de un usuario, el cual se ha equivocado escribiendo la palabra 迷迷糊糊 (*desorientado*) en lugar de 迷迷糊 “湖” (*lago confundido*), el sistema señala *lago* con una opinión negativa. En cuanto a *lago\_restaurante*, la puntuación negativa se debe a la influencia de otros elementos en el mismo comentario. Respecto a *posada\_lago*, descubrimos otro ruido relacionado con la palabra “客棧” que significa posada. El sistema de análisis sintáctico no detecta un nombre propio, por lo que esta palabra ha sido separada desde el nombre chino del foro Backpackers. Dicho nombre es “背包客棧”, que significa la posada de los mochileros. Sin embargo, la pareja *lago\_ruta* si que no tiene ningún error. La puntuación negativa se debe con frecuencia a la misma razón que con la pareja *lago\_ruta* del atributo carretera, es decir, la mala indicación en la ruta para llegar al lago.

## Playa

Este atributo ofrece 10 pares de palabras, siendo las parejas de palabras *playa\_puerto* (海邊\_港口), *playa\_designación del yo, si es mujer* (海邊\_老娘), *playa\_brisa marina* (海邊\_海風), *roca\_playa* (岩石\_海邊), *playa\_padre* (海邊\_爸爸), *cafetería\_playa* (咖啡館\_海邊)

y *playa\_zona rural* (海邊\_鄉間) las parejas de palabras más relevantes, ocupando cada una de estas parejas el 24%, el 11%, el 11%, el 9%, el 9%, el 9% y el 9% del total de comentarios del atributo playa respectivamente.

Los turistas de este foro expresan un sentimiento positivo relacionado con este atributo por el buen clima, el ambiente relajado y a su vez por la vida animada en la playa.

Por otro lado, la pareja que recibe más opiniones negativas es *zoo\_playa* (動物園\_海邊) en un 33% de los comentarios donde aparecen estas palabras. La razón por la que recibe puntuación negativa es por la influencia de otros elementos en el mismo comentario.

### ***Clima***

Observando 54 parejas de palabras con el atributo clima, la pareja *día o cielo\_color del cielo* u *hora de día* (天\_天色) ocupa el 7% de los comentarios totales relacionados con clima, *tiempo\_nublado* (天氣\_陰) el 7% y *tiempo\_clima* (天氣\_氣象) el 6% son las pares de palabras más destacadas.

Los sentimientos positivos sobre este atributo se deben a la información y la experiencia compartida de varios usuarios sobre el clima local. También existen comentarios que hablan sobre el bonito color del cielo o la precisión de la previsión de precipitaciones.

La pareja *fuerte nevada\_clima* (大雪\_氣候) es el par de palabras que se percibe con peor puntuación, con un 60% de comentarios negativos debido a las fuertes nevadas que interrumpen el viaje.

### ***Fauna***

En total existen 10 pares de palabras con este atributo. Las parejas de palabras *animal\_natural* (動物\_自然) y *animal\_entallar o escultura* (動物\_雕刻) son las más relevantes, ocupando el 17% y el 11% del total de comentarios relacionados con fauna respectivamente.

Los turistas poseen una imagen positiva relacionada con este atributo debido a la variedad de animales que se encuentran en una casa rural o las esculturas en forma de animales en la Sagrada Familia.

Por otro lado, las parejas que están representadas por mayor cantidad de opiniones negativas son *animal\_religión* (動物\_宗教) y *animal\_leche* (動物\_牛奶). *Animal\_religión* recibe una opinión negativa porque hay autores que relacionan las corridas de toros con la religión. Con respecto a *animal\_leche* se debe a los veganos quienes cuestionan la posibilidad de encontrar comida para ellos en España.

#### 4.2.1.3. La dimensión infraestructura turística

##### *Excursión*

En relación al atributo excursión, las parejas de palabras *festival\_viaje* (節\_遊), *naturaleza\_viaje* (大自然\_旅行), y *viaje\_excursión* (旅遊\_遊覽) son las que aparecen con más frecuencia en los comentarios relacionados con este atributo. Cada una de ellas cuenta con el 33%, el 6% y el 5% de ocupación en el total de los comentarios respectivamente.

Los sentimientos positivos de este atributo están influenciados por varios atributos y variables. Dado que la palabra excursión se puede entender igual que la palabra viaje en chino y los comentarios están limitados en los temas de viaje, no podemos averiguar la razón principal por la que este atributo recibe una buena puntuación.

Respecto a la pareja de palabras que recibe más comentarios negativos, encontramos la pareja *excursión\_hotel* (遊覽\_酒店). Los comentarios que mencionan esta pareja son textos sobre cómo organizar un itinerario o petición de recomendación sobre la planificación del viaje. El sistema dio mala puntuación a algunos comentarios por el uso de palabras. Sobre todo las palabras que dan sensación de un viaje muy precipitado. A continuación veremos un ejemplo capturado de un comentario.

DAY3 11/6 MADRID--SEGOVIA--MADRID--SEVILLA...

\*\*\*SCORE=0.0

.../0.0 一早/0.0 到/0.0 SEGOVIA/0.0 遊覽/0.05 , /0.0 之後/0.0 返回/0.0 MADRID/0.0 ; /0.0 深夜/0.0 //0.0  
凌晨/0.0 坐/0.0 巴士/0.0 到/0.0 SEVILLA/0.0 。 /0.0

... 一早到 SEGOVIA 遊覽, 之後返回 MADRID; 深夜/凌晨坐巴士到 SEVILLA。

\*\*\*SCORE=-0.275

(/0.0 宿/0.0 車/0.0 上/0.0 ) /0.0

(宿車上)

\*\*\*SCORE=0.0

Su traducción es:

Día 11 de Julio Madrid—Segovia—Madrid—Sevilla

Viaja a Segovia por la mañana muy temprano y después vuelva a Madrid. Por la noche o madrugada, coge el autobús para Sevilla. Duerme en el autobús.

### ***Alojamiento***

Este atributo ofrece 84 parejas de palabras y las parejas *tienda\_alojamiento* (店\_酒店), *dueño\_albergue* (主人\_民宿), y *cercano\_hotel* (近\_酒店) son las más relevantes, ocupando cada una de ellas el 16%, el 5% y el 5% de los comentarios totales relacionados con el atributo alojamiento respectivamente.

La buena imagen relacionada con este atributo se debe a los buenos tratos personales, la buena localización del alojamiento, así como de su limpieza y los servicios que ofrece.

Observando las parejas de palabras que reciben más comentarios negativos, observamos que las parejas *partido\_hotel* (球賽\_酒店), *callejón\_hostal* (巷子\_旅社), *hostal\_favorito* (旅社\_最愛), *mañana\_hotel* (上午\_酒店), *palabras pronunciadas en “Shì”\_hotel* (士\_酒店), *hostal\_albergue* (旅社\_民宿), *hostal\_calle* (旅社\_街道) y *precio\_albergue* (價位\_旅店) son las que reciben más del 50% de comentarios negativos. Entre estas, el par de palabras *partido\_hotel* obtiene un 100% de opiniones negativas.

Cuando revisamos los contenidos, observamos que la mayoría de comentarios negativos relacionados con la palabra *partido* se debe a las distancias entre el hotel y otros puntos de visita. Como excepción uno de ellos menciona el cierre del campo del fútbol en los días del partido y recomienda que no organice la visita del campo del fútbol en fecha de partido. Las parejas *callejón\_hostal* y *hostal\_albergue* reciben el 75% y el 50% del total de las opiniones negativas respectivamente. La razón principal es debida a los robos en callejones y algunas zonas peligrosas, por lo que recomiendan no elegir alojamiento situado en ese lugar. La pareja *hostal\_favorito* aparecen en casi los mismos comentarios que la pareja *callejón\_hostal*, por lo que recibe el 60% de comentarios negativos. Esta pareja aparece en los comentarios que hablan sobre la zona peligrosa, concretamente la zona Ciutat Vella. Además del nombre de la zona, los comentarios mencionan también que los hostales en aquella zona suelen ofrecer su habitación insistiendo en la localización central, que es lo que buscan los

turistas a la hora de elegir un alojamiento. Entre el 57% de comentarios negativos relacionados con la pareja *mañana\_hotel*, la mayoría de estos se producen porque desde hotel hasta otro punto de visita se requiere mucho tiempo. Tan sólo un comentario habla de la pérdida de tiempo en el chick-in. Respecto a la pareja *palabras pronunciadas en Shì\_hotel*, descubrimos un error en el análisis sintáctico por el que 士 (pronunciada en *shi*) está separada de 巴 (pronunciada en *Ba*). Los dos caracteres juntas significan autobús (巴士). La razón por la que la pareja *palabras pronunciadas en Shì\_hotel* obtiene el 50% de opiniones negativas se debe a la mala conexión de hotel. En relación al 50% de opiniones negativas de *hostal\_calle*, la causa también es por temas de seguridad y los usuarios recomiendan que no se pase por algunas calles concretas. Finalmente, parte del 50% de comentarios negativos sobre precio\_albergue se debe al ruido generado por otro atributo en los mismos comentarios. El resto son por la dificultad de reservar una habitación en un albergue que tenga buen precio y calidad.

### **Información turística**

Este atributo está representado por 101 pares de palabras, siendo las parejas de palabras *viaje\_sentido* o palabra pronunciada en *Yì* (旅行\_義) aparecidas en el 12% de comentarios de este atributo, *posada\_sentido* o palabra pronunciada en *Yì* (客棧\_義) en el 10%, *Francia\_sentido* o palabra pronunciada en *Yì* (法\_義) en el 10%, *Alemania\_sentido* o palabra pronunciada en *Yì* (德\_義) en el 8% y *sentido\_sentido* o palabra pronunciada en *Yì* (意義\_義) en el 6%. Sin embargo, destacamos la palabra “義” pronunciada *Yì* podría referirse a otra cosa, como por ejemplo Italia (“義大利” o “義”) o ideario (主義). Este error se debe al proceso de análisis sintáctico, y también a defectos de escritura de algunos usuarios, como por ejemplo la omisión de palabras en un contexto.

Con respecto a la declaración anterior, no podemos definir las causas de los sentimientos positivos sobre este atributo, ya que la mayoría de textos que reciben opiniones positivas contiene la palabra *Yì*, que se refiere al país Italia, y no a la información. Estos resultado son ruido y quedan eliminados de análisis.

Respecto a los pares de palabras que reciben más puntuaciones negativas están las parejas *información\_costumbre* (資料\_風俗), con un 67% de comentarios negativos seguida de *parte\_material* (份\_材料), que recibe el 40% de opiniones negativas. En el caso de *información\_costumbre*, la puntuación negativa se debe a, por ejemplo, un comentario que

anima a conocer las costumbres del destino en lugar de buscar solamente información básica. En cuanto a *parte\_material*, los comentarios relacionados con esta pareja involucran diversos temas como el tema de restaurantes como el de la gestión de los papeles.

### ***Bar y discoteca***

El atributo *bar y discoteca* está representado por 30 pares de palabras. En el 7% del total de comentarios relacionados con este atributo aparecen la pareja *bar\_estilo* o *viento* (酒館\_風), en el 6% *la vida\_bar* (生活\_酒館), en el 5% *discoteca\_aquella* (舞廳\_那), en el 5% *puesto\_bar* (攤\_酒吧), en el 5% *costumbre\_bar* (習慣\_酒吧), en el 5% *persona\_discoteca* (人\_舞廳) y en el 5% *arriba\_discoteca* (上\_舞廳).

Respecto a las opiniones positivas relacionadas con este atributo, los usuarios mencionan siempre la variedad de tapas y su precio junto con la palabra bar. También comentan el buen ambiente que se encuentra en estos establecimientos.

Por otro lado, la pareja de palabras que recibe más opiniones negativas es *hotel\_discoteca* (旅館\_舞廳) con un 50% de comentarios negativos. Existen muchas razones por la que obtiene mala puntuación, como por ejemplo el tema de seguridad, el tema de la estacionalidad como es el caso de Ibiza, el tema de ruido, etc.

### ***Restaurante***

Este atributo ofrece 29 pares de palabras. La pareja más relevante es *cafetería\_viento* (咖啡廳\_風), ocupando el 9% de comentarios del total de los comentarios relacionados con restaurante, seguida por *designación del yo, si es mujer\_restaurante* (老娘\_餐廳) el 6%, *almuerzo\_cafetería* (午餐\_咖啡館) el 5%, y *visita\_restaurante* (巡禮\_餐廳) el 5%.

Los sentimientos positivos que se muestran en los comentarios que contienen este atributo se deben al diseño y estilo del restaurante, la localización o el ambiente.

Respecto al par de palabras que recibe peores comentarios, encontramos la pareja *lago\_restaurante* (湖泊\_餐廳) que suma el 67% de opiniones negativas debido a la influencia de otros elementos en el mismo comentario.

### ***Centro de turistas***

Este atributo está representado por 11 pares de palabras. Dentro de ellas, las parejas de palabras *viaje\_oficina de turismo* (旅遊\_旅遊局) que aparece en el 35% de comentarios relacionados con centro de turísticas, *oficial\_oficina de turismo* (官方\_旅遊局) en el 19% y *oficina de turismo\_información* (旅遊局\_資料) en el 13% son las más relevantes.

Los usuarios que poseen una buena imagen relacionada con este atributo se deben a las buenas indicaciones de los puntos de interés colocados por el centro de turistas, la disponibilidad del sitio web oficial, y la cantidad de información actualizada y ofrecida en la web oficial de los centros de turistas,

Por otro lado, las parejas *oficina de turismo\_playa* (旅遊局\_海灘) y *oficina de turismo\_comunidad autónoma* (旅遊局\_自治區) son las que contienen más comentarios negativos y presentan un 60% y un 40% de opiniones negativas respectivamente. Respecto a *oficina de turismo\_playa*, la puntuación negativa surge porque en algunos comentarios recomiendan no permanecer en la zona de playa mucho tiempo y que se informe en la web de la oficina de turismo sobre otras actividades. En relación a *oficina de turismo\_comunidad autónoma*, la puntuación negativa se produce debido a la influencia de otro elemento en los mismos comentarios.

#### **4.2.1.4. La dimensión de ocio y entretenimiento turístico**

### ***Compras***

Contando con 17 *parejas de palabras*, este atributo compone principalmente los pares de palabras *castillo\_compras* (堡\_購物), ocupando el 13% del total de comentarios. A continuación le sigue *compras\_árabe* (購物\_阿拉伯) el 9%, *compras\_viaje* (採購\_旅行) el 8% y *peatón\_compras* (行人\_購物) el 7%.

La buena imagen que poseen los usuarios es debida a los pocos requisitos para la devolución del IVA, además del precio bajo y la variedad de productos ofrecidos en España.

Por otra parte, la pareja de palabras con más puntuación negativa es *lago\_compras* (湖\_購物), contando con el 67% de comentarios negativos. La razón por la que esta pareja

obtiene puntuación negativa es debido a la influencia de otros elementos en los mismos comentarios.

### ***Vida nocturna***

Este atributo ofrece 10 pares de palabras. Entre ellas, la pareja *vida nocturna\_día* (夜生活\_日) aparece en el 19% de comentarios totales de este atributo, seguidas por *vida nocturna\_calle* (夜生活\_街) con el 17%, *vida nocturna\_ruta* (夜生活\_路線) con el 10%, *vida nocturna\_calle* (夜生活\_街道) con el 10% y *vida nocturna\_paisaje* (夜生活\_風景) con el 10%.

A los turistas les impresiona la vida nocturna animada en España y se sorprenden de la cantidad de gente y las actividades en calles por la noche.

Por otro lado, las parejas *vida nocturna\_ruta* y *vida nocturna\_calle* (夜生活\_街道) reciben ambas el 25% de comentarios negativos por temas de seguridad.

### ***Casino***

No se detecta ninguna característica lingüística relacionada con este atributo.

### ***Zoo***

Este atributo tiene muy pocos pares de palabras lingüísticas, en concreto son 3 parejas. En el 36% de comentarios aparece la pareja *zoo\_posada* (動物園\_客棧), el 36% *zoo\_viaje* (動物園\_旅行) y el 27% *zoo\_playa* (動物園\_海邊).

Contemplando los comentarios que contienen este atributo, descubrimos que la visita del zoo es simplemente una actividad complementaria, y por tanto, los sentimientos positivos no son originados directamente por este atributo.

La pareja que recibe más opiniones negativas es *zoo\_playa* en un 33% de los comentarios totales. Sin embargo, visualizando el contenido, observamos que esta pareja obtiene la puntuación negativa debido a la influencia de otros elementos en los mismos comentarios.

### ***Deporte***

Este atributo está representado por 10 parejas de palabras. Las parejas más destacadas son *paisaje\_deporte* (景\_運動) en el 29% de comentarios totales, seguida de *deporte\_ciudad* (運動\_鎮) con el 16% y *religión\_deporte* (宗教\_運動) con el 10%.

Los sentimientos positivos de los usuarios sobre este atributo se deben a otros elementos. En algunos casos, encontramos usuarios que describen el ambiente relajado de gente local haciendo deporte.

La pareja *religión\_deporte* es la que tiene más opiniones negativas, con el 9% del total de comentarios conteniendo esta pareja. No obstante, el minúsculo porcentaje de comentarios negativos dificulta la percepción de la causa principal.

### ***Parque temático***

Este atributo está representado por 18 pares de palabras. Entre ellas, las parejas más relevantes son *persona\_parque temático* (人\_遊樂場) que ocupa el 8% del total de comentarios relacionado con parque temático, *museo\_parque temático* (博物館\_遊樂園) el 8%, *turista\_parque temático* (遊客\_遊樂園) el 8%, *dirección\_parque temático* (方向\_遊樂園) el 8%, *montaña\_parque temático* (山\_遊樂園) el 8% y *paisaje\_parque temático* (景\_遊樂園) el 6%.

La imagen positiva relacionada con este atributo se debe a los paisajes bonitos en los parques temáticos, los fabulosos diseños de la arquitectura, las atracciones divertidas, la gran magnitud de atracciones que componen ferias temporales y que se desplazan fácilmente de un lugar a otro.

Por otro lado, la pareja *persona\_parque temático* es la que recibe mayor porcentaje de comentarios negativos (40%). La puntuación negativa se debe a causas muy diversas como en un comentario que menciona el terremoto y la situación de cómo los habitantes refugiaron en el parque infantil.

#### 4.2.1.5. La dimensión de cultura, historia y arte

##### *Artesanía*

Este atributo cuenta con 20 pares de palabras y de las que las más relevantes son *a mano\_ artesanía* (手工\_藝品) apareciendo en el 8% de comentarios totales de artesanía, *viaje\_ artesanía* (旅行\_藝品) en el 7%, *posada\_ artesanía* (客棧\_藝品) en el 7%, *algunos/as\_ artesanía* (一些\_藝品) en el 6%, *ello, pronombre para casa sin vida\_ artesanía* (它\_藝品) en el 6%, *iglesia\_ artesanía* (教堂\_藝品) en el 6% y *arte\_ artesanía* (藝術\_藝術品) en el 6%.

Los sentimientos positivos que poseen los usuarios sobre este atributo principalmente son debidos al intercambio de información sobre lugares de exposición de productos de artesanía y sobre los productos típicos que pueden comprar para ser regalados.

Respecto a la pareja que recibe peores comentarios, encontramos la pareja *vida\_ artesanía* (生活\_藝品) con un 25% de opiniones negativas. Visualizando los contenidos de esta pareja, descubrimos que la puntuación negativa se debe principalmente a otro elemento en los mismos comentarios excepto uno de ellas que se encuentra relacionado con la peligrosidad del sitio donde se venden artesanías.

##### *Gastronomía*

En total existen 10 parejas de palabras con este atributo. La más llamativa es *sopa\_ gastronomía* (湯\_美食) que constituye el 16% del total de comentarios relacionados con gastronomía. Le sigue la pareja *buen vino\_ gastronomía* (美酒\_美食) con el 12%, *tenera\_ gastronomía* (牛肉\_美食) con el 12%, *filete\_ gastronomía* (牛排\_美食) con el 11% y *plato\_ gastronomía* (盤子\_美食) con el 11%.

La buena imagen que los turistas tienen respecto a este atributo es debido a la comida exquisita, la forma de combinar ingredientes, el precio, la forma de pedir la comida y el lugar para degustar gastronomía.

Respecto a las parejas de palabras que obtienen el porcentaje más alto de opiniones negativas, la pareja *visita\_ gastronomía* (巡禮\_美食) es la que recibe la mayor cantidad de votos negativos con un 25%. Revisando los textos, descubrimos que el resultado negativo está

acompañado de otros factores como por ejemplo la dificultad de entender la carta o la falta de tiempo que impide la degustación de la comida.

### ***Festival***

El atributo festival ofrece 35 pares de palabras. Las parejas *festival\_excursión* (節\_遊) que aparece en el 20% de comentarios relacionados con festival, *fuego\_festival* (火\_節) en el 15%, *festival\_fiesta* (節\_節慶) en el 15%, *toro\_festival* (牛\_節) en el 12% y *festival\_fiesta* (節\_節日) en el 11% son las palabras que aparecen con más frecuencia.

Los sentimientos positivos sobre este atributo se deben a las experiencias detalladas y compartidas por otros usuarios.

En relación a las parejas que reciben más comentarios negativos, encontramos las parejas *herencia\_fiesta* (傳承\_節慶), *fiesta\_costumbre* (節慶\_風俗), *personalidad\_fiesta* (個性\_節慶), *cuantioso\_festividad* (大部分\_慶典) y *particuidad\_fiesta* (特質\_節慶). Todos presentan un 40% de comentarios negativos. La mayoría de estas parejas están incluidas en los mismos comentarios que recomiendan conocer la cultura española de forma más profunda. Algunos comentarios mencionan el tema de limpieza en las fiestas. Incluso hay uno de estos comentarios que alude el tema de corrupción política con la organización de fiestas.

### ***Costumbre***

Este atributo ofrece 56 pares de palabras y cada una de ellas se distribuye uniformemente en los comentarios. Las parejas que aparecen con un porcentaje un poco más alto son *color del cielo* u *hora de día\_estilo* (天色\_風) en el 5% del total de comentarios relacionadas con costumbre, *campo\_estilo* (鄉村\_風) en el 4%, *estilo\_estilo* (風\_風貌) en el 4%, *zona rural\_estilo* (鄉間\_風) en el 3%, *costumbre\_bar* (習慣\_酒吧) en el 3%, *costumbre\_carretera* (習慣\_道路) en el 3%, *vocal\_estilo* (人聲\_風) en el 3% y *campo\_estilo* (原野\_風) en el 3%. Sin embargo, la palabra “風” aparte de poder significar estilo, también puede referirse a viento.

A los turistas generalmente les llaman la atención algunas costumbres como son por ejemplo el orden que los camareros tienen en un restaurante al preguntar primero por la bebida antes de la comida, dejar propina no es una obligación pero por costumbre mucha

gente lo hace, las diferencias de horarios de comida, la costumbre de comprar lotería en Navidad, etc.

Por otra parte, la pareja *tráfico\_folk* (交通\_民風) contiene un 67% de comentarios negativos. Visualizando los contenidos, observamos que varios autores mencionan la honestidad de los habitantes de algunos pueblos a pesar de que la falta de medios de transporte provoca problemas de acceso a dichos pueblos.

### **Concierto**

Este atributo ofrece 10 parejas de palabras. De entre ellas, *música\_concierto* (音樂\_音樂會), *sesión\_concierto* (場\_音樂會) y *actividad\_concierto* (活動\_音樂會) son las parejas que aparecen con más frecuencia, ocupando cada una de estas parejas el 31%, el 19% y el 9% de total de comentarios del atributo concierto.

La buena imagen relacionada con el atributo concierto se debe a la magnitud y lo lujoso del concierto, la gran cantidad de conciertos gratis, el ambiente y las diferentes músicas ofrecidas por los mismos.

Respecto a los pares de palabras que reciben opiniones positivas, sólo las parejas *sala de concierto\_concierto* (音樂廳\_音樂會), *calle\_concierto* (街\_音樂會) y *museo\_concierto* (博物館\_音樂會) cuentan con el 20% de comentarios negativos. El resto recibe comentarios positivos. Este porcentaje de opiniones negativas no es un valor elevado. Observando el contenido de los comentarios contenidos con opiniones negativas, observamos algunos que indican problemas de reservar tickets y dudas sobre cómo poder llegar al auditorio sin tener que preocuparse por los robos.

### **Arte**

Este atributo cuenta con 19 pares de palabras lingüísticas. Las parejas de palabras más llamativas son *iglesia\_arte* (教堂\_美術) que ocupa el 16% de comentarios de este atributo, *museo\_arte* (博物館\_美術) el 14%, *arte\_arte* (美術\_藝術) el 13% y *arquitectura\_arte* (建築\_美術) el 13%.

Los sentimientos positivos relacionados con este atributo son fruto de un buen itinerario de visita que combina las actividades de arte con otras actividades turísticas.

Este atributo no ha recibido muchas opiniones negativas. La pareja que obtiene peor puntuación es *punteo\_arte* (橋樑\_藝術) con el 33% de comentarios negativos, principalmente causados por un comentario que critica el museo Guggenheim.

### **Museo**

Existen en total 19 parejas de palabras con este atributo. Las parejas de palabras más relevantes son *museo\_hospital* (美術館\_醫院) que contabiliza el 16% del total de comentarios relacionados con el atributo museo, *museo\_Sofía* (美術館\_蘇菲) el 11%, *museo\_teatro* (博物館\_戲劇) el 11% y *colección\_museo* 收藏\_美術館 el 9%.

La razón por la que este atributo recibe una imagen positiva es debido a los intercambios de información relacionados con museos, como son el horario de apertura, el precio de la entrada o los transportes.

Entre las parejas de este atributo, las parejas que obtienen peor puntuación son *museo\_hospital* (美術館\_醫院) y *museo\_Sant Pau* (美術館\_聖保祿), contando con el 38% de comentarios negativos cada una. Sin embargo, revisando el contenido de estos, descubrimos la pareja *hospital\_Sant Pau* se refiere al hospital Sant Pau. La razón por la que esta pareja obtiene opiniones negativas es debida a otros elementos en los mismos comentarios.

### **Monumento**

En el caso de monumento, este atributo está representado por 38 pares de palabras. Las parejas de palabras más frecuentes son *iglesia\_capilla* (教堂\_禮拜堂) ocupando el 9% del total de comentarios de monumento, seguida de la pareja *naturaleza\_iglesia* (大自然\_教堂) con el 8%, *monumento\_paisaje* (古蹟\_景) con el 7%, *color del cielo u hora de día\_iglesia* (天色\_教堂) con el 5%, *castillo\_Blancanieves* (城堡\_白雪公主) con el 4% y *siglo\_monumento* (世紀\_古蹟) con el 4%.

Los turistas poseen una imagen positiva sobre este atributo debido a la historia, decoración lujosa y el diseño artístico de la arquitectura, etc.

Entre las 38 parejas de palabras, *monumento\_yacimiento* (古蹟\_遺跡) es la pareja que cuenta con más porcentaje de opiniones negativas, en el 50% de comentarios. Estos

comentarios negativos están en relación con la lentitud de reformar o reconstruir de los yacimientos.

### **Religión**

El atributo religión está representado por 10 parejas de palabras. Dentro de ellas, la pareja de palabras *creencia\_religión* (信仰\_宗教) constituye el 16% del total de los comentarios que incluye este atributo, seguida de *religión\_gobernar* (宗教\_統治) con el 16% y *gente\_religión* (世人\_宗教) con el 12%.

Los turistas de este foro prestan mucha atención a la historia, las costumbres y las ceremonias relacionado con este atributo. También se sienten atraídos por el camino de Santiago, la arquitectura de los templos y sus símbolos. Todo esto deja una buena imagen a los turistas.

La pareja de palabras que recibe opiniones negativas con el porcentaje más alto es *animal\_religión* (動物\_宗教), con el 20% y se debe comentarios anti-taurinos.

#### **4.2.1.6. La dimensión de factores políticos y económicos**

### **Política**

Observando los 10 pares de palabras con este atributo, comprobamos que las parejas de palabras más relevantes, la pareja *política\_camino* (政治\_道), *política\_Roma* (政治\_羅馬), *política\_etapa* (政治\_時期), *política\_pan* (政治\_麵包) ocupan el 17%, 15%, 15% y 11% del total de comentarios del atributo Política respectivamente.

Este atributo está relacionado siempre con otros atributos como son la religión y la economía. A los turistas les llama la atención los poderes de la religión católica sobre la política. Ellos relacionan la situación política con el paro y la situación económica, y también mencionan los corruptos políticos y los salarios que cobran. No obstante, no percibimos sentimientos negativos sobre ninguna de estas opiniones.

### **Precio**

En total existen 63 parejas de palabras con este atributo. Las parejas de palabras que aparecen con más frecuencia son *valor\_precio* (值\_價), que contabilizada el 24% de los

comentarios totales de precio, *precio\_nivel de precio* (價\_價位) el 13%, *precio\_paisaje* (價\_風景) el 6% y *precio\_defecto* (價\_缺點) el 5%.

La puntuación positiva de este atributo se debe a las recomendaciones sobre restaurantes con precios razonables y buen servicio, las recomendaciones sobre los transportes, tanto de compañías de alquiler de coches como compañías aéreas de bajo coste.

Entre los 63 pares de palabras, las parejas que reciben más porcentaje de opiniones negativas son *precio\_hotel* (價位\_旅店) y *transporte\_precio* (交通\_行情). Ambas reciben el 50% de comentarios negativos. La puntuación negativa de *precio\_hotel* se debe principalmente a la dificultad de reservar un hotel bueno a un precio razonable. Respecto a la pareja *transporte\_precio*, los comentarios negativos surgen principalmente después de que los turistas compararan los precios de un mismo vuelo y descubrieron la gran diferencia de precios entre varias compañías aéreas.

### ***Economía***

Este atributo ofrece 11 pares de palabras, siendo las parejas de palabras más relevantes la pareja *economía\_información* (商\_資料) ocupando el 46% de los comentarios totales de economía, seguida de *crisis\_economía* (危機\_經濟) con el 8%.

Visualizando los comentarios, este atributo se encuentra emparejado con otros atributos como son información turística, compras y precios. Las parejas de palabras como *precio de productos\_economía* (物價\_經濟) y *economía\_información* (商\_資料) son las que reciben mejor puntuación. La buena imagen de este atributo tiene como origen la información compartida entre usuarios.

Por otro lado, las parejas que reciben peor puntuación con este atributo son *tasa de desempleo\_economía* (失業率\_經濟) y *economía\_móvil* (經濟\_行動電話). Ambas parejas contabilizan el 33% del total de comentarios negativos. En cuanto a las opiniones negativas que surgen de la pareja *desempleo\_economía*, los usuarios mencionan la grave situación de desempleo en España debida a la crisis. En relación a la pareja *economía\_móvil*, todos los comentarios citan el número de emergencias de la Oficina Económica y Cultural de Taipéi para los casos de siniestros o robos.

## ***Seguridad***

Este atributo está representando por 10 pares de palabras. Entre ellas, la pareja de palabras *seguridad\_dinero* (治安\_財) constituye el 27% de comentarios con este atributo, seguida de la pareja *robo\_seguridad* (偷竊\_治安) con el 13% y *violencia\_seguridad* (暴力\_治安) con el 11%. Estas son las palabras que aparecen con más frecuencia en los comentarios relacionados con el atributo seguridad.

La buena imagen que reciben los turistas sobre este atributo se debe a la presencia de policías en varias partes y la advertencia de otros usuarios sobre las zonas peligrosas y precauciones a tomar para evitar los robos.

Las parejas de palabras *violencia\_seguridad* (暴力\_治安) y *ladrón\_seguridad* (搶匪\_治安) son las que reciben más comentarios negativos. Concretamente, aparecen en el 67% de comentarios negativos de ambas parejas. Debido a varios comentarios que relatan sucesos de robos y advertencias de visita a España, generando las opiniones negativas relacionadas con estas parejas de palabra.

### **4.2.1.7.La dimensión de *entorno natural***

#### ***Congestión***

No se detecta ninguna característica lingüística acerca de este atributo en este foro.

#### ***Aglomeración***

No existe ninguna característica lingüística relacionada con este atributo en este foro.

#### ***Contaminación***

No existe ninguna característica lingüística acerca del atributo contaminación en este foro.

#### ***Ciudad y pueblo***

Ocupando el tercer puesto de atributos, con mayor cantidad de comentarios, el atributo “ciudad y pueblo” ofrece 139 pares de palabras lingüísticas. Entre estos pares de palabras, la pareja *establecimiento* o *pueblo\_vino* (莊\_酒) suma el 6% del total de comentarios en relación

a ciudad y pueblo, la pareja *paisaje\_pueblo* (景\_莊) el 4%, *ciudad o castillo\_pueblo* (城\_莊) el 4%, *montaña\_pueblo* (山\_莊) el 3% y *trozo\_pueblo* (片\_莊) el 3%. Sin embargo, destacamos que la palabra “莊” puede referirse a un pueblo, un sitio, una granja o también a un comercio dependiente de los caracteres acompañados de la palabra “莊”.

A los turistas les impresionan el paisaje, la larga historia, el sistema de defensa, el color y la tranquilidad de los pueblos y ciudades españoles.

Por otra parte, entre las parejas de este atributo, las parejas *arriba\_pueblo* (上\_村子), *unidad\_pueblo* (個\_村子) y *ciudad\_pueblo* (城市\_鄉鎮) son las que reciben la mayoría de opiniones negativas, contando cada una de estas parejas con el 57%, el 50% y el 40% de comentarios negativos respectivamente. Los comentarios negativos de *arriba\_pueblo* y *unidad\_pueblo* principalmente son debidos a la dificultad de llegar pueblos de montaña.

### **Limpieza**

Este atributo está representado por 19 parejas de palabras. Las parejas de palabras que aparecen con más frecuencia son la pareja *viaje\_limpieza* (旅行\_衛生) que cuenta con el 9% de comentarios totales del atributo limpieza, seguida por la pareja *año\_limpieza* (年\_衛生), *tres\_limpieza* (三\_衛生), *habitación\_limpieza* (房\_清潔), *posada\_limpieza* (客棧\_衛生) y *todo el mundo\_limpieza* (大家\_衛生) ocupando cada una de estas parejas el 7% de comentarios.

Generalmente, los turistas que tienen una buena imagen sobre la limpieza se debe a que los sitios donde se alojan se encuentran en un estado limpio. Por lo que percibimos, las parejas de palabras que reciben mejor puntuación son *habitación\_limpieza* (房\_清潔), y *posada\_limpieza* (客棧\_衛生).

Respecto a los pares de palabras con más opiniones negativas, la pareja *viaje\_limpieza* (旅遊\_衛生) y *personal\_limpieza* (人員\_清潔) son las que reciben peores puntuaciones, con el 40% y el 38% de comentarios negativos respectivamente. Revisando los comentarios relacionados con *viaje\_limpieza*, observamos que la puntuación negativa se debe al tema de robos. Curiosamente, descubrimos que en varios comentarios recomiendan dejar basura en bolsillos donde los ladrones suelen robar. En cuanto a la pareja *personal\_limpieza*, algunos

comentarios mencionan la suciedad que dejó la gente después de fiestas. Otros comentarios citan al personal de la limpieza de los alojamientos.

### ***Paisaje***

El atributo paisaje ofrece 107 pares de palabras. Los pares de palabras lingüísticas están distribuidos uniformemente. Las parejas de palabras *color del cielo* u *hora de día\_paisaje* (天色\_景) y *paisaje\_natural* (景觀\_自然) ocupan cada una de ellas el 3% de comentarios totales y las parejas *paisaje\_rapsodia* (景\_狂想曲), *paisaje\_montaña alta* (景\_高山), *viento* o *estilo\_paisaje* (風\_風貌), *fondo\_paisaje* (底部\_景), *arista\_paisaje* (芒\_風景), *rotación\_paisaje* (旋轉\_景) y *color del cielo* u *hora de día\_paisaje* (天色\_景色) ocupan cada una el 2% de comentarios totales. Estas parejas son las que aparecen con más frecuencia.

A los usuarios de este foro les impresionan los paisajes naturales no solamente en las zonas rurales sino también en las grandes ciudades.

Por otro lado, la pareja de palabras *paisaje\_pueblo* (景觀\_鄉間) es la que recibe más comentarios negativos y cuenta con el 33% de comentarios negativos. Principalmente, la puntuación negativa se debe a la lluvia y a la dificultad de circular con el coche en un pueblo.

#### **4.2.1.8.La dimensión de entorno social**

### ***Comunicación***

Este atributo está representado por 21 parejas de palabras. Las parejas *comunicación\_vía* o *camino* (溝通\_道), *conmutar\_auto ayuda* (交流\_自助), *conmutar\_viaje* (交流\_旅行), *conmutar\_posada* (交流\_客棧), *conmutar\_foro* (交流\_論壇) y *comutar\_texto original* (交流\_原文) son las que aparecen con más frecuencia, constituyendo el 11%, 10%, 10%, 9% y 9% del total de los comentarios relacionados con el atributo comunicación.

Los turistas mencionan mucho el problema de comunicación con la gente local ya que no suelen manejar idiomas extranjeros. Sin embargo mencionan siempre la simpatía y la amabilidad de la gente local que les prestan ayuda. Estas son la razones por las que los turistas tienen una buena imagen respecto a este atributo.

Por otro lado, la parejas de palabra *móvil\_comunicación* (手機\_溝通) es la que recibe más comentarios negativos, suponiendo el 38% del total de comentarios que aparece esta pareja. Existe una parte de opiniones negativas debido al tema de robos y otro parte debido al tema del idioma.

### ***Habitante***

Este atributo ofrece 10 parejas de palabras. Entre ellas, las parejas de palabras que aparecen con más frecuencia son la pareja *habitante\_valle* (居民\_山谷) que supone el 13% del total de los comentarios de habitante, seguida de la pareja *habitante\_obturación* (居民\_快門) con el 13%, *habitante\_plataforma* (居民\_平台) con el 13%, *tierra natal\_habitante* (家鄉\_居民) con el 11%, *habitante\_llanura* (居民\_平原) con el 11% y *habitante\_montaña* (居民\_高山) con el 11%. La simpatía y la amabilidad de la gente local es la causa de que los turistas poseen una buena imagen sobre este atributo. No percibimos ninguna pareja de palabras que reciba opiniones negativas.

### ***Calidad de vida***

Observando 15 pares de palabras relacionadas con el atributo Calidad de vida, las parejas de palabras *vida\_arista* (生活\_芒), *vida\_designación del yo, si es mujer* (生活\_老娘) y *zona rural\_vida* (原野\_生活) son las que aparecen más frecuentemente, constituyendo cada una de estas el 15%, el 12% y el 9% del total de comentarios sobre el atributo calidad de vida.

La imagen positiva de este atributo se debe principalmente al estilo de vida que los turistas tienen en España y no a la calidad de vida que los habitantes tienen. Es decir, los usuarios comparten sus experiencias de viajar en España contando su vida tranquila y relajada.

Respecto a los pares de palabras que reciben más opiniones negativas, observamos que las parejas de este atributo obtienen generalmente buenas puntuaciones, excepto las parejas *unidad\_vida* (個\_生涯) y *uno\_vida* (一\_生涯) que poseen la mayoría de comentarios negativos (el 20% en ambos casos). La razón por la que estas parejas reciben puntuación negativa es debida a un comentario que cuenta la triste vida de un torero.

#### 4.2.2. Visualización de los comentarios en Qyer

Examinando el conjunto de resultados sobre los atributos del foro Qyer en la tabla 8, comprobamos que temas relacionados con la dimensión de infraestructura general, infraestructura turística y entorno natural son los más hablados. Estas tres dimensiones suman el 23%, 18% y 20% de los comentarios totales respectivamente. Comparándolo con el tema más hablado del foro de Backpackers, la única diferencia es que a los usuarios de Backpackers les importan más los temas de historia, cultura y arte, mientras que los de Qyer se interesan más por el entorno natural. Al igual que el caso de Backpackers, la dimensión que obtiene más comentarios también es el que genera más opiniones negativas.

En el foro de Qyer, los atributos zona rural y casino reciben el 100% de comentarios positivos. Tan solo el atributo contaminación obtiene comentarios negativos, llegando al 30% de sus comentarios totales.

Tabla 8. Resultado de Qyer

Dimensión	Atributos	qyall_content		
		Positivo	Neutral	Negativo
Infraestructura_general	servicio_sanitario	76	0	16
	infraestructura_comercial	370	1	44
	aeropuerto	153	1	29
	puente	114	2	21
	medio_de_transporte	180	0	47
	telecomunicaciones	188	0	34
	puerto	329	3	61
	carretera	1606	9	210
Recursos_naturales	reserva_natural	179	0	10
	flora	116	0	7
	zona_rural	15	0	0
	desierto	48	0	12
	montaña	152	0	20
	lago	111	0	2
	playa	127	0	27
	clima	350	5	55
	fauna	61	0	7
Infraestructura_turística	excursión	903	7	129
	alojamiento	509	2	92
	información_turística	590	1	79
	bar_discoteca	100	2	16
	restaurante	307	0	51
	centro_de_turistas	50	0	9
Ocio_y_entretenimiento_turístico	compras	54	2	4
	vida_nocturna	86	2	12
	casino	43	0	0
	zoo	46	0	16
	deporte	339	0	65
	parque_temático	44	0	4
Cultura_historia_y_arte	artesano	176	1	26
	gastronomía	69	0	2
	festival	313	1	67
	costumbre	399	0	42
	concierto	46	0	10
	arte	262	1	13
	museo	186	0	9
	monumento	235	1	9
	religión	70	0	3
Factores_políticos_y_económicos	política	59	0	1
	precio	1179	14	118
	economía	66	0	1
	seguridad	57	2	17
Entorno_natural	congestión	0	0	0
	aglomeración	0	0	0
	contaminación	9	0	8
	ciudad_pueblo	1278	3	272
	limpieza	211	0	39
	paisaje	1002	4	123
Entorno_social	comunicación	375	0	54
	habitante	52	1	3
	calidad_de_vida	135	0	7

En los párrafos que siguen a continuación exploramos cada dimensión a través de sus atributos. Visualizando el resultado de los pares de palabras lingüísticas de cada atributo, observamos los comentarios de aquellas parejas de palabras que reciben más comentarios positivos/negativos y buscamos las razones por las que reciben opiniones positivas/negativas mediante la revisión de los contenidos.

#### 4.2.2.1. La dimensión de infraestructura general

##### *Carretera*

Este atributo ofrece 126 pares de palabras. Las parejas de palabras que aparecen con más frecuencia son la parejas de palabras *momento\_camino* (時候\_途) que aparece en el 10% del total de comentarios relacionados con este atributo, seguida por *uno mismo\_camino* (自己\_途) en el 10%, *ciudad\_camino* (城市\_途) en el 9%, *avenida\_carretera* (道\_道路) en el 7%, *cosa\_camino* (東西\_途) en el 5% y *Rambla\_calle* (蘭布\_街) en el 5%.

Los usuarios de este foro poseen una buena imagen sobre este atributo debido al buen estado de los carriles de bicicletas, las interesantes callejuelas distribuidas como un laberinto, etc. También los turistas relacionan siempre los paisajes y edificios con este atributo.

Por otra parte, de los 126 pares de palabras, las parejas *calle\_lado pequeño* (街\_邊小), *sección\_palabras que se pronuncian con "Di"* (段\_第) y *zoo\_calle* (動物園\_街頭) son las que reciben peor puntuación, contando con el 60%, el 50% y el 50% de comentarios negativos respectivamente. Debido a un error de análisis sintáctico, la pareja de palabras *calle\_lado pequeño* está separada de una a otra. Se refiere realmente a pequeños comercios en la calle, tanto las tiendas, restaurantes como puestos de venta. Respecto a los comentarios negativos relacionados con la pareja *calle\_lado pequeño*, el tema de los precios es la principal razón por la que esta pareja obtiene puntuación negativa. El mismo error de análisis sintáctico aparece igualmente en la pareja *sección\_ "Di"*, sin embargo en este caso no podemos encontrar el origen de las opiniones negativas debido al amplio significado de estos caracteres chinos. La pareja *zoo\_calle* recibe comentarios negativos por la falta de conexiones de transporte directo para llegar al zoo.

## ***Puerto***

Observando las 28 parejas de palabras del atributo puerto, descubrimos que la pareja de palabras *puerto\_puerto* (港\_港口) aparece en el 49% de los comentarios totales de este atributo como la pareja más relevante.

A los usuarios les impresionan los paisajes que se encuentran junto a los puertos, el ambiente, los monumentos que se encuentran junto a este atributo, la historia relacionada con el desarrollo de puertos, las buenas comidas de los restaurantes cercanos a los puertos, etc.

Por otro lado, las parejas que obtienen mayor número de opiniones negativas son *puerto\_terminal* (港\_航站) y *paisaje\_puerto* (景色\_港灣), contabilizando un 50% de los comentarios negativos. La mala puntuación de *puerto\_terminal* se debe al tema de los precios y la localización. En relación a las opiniones negativas de *paisaje\_puerto*, estos tienen como origen la comparación de estos dos elementos con los de otro destino.

## ***Telecomunicaciones***

En total existen 30 características lingüísticas que han sido emparejadas con este atributo. Las parejas de palabras que aparecen con más frecuencia son la pareja *fax\_teléfono* (傳真\_電話) que suma el 8% de los comentarios totales de telecomunicaciones, seguida de *móvil\_cremallera* (手機\_拉鏈) con el 7%, *fax\_documento original* (傳真\_原件) con el 5%, *paquete\_móvil* (小包\_手機) con el 5%, *móvil\_programa* (手機\_軟件) el con 5%, *carta de certificación\_teléfono* (證明信\_電話) con el 4%, *fax\_solicitante* (傳真\_申請人) con el 4%, *fax\_sello oficial* (傳真\_公章) con el 4%, *fax\_puesto de trabajo* (傳真\_職位) con el 4%, *doble cara\_teléfono* (正反\_電話) con el 4% y *fax\_fecha reciente* (傳真\_近期) con el 4%.

La imagen positiva que poseen los turistas relacionada con este atributo se debe al intercambio de información sobre número de teléfono de administraciones públicas o la embajada, tanto de China como de España. Visualizando los comentarios, comprobamos que la mayoría de sentimientos positivos tienen su origen en la información detallada sobre el procedimiento de solicitud de visado.

Por otra parte, la pareja de palabras *móvil\_rñonera* (手機\_腰包) contabiliza el 67% de comentarios negativos, la pareja *móvil\_cremallera* (手機\_拉鏈) el 63% y *paquete\_móvil*

(小包\_手機) el 55%, siendo las parejas que reciben más opiniones negativas. Casi todos los comentarios de estas tres parejas reciben esta puntuación por el tema del robo del teléfono móvil.

### **Medio de transporte**

Este atributo está representado por 19 pares de palabras. Entre todas, la pareja *transporte\_medio* o *herramienta* (交通\_工具) constituye el 47% de los comentarios totales relacionados con medio de transporte.

La imagen positiva respecto a este atributo tiene su origen en la comodidad y gran cobertura del transporte público.

Respecto a las parejas que reciben más comentarios negativos, observamos las parejas *transporte\_contaminación* (交通\_汙染) con el 67% de comentarios negativos en los totales y *transporte\_conocimiento* (交通\_見聞) con el 67%. El caso de las opiniones negativas de *transporte\_contaminación* es debido a ciertas recomendaciones sobre un menor uso del vehículo alquilado para disminuir la contaminación del aire. Respecto a la pareja *transporte\_conocimiento*, las opiniones negativas son debidas a la mala organización del viaje causada por no conocer los medios de transporte del destino.

### **Puente**

En total existen 14 parejas de palabras con este atributo. Las parejas de palabras *puente\_grande\_puente* (大橋\_橋), *Cristo\_puente* (基督\_橋), *antigua\_Roma\_puente* (古羅馬\_橋) y *otro\_lado\_puente* (對岸\_橋) son las que aparecen con más frecuencia y suponen respectivamente el 22%, 14%, 9% y 9% de los comentarios totales del atributo puente.

Este atributo acompañado de otros elementos proyecta una buena imagen a los turistas. Estos elementos pueden ser monumentos, museos, ciudades, etc. La razón principal por la que estas combinaciones proyectan buena imagen a los turistas es debido a los bonitos paisajes que ofrecen.

Respecto a las parejas con más puntuaciones negativas, la pareja *arriba\_“Hóngqiáo”* (上\_虹橋), *tiempo\_“Hóngqiáo”* (時間\_虹橋) y *2\_“Hóngqiáo”* (2\_虹橋) reciben cada una de estas el 40% de comentarios negativos por la mala opinión del aeropuerto Hongqiao en

Shanghái. Los comentarios relacionados con el aeropuerto Hongqiao aparecen con el atributo puente debido a que la palabra “橋” significa puente. Estos resultados son ruido y quedan eliminados del análisis.

### ***Aeropuerto***

Este atributo ofrece 20 pares de palabras. Entre ellas, las parejas de palabras más relevantes son *planta\_terminal* (樓\_航站) que constituye el 38% de comentarios totales de este atributo, *iglesia\_aeropuerto* (教堂\_飛機場) con el 7%, *tienda libré de impuesto\_terminal* (免稅店\_航站) con el 6% y *tarjeta de embarque\_terminal* (登機牌\_航站) con el 5%.

La buena imagen que poseen los turistas relacionada con este atributo tiene su origen en los intercambios de información sobre las infraestructuras y servicios disponibles en los aeropuertos. Los atributos que se relacionan frecuentemente con aeropuerto son medio de transporte e infraestructura comercial.

Por otro lado, la pareja *puerto\_terminal* (港\_航站) es la que recibe peor puntuación, con un 50% de comentarios negativos. Las razones son similares a las que mencionamos en el caso del atributo puerto, debido al tema de los precios y la localización.

### ***Infraestructura comercial***

En total se han localizado 41 pares de palabras con este atributo. Las parejas de palabras más llamativas son la pareja *día\_puesto* (天\_攤) que suma el 17% de los comentarios totales de este atributo, *tienda\_tienda* (店\_店鋪) el 14%, *tienda\_restaurante* (店\_飯館) el 7%, *viaje\_establecimiento* (旅遊\_莊) el 6% y *tienda\_todo* (店鋪\_都) el 5%.

A los turistas les llaman la atención los quioscos en las calles no solo por su apariencia, sino también por los productos que venden. Además prestan mucha atención a los mercados de artesanía. Les gusta el ambiente en los mercados tradicionales e intercambian información entre ellos sobre precios, servicios y productos de las tiendas de souvenir.

Por otra parte, la pareja *pequeño puesto\_restaurante* (小攤\_餐廳) recibe el 33% de comentarios negativos y es la peor valorada. La razón principal se debe a la congestión de gente.

### *Servicio sanitario*

En total 10 parejas de palabras han sido localizadas con el atributo servicio sanitario. Las parejas de palabras que aparecen con más frecuencia son *Pablo\_hospital* (保羅\_醫院) que suma el 20% de comentarios de este atributo, *puerto\_hospital* (港\_醫院) con el 18% y *Sant Pau\_hospital* (聖保羅\_醫院) con el 14%.

La buena imagen que poseen los turistas respecto a este atributo tiene como origen el Hospital de la Santa Cruz y San Pablo en Barcelona, debido a su belleza y su gran diseño arquitectónico, además de situarse de camino entre diversos puntos de interés como la Sagrada Familia.

Por otro lado, la pareja *hospital\_concierto* (醫院\_音樂會) es la que recibe la peor puntuación, contando con el 50% de comentarios negativos. El precio de los conciertos y el horario de apertura del hospital son las razones principales por la que esta pareja obtiene puntuación negativa.

#### **4.2.2.2.La dimensión de recursos naturales**

### *Reserva natural*

Este atributo está presenta a través de 24 pares de palabras. Entre ellas, las parejas de palabras más relevantes son *naturaleza\_natural* (大自然\_自然), *paisaje\_natural* (景觀\_自然), *animal\_natural* (動物\_自然), *zona protegida\_natural* (保護區\_自然) y *composición\_natural* (組成\_自然), que suman, respectivamente, el 22%, el 11%, el 8%, el 5% y el 5% de los comentarios totales relacionados con el atributo reserva natural.

La admiración que tienen los turistas en relación a este atributo se debe a la gran belleza de varios sitios turísticos como son las islas Canarias, las playas, los parques nacionales, los pueblos o ciudades, además de esculturas y monumentos.

Respecto a las parejas con más opiniones negativas, las parejas *animal\_natural* (動物\_自然), *zona protegida\_natural* (保護區\_自然), *naturaleza\_paisaje* (大自然\_風光) y *naturaleza\_precipicio* (大自然\_懸崖) reciben cada una un 20% de comentarios negativos. Dado que el porcentaje de las puntuaciones negativas no es alto, no podemos determinar las razones obvias de estas palabras que justifican las opiniones negativas.

### **Zona rural**

En cuanto al atributo zona rural, existen sólo 5 pares de palabras. Estos son *cielo o día\_zona rural* (天\_曠野), *ciudad\_zona rural* (城市\_荒地), *zona rural\_paisaje* (曠野\_風景), *iglesia\_zona rural* (教堂\_荒地), e *iglesia\_zona rural* (教堂\_曠野). Todas estas parejas de palabras proporcionan la misma cantidad de comentarios (el 20%) y no reciben ninguna opinión negativa.

Los turistas de este foro poseen una buena imagen sobre este atributo principalmente por los paisajes que ofrece.

### **Flora**

En total existen 22 parejas de palabras con este atributo, siendo las parejas de palabras *planta\_pueblo* (植物\_鎮), *planta\_tropical* (植物\_熱帶) y *animal\_planta* (動物\_植物) las parejas más relevantes, ocupando cada una de ellas el 12%, el 8% y el 7% del total de comentarios relacionados con flora.

La buena imagen que tienen los turistas relacionada con este atributo se debe a la gran variedad de plantas tanto comestibles como de decoración. Los usuarios mencionan también las plantas situadas en el interior de los monumentos o sitios turísticos.

De estas parejas de palabras, las que reciben mayor porcentaje de comentarios negativos son *planta\_pueblo* (植物\_鎮), *planta\_vegetación* (植物\_植被) y *planta\_volcán* (植物\_火山). Cada una de estas parejas cuenta con un 20% de comentarios negativos. La única razón por la que recibe puntuaciones negativas es debida a la dificultad de llegar a destinos naturales con dicha vegetación.

### **Desierto**

En total existen 10 pares de palabras relacionadas con este atributo. Las parejas de palabras más relevantes son la pareja *isla\_desierto* (島\_沙漠) que aparece en el 20% del total de comentarios de este atributo, seguida de la pareja *desierto\_camino* (沙漠\_道) con el 15%, *desierto\_característica* (沙漠\_特色) con el 12% y *desierto\_archipiélago* (沙漠\_群島) con el 12%.

Los sentimientos positivos que poseen los usuarios sobre este atributo se deben principalmente a las dunas existentes en Gran Canaria. Los turistas se sienten atraídos por la orografía del terreno y el espectacular paisaje que componen las dunas.

Por otro lado, la pareja *desierto\_playa* (沙漠\_沙灘) es la que recibe peor puntuación con un 40% de comentarios negativos. Sin embargo, la razón por la que esta recibe la puntuación negativa es debido a la influencia de otros elementos dentro de los mismos comentarios. Es decir, esto constituye ruido y para dar credibilidad a los resultados, este queda eliminado del análisis.

### **Montaña**

Este atributo está representado por 28 parejas de palabras. La pareja *montaña\_montaña alta* (山\_高山) es la más llamativa y suma el 12% de comentarios totales relacionados con el atributo montaña. A continuación le siguen *montaña\_monte* (山\_山地) con el 10%, *montaña\_puerto* (山\_港灣) con el 6%, *ruta\_montaña* (路\_高山) con el 6%, *montaña\_obra de arte* (山\_藝品) con el 6% y *ciudad o pueblo\_monte* (城\_山地) con el 5%.

La buena imagen de este atributo tiene como origen la espectacularidad del paisaje. A los turistas les gusta contemplar la vista paronímica desde sitios elevados.

Entre los pares de palabras de este atributo, la pareja *montaña\_puerto* (山\_港灣) es la que obtiene mayor porcentaje de comentarios negativos con el 40%. Esto se debe a la falta de transporte.

### **Lago**

En total existen 20 parejas de palabras con este atributo. Las parejas más relevantes en todos comentarios relacionados con este atributo son *lago\_natural* (湖\_自然) que suma el 13% de comentarios, *paisaje\_lago* (景色\_湖) el 11%, *palabras pronunciadas con “Zhì”\_lago* (制\_湖) el 9%, *montaña\_lago* (山\_湖泊) el 5%, *lago\_museo* (湖\_美術館) el 5% y *significado\_lago* (意義\_湖) el 5%. Los turistas poseen una buena imagen relacionada con este atributo debido al paisaje natural. Ninguna de las 20 parejas recibe más del 10% de comentarios negativos.

## **Playa**

Este atributo presenta 10 pares de palabras. Entre ellas, las parejas de palabras que aparecen con más frecuencia son *playa\_deporte* (海邊\_運動) que suma el 20% de los comentarios de playa, *surf\_playa* (冲浪\_海邊) el 19%.

A los turistas les llama la atención las actividades que se pueden practicar en la playa en un ambiente animado. También mencionan la comida de los chiringuitos en la playa.

Por otro lado, la pareja de palabras que obtiene peor opinión es *playa\_deporte* (海邊\_運動) con un 29% de comentarios negativos. La razón por la que ocurre esto es debido al tema del clima.

## **Clima**

En total se han localizados 39 parejas de palabras con este atributo. Siendo las parejas de palabras más relevantes en los comentarios de clima la pareja *clima\_lluvia* (天氣\_雨) manifestándose en el 21% comentarios, seguida de *día o cielo\_puesto* (天\_攤) en el 18%, *clima\_pronóstico* (天氣\_預報) en el 9% y *clima\_día nublado* (天氣\_陰天) en el 4%.

La buena imagen de este atributo tiene su origen en la información útil compartida por los usuarios del foro. Visualizando los comentarios, entendemos que los turistas de este foro organizan su itinerario de viaje día a día dependiendo de la previsión meteorológica en tiempo real.

Entre las 39 parejas de palabras, en los comentarios que contienen la pareja *cafetería\_día o cielo* (咖啡廳\_天) existe un 40% de comentarios negativos, seguida de *clima\_tormenta* (天氣\_暴雨) con el 38%. Respecto a la pareja *clima\_tormenta*, la puntuación negativa se genera por las fuertes lluvias que perjudican la experiencia del viaje. La razón por la que la pareja *cafetería\_día o cielo* recibe mala puntuación es debido al tema de seguridad, y en cuyos comentarios se advierte a los turistas tener precaución con la gente en la cafetería. Se trata por tanto de comentarios negativos de otro atributo, por lo que pasa a interpretarse en otra dimensión.

## ***Fauna***

Este atributo ofrece 10 pares de palabras. Como parejas de palabras que aparecen con más frecuencias de encuentran la pareja *animal\_natural* (動物\_自然), suponiendo el 22% de comentarios relacionados con el atributo fauna, *animal\_mil* (動物\_千) el 19% y *animal\_planta* (動物\_植物) el 13%.

La buena imagen de este atributo está relacionada siempre con otros elementos como las zonas rurales, las montañas, la comida, las obras de arte o las corridas de toros. Principalmente esta buena imagen tiene lugar debido a la variedad de animales en España y el realismo de muchas piezas de arte.

Entre todas las parejas de este atributo, la pareja *animal\_planta* es la que recibe peor puntuación. La razón de esta mala percepción se debe a una expresión usada por un autor que describe la situación de robos en España. En este comentario afirma que España es muy bonita pero los lugares más bellos esconden más oscuridades que no se ven. Al igual que las plantas y los animales venenosos, siempre son glamurosos.

### **4.2.2.3.La dimensión de infraestructura turística**

## ***Excursión***

En total existen 57 pares de palabras relacionados con este atributo. Las parejas de palabras más relevantes son la pareja *viaje\_paisaje* (遊\_風景) que suma el 14% del total de los comentarios de este atributo, *economía\_viaje* (經濟\_遊) el 11%, *profundidad\_viaje* (深度\_遊) el 8%, *precio\_viaje* (物價\_遊) el 5%, *viaje\_panorama* (遊\_風光) el 4% y *viaje\_plan* (出遊\_計劃) el 4%.

Este atributo aparece en los textos en que los turistas comparten sus buenas experiencias de viaje, es decir, sus notas de viaje. Por lo tanto, los sentimientos expresados a través de sus contenidos son positivos.

Por otra parte, la pareja de palabras *cafetería\_turismo* (咖啡廳\_觀光) es la que recibe mayor porcentaje de comentarios negativos con un 67%. La razón principal es debido al tema de seguridad.

### ***Alojamiento***

Este atributo está representado por 86 parejas de palabras. Las parejas de palabras se encuentran uniformemente distribuidas en todos los comentarios. Las que aparecen con más frecuencia son *hostal\_juventud* (旅社\_青年) que suma el 6% del total de los comentarios de este atributo, *hostal\_joventud* (旅舍\_青年) el 3% y *albergue\_hostal* (旅店\_旅社) el 3%. A continuación le siguen *hotel* (酒店), *hotel\_albergue* (旅館\_民宿), *zona\_hostal* (區\_旅舍), *cafetería\_hotal* (咖啡館\_旅館), *albergue\_espacio* (民宿\_間), *hotel\_carretera* (賓館\_馬路), *castillo\_hostal* (城堡\_旅社), *hotel\_ropa de cama* (旅館\_被子), *montaña\_hostal* (山\_旅社), *hostal\_cena* (旅社\_晚餐), *paisaje\_hotel* (景象\_酒店), *hotel\_limpieza* (旅店\_衛生) y *sótano\_hotel* (窖\_酒店), apareciendo cada una de ellas en el 2% de comentarios.

La buena imagen que poseen los usuarios sobre este atributo tiene su origen tanto en la buena limpieza, como la localización cercana al punto de visita, el precio razonable y el buen trato personal.

Por otro lado, las parejas que reciben mayor porcentaje de comentarios negativos son *puerto\_hotel* (港灣\_飯店) con un 67% de comentarios negativos, *habitación en oferta\_hotel* (特價房\_酒店) el 60%, *hotel\_albergue* (旅館\_民宿) el 57% y *entero\_albergue* (整\_民宿) el 50%. En el caso de *puerto\_hotel*, la opinión negativa es debido a la influencia de otros elementos en los mismos comentarios. En cuanto a *hotel\_albergue*, la dificultad de reservar habitación en la temporada alta es la primera causa. El alto precio del alojamiento es el motivo principal que la pareja *habitación en oferta\_hotel* y la pareja *entero\_albergue* reciben mala puntuación. Sin embargo, después de revisar los comentarios, descubrimos que la palabra “entero (整)” en la pareja *entero\_albergue* no es significativa para la palabra *albergue*.

### ***Información turística***

En total 99 pares de palabras están relacionados con este atributo. Las parejas de palabras se encuentran distribuidas uniformemente en los comentarios. Las más relevantes son *precio\_información* (價\_消息) e *información\_viento o estilo* (消息\_風), ocupando cada una de ellas el 5% y el 3% de comentarios totales de este atributo respectivamente.

La información compartida por los usuarios sobre el procedimiento de solicitud de visado y la historia de los puntos de visita son las causas principales por lo que los turistas poseen una buena imagen respecto a este atributo.

Por otro lado, las parejas que obtienen más comentarios negativos son *tiempo\_saber* (時間\_見聞) con el 67% de opiniones negativas, *tráfico\_saber* (交通\_見聞) el 67%, *dentro\_saber* (裏\_見聞) el 60%, *cielo o día\_saber* (天\_見聞) el 60%, *yo\_saber* (我\_見聞) el 57% y *vez\_saber* (次\_見聞) el 50%. En relación a *tiempo\_saber*, las opiniones negativas son generadas por la falta de tiempo. La pareja *tráfico\_saber* se debe a la influencia de otros elementos. La pareja *yo\_saber* es ruido por el caculo de PMI. Sin embargo, en otras parejas como *dentro\_saber*, *cielo o día\_saber* y *vez\_saber* son los errores causados por el análisis sintáctico, por lo tanto quedan eliminados del análisis.

### **Bar y discoteca**

Este atributo está representado por 20 parejas de palabras, siendo una de las parejas de palabras con más frecuencia la pareja *camino\_bar* (道\_酒館) que suma el 12% del total de los comentarios de bar y discoteca. Le sigue de la pareja *vida nocturna\_bar* (夜生活\_酒吧) con el 11%, *cafetería\_bar* (咖啡館\_酒吧) con el 7%, *bar\_restaurante* (酒館\_餐廳) con el 7% y *comida rápida\_bar* (快餐\_酒吧) con el 6%.

Los turistas se sienten atraídos por el ambiente y la decoración de los bares. Les gusta catar vino y participar en actividades organizadas por los bares y discotecas.

La pareja que obtiene más comentarios negativos es *cafetería\_bar* (咖啡廳\_酒吧) con un 67%. El tema de seguridad es la primera causa por la que esta pareja obtiene tantas opiniones negativas.

### **Restaurante**

Este atributo ofrece 60 características. Entre estas, la pareja de palabras que aparecen con más frecuencia en los comentarios son *tienda\_restaurante* (店\_飯館) supone el 8% del total de comentarios relacionados con restaurante, seguida de las parejas *calle\_restaurante* (街\_飯館) con el 7%, *pan\_restaurante* (面包\_飯館) con el 3% y *paisaje\_restaurante* (景象\_餐廳) el 3%.

Los turistas tienen una buena imagen relacionada con este atributo principalmente debido a las comidas deliciosas. Revisando los comentarios, observamos que a los usuarios les extraña mucho el acompañamiento de pan y piquitos en cada comida,

Por otra parte, las parejas que reciben más opiniones negativas son *cafetería\_bar* (咖啡廳\_酒吧) con el 67% de comentarios negativos, *cafetería\_turismo* (咖啡廳\_觀光) con el 67% y *cafetería\_ciudad* (咖啡廳\_城市) con el 50%. Todas se deben a temas de seguridad.

### ***Centro de turistas***

En total existen 10 parejas de palabras relacionadas con este atributo. Las parejas *zona\_centro de turistas* (區\_旅遊局), *centro de turistas\_calle* (旅遊局\_街), *museo\_centro de turistas* (博物館\_旅遊局), *sitio\_centro de turistas* (位子\_旅遊局) y *sentimiento\_centro de turistas* (感受\_旅遊局) son las que aparecen con mayor frecuencia en los comentarios relacionados con este atributo. Cada una de estas parejas suma el 19%, el 17%, el 12%, el 10% y el 10% de los comentarios totales.

Visualizando los contenidos, descubrimos que los centros de turistas son sitios donde los turistas inician su viaje a través de los mapas que allí les facilitan. Los turistas poseen una buena imagen sobre este atributo debido a los servicios ofrecidos como son la consulta de información turística, la venta de billetes y la recogida de mapas y publicadas.

Entre estas 10 parejas, *autopista\_centro de turistas* (公路\_旅遊局) es la que recibe el comentario negativo. La razón por la que recibe una puntuación negativa se debe al uso de palabras. A continuación capturamos la frase que el sistema CopeOpi considera negativa.

嚴重推薦他們的旅遊局網站 GRANCANARIA.COM, 甚至可以下載 AUDIOGUIDE 到你的 IPOD 裏。

\*\*\*SCORE=-0.5573313

Su traducción en español es:

Recomienda gravemente su página web oficial de turismo grancanaria.com, incluso puede descargar el audio guía en tu iPod.

La palabra “嚴重” subrayada igual que su traducción suele usarse en situaciones que pone en peligro la vida. Se trata de un error semántico causado por el propio autor. También es un problema que se encuentra frecuentemente en el procesamiento del lenguaje natural. Consideramos este resultado un ruido, por lo que no contabilizamos este en nuestro análisis.

#### 4.2.2.4. La dimensión de ocio y entretenimiento turístico

##### *Compras*

Este atributo ofrece 10 parejas de palabras. Las parejas de palabras que aparecen con más frecuencia son *gran almacén\_compras* (百貨公司\_購物) que suma el 13% de comentarios relacionados con el atributo compras, *Lafayette\_compras* (老佛爺\_購物) el 12%, *billete pagueño\_compras* (小票\_購物) el 12%, *edificio principal\_compras* (主樓\_購物) el 10%, *tarjeta de descuento\_compras* (折扣卡\_購物) el 10% y *precio por unidad\_compras* (單價\_購物) el 10%.

Los usuarios intercambian información de compras como son los lugares para comprar, la mejor forma de pago, productos y servicios. Estas son las razones por las que este atributo recibe una buena imagen.

Respecto a la pareja que recibe mayor porcentaje de opiniones negativas, la pareja *tarjeta de descuento\_compras* (折扣卡\_購物) es la que recibe más comentarios negativos, concretamente el 33% de todos comentarios relacionados con esta pareja. La razón se debe a la diferente forma del uso de la tarjeta de descuento dependiendo de la nacionalidad del usuario o de la normativa de ventas en las mismas cadenas.

##### *Vida nocturna*

En total existen 10 pares de palabras relacionados con este atributo. Las parejas de palabras *vida nocturna\_camino* (夜生活\_道), *vida nocturna\_paisaje* (夜生活\_景) y *vida nocturna\_bar* (夜生活\_酒吧) son las que aparecen con más frecuencia en los comentarios. Cada una de ellas supone el 30%, el 25% y el 13% de los comentarios relacionados con vida nocturna respectivamente.

A los turistas les llama la atención la animada vida nocturna y las actividades que pueden hacerse como tomar una copa, bailar, etc. También tienen una buena imagen de los paisajes y las luces nocturnas.

Por otra parte, las parejas que obtienen más opiniones negativas son *vida nocturna\_camino* (夜生活\_道) y *cafetería\_vida nocturna* (咖啡館\_夜生活). Cada una de ellas cuenta con un 20% de comentarios negativos. Analizando los comentarios de estas dos parejas, no percibimos la razón clave de estas opiniones.

### **Casino**

Este atributo ofrece 10 pares de palabras. Las parejas de palabras más relevantes son la pareja *montaña\_casino* (山\_賭場) que suma el 16% del total de comentarios de casino, la pareja *casino\_restaurante* (賭場\_餐廳) el 14%, *norte\_casino* (北\_賭場) el 14% y *almuerzo\_casino* (午餐\_賭場) el 12%. La imagen positiva de este atributo principalmente es debida al ambiente en los casinos y también a la localización de estos cercanos a algunos puntos de visita. No se detectan comentarios negativos relacionados con las parejas de palabras con este atributo.

### **Zoo**

En total existen 10 pares de palabras con este atributo. Las parejas de palabras más frecuentes son *zoo\_calle* (動物園\_街), presente en el 19% de comentarios, *zoo\_montaña* (動物園\_山) en el 18%, *zoo\_compras* (動物園\_購物) en el 10% y *zoo\_castillo* (動物園\_城堡) en el 10%.

El zoo que los usuarios mencionan es el de Barcelona. Ya que este zoo se encuentra de paso entre otros puntos de interés, los usuarios lo visitan como un punto de descanso. Esta es la razón por la que este atributo recibe buena puntuación.

Por otra parte, la pareja *zoo\_calle* (動物園\_街頭) recibe el 50% de comentarios negativos. Representa la pareja con mayor número de opiniones negativas con este atributo. La razón por la que recibe esta imagen es debida a la influencia de otros elementos que aparecen en los mismos comentarios. Estos resultados son ruido y quedan eliminados del análisis.

## ***Deporte***

Observando 20 pares de palabras con este atributo, la pareja más relevante es *calle\_deporte* (路\_體育) que suma el 8% de los comentarios de este atributo. A continuación le siguen las parejas *surf\_deporte* (冲浪\_運動), *villa\_deporte* (別墅\_運動), *planificar\_deporte* (策劃\_運動), *inmigrante origen de china\_deporte* (華僑\_運動). Cada una de estas parejas suma el 7% del total de comentarios.

Los usuarios suelen mencionar el atributo deporte junto con otros elementos como estadio, partido de fútbol o concierto, por lo que entendemos que estas son las actividades o sitios que a estos turistas les interesan.

Entre las parejas de palabras, las parejas *surf\_deporte* (冲浪\_運動), *villa\_deporte* (別墅\_運動), *planificación\_deporte* (策劃\_運動), *inmigrante\_deporte* (華僑\_運動) son las que reciben mayor porcentaje de comentarios negativos. Sin embargo, revisando los comentarios de estas parejas, observamos que se tratan de un mismo contenido pero está publicado varias veces.

## ***Parque temático***

En total existen 10 parejas de palabras relacionadas con este atributo. La pareja *pueblo\_parque temática* (城\_遊樂場) ocupa el 17% de los comentarios totales de parque temático, seguida de *playa\_parque temático* (海邊\_遊樂場) con el 13% y *montaña\_parque temático* (山\_遊樂場) con el 13%. Sin embargo, descubrimos que la palabra “遊樂場” en este foro también puede referirse a un parque infantil.

A los turistas les gusta el diseño de los parques temáticos, los colores y las atracciones relacionadas con este atributo.

Con este atributo, la pareja de palabras que recibe la peor puntuación es *habitante\_parque temática* (居民\_遊樂場) con un 33% de comentarios negativos. La razón por la que recibe opiniones negativas es debida a comentarios muy diversos como un comentario que menciona que la crisis en el pueblo obliga a los habitantes a salir del mismo, quedando como un pueblo fantasma.

### 1.2.2.5. La dimensión de cultura, historia y arte

#### *Artesanía*

Este atributo cuenta con 30 parejas de palabras. Entre ellas, las parejas de palabras más relevantes son *arte\_obra de arte* (藝術\_藝術品) que suma el 13% de comentarios de este atributo, *pueblo o castillo\_obra de arte* (城\_藝術品) el 6%, *obra de arte\_estilo* (藝術品\_風格) el 5%, *montaña\_obra de arte* (山\_藝術品) el 5%, *manual\_obra de arte* (手工\_藝術品) el 4%, *óleo\_obra de arte* (油畫\_藝術品) el 4% y *obra de arte\_estilo o viento* (工藝品\_風) el 4%.

La buena imagen de este atributo se debe al buen diseño de las obras arquitectónicas, los famosos pintores españoles, los museos donde se exponen las obras, los fabulosos espectáculos en las calles, etc.

La pareja de palabras que obtiene más opiniones negativas es *artesanía\_arte* (工藝品\_藝術) con un 50% del total de comentarios negativos relacionados con esta pareja. La causa principal es debida a la hora de apertura de las tiendas.

#### *Gastronomía*

Este atributo ofrece 10 pares de palabras. Las parejas más frecuentes son la pareja *vino bueno\_gastronomía* (美酒\_美食) que suma el 24% de comentarios, seguida de las parejas *gastronomía\_Nápoles* (美食\_那不勒斯), *año de experiencia de conducción\_gastronomía* (年駕齡\_美食) y *cúpula\_gastronomía* (穹頂\_美食) suponiendo cada una de ellas su presencia en un 11% de comentarios.

Los turistas que poseen una buena imagen sobre este atributo suelen citar la paella, las tapas o el vino junto a este atributo.

Casi todas las parejas con este atributo obtienen opiniones positivas, excepto la pareja *cúpula\_gastronomía* (穹頂\_美食). Este par de palabras es lo que recibe más comentarios negativos con el 13%. Estos sentimientos negativos están provocados principalmente por el tiempo de espera de un ascensor situado bajo una cúpula.

## ***Festival***

En total 47 parejas de palabras han sido localizadas y asociadas con festival. Las parejas de palabras *festival\_natural* (節\_自然), *festival\_calle* (節\_街頭), *festival\_hospital* (節\_醫院), *festival\_fiesta* (節\_節慶), *festival\_vagón* (節\_車廂) y *festival\_terminal* (節\_航站) son las que aparecen con más frecuencia, sumando respectivamente el 16%, el 8%, el 5%, el 5%, el 4% y el 4% de comentarios totales de este atributo.

Los turistas poseen una buena imagen sobre este atributo debido a las bonitas decoraciones, al buen ambiente y las actividades que se desarrollan durante las fiestas.

Por otro lado, las parejas que reciben mayor porcentaje de comentarios negativos son *domingo\_fiesta* (周日\_節假日) con el 63%, *fiesta\_calle* (節假日\_街) con el 60% y *lunes\_fiesta* (周一\_節假日) con el 50%. Las razones por las que reciben opiniones negativas pueden ser por los comercios cerrados durante los días de fiesta, por larga cola para comprar entradas durante la fiesta, por los monumentos cerrados los lunes, etc.

## ***Costumbre***

Este atributo ofrece 65 parejas de palabras. Las parejas aparecen con más frecuencia en los comentarios son *costumbre\_color* (習慣\_色) en el 8%, *política\_estilo* (政治\_風) en el 5%, *vida\_costumbre* (生活\_習慣) en el 5% y el 3% por cada una de las siguientes parejas: *humano\_costumbre* (人情\_風土), *costumbre\_pan* (習慣\_面包), *este momento\_estilo* (此刻\_風), y *costumbre\_estilo* (世俗\_風).

A los turistas les llama la atención la costumbre de la siesta o la costumbre de casarse en la iglesia y más tarde celebrar la boda en un hotel de lujo. Estas son las causas por las que este atributo recibe una buena imagen.

Las parejas *previo\_costumbre* (前者\_習慣) y *costumbre\_fiesta* (世俗\_節日) son las que reciben peor puntuación y cuentan cada una de ellas con un 33% de comentarios negativos. Las opiniones negativas de las parejas anteriores son debidas a las costumbres originales de los turistas no por las costumbres de receptores y la de la siguiente se debe a la historia de la tradición católica como puede ser el tema de la inquisición.

### **Concierto**

Este atributo cuenta con 10 parejas de palabras, siendo las parejas de palabras que más aparecen, la pareja *paisaje\_concierto* (景\_音樂會) que suma el 21% de comentarios de este atributo, seguida de *arte\_concierto* (藝術\_音樂會) con el 14% y *Picasso\_concierto* (畢加索\_音樂會) con el 13%.

La buena imagen tiene su origen en los conciertos gratis y las fabulosas salas de concierto. A los turistas les sorprenden las largas colas que hay que hacer para comprar las entradas a algunos conciertos de pago.

La pareja *hospital\_concierto* (醫院\_音樂會) obtiene el 50% de comentarios negativos y es la que recibe mayor porcentaje de opiniones negativas de todas las parejas relacionadas con concierto. La razón por la que recibe las opiniones negativas es debida al horario de apertura del hospital y del tiempo que dura para restaurarlo.

### **Arte**

En total existen 19 parejas de palabras relacionadas con el atributo arte. Las parejas que aparecen con más frecuencia en los comentarios de este atributo son *arte\_arte* (美術\_藝術) que suman el 17% de los comentarios totales, *arte\_museo* o *galería de arte* (美術\_美術館) el 17%, *museo\_arte* (博物館\_美術) el 12% y *pintura\_arte* (畫\_美術) el 10%.

La buena imagen de este atributo se debe a la belleza de los museos y a las entradas gratis para visitar las fabulosas obras.

Entre todas parejas de este atributo, la pareja *puerto\_arte* (港口\_美術) es la que recibe más comentarios negativos sumando el 25% de comentarios totales. Visualizando los comentarios, observamos que esto se debe a las influencias de otros elementos.

### **Museo**

En total existen 20 parejas de palabras con el atributo museo. Las parejas de palabras más relevantes son *arte\_museo* o *galería de arte* (美術\_美術館) que aparece en el 24% de los comentarios, *museo\_arte* (博物館\_美術) en el 16%, *museo\_obra de arte* (博物館\_藝術品) en el 9% y *museo\_capilla* (博物館\_禮拜堂) en el 5%.

Las entradas gratis o con descuento, las infraestructuras cómodas de los museos, sus decoraciones, sus jardines, exposiciones y las auto-guías son las causa principales por la que los turistas tienen una buena imagen de este atributo.

El par de palabras *museo\_capilla* (博物館\_禮拜堂) que cuenta con el 20% de comentarios negativos del total de comentarios es la que resulta con peor puntuación. La razón por la que recibe opiniones negativas es porque los turistas se pierden con frecuencia en la Sagrada Familia.

### **Monumento**

Este atributo está representado por 40 pares de palabras. Entre ellas, las parejas de palabras más llamativas son *iglesia\_obra de arte* (教堂\_藝術品) que supone el 7% de comentarios, *sendero\_iglesia* (小道\_教堂) el 6%, *iglesia\_capilla* (教堂\_禮拜堂) el 5%, *castillo\_St. Georges* (城堡\_聖喬治) el 4%, *siglo\_monumento* (公元\_古跡) el 4% y el 3% de cada una de las parejas *monumento\_terreno* (古跡\_土地), *monasterio\_capilla* (修道院\_禮拜堂), *monumento\_gobernar* (古跡\_統治), *monumento\_famoso* (古跡\_名勝), *artesanía\_iglesia* (工藝品\_教堂), *capilla\_arte* (禮拜堂\_藝術), *rey\_capilla* (國王\_禮拜堂), *palacio\_capilla* (王宮\_禮拜堂) y *monumento\_reino* (古跡\_王國).

Los turistas poseen una buena imagen relacionada con este atributo debido a las decoraciones espectaculares, la arquitectura y la historia que lo envuelve.

Entre todas las parejas, *castillo\_St. Georges* (城堡\_聖喬治) es la que obtiene mayor porcentaje de comentarios negativos, con el 22%. La causa principal es debido a temas de transporte.

### **Religión**

Con este atributo tenemos 10 parejas de palabras. Las parejas de palabras *religión\_política* (宗教\_政治) que suma el 21% de comentarios del atributo religión, *religión\_nivel* (宗教\_階層) el 12%, *rey\_religión* (君王\_宗教) el 11% y *religión\_oro y plata* (宗教\_金銀) el 10% son las que aparecen con más frecuencia.

Los turistas tienen una buena imagen de este atributo debido a las interesantes historias acontecidas entre católicos y árabes, el gran poder de la religión sobre la política y diversas ceremonias religiosas.

Entre estas 10 parejas de palabras, la pareja *religión\_hereje* (宗教\_異教徒) es la que recibe mayor porcentaje de comentarios negativos, con un 20%. La razón por la que recibe opiniones negativas es debida entre otras a las persecuciones producidas en el siglo XV por temas de religión.

#### 1.2.2.6. La dimensión de factores políticos y económicos

##### *Política*

Este atributo ofrece 10 parejas de palabras. Las parejas de palabras *voluntad\_política* (意志\_政治), *política\_democracia* (政治\_民主), *más tarde\_política* (後世\_政治), *guerra\_política* (戰役\_政治) y *política\_sistema judicial* (政治\_法制) son las que aparecen con más frecuencia en los comentarios. Cada una de ellas está presente en el 15%, el 13%, el 13% y 10% de los comentarios totales de este atributo.

Los turistas se encuentran atraídos por la historia política de España como es la política en la época romana, árabe, de los Reyes Católicos, de la Guerra Civil hasta la situación de Cataluña en la actualidad.

La pareja *política\_hereje* (政治\_異教徒) contiene un 20% de comentarios negativos, siendo la que recibe peor opinión comparada con otras parejas con este atributo. La causa es similar a la pareja *religión\_hereje* del atributo religión, debido a lo que ha pasado en la historia.

##### *Precio*

Este atributo está representado por 85 pares de palabras. Las parejas de palabras *valor\_montaña* (值\_山) que contabiliza el 13% de los comentarios relacionados con el precio, *precio\_valor original* (價\_原價) el 7% y *cosa\_precio* (物\_物價) el 6% son las que aparecen con más frecuencia.

La buena imagen relacionada con este atributo principalmente se debe a los intercambios de información sobre las compras de billetes de transporte y entradas. Los

usuarios comparten sus experiencias de compra y muestran a otros la forma de conseguir un precio más bajo sobre estos.

Por otra parte, la pareja *precio\_callejón* (價錢\_巷子) recibe el mayor porcentaje de opiniones negativas, concretamente un 27%. Revisando los contenidos, observamos que los temas de seguridad e idioma suponen las principales causas.

### ***Economía***

En total existen 10 parejas de palabras con este atributo, siendo las parejas de palabras que aparecen en más comentarios las siguientes: *política\_economía* (政治\_經濟) en el 25%, *economía\_oro y plata* (經濟\_金銀) en el 9%, *cura\_economía* (教士\_經濟) en el 9%, *sistema judicial\_economía* (法制\_經濟) en el 9%, *economía\_doble rey* (經濟\_雙王) en el 9% y *economía\_gasto militar* (經濟\_軍費) en el 9%.

Los turistas relacionan el atributo economía con la política, relacionándolos en la historia de ambas. Visualizando los contenidos, observamos que los turistas piensan que la existencia de monumentos lujosos se debe a un buen estado económico por lo que este atributo recibe una buena imagen, Además, no aparece ninguna pareja que reciba comentarios negativos superando el 10 % de los comentarios.

### ***Seguridad***

En total existen 10 parejas de palabras con este atributo. Este atributo cuenta con el 14% de comentario donde aparece la pareja de palabras *seguridad\_alerta* (治安\_警惕), el 13% con *callejón\_seguridad* (巷子\_治安), el 13% con *cremallera\_seguridad* (拉鏈\_治安), el 13% con *bosillo\_seguridad* (口袋\_治安), etc.

La buena imagen que poseen los turistas se genera a través de la información compartida entre los usuarios que advierten de zonas peligrosas y consejos para evitar robos. También se debe gracias a la presencia de policía en las calles turísticas.

Por otro lado, la pareja *seguridad\_alerta* (治安\_警惕) es la que recibe mayor porcentaje de comentarios negativos con el 36% y principalmente se debe a temas de robo.

### 1.2.2.7. La dimensión de *entorno natural*

#### ***Congestión***

No existe ninguna característica lingüística con este atributo.

#### ***Aglomeración***

No existe ninguna característica lingüística con este atributo.

#### ***Contaminación***

Se localizan 4 pares de palabras con el atributo contaminación. Las parejas de palabras que aparecen con más frecuencia en estos comentarios son *uno\_contaminación* (一\_污染) y *unidad\_contaminación* (個\_污染), cada una de ellas presente en el 29% de comentarios totales.

La buena imagen que obtienen los turistas se debe principalmente a la cantidad de plantas que se encuentran.

Respecto a la pareja de palabras que recibe mayor porcentaje de comentarios negativos, encontramos la pareja *tráfico\_contaminación* (交通\_污染) con el 67%. Esto es debido a comentarios que recomiendan el uso de transportes públicos.

#### ***Ciudad y pueblo***

Este atributo ofrece 124 pares de palabras. Las parejas de palabras que aparecen con más frecuencia son *pueblo\_límite* (城\_際) que constituye el 10% del total de comentarios de ciudad y pueblo, seguido de *tienda\_pueblo* (店\_鎮) con el 10%, *montaña\_pueblo* (山\_鎮) con el 9%, *playa\_pueblo* (海邊\_鎮) con el 7%, *blanco\_pueblo* (白色\_鎮) con el 4% y *ciudad\_deporte* (城市\_運動) con el 3%

A los turistas les llama la atención la gran belleza de los pueblos de montaña y costeros. Les impresionan la belleza de los paisajes y les gusta mencionar la historia de las ciudades y pueblos.

Por otra parte, las parejas *habitante\_pueblo* (居民\_村莊), *pueblo\_carretera* (村子\_路), *mar\_pueblo* (大海\_村落), *pueblo\_puerto* (小鎮\_港灣) y *cafetería\_ciudad* (咖啡廳\_城市)

son las que ofrecen mayor porcentaje de opiniones negativas. Cada una de ellas recibe respectivamente el 67%, el 60%, el 50%, el 50% y el 50% de comentarios negativos. La puntuación negativa de la pareja *habitante\_pueblo* se debe principalmente a la historia sobre la religión y el precio del peaje en autopistas. La pareja *pueblo\_carretera* también se debe al precio del peaje. En relación a la pareja *mar\_pueblo*, las opiniones negativas se deben a los temas de distancia, tiempo y clima. Respecto a la pareja *pueblo\_puerto*, la falta de tiempo impide a visitar pueblos costeros es la causa por lo que recibe opiniones negativas. Los comentarios negativos de la última pareja *cafetería\_ciudad* son debido a temas de seguridad.

### **Limpieza**

Observando los 10 pares de palabras encontrados con este atributo, las parejas de palabras *limpieza\_aseo* (衛生\_衛生間) que constituye el 30% de comentarios totales de este atributo, *limpieza\_viento* (衛生\_風) con el 14% y *cama\_limpieza* (床\_衛生) con el 12% son las que aparecen con más frecuencia.

La imagen positiva de este atributo se debe principalmente a la limpieza de los alojamientos donde los turistas se alojan. Ellos lo que valoran más es la limpieza de la cama y del aseo.

Por otro lado, la pareja *albergue\_limpieza* (旅店\_衛生) que recibe la peor puntuación. La causa principal es debida a comentarios que dudaba sobre la limpieza del albergue. Por ejemplo, a un caso que habla de que el dueño es chino.

### **Paisaje**

En total existen 98 pares de palabras, siendo las parejas de palabras más relevantes las siguientes: *castillo\_paisaje* (城堡\_景) con un 15% de comentarios totales de este atributo, *paisaje\_paisaje* (景\_景色) con el 11%, *paisaje\_excursión* (景\_遊覽) con el 10%, *panorama\_paisaje* (全景\_景) con el 6%, *paisaje\_vista* (景\_景象) con el 4% y *paisaje\_hospital* (景\_醫院) con el 4%.

Los turistas poseen una buena imagen sobre este atributo debido a los bonitos paisajes que rodean a los monumentos, a los sitios rurales, a pueblos y ciudades, a la playa y a los puntos elevados como son las montañas.

Por otra parte, la pareja *paisaje\_puerto* (景色\_港灣) recibe un 50% de comentarios negativos debidos a la influencia de otros elementos, como son por ejemplo los transportes y visados.

### 1.2.2.8. La dimensión de *entorno social*

#### *Comunicación*

Observando 30 pares de palabras con este atributo, las parejas más relevantes en los comentarios relacionados con el atributo comunicación son *comunicación\_punto* (溝通\_點) que contabiliza el 9% de comentarios totales, *central\_comunicación* (中\_溝通) el 8%, *él\_comunicación* (他\_溝通) el 7% con un 6% cada una de las siguientes: *comunicación\_viaje* (交流\_旅遊), *comunicación\_viaje* (交流\_旅行), *comunicación\_tráfico* (交流\_交通), y con un 5% la pareja *comunicación\_primero* (溝通\_第一).

La buena imagen sobre este atributo se debe a la buena actitud y simpatía de la gente local, a pesar de que la mayoría de esta no maneja bien el idioma inglés.

Por otro lado, la pareja que obtiene más opiniones negativas es *comunicación\_teléfono* (溝通\_電話) contando con el 40% de comentarios negativos. Unos comentarios negativos son debido a la dificultad de comunicar con el personal de Hertz, compañía de alquiler de coches. Otros son por los malos servicios y actitud en la embajada española.

#### *Habitante*

Con este atributo tenemos 10 pares de palabras relacionadas. Las parejas de palabras *habitante\_color* (居民\_彩色) que suma el 25% del total de los comentarios relacionados con habitante, *habitante\_aviso* (居民\_須知) el 11%, *edificio\_principal\_habitante* (主樓\_居民) el 11% y *habitante\_Bizantino* (居民\_拜占庭) el 11% son las más relevantes.

La buena imagen de este atributo se genera principalmente por el tema del IVA. Visualizando los comentarios, contemplamos que a los usuarios les importa mucho el tema de la devolución del IVA por lo que expresan un buen sentimiento y conocen muy bien que para la devolución del mismo es requisito no ser habitante local.

Por otro lado, la pareja *habitante\_parque temático* (居民\_遊樂場) es la que recibe mayor número opiniones negativas, contabilizando un 33% de comentarios negativos debidos a comentarios diversos como el que menciona la crisis en el pueblo que obliga a los habitantes a salir del pueblo dejando un pueblo fantasma (véase 居民\_遊樂場 3.txt).

### **Calidad de vida**

En total existen 20 pares de palabras. Dentro de ellas, las parejas de palabras *danza de Flamenco\_vida* (弗朗明哥舞\_生活) constituye el 14% del total de los comentarios que incluye el atributo calidad de vida como uno de sus elementos, *noche en Barcelona\_vida* (夜巴薩羅\_生活) el 13%, *política\_vida* (政治\_生活) el 12%, *seglar\_vida* (世俗\_生活) el 5% y *persona común y corriente\_vida* (常人\_生活) el 5%.

A los turistas les admira el estilo de vida español. Visualizando los comentarios, los turistas mencionan que en cualquier lugar o momento pueden encontrarse con gente cantando o bailando flamenco. También se sienten muy a gusto paseando por los jardines o las playas como lo hace la gente local.

Entre las parejas con este atributo, la pareja *vida\_hereje* (生活\_異教徒) es la que recibe más opiniones negativas, con un 20% de comentarios negativos de los comentarios donde aparecen estas palabras. La razón principal es debida a comentarios sobre condiciones de vida como el que menciona la dura vida de los gitanos a lo largo de la historia.

### **4.3.Evaluación de las técnicas aplicadas**

Para poder conocer el rendimiento de este método, calculamos el valor-F de los comentarios en cada. Esta forma de evaluar el rendimiento del sistema se usa generalmente para valorar los sistemas de recuperación y extracción de información (Ku & Chen, 2010; Fermín L, 2013).

Las métricas empleadas en el valor-F son el cálculo de la precisión (*precision*) y de la exhaustividad (*recall*). Se calculan a través de los números como podemos ver en la tabla a continuación:

Tabla 9. Los componentes para el cálculo del valor- F

		Condición verdadera	
		Correct	Not correct
Condición predictiva	Selected	True positive	False positive
	Not selected	False negative	True negative

En esta tabla tenemos los conceptos verdadero positivo (*True positive*) que representa el número de elementos correctamente reconocidos, falso positivo (*False positive*), que es el número de elementos incorrectamente reconocidos, falso negativo (*False negative*) como el número de elementos reales no reconocidos y verdadero negativo (*True negative*) siendo el número de elementos reales no conocidos. En otras palabras, en nuestro caso, *true positive* se da cuando un comentario es positivo y el sistema lo clasifica positivo correctamente. Por el contrario, si el sistema se equivoca y lo clasifica como neutral o negativo será *False negative*; *true negative* es un comentario neutral o negativo no ha sido clasificado por el sistema como comentario positivo, justo al contrario que *False positive*. Finalmente, *True positive* y *True negative* son los resultados que se procesan correctamente.

La forma de calcular la precisión es medir la proporción de los elementos que fueron clasificados correctamente sobre el total de los clasificados en una misma clasificación como se muestra a continuación:

$$precision = \frac{verdadero\ positivo}{verdadero\ positivo + falso\ positivo}$$

Un valor de precisión igual a uno significa que todos los resultados interpretados son relevantes. A mayor proximidad obtenida a uno, mejor interpretación se obtendrá.

De forma similar, la exhaustividad (*recall*) mide la proporción de los elementos clasificados correctamente entre los que efectivamente pertenecen a la misma.

$$recall = \frac{verdadero\ positivo}{verdadero\ positivo + falso\ negativo}$$

Un mayor ratio de *recall*, una mayor relevancia en la interpretación de los elementos. Es decir, cuanto mayor sea este valor, más cantidad de resultados procesará.

De esta manera, podemos medir el rendimiento del resultado con la medida del valor-F que combina los valores de precisión y exhaustividad, calculándose como sigue:

$$F_{\beta} = \frac{(1 + \beta^2)precision * recall}{\beta^2 * precision + recall}$$

Cuando asignamos un valor de  $\beta$  igual que uno, significa que damos la misma importancia a la precisión que a la exhaustividad. En nuestro este estudio, determinamos  $\beta = 1$  para dar la misma importancia a una y la otra.

En las evaluaciones, tanto en el caso de Backpackers como en el caso de Qyer, escogemos 50 comentarios aleatorios para calcular el rendimiento del resultado. Analizamos 100 comentarios, de los cuales 50 son de Backpackers y 50 de Qyer, y escogemos 5 chino-parlantes para votar el sentimiento de cada comentario ya sea positivo, neutral o negativo. Escogemos este resultado como condición verdadera para poder comparar con los resultados finales, calculando las veces de la interpretación correcta e incorrecta, consiguiendo el valor de la precisión, *recall* y  $F_1$ .

Como resultado de la evaluación, este método del trabajo logró el 60% y el 56% de exactitud sobre las determinaciones de opiniones en el caso de Backpackers y Qyer respectivamente. Obtuvimos una precisión de 0.833, una *recall* 0.5 y un valor- $F_1$  0.624 en el resultado de Backpackers. En cuanto al Qyer, la precisión, la *recall* y el valor- $F_1$  alcanzaron al 0.8, 0.47 y 0.592 respectivamente.

Observamos que en ambos casos obtuvimos un alto valor de precisión y un bajo valor en la *recall*. Esto significa que el sistema devuelve pocos resultados pero la mayoría de ellos se encuentran definidos correctamente. Por lo tanto, consideramos que con este método obtenemos nuestros objetivos con un resultado aceptable.

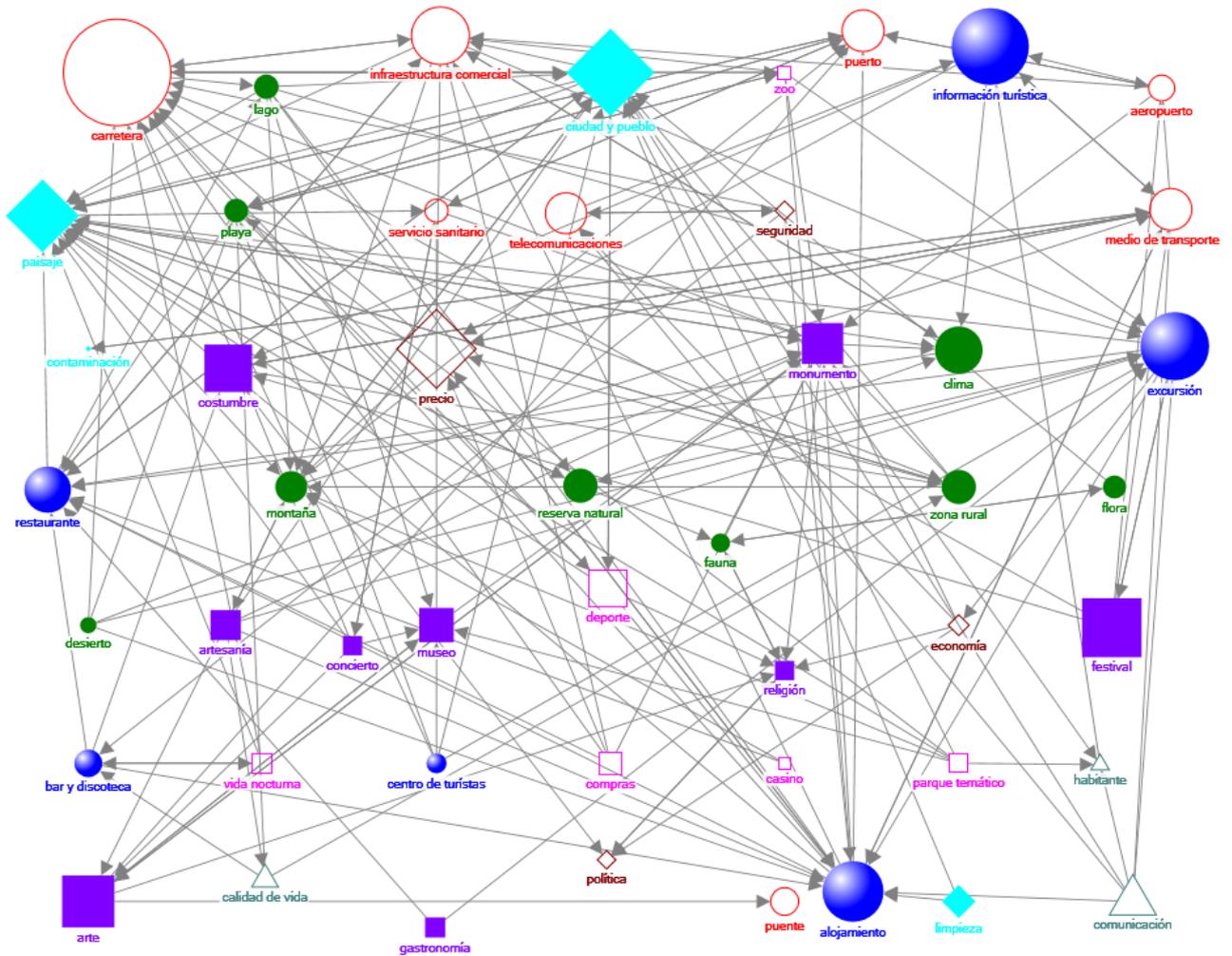
#### **4.4. Imagen de España según los comentarios en las RRSs chino parlantes**

Analizando los resultados anteriores de forma conjunta podemos decir que España se percibe como un país con gran belleza en su paisaje tanto de los recursos naturales como de los sitios turísticos. Aparte de visitar los monumentos, a los turistas les gusta realizar otras actividades en este destino como son visitar los pueblos y ciudades, hacer las compras, participar en los conciertos, ver los bailes y cantos de flamenco. Lo que más les preocupan son el tema de robo, la conexión de transportes y la dificultad de comunicación. Observamos que los turistas chino-parlante se encuentran interesados en atributos relacionados con, en primer lugar, las infraestructuras generales, seguidas de las infraestructuras turísticas, el entorno natural, la cultura, la historia y el arte, los factores políticos y económicos, los recursos naturales, el entorno social y por último, el ocio y el entretenimiento turístico. Es decir, no todas las dimensiones que componen la imagen reciben igual atención por parte de estos turistas. Esto es interesante pues para la valoración global del destino sabemos que un fallo, por ejemplo, en los atributos que componen la dimensión infraestructuras se paga más que en los atributos relacionados con el ocio.

Por otro lado, debemos destacar que si para unos atributos las valoraciones están muy definidas en varias ocasiones nos resulta difícil determinar claramente cuáles son los atributos principales que provocan buena o mala imagen a los turistas. Visualizando los comentarios, comprobamos que en un mismo comentario habitualmente se compaginan varios atributos, y además que les pueden proporcionar carácter positivo o negativo según se relacionen.

Como ya se apunta en la revisión literaria los atributos que componen la imagen del destino no son independientes sino que están interconectados de forma que las apreciaciones y valoraciones de uno influyen en las percepciones y valoraciones de los otros. En la figura 26 se muestran las relaciones existentes entre los distintos atributos que aparecen en los comentarios.

Figura 26. Las interacciones entre los atributos



Cada atributo está representado por una figura geométrica que refleja la dimensión a la que pertenece. De este modo los círculos con línea roja representan a la dimensión de Infraestructura General, los que tienen relleno azul representan a la dimensión de Infraestructura Turística, los rellenos de color verde representan a la dimensión de Recursos Naturales, el cuadrado con línea rosa representa a la dimensión de Ocio y Entretenimiento turístico, el relleno de color violeta representa a la dimensión de Cultura, Historia y Arte, los triángulos representan a la dimensión de Entorno Social, los rombos rellenos de color celeste representa a la dimensión Entorno Natural, con línea marrón representa a la dimensión de Factores Políticos y Económicos.

El tamaño de la figura que representa cada atributo magnifica la cantidad de comentarios donde aparece dicho atributo, es decir, refleja el peso de ese atributo o lo que es lo mismo la importancia de este para el turista.

Por último, cada flecha indica el origen de la influencia que cada atributo recibe de otro atributo. Analizando las cantidades de flechas que reciben, podemos decir que el atributo más influyente es el paisaje, afectando a 14 atributos. El atributo más influenciado por otros es el alojamiento según las cantidades de flechas lanzadas.

A través de este gráfico, confirmamos la siguiente teoría sobre la imagen de destino que es “la imagen de destino es un sistema de interacción entre pensamientos, opiniones, sentimientos, visualizaciones y comportamientos que tienen los visitantes respecto a un lugar determinado (Tasci, Gartner, & Cavusgil, 2007, pág. 200)”. Con esto abordamos el último objetivo específico del segundo objetivo general.

# **CONCLUSIONES, LIMITACIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN**

## 5. CONCLUSIONES, LIMITACIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN

**Resumen:** En este último capítulo, destacamos las ventajas de este método de trabajo y su contribución en la parte teórica de los estudios de la imagen de destino y las RRSS chino-parlantes. Además, ofrecemos algunas aportaciones para mejorar el rendimiento del modelo propuesto, planificando tareas para una futura línea de investigación.

En el presente trabajo se diseña una técnica de extracción y análisis de los comentarios en redes sociales sobre todo de los chino-parlantes. Usamos un método novedoso y automático que se basa en la implementación de módulos de programación desarrollados con el lenguaje Python y que integra la teoría de la imagen de destino, el tesoro de la OMT y la base de datos léxicas del idioma chino E-howNet a nuestros parámetros. De esta forma conseguimos, en primer lugar, extraer los comentarios de forma automática desde determinadas RRSS, construir una lista de atributos para medir la imagen de destino, crear listas de palabras claves en chino que corresponden a cada atributo, detectar cuáles son las características lingüísticas más relacionadas, y por el último, lograr extraer los comentarios relacionados con la imagen de destino.

En el último paso, mediante la conexión con el sistema de CopeOpi, conseguimos la valoración individual de cada comentario, la valoración conjunta a través de las parejas de palabras y la valoración global de los atributos. Gracias a esto descubrimos la opinión y la imagen que los usuarios o turistas chino-parlantes tienen sobre un destino.

Por todo ello, nuestra propuesta, se trata de un flujo de trabajo novedoso en varios aspectos:

Primero, la forma de estructurar el mapa de la imagen se basa en los comentarios de los foros. Esto cumple nuestro objetivo de obtener respuestas espontáneas de los turistas y de conocer qué es lo que realmente les importa cuando viajan a un destino con representatividad en los resultados.

Segundo, se propone una metodología automática lo que permite respuestas ilimitadas y menos posibilidades de sesgo y gran ahorro de tiempo. Debido a que es automática y secuencial, nuestra técnica propuesta es mucho más rápido que las técnicas manuales o semiautomáticas, que requieren introducir datos a los diferentes programas a cada paso, y disminuye además la propensión de sesgo subjetivo. Además, el método propuesto elimina la necesidad de leer a través de contenido publicado y clasificar diversas palabras.

Tercero, este método de trabajo nos permite analizar detalles concretos de los comentarios. Ahorra mucho tiempo en la visualización de cada comentario y la clasificación de palabras pero a la vez, es flexible y, si queremos conocer más detalle, tan solo necesitamos contemplar los comentarios que nos interesan (que sea con opinión positiva o negativa), en lugar de todos los comentarios relacionados con estas parejas de palabras. Incluso, podemos estudiar la nota individual de cada comentario para conocer los puntos fuertes y débiles de nuestros atributos. Así, los investigadores con el modelo propuesto sólo necesitan estudiar el resultado del último paso para obtener una imagen completa.

Cuarto, al trabajar con comentarios y pares de palabras, es decir palabras relacionadas, se revela como un método útil para rastrear no solo valoraciones de los atributos sino también sentimientos y opiniones sobre esos atributos y la importancia que le dan a cada uno de ellos.

Quinto, se trata de un método que da muy buenos resultados para trabajar con la lengua china. Lengua que hasta ahora había sido un hándicap en el estudio del lenguaje natural. La técnica propuesta identifica correctamente las opiniones y el sentimiento general de que los usuarios tienen con respecto a un destino en particular.

Independientemente de las aportaciones metodológicas su contrastación empírica sobre el destino España supone de uno de los primeros estudios que analizan la imagen de un destino para los turistas chino-parlantes.

Esta contrastación empírica también nos ha permitido constatar que la imagen de un destino es un sistema de interacción entre pensamientos, opiniones, sentimientos, visualizaciones y comportamientos que tienen los visitantes respecto a un lugar determinado tal y como Tasci, Gartner, & Cavusgil, comentaron (2007).

Sin embargo, esta metodología novedosa es susceptible de mejora, como es el caso de la implementación del sistema CKIP. Este programa nos facilita enormemente el trabajo del análisis sintáctico de los textos chinos, sin embargo, este sistema no es capaz de separar los caracteres correctamente a la hora de analizar los nombres propios, como por ejemplo los nombres de lugar. Poniendo como ejemplo Barcelona, podemos encontrar en el foro de Backpackers y Qyer varias palabras en chino que utilizan distintos caracteres y que todas ellas significan Barcelona como podemos ver a continuación:

巴塞羅(NB)那(NEP),

巴賽羅納(NB),

巴塞(NB),

巴薩(NB),

巴塞隆納(NC),

BARCELONA (FW)

Cada una de estas palabras significa Barcelona pero están compuestas por diferentes caracteres tanto en idioma chino como en el original. La razón principal es debido a que los chino-parlantes suelen traducir fonéticamente un nombre propio extranjero a su idioma. La traducción de dicho nombre podría variar dependiendo del autor de cada comentario, puesto que existen muchos caracteres chinos que se pronuncian con el mismo sonido. Por otra parte, el sistema asigna una etiqueta distinta a cada palabra, como podemos ver en el ejemplo anterior 巴塞羅 (Nb) 那 (Nep). Esta palabra se encuentra dividida en dos partes porque en la base de datos no aparece 巴塞羅那 como una palabra en sí. Nb significa nombre propio y Nep significa pronombre demostrativo. Por lo tanto, el trabajo de buscar y clasificar palabras de este tipo se ha convertido en una de las necesidades de mejora del rendimiento de esta metodología.

En el proceso de extracción de características, creamos la primera estructura de la imagen de destino en chino con el fin de extraer la opinión y sus características. Sin embargo, a la hora de asignar las palabras chinas y sus sinónimos a los atributos, detectamos un obstáculo. En muchos casos, un único carácter chino cambia su significado cuando se une a otro carácter. Por ejemplo, en el caso del carácter “莊” que aparece en el atributo *ciudad pueblo* y también lo de *infraestructura comercial*. Cuando añadimos el carácter “布” (significa tela) delante de “莊”, el conjunto de estos dos forma una palabra que significa tienda de tela (布莊); si añadimos el carácter “農” (que significa agricultura) delante de “莊”, entonces toma el significado de granja o finca (農莊); y si ponemos “村” delante de “莊”, los dos juntos significan pueblo (村莊). Debido a esto, recomendamos especialmente tener precaución con las palabras de un único carácter.

Aparte de este caso, detectamos algunos errores de interpretación de opinión sobre todo en los comentarios largos que contienen varios atributos. Se supone que los comentarios de los dos foros mencionados se encuentran clasificados por temas, sin embargo, en algunos casos sus contenidos podrían estar involucrados en múltiples temas, como por ejemplo en las notas del viajero. Este tipo de comentario suele ser muy largo y estar asociado a muchos atributos. Calculamos el valor de PMI de los *features* basándonos en comentarios en lugar de las frases individuales que los componen, ya que los *features* que obtienen un alto valor de PMI no se encuentran siempre en la misma frase junto a su atributo. En ocasiones podemos localizar un atributo en una parte del texto y su característica en otra. Estos dan a una pareja de palabras la posibilidad de ser afectada por otros atributos y sus características. Es por ello que, en algún caso observamos que CopeOpi detecta un comentario que contiene una pareja de palabras negativo, pese a que realmente los contenidos donde se sitúa la pareja de palabras no demuestran opiniones negativas. Para un futuro trabajo, recomendamos que se pruebe a calcular el valor de PMI basándose en las frases o párrafos para examinar si a través de este método se puede aumentar la precisión de esta metodología.

Por otra parte, en ocasiones el elevado valor de PMI ocurre cuando se haya la existencia de alguna palabra poco usada. Palabras que aparecen en pocos comentarios pero coocurren siempre juntas con los atributos que definimos pueden resultar con una puntuación alta de PMI. Para ello, recomendamos que se cuente la frecuencia de palabras antes de emparejar un atributo con característica en un futuro trabajo.

Estas revelaciones sobre los errores cometidos son pasos necesarios para tener éxito y nos ofrecen pistas para la mejora del rendimiento de este método de trabajo. También nos permite desarrollar una futura línea de investigación. En vista de que no existe ninguna referencia turística en chino, observamos la necesidad de crear un corpus turístico chino basado en los comentarios o materiales on-line. Más adelante, desde ahí podremos crear un tesoro chino de la imagen de destino donde se detallan los atributos turísticos, sus características y los subordinantes de las características, incluyendo los sinónimos de las palabras, así como una ontología de la imagen de destino.

# **CONCLUSIONS, LIMITATIONS AND FUTURE INVESTIGATIONS**

## 5. CONCLUSIONS, LIMITATIONS AND FUTURE INVESTIGATIONS

**Summary:** In this final chapter, we highlight the advantages of this proposed method and its contribution to the studies of destination images and social-networks analyses in Chinese. We also provide recommendations to improve the performance of the proposed model and plan tasks for future research.

The large number of Chinese visitors and the enormous amount of revenue raised by them have promoted great interest in knowing and studying the behaviour of these tourists. Accompanying by the development of new technologies, online comments have become a new source of information to reach real feelings and direct opinions of these tourists. However, there are few researchers who devote to studies related to these topics because of the difficulty in interpreting Chinese language and the use of different social-networks platforms.

To resolve these problems, this thesis proposed a novel and automatic method for extracting opinions of tourists from social media. Based on the modules that we developed with Python, we incorporated the destination image theory, the thesaurus built by UNWTO and the Chinese lexical database E-how Net through which we successfully extracted comments from certain social networks automatically, build a list of attributes to measure the image, created lists of keywords in Chinese that correspond to each attribute, detected significant tourism-related attributes and associated features related to the destination image.

Finally, with the aid of CopeOpi, an individual assessment of each comment, the joint assessment by the word-pairs and a global assessment of the building attributes were provided. The proposed technique correctly identifies the opinions and the overall image that Chinese tourists have regarding Spain.

Through this workflow, we can work with massive Chinese data to correctly identify tourism attributes according to parameters that were determined a priori. With the proposed method, we save a lot of time in gathering online comments, viewing each context and classifying keywords. To have more details of each attribute, we merely need to analyze comments with extreme views, as we did in the section 4.3. and 4.4. Also, this method allows us to study the individual score of each comment for understanding the original of bad images and vice versa.

Despite these advantages, this novel method has some limitations.

The implementation of CKIP system greatly facilitates our work of parsing Chinese texts; however, this system is not support the parsing of proper names, such as place names. For example, we found the term Barcelona was introduced in different Chinese characters on Qyer and Backpackers forums as we can see below:

*巴塞羅(NB)那(NEP),*

*巴賽羅納(NB),*

*巴塞(NB),*

*巴薩(NB),*

*巴塞隆納(NC),*

*BARCELONA (FW)*

Each of these words means the same Barcelona, but is displayed by different Chinese characters or foreign language. It is because that Chinese users use to translate foreign names phonetically to their language. The translation of these terms may vary depending on the author of each comment since many Chinese characters are different but are pronounced in the same tone. Moreover, CKIP assigns different post-tags to each word, as we see in the example 巴塞羅 (Nb) 那 (NEP) above. This word is divided into two parts because the word 巴塞羅那 is not in CKIP database. Therefore, CKIP considers 巴塞羅 as Nb (means proper name) and 那 as Nep (means demonstrative pronoun). Regarding above statement, we recommend future researchers keep working on categorizing proper words related to the destination to improve the performance of this method.

In the process of feature extraction, we created the first image structure in Chinese for analyses of destinations. Notwithstanding, in the process of attribute assignment, we encountered obstacles related to synonyms of attributes, which contain only one letter or character. In many cases, a single Chinese character, such as “莊” changes its meaning when attached to another character. For example, when we add the character “布” (means cloth) before “莊”, the set of these two forms a word which means fabric (布莊); if we add the character “農” (means agriculture) in front of “莊” then takes the meaning farm (農莊); and if we “村” before “莊” the two together mean town (村莊). Because of this, we strongly recommend caution in assigning the words of one character as attributes.

Apart from this problem, we detected some misinterpretation of opinion especially in long comments that contain multiple attributes. It is assumed that the comments of these two forums are classified by subject, however, in some cases contents could be involved in many issues, such as notes of the traveller. This type of comment is often very long and be associated with many attributes. We calculate the value of PMI based on posts rather than sentences, and therefore the features they get a high value of PMI is not always in the same sentence with his attribute. Occasionally, we found attribute and its characteristic in separating paragraphs. These give word-pairs the possibility of being affected by other attributes or characteristics. That is why, in some cases we see that CopeOpi detects a comment that contains a word-pair negative, however, when we pass this comment to Chinese speakers, no one considers it a negative opinion. For future study, we recommend that investigator calculate the value of PMI based on phrases or paragraphs to examine whether through this change can improve the accuracy of the methodology of this thesis.

Moreover, sometimes the high value of PMI occurred when there is a no frequent word. Words that appear in few comments but always co-occur with attributes may have a high PMI value. To do this, we recommend consulting word frequency before creating word-pairs with attributes in future study.

These revelations about mistakes are necessary steps to succeed and offer clues to improving the performance of this method of work and these are entrenched problems that researchers face in this field. All statements mentioned above allow us to develop a future line of research. There is always a preliminary step that involves refining the databases used for parsing. A sufficiently adequate Chinese corpus of tourism destinations and terms would

considerably improve the performance of the proposed method. As a basis for future work, we strongly urge scholars to build a Chinese tourism thesaurus oriented toward research in analysing opinions of tourism destinations. Such a thesaurus should contain details of tourism-related attributes and products, along with their synonyms, as well as web ontology for a deep and integrated understanding of opinions with regard to tourism destinations.

# BIBLIOGRAFÍA

## BIBLIOGRAFÍA

- Albacete Sáez, C. A., & Herrera Fernández, M. (2012). *Las tecnologías de la información y comunicación en el sector turístico : una aproximación al uso de las redes sociales*. Madrid: Centros de Estudios Ramón Areces.
- Alcázar Martínez, B. d. (2002). *Los canales de distribución en el sector turístico*. Madrid: ESIC,2002.
- Aldebert, B., Dang, R. J., & Longhi, C. (2011). Innovation in the tourism industry: The case of Tourism@. *Tourism Management* , 32, 1204-1213.
- Alén González, M., Comesaña Rodríguez, L., & Fraiz, J. (2009). Análisis de los determinantes del uso de fuentes de información previas al viaje y de las oficinas de turismo en destino. *Estudios y Perspectivas en Turismo* , 18 (5), 546-566.
- Baidewei. (5 de mayo de 2014). *El turismo chino: ¿qué busca?* Recuperado el 1 de Octubre de 2015, de <http://www.baidewei.cat/es/turismo-chino/>
- Baloglu, S., & McCleary, K. W. (1999). A model of destination image. *Annals of Tourism Research* , 26 (4), 868-897.
- Banyai, M. (2011). Dracula's image in tourism: Western bloggers versus tour guides. *Journal of Travel & Tourism Marketing* , 28 (7), 689-719.
- Berli, A., & Martín, J. D. (2004a). Factores influencing destination image. *Annals of Tourism Research* , 31 (3), 657-681.
- Berli, A., & Martín, J. D. (2004b). Tourists' characteristics and the perceived image of tourist destinations: a quantitative analysis—a case study of Lanzarote, Spain. *Tourism Management* , 25, 623-636.
- Bell, C. (2008). 100% PURE New Zealand: Branding for back-packer. *Journal of vacation Marketing* , 14 (4), 345-355.
- Betancort García, B., Ocón Carreras, A., Galán Moreno, M., & Rubio Royo, E. (2008). Estudio de los portales turísticos de canarias como organizaciones de marketing de destinos. *Turitec 2008. VII Congreso Nacional Turismo y Tecnologías de la*

- Información y las Comunicaciones* (págs. 201-215). Málaga: Escuela Universitaria de Turismo.
- Bieger, T., & Laesser, C. (2004). Information sources for travel decisions: Toward a source process model. *Journal of Travel Research* , 42 (4), 357-371.
- Boo, S., & Busser, J. A. (2005). The hierarchical influence of visitor characteristics on tourism destination images. *Journal of Travel & Tourism Marketing* , 19 (4), 55-67.
- Bull, A. (1994). *La Economía del Sector Turístico*. Madrid: Alianza Editorial.
- Burgueño, P. F. (2 de marzo de 2009). *Blog personal de Pablo Burgueño*. Recuperado el 25 de marzo de 2014, de Clasificación de Redes Sociales: <http://www.pabloburgueno.com/2009/03/clasificacion-de-redes-sociales/>
- Camprubí, R., Guia, J., & Comas, J. (2008). Destination Networks and Induced Tourism Image. *Tourism Review* , 63 (20), 47-58.
- Castañeda, L. Q., & Gutiérrez, I. P. (2010). Redes Sociales y otros Tejidos Online para Conectar Personas. En L. Q. Castañeda, *Apredizaje con Redes Sociales: Tejidos Educativos para los Nuevos Entornos* (págs. 17- 39). Sevilla: MAD, S. L.
- Chen, C. M., Chen, S. H., & Lee, H. T. (2010). Assessing Destination Image Through Combining Tourist Cognitive Perceptions with Destination Resources. *International Journal of Hospitality & Tourism Administration* , 11, 59-75.
- Chen, F.-W. (2010). *Análisis de "Buenas prácticas" para la Promoción del Destino a Través de la Técnica del Benchmarking*. Málaga: Universidad de Málaga.
- Chen, H.-J., Chen, P.-J., & Wang, M.-H. (2012). Application of Perception-Promotion Matrix Model—The Case of Kaohsiung City. *Journal of Travel & Tourism Marketing* , 29 (2), 163-184.
- Chen, J. S., Chang, L.-L., & Cheng, J.-S. (2010). Exploring the Market Segments of Farm Tourism in Taiwan. *Journal of Hospitality Marketing & Management* , 19, 309-325.
- Chen, K.-J. (2013). *Lexical Semantic Representation and Semantic Composition. An Introduction to E-HowNet*. Taipei: Institute of Information Science, Academia Sinica.

- Chen, K.-J., & Liu, S.-H. (1992). Word identification for Mandarin Chinese sentences. *Proceedings of the 14th conference on Computational linguistics - Volume 1* (págs. 101-107). Nantes, France: Association for Computational Linguistics.
- Choi, S., Lehto, X. Y., & Morrison, A. M. (2007). Destination image representation on the web: Content analysis of Macau travel related websites. *Tourism Management* , 28 (1), 118-129.
- Chon, K. (1990). The role of destination image in tourism: a review and discussion. *The Tourist Review* , 45 (2), 2-9.
- Church, K. W., & Hanks, P. (1990). Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics* , 16 (1), 22-29.
- Cooper, C., Fletcher, J., Gilbert, D., & Wanhill, S. (1993). *Tourism: Principles and Practice*. London: Pitman Publishing.
- Cornelissen, S. (2005). Producing and imaging 'place' and 'people': the political economy of South African international tourist representation. *Review of International Political Economy* , 12 (4), 674-699.
- Correia Loureiro, S. M., & Miranda González, F. J. (2008). The importance of Quality, Satisfsaction, Trust, and Image im relation to Rural Touris Loyalty. *Journal of Travel & Tourism Marketing* , 25 (2), 117-136.
- Cousin, S. (2008). The nation state as an identifying image : Traditions and stakes in tourism policy, Touraine, France. *Tourist Studies* 8:193 , 8 (2), 193-209.
- Cox, C., Burgess, S., Sellitto, C., & Buultjens, J. (2009). The Role of User-Generated Content in Tourists' Travel Planning Behavior. *Journal of Hospitality Marketing & Management* , 743-764.
- Crespo, R. M., & García, J. J. (2010). La Madeja Tecnificada. En L. Castañeda Quintero, *Aprendizaje con Redes Sociales: Tejidos educativos para los nuevos entornos* (págs. 63-90). Sevilla: MAD, S.L.

- Crompton, J. L. (1979). An assessment of the image of Mexico as a vacation destination and the influence of geographical location upon that image. *Journal of Travel Research* , 17 (4), 18-24.
- Djafarova, E., & Anderson, C. (2008). Contribution of figurative devices to representataion of tourism images. *Journal of Vacation Marketin. Vol. 14 Num. 4* , 14 (4), 291-303.
- Dolnicar, S., & Grün, B. (2013). Validly measuring destination image in survey studies. *Journal of Travel Research* , 52 (1), 3-14.
- Dong, Z., & Dong, Q. (1993). *HowNet Knowledge Database*. Recuperado el 30 de agosto de 2014, de <http://www.keenage.com/>
- Dorwart, C. E., Moore, R. L., & Leung., Y.-F. (2006). Visitor Employed Photography: Its potential and use in evaluating visitors' perceptions of resource impacts in trail and park settings. *Proceedings of the 2006 Northeastern Recreation Research Symposium* , GTR-NRS-P-14, 307-315.
- Echtner, C. M., & Ritchie, J. B. (2003). The meaning and measurment of destination image. *The Journal of Tourism Studies* , 14 (1), 37-48.
- Echtner, C. M., & Ritchie, J. R. (1993). The measurement of destination image: An empirical assessment. *Journal of Travel Research* , 31 (3), 3-13.
- Ekinci, Y., & Hosany, S. (2006). Destination Personality: An Application of Brand Personality to Tourism Destinations. *Journal of Travel Research* , 45, 127-139.
- Elliot, S., Papadopoulos, N., & Kim, S. S. (2011). An integrative model of place image : Exploring relationships between destination, product, and country images. *Journal of Travel Research* , 50 (5), 520-534.
- Feighey, W. (2003). Negative Image? Developing the Visual in Tourism Research. *Current Issues in Tourism* , 6 (1), 76-85.
- Fermín L, C. M. (2013). *Extracción de Opiniones sobre Características: Un Enfoque Práctico Adaptable al Dominio*. Sevilla: Universidad de Sevilla.
- Font, X. (1997). Managing the Tourist Destination's Image. *Journal of Vacation Marketing* , 3 (2), 123-131.

- Forga, J. M., & Valiente, G. C. (enero-junio de 2013). La participación en redes sociales y su incidencia sobre el comportamiento y satisfacción de los consumidores del turismo. Un estudio comparativo en diferentes recursos de turismo industrial en Cataluña, Alsacia y Escocia. *Investigaciones Turísticas Nº 5* , 29-59.
- Frochot, I. (2003). An analysis of regional positioning and its associated food images in French tourism regional brochures. *Journal of Travel & Tourism Marketing* , 14 (3-4), 77-96.
- Gallarza, M. G., Saura, I. G., & García, H. C. (2002). Destination Image Towards a Conceptual Framework. *Annals of Tourism Research* , 29 (1), 56-78.
- García, B. B., Carreras, A. O., Álvarez, L. A., & Royo, E. R. (2010). Estudio sobre la utilización de plataformas de red social por las OMDs. *TuriTec 2010, VIII Congreso Nacional Turismo y Tecnologías de la Información y las Comunicaciones* (págs. 47-59). Málaga: Escuela Universitaria de Turismo.
- Garrod, B. (2009). Understanding the Relationship between Tourism Destination Imagery and Touris Photography. *Journal of Travel Research* , 47 (346), 346-358.
- Goodrich, J. N. (1978). A New Approach to Image Analysis through Multidimensional Scaling. *Journal of Travel Research* , 17 (2), 2-7.
- Govers, R., & Go, F. M. (2004). Cultural identities constructed, imagined and experienced: A 3-gap tourism destination image mode. *Tourism* , 52 (2), 165-182.
- Govers, R., & Go, F. M. (2005). Projected destination image online: Website content analysis of pictures and text. *Information Technology & Tourism* , 7 (2), 73-89.
- Govers, R., Go, F. M., & Kumar, K. (2007). Promoting Tourism Destination Image. *Journal of Travel Research*.46:15 , 46, 15-23.
- Gretzel, U. (2006). Consumer Generated Content - Trends and Implications for Branding. *e-Review of Tourism Research* , 4 (3), 9-11.
- Guevarra Plaza, A. (2008). La Web 2.0 y su aplicación al sector turístico. *Travel 2.0. Bit* (170), 38-40.

- Gunn, C. A., & Var, T. (2002). *Tourism Planning: Basic, Concepts, Cases. 4th Ed.* New York: Routledge.
- Gunn, C. (1972). *Vacationscape. Designing Tourist Regions.* Washington DC: Taylor and Francis/University of Texas.
- Gursoy, D., & McCleary, K. W. (2004). An integrative model of tourists' information search behavior. *Annals of tourism research* , 31 (2), 353-373.
- Hallmann, K., & Breuer, C. (2010). Image Fit between Sport Events and their Hosting Destinations from an Active Sport Tourist Perspective and its Impact on Future Behaviour. *Journal of Sport & Tourism* , 15 (3), 215-237.
- Hernández-Lobato, L., Solis-Radilla, M. M., Moliner-Tena, M. A., & García, J. S. (2006). Tourism Destination Image, Satisfaction and Loyalty: A Study in Ixtapa-Zihuatanejo, Mexico. *Tourism Geographies: An International Journal of Tourism Space, Place and Environment* , 8 (4), 343-358.
- Hu, Y., & Ritchie, J. (1993). Measuring Destination Attractiveness: A Contextual Approach. *Journal of Travel Research* , 32 (2), 25-34.
- Hudson, S., Wang, Y., & Gil, S. M. (2011). The Influence of a Film on Destination Image and the Desire to Travel: a Cross-Cultural Comparison. *International Journal of Tourism Research* , 13, 177-190.
- Huertas, A. (2012). ¿Web sites o Social Media? ¿Dónde se busca la información turística de los destinos? *TuriTec 2012, IX Congreso Nacional Turismo y Tecnologías de la Información y las Comunicaciones* (págs. 75-93). Málaga: Escuela Universitaria de Turismo, Universidad de Málaga.
- Hunt, J. D. (1975). Image as a factor in tourism development. *Journal of Travel Research* , 13 (3), 1-17.
- Hunter, W. C. (2010). Groomed Spaces on Jeju Island: A Typology of Photographic Representations for Tourism. *International Journal of Research* , 12, 680-695.
- Instituto Nacional de Estadística, I. (enero de 2014). Recuperado el 25 de septiembre de 2015, de El comercio electrónico y el uso de las nuevas tecnologías. Compras por Internet.:

[http://www.ine.es/ss/Satellite?L=es\\_ES&c=INECifrasINE\\_C&cid=1259943296411&p=1254735116567&pagename=ProductosYServicios%2FINECifrasINE\\_C%2FPYSDetalleCifrasINE](http://www.ine.es/ss/Satellite?L=es_ES&c=INECifrasINE_C&cid=1259943296411&p=1254735116567&pagename=ProductosYServicios%2FINECifrasINE_C%2FPYSDetalleCifrasINE)

*Internet Live Stats*. (s.f.). Recuperado el 25 de 09 de 2015, de [www.internetlifestats.com](http://www.internetlifestats.com):  
<http://www.internetlifestats.com/internet-users/#definitions>

Ivanov, S. H., Illum, S. F., & Liang, Y. (2010). Application of destination brand molecule on destination image and brand perception: An exploratory study. *Original Scientific Paper* , 58 (4), 339-360.

Jani, D., & Hwang, Y.-H. (2011). User-generated Destination Image through Weblogs:. *Asia Pacific Journal of Tourism Research* , 16 (3), 339-356.

Jetter, L. G., & Chen, R. J. (2011). Destination Branding and Images: Perceptions and Practices from Tourism Industry Professionals. *International Journal of Hospitality & Tourism Administration* , 12, 174-187.

Jönsson, C., & Devonish, D. (2008). Does Nationality, Gender, and Age Affect Travel Motivation? A Case of Visitors to the Caribbean Island of Barbados . *Journal of Travel & Tourism Marketing* , 25 (3-4), 398-408.

Joppe, M., Martin, D. W., & Waalen, J. (2001). Toronto's image as a destination: A comparative importance-satisfaction analysis by origin of visitor. *Journal of Travel Research* , 39, 252-260.

Karahanna, E., Evaristo, J. R., & Srite, M. (2005). Levels of culture and individual behavior: An integrative perspective. *Journal of International Management* , 13 (2), 1-20.

Karim, S. A., & Chi, C. G.-Q. (2010). Culinary tourism as a destination attraction: An empirical examination of destinations' food image. *Journal of Hospitality Marketing & Management* , 19, 531-555.

Kim, J. (., & Tussyadiah, I. P. (2013). Social Networking and Social Support in Tourism Experience: The Moderating Role of Online Self-Presentation Strategies. *Journal of Travel & Tourism Marketing* , 30 (1-2), 78-92.

- Kneesel, E., Baloglu, S., & Millar, M. (2010). Gaming Destination Images: Implications for Branding. *Journal of Travel Research* , 49 (1), 68-78.
- Ku, L.-W., Liu, I.-C., Lee, C.-Y., Chen, K.-h., & Chen, H.-H. (2008). Sentence-Level Opinion Analysis by CopeOpi in NTCIR-7. *Proceedings of NTCIR-7 Workshop Meeting* (págs. 260-267). Tokyo: NTCIR.
- Ku, W.-L., & Chen, H.-H. (2010). 中文意見分析之概況.技術與應用[An overview of Chinese opinion analysis, technology and application.]. *ACLCLP (The Association for Computational Linguistics and Chinese Language Processing)* , 20 (5), 5-20.
- Lange-Faria, W., & Elliot, S. (2012). Understanding the role of social media in destination marketing. *Tourism: An International Multidisciplinary Journal of Tourism* , 7 (1), 193-211.
- Langheinrich, M., & Karjoth, G. (2010). Social networking and the risk to companies and institutions. *Information Security Technical Report* , 15, 51-56.
- Law, R., & Cheung, S. (2010). The Perceived Destination Image of Hong Kong as Revealed in the Travel Blogs of Mainland Chinese Tourists. *International Journal of Hospitality & Tourism Administration* , 11, 303-327.
- Lee, G., & Lee, C.-K. (2009). Cross-cultural comparison of the image of Guam perceived by Korean and Japanese leisure travelers: Importance–performance analysis. *Tourism Management* , 30, 922-931.
- Lee, R., & Lockshin, L. (2011). Halo effects of tourists' destination image on domestic product perceptions. *Australasian Marketing Journal* , 19 (1), 7-13.
- Lee, R., & Lockshin, L. (2012). Reverse Country-of-Origin Effects of Product Perceptions on Destination Image. *Journal of Travel Research* , 51 (4), 502-511.
- Lee, S.-H., & Sparks, B. (2007). Cultural influences on travel lifestyle: A comparison of Korean Australian and Korean in Korea. *Tourism Management* , 28, 505-518.
- Lee, T. H. (2009). A structural model to examine how destination image, attitude, and motivation affect the future behavior of tourists. *Leisure Sciences: An Interdisciplinary* , 31 (3), 215-236.

- Leiva-Olivencia, J. L., Guevara-Plaza, A., Rossi-Jimenez, C. M., & Aguayo-Maldonado, A. (2014). Realidad Aumentada y Sistemas de Recomendación Grupales. Una nueva perspectiva en sistemas de destinos turísticos. *Estudios y perspectivas en*, 23, 40-59.
- Li, X. R., & Stepchenkova, S. (2012a). Chinese Outbound Tourists' Destination Image of America : Part I. *Journal of Travel Research* , 51 (3), 250-266.
- Li, X. R., & Stepchenkova, S. (2012b). Chinese Outbound Tourists' Destination Image of America : Part II. *Journal of Travel Research* , 51 (6), 687-703.
- Li, X., & Wang, Y. (2011). China in the Eyes of Western Travelers as Represented in Travel Blogs. *Journal of Travel & Tourism Marketing* , 28 (7), 689-719.
- Lim, Y., Chung, Y., & Weaver, P. A. (2012). The impact of social media on destination branding: Consumer-generated videos versus destination marketer-generated videos. *Journal of Vacation Marketing* , 18 (3), 197-206.
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism management* , 29 (3), 458-468.
- López, M., & Sicilia, M. (2012). La participación del individuo en el boca a boca electrónico como. *TuriTec 2012, IX Congreso Nacional Turismo y Tecnologías de la Información y las Comunicaciones* (págs. 381-393). Málaga: Escuela Universitaria de Turismo, Universidad de Málaga.
- MacKay, K. J., & Couldwell, C. M. (2004). Using Visitor-Employed Photography to Investigate Destination Image. *Journal of Travel Research* 42: 390 , 42, 390-396.
- Manuzarc. (02 de diciembre de 2014). *Los viajeros*. Recuperado el 16 de febrero de 2015, de <http://www.losviajeros.com/foros.php?p=4586326#4586326>
- Mariné-Roig, E. (2010). Los "Travel Blogs" como objetos de estudio de la imagen percibida de un destino. [The travel blogs as objects of study of the perceived image of a destination]. *TuriTec 2010, VIII Congreso Nacional Turismo y Tecnologías de la Información y las Comunicaciones* (págs. 61-75). Málaga: Escuela Universitaria de Turismo.

- Martín Gutiérrez, H. S. (2005). *Estudio de la Imagen de Destino y el Proceso Global de Satisfacción: Adopción de un Enfoque Integrado*. Santander: Departamento de Administración de Empresas, Universidad de Cantabria.
- Martín, A. V. (2012). *Turismo 2.0: iniciando el despegue*. España: Septem Ediciones.
- Martín, H. S., & Rodríguez del Bosque, I. A. (2008). Exploring the cognitive–affective nature of destination image and the role of psychological factors in its formation. *Tourism Management* , 29, 263-277.
- Mata, F. L. (2011). *Extracción de Opiniones sobre Características: Un Enfoque Práctico Adaptable al Dominio [Opinion Extractions about features: A practical approach adaptable to the domain]*. Sevilla: Universidad de Sevilla.
- Mayo, E. J. (1973). Regional images and regional travel behaviour. *The Travel Research Association Fourth Annual Conference Proceedings*. Sun Valley: Idaho.
- McCartney, G. (2008). Does one culture all think the same? An investigation of destination image perceptions from several origins. *Tourism Review* , 63 (4), 13-26.
- McCartney, G., Butler, R., & Bennett, M. (2008). A Strategic Use of the Communication Mix in the Destination Image-Formation Process. *Journal of Travel Research.* , 47 (2), 183-196.
- McCartney, G., Butler, R., & Bennett, M. (2009). Positive tourism image perceptions attract travellers – fact or fiction? The case of Beijing visitors to Macao. *Journal of Vacation Marketing* , 15 (2), 179-193.
- Míguez, M. I., Roig, E. M., & Huertas, A. (2014). ¿Utilizan los destinos turísticos los medios sociales para creae diálogo con sus públicos? Estudio de los top post de facebook y twitter de los destinos turísticos españoles. *TutiTec 2014. X Congreso de Turismo y Tecnologías de la Información y las Comunicaciones* (págs. 108-121). Málaga: Escuela Universitaria de Turismo.
- Miller, M. M. (1993). *User's guide to VBPro: A program for qualitative and quantitative analysis of verbatim text*. Knoxville, TN: M. Mark Miller.

- Millet, O. F. (2011). *La imagen de un destino turístico como herramienta de marketing*. Málaga, España: Universidad de Málaga (UMA), Grupo EUMEDNET.
- Milman, A. (2011). The Symbolic Role of Postcards in Representing a Destination Image: The case of Alanya, Turkey. *International Journal of Hospitality & Tourism Administration* , 12, 144-173.
- Mohan, L. J. (2010). Effect of destination image on attendance at team sporting events. *Tourism and Hospitality Research* , 10, 157-170.
- Muñoz Vera, G., & Elósegui, T. (2011). En G. & Muñoz Vera, *El arte de medir: Manual de analítica web*. Barcelona: Profit.
- Murphy, P., Pritchard, M. P., & Smith, B. (2000). The destination product and its impact on traveller perceptions. *Tourism Management* , 43-52.
- Nezakati, H., Amidi, A., Jusoh, Y. Y., Moghadas, S., Aziz, Y. A., & Sohrabinezhadtalemi, R. (2015). Review of Social Media Potential on Knowledge Sharing and Collaboration in Tourism Industry. *Procedia - Social and Behavioral Sciences* , 172, 120-125.
- Oliveira, E., & Panyik, E. (2015). Content, context and co-creation: Digital challenges in destination branding with references to Portugal as a tourist destination. *Journal of Vacation Marketing* , 2 (1), 53-74.
- Organización Mundial del Turismo (1998). *Introducción al Turismo*. Madrid: Organización Mundial del Turismo.
- Organización Mundial del Turismo (2015). *Panorama OMT del Turismo Internacional, 2015 Edition*. Organización Mundial del Turismo.
- Organización Mundial del Turismo (2001). *Tesaurus de Turismo y Ocio*. Paris: Organización Mundial del Turismo.
- Pan, B., MacLaurin, T., & Crotts, J. C. (2007). Travel Blogs and the Implications for Destination Marketing. *Journal of Travel Research* , 46, 35-45.
- Park, Y., & Njite, D. (2010). Relationship between Destination Image and Tourists' Future Behavior: Observations from Jeju Island, Korea. *Asia Pacific Journal of Tourism Research* , 15 (1), 1-20.

- Parra-López, E., Bulchand-Gidumalb, J., Gutiérrez-Tañoa, D., & Díaz-Armas, R. (2011). Intentions to use social media in organizing and taking vacation trips. *Computers in Human Behavior* , 27 (2), 640-654.
- Phillips, W. J., & Jang, S. (2010). Destination Image Differences between Visitors and Non-Visitors: A Case of New York City. *International Journal Of Tourism Research* , 12, 642-645.
- Phillips, W. J., Wolfe, K., & Hodur, N. (2013). Tourist Word of Mouth and Revisit Intentions to Rural Tourism Destinations: a Case of North Dakota,USA. *International Journal of Tourism Research* , 15, 93-104.
- Pike, S. D. (2002). Destination Image Analysis: A Review of 142 Papers from 1973-2000. *Tourism Management* , 23 (5), 541-549.
- Pizam, A., & Sussmann, S. (1995). Does nationality affect tourist behavior? *Annals of Tourism Research* , 22 (4), 901-917.
- Placido, A. D. (2010). Interactividad usuario-usuario y redes sociales online en el sector turístico. Análisis de las páginas web turísticas oficiales de las administraciones andaluzas. *Turitec 2010. VIII Congreso Turismo y Tecnologías de la Información y las Comunicaciones* (págs. 77-91). Málaga: Escuela Universitaria de Turismo.
- Polo Peña, A. I., Frías Jamilena, D. M., & Rodríguez Molina, M. Á. (2012). Validation of cognitive image dimensions for rural tourist destinations: A contribution to the management of rural tourist destinations. *Journal of Vacation Marketing* , 18 (4), 261-273.
- Popescu, A.-M., & Etzioni, O. (2007). Extracting Product Features and Opinions from Reviews. En A. Kao, & S. R. Poteet, *Natural Language Processing and Text Mining* (págs. 9-28). Nueva York: Springer Science & Business Media.
- Prebensen, N. K. (2005). Country as Destination—Norwegian Tourists' Perceptions and Motivation. *Journal of Hospitality & Leisure Marketing* , 12 (3), 63-85.
- Ramkissoon, H., & Nunkoo, R. (2011). City Image and Perceived Tourism Impact: Evidence from Port Louis, Mauritius. *International Journal of Hospitality & Tourism Administration* , 12, 123-143.

- Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*, 41 (3), 647-656.
- Reisinger, Y., & Turner, L. W. (2002). Cultural differences between Asian tourist markets and Australian hosts, Part 1. *Journal of Travel Research*, 40, 295-215.
- Rezende-Parker, A. M., Morrison, A. M., & Ismail, J. A. (2003). Dazed or confused? An exploratory study of the image of Brazil as a travel destination. *Journal of Vacation Marketing*, 9, 243-259.
- Richardson, S. L., & Crompton, J. L. (1988). Cultural variations in perceptions of vacation attributes. *Tourism Management*, 9 (2), 128-136.
- Richardson, S. L., & Crompton, J. (1988). Vacation Patterns of French and English Canadians. *Annals of Tourism Research*, 15 (3), 430-435.
- Rodríguez Ruibal, A. (2010). Utilización de las redes sociales como medio de promoción en el sector turístico. Opinión, valoración e interpretación de los comportamientos de los turistas. *TuriTec* (págs. 1-11). Málaga: Escuela Universitaria de Turismo Campus de Teatinos s/n.
- Roig, E. M. (2010). Los "Travel Blogs" como objetos de estudio de la imagen percibida de un destino. *TuriTec 2010, VIII Congreso Nacional Turismo y Tecnologías de la Información y las Comunicaciones* (págs. 61-75). Málaga: Escuela Universitaria de Turismo, Universidad de Málaga.
- Ruibal, A. R. (2010). Utilización de las redes sociales como medio de promoción en el sector turístico. Opinión, valoración e interpretación de los comportamientos de los turistas. *TuriTec 2010, VIII Congreso Nacional Turismo y Tecnologías de la Información y las Comunicaciones* (págs. 1-11). Málaga: Escuela Universitaria de Turismo, Universidad de Málaga.
- Ryan, C., & Aicken, M. (2010). The destination image gap – visitors' and residents' perceptions of place: evidence from Waiheke Island, New Zealand. *Current Issues in Tourism*, 13 (6), 541-561.

- Ryan, C., & Cave, J. (2005). Structuring destination image: A qualitative approach. *Journal of Travel Research* , 44, 143-150.
- Schiffman, L., & Kanuk, L. (1997). *Comportamiento del Consumidor. 5ª Edición.* . México: Prentice Hall Hispanoamericana, S.A.
- Shani, A., & Wang, Y. (2011). Destination Image Development and Communication. En Y. Wang, A. Pizam, & D. Breiter, *Destination Marketing and Management : Theories and Applications* (págs. 130-148). Cambridge: CABI Publishing.
- Shilbury, D., Westerbeek, H., Quick, S. F., & Karg, A. (2014). *Strategic sport marketing (4th ed.)*. Sydney: Allen & Unwin.
- Stabler, M. (1995). The Image of Destination Regions: Theoretical and Empirical Aspects. En B. Goodall, & .Ashworth *Marketing in Tourism Industry: The Promotion of Destination Regions* (págs. 133-159). London: Routledge.
- Stepchenkova, S., & Mills, J. E. (2010). Destination Image: A Meta-Analysis of 2000–2007 Research. *Journal of Hospitality Marketing & Management* , 19, 575-609.
- Stepchenkova, S., & Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tourism Management* , 36, 590-601.
- Stern, E., & Krakover, S. (1993). The Formation of a Composite Urban Image. *Geographical Analysis* , 25, 130-146.
- Tang, L., Choi, S., Morrison, A. M., & Lehto, X. Y. (2009). The many faces of Macau: A correspondence analysis of the images communicated by online tourism information sources in English and Chinese. *Journal of Vacation Marketing* , 15 (1), 79-95.
- Tasci, A. D. (2007). Methodology Influences on Destination Image: The Case of Michigan. *Current Issues in Tourism* , 10 (5), 480-501.
- Tasci, A. D., & Gartner, W. C. (2007). Destination image and its functional relationships. *Journal of Travel Research* , 45 (4), 413-425.

Tasci, A. D., Gartner, W. C., & Cavusgil, S. T. (2007). Conceptualization and operationalization of destination image. *Journal of Hospitality & Tourism Research* , 31 (2), 194-223.

*The State of Chinese Social Media in 2015: What You Need to Know Six Takeaways about WeChat, Weibo and Up-and-Coming Platforms.* (04 de Junio de 2015). Recuperado el 02 de septiembre de 2015, de [http://www.ciccorporate.com/index.php?option=com\\_content&view=article&id=1296:the-state-of-chinese-social-media-in-2015-what-you-need-to-know-six-takeaways-about-wechat-weibo-and-up-and-coming-platforms&catid=99:archives-2014&Itemid=208&lang=en](http://www.ciccorporate.com/index.php?option=com_content&view=article&id=1296:the-state-of-chinese-social-media-in-2015-what-you-need-to-know-six-takeaways-about-wechat-weibo-and-up-and-coming-platforms&catid=99:archives-2014&Itemid=208&lang=en)

Tse, T. S., & Zhang, E. Y. (2013). Analysis of Blogs and Microblogs: A Case Study of Chinese Bloggers Sharing Their Hong Kong Travel Experiences. *Asia Pacific Journal of Tourism Research* , 18 (4), 314-329.

Turney, P. D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *40th Annual Meeting of the Association for Computational Linguistics (ACL'02)* (págs. 417-424). Philadelphia, Pennsylvania, USA: National Research Council of Canada.

Turney, P. D. (8-10 de July de 2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)* , 417-424.

Um, S., & Crompton, J. (1990). Attitude Determinants in Tourism Destination Choice. *Annals of Tourism Research* , 17, 432-448.

Unión Internacional de Telecomunicaciones (2014). *Informe sobre Medición de la Sociedad de la Información 2014*. Ginebra: Unión Internacional de Telecomunicaciones.

Unión Internacional de Telecomunicación (2015). *ITU ICT Facts and Figures – The world in 2015*. Geneva: UIT.

Verlegh, P. W., & Steenkamp, J.-B. E. (1999). A review and meta-analysis of country-of-origin research. *Journal of Economic Psychology* , 20, 521-546.

- Wang, H.-Y. (2012). Investigating the determinants of travel blogs influencing readers' intention to travel. *The Service Industries Journal* , 32 (2), 231-255.
- Wang, Y. (2008). Collaborative destination marketing: Roles and strategies of convention and visitors bureaus. *Journal of Vacation Marketing* , 14 (3), 191-209.
- Wang, Y. (2011). Destination Marketing and Management: Scope, Definition and Structures. En Y. Wang, A. Pizam, & D. Breiter, *Destination Marketing and Management: Theories and Applications* (págs. 1-20). Cambridge: CABI Publishing.
- We Are Social Singapore, s. (Enero de 2015). *slideshare*. Recuperado el 25 de septiembre de 2015, de Digital, Social & Mobile in 2015: <http://www.slideshare.net/wearesocialsg/digital-social-mobile-in-2015>
- Woodside, A., & Lysonsky, S. (1989). A General Model of Traveler Destination Choice. *Journal of Travel Research* , 27 (4), 8-14.
- Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management* , 31 (2), 179-188.
- Xiao, H., & Mair, H. L. (2006). "A paradox of images": Representation of China as a tourist destination. *Journal of Travel & Tourism Marketing* , 20 (2), 1-14.
- Yang, J., He, J., & Gu., Y. (2012). The implicit measurement of destination image: The application of Implicit Association Tests. *Tourism Management* , 33, 50-52.
- Yao, Y., Chen, Z., & Jia, Y. (2013). Tourist Trustworthiness of Destination: Dimension and Its Consequence. *Tourism Tribune* , 28 (4), 48-55.

# ANEXO

# ANEXO

## A. 1. Las dimensiones de imagen de destino

Natural Resources	General Infrastructure	Tourist Infrastructure
Weather Temperature Rainfall Humidity Hours of sunshine	Development and quality of roads, airports and ports Private and public transport facilities Development of health services	Hotel and self-catering accommodation Number of beds Categories Quality
Beaches Quality of seawater Sandy or rocky beaches Length of the beaches Overcrowding of beaches	Development of telecommunications Development of commercial infrastructures Extent of building development	Restaurants Number Categories Quality Bars, discotheques and clubs
Wealth of countryside Protected nature reserves Lakes, mountains, deserts, etc. Variety and uniqueness of flora and fauna		Ease of access to destination Excursions at the destination Tourist centers Network of tourist information
<b>Tourist Leisure and Recreation</b> Theme parks Entertainment and sports activities Golf, fishing, hunting, skiing, scuba diving, etc. Water parks Zoos Trekking Adventure activities Casinos Night life Shopping	<b>Culture, History and Art</b> Museums, historical buildings, monuments, etc. Festival, concerts, etc. Handicraft Gastronomy Folklore Religion Customs and ways of life	<b>Political and Economic Factors</b> Political stability Political tendencies Economic development Safety Crime rate Terrorist attacks Prices
<b>Natural Environment</b> Beauty of the scenery Attractiveness of the cities and towns Cleanliness Overcrowding Air and noise pollution Traffic congestion	<b>Social Environment</b> Hospitality and friendliness of the local residents Underprivilege and poverty Quality of life Language barriers	<b>Atmosphere of the Place</b> Luxurious Fashionable Place with a good reputation Family-oriented destination Exotic Mystic Relaxing Stressful Fun, enjoyable Pleasant Boring Attractive or interesting

*Fuente:* La tabla está extraída de “Factors influencing destination image,” por Beerli & Martín, 2004, *Annals of Tourism Research*, Vol. 31, No. 3, pp. 659.





## A. 3. Etiqueta usado por CKIP

Etiquetas de CKIP	Significado	Ejemplo
<b>A</b>	Adjetivo	空心
<b>Caa</b>	Conjunción Coordinantes	和、跟
<b>Cab</b>	Conjunción para dar ejemplos	等、等等
<b>Cba</b>	Conjunción correlativa	因為、即使
<b>Cbb</b>	Conjunción subordinantes	就是、不但
<b>Da</b>	Adverbio de cantidad	一共、都
<b>Dfa</b>	Adverbio comparativo antes de predicado	非常、很、極了
<b>Dfb</b>	Adverbio comparativo después de predicado	得很、極了
<b>Di</b>	Adverbio de tiempo	著、過、起
<b>Dk</b>	Adverbio sentencial	總之、老實說、據說
<b>D</b>	Adverbio	一共、至少、大約
<b>Na</b>	Nombre común	泥土、鹽、水、衣服
<b>Nb</b>	Nombre propio	詩經、雙魚座
<b>Nc</b>	Palabra de lugar	郵局、市場
<b>Ncd</b>	Palabra en explicación de posición	上頭、中間、左方
<b>Nd</b>	Palabra de tiempo	唐朝、西漢
<b>Neu</b>	Modificador de número	三、三千五百、幾、好幾
<b>Nes</b>	Modificador específico	某、該、本、同
<b>Nep</b>	Pronombre demostrativo	這、那、哪、什麼
<b>Neqa</b>	Modificador de cantidad	許多、百分之五十
<b>Neqb</b>	Modificador detrás del número	三點 <b>正</b> 、五十歲 <b>出頭</b>
<b>Nf</b>	Clasificador	一"張"桌子、一"個"杯子
<b>Ng</b>	Posposición	街上
<b>Nh</b>	Pronombre	我、你、他
<b>I</b>	Interjección	哇、喔、嗯
<b>P</b>	Preposición	他 <b>趁此機會</b> 大發牢騷
<b>T</b>	Partícula	了、喔
<b>VA</b>	Predicado intransitivo de acción	上升、坐
<b>VB</b>	Predicado verbal de acción con un objeto delante de verbos	求婚、拜年、說媒
<b>VC</b>	Predicado verbal de acción con un objeto que sea sustantivo	闖入、奔赴、逃離、住
<b>VCL</b>	Verbo que conecta con objeto de localización	到
<b>VD</b>	Predicado con dos objetos	送、交、傳、借
<b>VE</b>	Predicado conectado con una frase con mismo sujeto	大聲疾呼、自言自語
<b>VF</b>	Predicado conectado con una frase	答應、勸
<b>VG</b>	Predicado para conectar	稱呼、喊、命名
<b>VH</b>	Predicado intransitivo de estado	浪漫、特別
<b>VI</b>	Predicado transitivo de estado con un solo objeto	心動、灰心
<b>VJ</b>	Predicado transitivo de estado con un solo objeto y dos	迎合、久仰、皈依

	argumentos	
<b>VK</b>	Predicado transitivo de estado para conectar un argumento en la forma de frases	清楚、明白、期望
<b>VL</b>	Predicado transitivo de estado necesita dos o tres argumentos	樂於、甘願
<b>V_2</b>		有
<b>DE</b>		的.之.得.地
<b>SHI</b>		是
<b>FW</b>	Lengua extranjera	Madrid
<b>COLONCATEGORY</b>	Dos puntos	< : >
<b>COMMACATEGORY</b>	La coma	< , >
<b>DASHCATEGORY</b>	Guion	< - >
<b>ETCCATEGORY</b>	Puntos suspensivos	< ... >
<b>EXCLAMATIONCATEGORY</b>	Signo de exclamación	< ! >
<b>PARENTHESISCATEGORY</b>	Los parénTesis	< ( ) >
<b>PAUSECATEGORY</b>	La coma china	< 、 >
<b>PERIODCATEGORY</b>	El punto	< 。 >
<b>QUESTIONCATEGORY</b>	Signo de interrogación	< ? >
<b>SEMICOLONCATEGORY</b>	El punto y coma	< ; >
<b>SPCHANGECATEGORY</b>	Dos barras	<    >

*Fuente:* La tabla está traducido y editado basado en la publicación de la Academia Sinica en su web: [ckipsvr.iis.sinica.edu.tw/papers/category\\_list.doc](http://ckipsvr.iis.sinica.edu.tw/papers/category_list.doc)

## A. 4. Stanford Parser Post Tags Stanford Parser Post Tags

POS tags	Significado	Ejemplo
AD	adverb	还
AS	aspect marker	着
BA	把 in ba-construction	把,将
CC	coordinating conjunction	和
CD	cardinal number	一百
CS	subordinating conjunction	虽然
DEC	的 in a relative-clause	的
DEG	Associative 的	的
DER	得 in V-de const. and V-de-R	得
DEV	地 before VP	地
DT	determiner	这
ETC	for words 等,等等	等,等等
FW	foreign words	ISO
IJ	Interjection	啊
JJ	other noun-modifier	男,共同
LB	被 in long bei-const	被,给
LC	localizer	里
M	measure word	个
MSP	other particle	所
NN	common noun	书
NR	proper noun	美国
NT	temporal noun	今天
OD	ordinal number	第一
ON	onomatopoeia	哈哈,哗哗
P	preposition excl. 被 and 把	从
PN	Pronoun	他
PU	Punctuation	、?。
SB	被 in short bei-const ,	被,给
SP	sentence-final particle	吗
VA	predicative adjective	红
VC	是	是
VE	有 as the main verb	有
VV	other verb	走

Fuente: La tabla está extraída de Xia, Fei. (2000). The Part-Of-Speech Tagging Guidelines for the Penn Chinese Treebank (3.0). Retrieved from: <http://www.cis.upenn.edu/~chinese/ctb.html>

## A. 5. Glosario de términos para la medición de la IDT

IG	RN	IT	ODT	CHA	FPE	EN	ES
servicio_sanitario :	reserva_natural :	excursión :	compras :	artesano :	político :	congestión :	comunicación :
病院	大自然	出遊	買進	工藝品	政治	阻塞	交流
衛生院	自然	巡禮	瞎拼	藝品	precio :	封死	你來我往
醫院	保護區	郊遊	選購	藝術品	市價	堵	相通
infraestructura_comercial :	flora :	旅行	購	gastronomía :	售價	堵住	通訊
合作社	花木	野遊	購入	美食	單價	堵塞	溝通
百貨商店	花樹	遊	購得	festival :	牌價	淤	habitante :
店	草木	遊山玩水	購買	節日	價目	閉塞	住民
店鋪	葫蘆	遊逛	購置	節假日	價位	塞住	居民
店鋪	植物	遊園	置	喜慶	價碼	壅	calidad_de_vida :
門市	zona_rural :	遊歷	辦	慶典	標價	梗	生活
鋪子	山野	遊覽	收	節慶	貴賤	aglomeración :	生涯
鋪	郊野	遠足	收買	節	價	摩肩接踵	溫飽
莊	原野	踏青	收購	costumbre :	行情	摩肩擦踵	
小攤	荒	遨遊	沽	世俗	底價	擁擠	
地攤	荒地	環遊	採買	風俗	物價	擠	
路邊攤	荒郊	觀光	採購	習俗	訂價	擠滿	
攤子	荒郊野外	旅遊	頂下	俗	值	人滿為患	
攤	荒原	alojamiento :	買入	民俗	起價	水泄不通	
aeropuerto :	荒野	客棧	買下	民風	價錢	水洩不通	
飛機場	荒漠	旅店	買下來	風	economía :	軋	
航站	曠野	旅社	買去	風土	商	盈門	
puente :	野外	旅舍	買來	習慣	經濟	擠人	
吊橋	desierto :	飯店	購物	concierto :	seguridad :	contaminación :	
虹橋	沙漠	會館	購	音樂會	治安	汙染	
橋	montaña :	賓館	代購	演奏會		ciudad_pueblo :	
橋樑	山地	酒店	代買	arte :		城	
medio_de_transporte :	山陵	民宿	買	美術		都市	
輸	嶺	旅館	買	藝術		小鎮	
捎	高山	información_turística :	夜生活	museo :		屯子	
捎來	山	含意	casino :	文物館		市鎮	
送來	lago :	含義	賭場	博物院		田園	
搭載	水潭	見聞	zoo :	美術館		村子	
搬運	湖泊	定義	動物園	博物館		村莊	
載運	湖	信息	deporte :	monumento :		村落	
運	playa :	界定	瑜伽	城堡		村鎮	
運載	灘	音訊	體育	古跡		城鄉	
運輸	海邊	風信	運動	遺蹟		城鎮	
輸送	clima :	風聲	parque_temático :	跡		鄉下	
駁	天	消息	遊樂場	天主堂		鄉野	
交通	天色	涵義	遊樂園	禮拜堂		鄉間	
telecomunicaciones :	天候	意味		教堂		鄉鎮	
大哥大	天氣	意思		古蹟		鎮	
外線	氣候	意涵		religión :		莊	
資訊網	氣象	意義		教		鄉	
網際	陰晴	概念		宗教		城市	
電訊	fauna :	資料				鄉村	
傳真	野獸	資訊				limpieza :	
網路	動物	義				清清爽爽	
wifi		材料				清新	
行動電話		風				清潔	
手機		訊息				淨	
電話		bar_discoteca :				窗明几淨	
電信		酒館				潔	
puerto :		舞場				潔淨	
自由港		酒吧				衛生	
港		bar				清爽	
港灣		舞廳				明淨	
碼頭區		restaurante :				純淨	
避風港		食堂				乾乾淨淨	
港口		菜館				paisaje :	
carretera :		飯館				山水	
小道		飯廳				地景	
公路		館子				背景	
出路		酒家				風光	
巷子		餐廳				風貌	
段		咖啡廳				景	
省道		咖啡館				景色	
馬路		centro_de_turistas :				景物	
途		觀光局				景致	
陸路		旅遊局				景象	
街						景像	
街路						景緻	
街道						景觀	
街頭						觀瞻	
路						風景	
路徑							
路線							
道							
道路							

## A. 6. Parejas de palabras obtenidas con sus respectivos valores de PMI

### A. 6. 1. Backpackers

病院 no matching pairs	沙漠 生活 4.9798858574	市鎮 旅行 3.07215569341
衛生院 no matching pairs	沙漠 藝術 4.54762255614	動物園 旅行 3.07215569341
聖保羅 醫院 5.2484917828	教堂 沙漠 3.93256686312	客棧 旅行 3.02579907914
聖保祿 醫院 5.2484917828	阻塞 no matching pairs	野遊 no matching pairs
莫 醫院 4.77848815355	封死 no matching pairs	節 遊 4.00966826499
卡爾 醫院 4.7560152977	堵 no matching pairs	日 遊 2.50705178885
瑪利亞 醫院 4.60663789662	堵住 no matching pairs	旅遊 遊 2.4119779179
工藝品 no matching pairs	堵塞 no matching pairs	11 遊 2.37856885577
手工 藝品 6.26527405322	淤 no matching pairs	客棧 遊 2.33866156518
生活 藝品 3.99905660439	閉塞 no matching pairs	遊山玩水 no matching pairs
教堂 藝品 3.23941968256	塞住 no matching pairs	遊逛 no matching pairs
藝品 餐廳 2.97740535686	壅 no matching pairs	遊園 no matching pairs
它 藝品 2.82967524484	梗 no matching pairs	遊歷 no matching pairs
藝術 藝術品 4.87612662311	沙漠 生活 4.9798858574	港 遊覽 4.34096023101
作品 藝術品 4.80201865096	原野 生活 4.84635446478	景色 遊覽 3.60626404754
博物館 藝術品 4.06093200811	生活 田園 4.73302577947	遊覽 酒店 3.49626315232
教堂 藝術品 3.75024530633	橋樑 生活 4.69220378495	城堡 遊覽 3.31134081383
藝術品 資料 3.3865773269	生活 芒 4.50988222816	手機 遊覽 3.28992671933
觀光局 資料 3.50436036256	教堂 生涯 3.64488479067	遠足 no matching pairs
古蹟 旅遊局 3.91338872208	旅行 生涯 2.84901214209	踏青 no matching pairs
官方 旅遊局 3.69402589361	客棧 生涯 2.60102582965	遨遊 no matching pairs
旅遊局 自治區 3.49666645945	一 生涯 1.40828956391	環遊 no matching pairs
旅遊局 藝術 3.0569681797	個 生涯 1.17332213794	觀光 觀景 3.98252730244
旅遊局 風景 3.02969915158	溫飽 no matching pairs	總論 觀光 3.98252730244
村莊 酒館 4.44534370702	段 遊樂場 3.92673594281	山坡地 觀光 3.98252730244
服務生 酒館 3.92209556326	教堂 遊樂場 3.64488479067	原野 觀光 3.98252730244
居民 酒館 3.83354216592	路 遊樂場 3.47006367667	巡禮 觀光 3.72016303797
生活 酒館 3.8167350476	城市 遊樂場 3.00580483138	旅遊 港灣 3.8783149867
酒館 風 3.8043125276	客棧 遊樂場 2.60102582965	旅遊 沙漠 3.65517143538
舞場 no matching pairs	遊樂園 風 4.79756430061	原野 旅遊 3.65517143538
景像 酒吧 5.49814645966	景 遊樂園 4.68115394877	旅遊 田園 3.40831135745
信仰 酒吧 4.46852704248	遊樂園 馬路 4.54762255614	巡禮 旅遊 3.39280717092
鄉間 酒吧 4.45669258483	城堡 遊樂園 4.15506085287	夜生活 節日 4.51173579681
攤 酒吧 4.45669258483	遊樂園 風景 4.05034989878	夜生活 街道 4.16944959032
天色 酒吧 4.39953417099	出遊 no matching pairs	夜生活 手機 3.63287147045
bar no matching pairs	巡禮 鄉間 6.09643584676	夜生活 街 3.61107681063
巷子 舞廳 4.79756430061	天色 巡禮 5.75159536047	夜生活 海邊 3.32215172993
舞廳 飯店 3.38989507991	巡禮 村莊 5.46391328802	夜生活 節日 4.51173579681
旅館 舞廳 2.95613322159	巡禮 道路 5.29734008819	節慶 節日 4.07641772555
交通 舞廳 2.19533196711	巡禮 風光 5.07950158911	節 節日 3.78140961301
客棧 舞廳 1.90787864909	郊遊 no matching pairs	宗教 節日 3.75263064846
沙漠 都市 5.74202590945	旅行 荒漠 3.07215569341	節日 街道 3.08098967312
橋 沙漠 5.37430112933	旅行 景像 3.07215569341	節假日 no matching pairs

喜慶 no matching pairs	意思 景象 3.75349945824	博物館 湖泊 3.77324993566
慶典 遊行 4.38471908521	意涵 no matching pairs	湖泊 餐廳 3.48823098063
慶典 節慶 4.25955594225	原野 意義 5.50313400117	教堂 湖泊 3.46256323388
宗教 慶典 4.11809042195	意義 材料 5.09766889306	湖泊 資訊 3.20305203839
慶典 節 3.5590827216	意義 高度 4.71467664081	湖 購物 4.44187749905
慶典 活動 3.43316021449	意義 義 4.62766526382	手機 湖 4.17838265221
節慶 風俗 5.70647492519	咖啡廳 意義 4.47351458399	山 湖 3.93576886516
民俗 節慶 5.70647492519	植物 概念 5.29004078571	湖 風景 3.51972164772
傳承 節慶 5.30100981708	概念 狀 5.20302940872	小鎮 湖 3.48765971526
特質 節慶 5.01332774463	政治 概念 5.20302940872	畫 美術 4.63393487683
煙火 節慶 4.38471908521	大自然 概念 4.89764775917	美術 藝術 3.95983589124
節 聖體 4.60053659643	概念 高級 4.79756430061	作品 美術 3.88572791909
節 蕃茄 4.16381894481	觀光局 資料 3.50436036256	美術 美術館 3.65016184777
節 遊 4.00966826499	藝術品 資料 3.3865773269	天氣 美術 3.63441349081
牛 節 3.89188522933	沙子 資料 2.91657369765	藝術 藝術品 4.87612662311
節 節日 3.78140961301	旅遊局 資料 2.82018185198	沙漠 藝術 4.54762255614
天 飛機場 1.7367149696	資料 風俗 2.811213182	橋樑 藝術 4.54762255614
航站 道路 3.82510329727	湖泊 資訊 3.20305203839	採購 藝術 4.54762255614
公路 航站 3.21464783255	景緻 資訊 3.20305203839	情境 藝術 4.36530099935
航站 處 2.46290648773	小子 資訊 3.20305203839	灘 no matching pairs
機場 航站 2.34263333007	天色 資訊 3.20305203839	動物園 海邊 5.02689982217
航站 路線 2.32633095272	大圖 資訊 3.20305203839	海邊 爸爸 4.28496247744
含意 no matching pairs	意義 義 4.62766526382	海邊 港口 4.24194509236
含義 no matching pairs	義 醫院 3.86219742168	海邊 老娘 4.21596960596
旅行 見聞 2.84901214209	法 義 3.52905297515	原野 海邊 4.1106090903
客棧 見聞 2.60102582965	德 義 3.52905297515	交流 語言 3.61788418885
定義 no matching pairs	生活 義 3.41126993949	交流 民宿 3.58694695357
信息 酒吧 4.24538349116	意義 材料 5.09766889306	交流 旅遊 2.77970269803
信息 電話 3.25711925966	材料 自然 4.99669297572	交流 客棧 2.70638634531
信息 生活 3.25711925966	材料 道 4.41710049471	交流 城市 2.70570023893
信息 民宿 3.02733116563	材料 海鮮 4.06804947588	你來我往 no matching pairs
信息 地址 2.94474457462	材料 水 3.93451808326	相通 no matching pairs
界定 no matching pairs	原野 風 5.08524637306	網路 通訊 3.26989898631
音訊 no matching pairs	人聲 風 5.02070785193	溝通 物價 4.15506085287
風信 no matching pairs	農舍 風 4.97191768776	溝通 習慣 4.10195102756
風聲 no matching pairs	天色 風 4.90292481627	居民 溝通 4.0848565942
消息 清潔 5.14896218745	鄉村 風 4.79756430061	溝通道 4.08062825809
消息 通知 4.27349345009	專頁 訊息 3.17780942121	溝通 酒吧 3.95770141871
消息 物價 3.15653202276	粉絲 訊息 2.80311597176	大哥大 no matching pairs
新聞 消息 2.88719908897	臉 訊息 2.72657589464	外線 no matching pairs
消息 經濟 2.70951791174	訊息 那兒 2.60102582965	資訊網 no matching pairs
涵義 no matching pairs	訊息 錯誤 2.49445024701	網際 no matching pairs
意味 no matching pairs	動物園 海邊 5.02689982217	電訊 no matching pairs
意思 河岸 4.05668571723	動物園 客棧 3.11185145341	傳真 經濟 4.84302667469
巡禮 意思 4.04118153069	動物園 旅行 3.07215569341	傳真 電話 4.79756430061
意思 頭頂 4.01586372271	水潭 no matching pairs	傳真 簽證 3.6554669
五顏六色 意思 3.98769284574	湖泊 路線 4.06093200811	傳真 護照 3.49103490796

傳真 資訊 3.08526900274	特質 背景 5.54395599569	景觀 鄉村 4.52386847013
網路 通訊 3.26989898631	民俗 背景 5.54395599569	景觀 鄉間 4.39870532718
網路 連線 3.08757742952	植物 背景 5.22550226457	咖啡館 景觀 4.34154691334
1.15 網路 3.08757742952	景象 背景 5.05844817991	觀瞻 no matching pairs
原野 網路 3.046755435	田園 風光 5.85269147734	原野 風景 5.03117915179
網路 高山 3.02477652828	原野 風光 5.62954792603	橋樑 風景 4.74349707934
wifi no matching pairs	標示牌 風光 5.38268784809	芒 風景 4.56117552255
經濟 行動電話 5.2484917828	鄉村 風光 5.34186585357	羊腸小徑 風景 4.56117552255
行動電話 電話 5.02070785193	風光 風貌 5.24655567377	田園 風景 4.56117552255
行動電話 路 3.28774211988	風光 風貌 5.24655567377	小道 小鎮 4.08728083781
客棧 行動電話 2.70638634531	風 風貌 4.70225412081	天 小道 2.24754059337
旅行 行動電話 2.666905853	風景 風貌 3.95503971898	公路 鄉間 5.11176781744
手機 智慧型 4.93215445458	生活 風貌 3.90374642459	公路 國道 4.94924888794
手機 泳 4.59568221796	天氣 風貌 3.83920790345	公路 高山 4.80948694557
手機 電信 4.46215082534	景 遊樂園 4.68115394877	公路 原野 4.72610533663
手機 湖 4.17838265221	原野 景 4.68115394877	公路 道路 4.67731517246
手機 視線 3.89070057976	山洞 景 4.60419290763	出路 no matching pairs
行動電話 電話 5.02070785193	景 高山 4.54139200639	巷子 狗狗 5.12606836758
傳真 電話 4.79756430061	景 頭頂 4.49883239197	巷子 老娘 4.90292481627
救助 電話 4.50988222816	景色 橋樑 5.71849541204	巷子 舞廳 4.79756430061
急難 電話 4.50988222816	景色 田園 5.47163533411	巷子 路燈 4.61524274382
電話 電話亭 4.17340999154	原野 景色 5.2484917828	巷子 景象 4.53520003614
業者 電信 5.3616344389	景色 老娘 5.13070874714	段 荒漠 4.43756156658
中華 電信 5.18728105175	景色 農場 5.09434110297	段 遊樂場 3.92673594281
訊號 電信 4.75549863533	景物 no matching pairs	景像 段 3.74441438602
簡訊 電信 4.71104687275	景致 飯店 3.20757352311	攤子 段 3.74441438602
手機 電信 4.46215082534	其中 景致 3.10052259411	原野 段 3.74441438602
自由港 no matching pairs	教堂 景致 3.05709812577	省道 no matching pairs
港 貝爾 5.44419146554	旅館 景致 2.99695521611	橋樑 馬路 5.2407697367
卡爾 港 4.75104428498	景致 阿 2.78711563042	原野 馬路 5.05844817991
港 神殿 4.59689360515	景象 線條 5.80288865486	峭壁 馬路 4.72197594329
港 碼頭 4.59689360515	大自然 景象 5.80288865486	土黃色 馬路 4.72197594329
奧運 港 4.59689360515	天色 景象 5.75159536047	網誌版 馬路 4.6529830718
旅遊 港灣 3.8783149867	景象 牆面 5.63381232481	城市 途 2.19487461517
教堂 港灣 3.86802834199	景象 氣息 5.34613025236	四 途 2.08507950244
港灣 路 3.69320722799	大自然 景像 6.06525291932	客棧 途 2.01323916475
客棧 港灣 2.82416938096	景像 酒吧 5.49814645966	旅行 途 1.97354340474
旅行 港灣 2.78447362096	故事 景像 5.066170226	朋友 途 1.82886069944
碼頭區 no matching pairs	景像 生活 4.50988222816	陸路 no matching pairs
避風港 no matching pairs	店 景像 4.31834401626	住戶 街 3.95770141871
港口 高山 4.72151817262	景緻 資訊 3.20305203839	尾端 街 3.92953054174
港 港口 4.54282638388	旅館 景緻 3.1792767729	嘴巴 街 3.86053767026
咖啡廳 港口 4.52480787838	交通 景緻 2.70615759087	老娘 街 3.81174750609
海邊 港口 4.24194509236	天 景緻 2.06521903657	外型 街 3.81174750609
港口 道路 4.22470328593	上 景緻 1.44244964044	街路 no matching pairs
山水 no matching pairs	景觀 田園 5.15247712956	橋樑 街道 5.65105413125
地景 no matching pairs	景觀 自然 4.62071829612	原野 街道 4.95790695069
背景 風俗 5.54395599569		

峭壁 街道 4.84457826538	特質 風俗 7.15339390812	客棧 採購 2.8887079021
老娘 街道 4.77558539389	個性 風俗 6.18831301208	採購 旅行 2.84901214209
天色 街道 4.77558539389	節慶 風俗 5.70647492519	我們 採購 2.17021546026
小攤 街頭 5.49071148117	生活 習俗 3.90374642459	天 採購 1.7367149696
藝人 街頭 5.23142888324	教堂 習俗 3.14410950276	頂下 no matching pairs
拼貼 街頭 4.72060325947	段 習俗 3.13827858245	買入 no matching pairs
異國 街頭 4.54624987233	此 習俗 2.71590300219	買下 no matching pairs
景象 街頭 4.53520003614	習俗 路 2.68160631631	買下來 no matching pairs
荒漠 路 3.98088930044	俗 no matching pairs	買去 no matching pairs
路 路邊攤 3.75774574912	民俗 風俗 7.33571546492	買來 no matching pairs
橋樑 路 3.75774574912	民俗 節慶 5.70647492519	湖 購物 4.44187749905
原野 路 3.75774574912	民俗 背景 5.54395599569	事宜 購物 4.20239752841
港灣 路 3.69320722799	民俗 生活 4.22220015571	咖啡廳 購物 3.94288633293
大自然 路徑 5.37210573876	文化 民俗 4.21665997533	空中 購物 3.87389346144
路徑 都市 4.57887509965	城市 民風 3.2289483827	便利 購物 3.76056477613
路徑 道路 4.57887509965	交通 民風 3.11162269898	購 no matching pairs
行人 路徑 4.55070422268	原野 風 5.08524637306	代購 no matching pairs
巷子 路徑 4.32756067137	人聲 風 5.02070785193	代買 no matching pairs
湖泊 路線 4.06093200811	農舍 風 4.97191768776	買 no matching pairs
大圖 路線 3.77324993566	天色 風 4.90292481627	摩肩接踵 no matching pairs
原野 路線 3.77324993566	鄉村 風 4.79756430061	摩肩擦踵 no matching pairs
○ 路線 3.77324993566	交通 風土 2.55200691105	擁擠 no matching pairs
巡禮 路線 3.69320722799	旅行 風土 2.22485783302	擠 no matching pairs
米飯 道 5.33339122658	天 風土 1.68792480543	擠滿 no matching pairs
原野 道 5.11024767527	一 風土 1.07181732728	人滿為患 no matching pairs
布丁 道 5.0820767983	我 風土 0.388841641463	水泄不通 no matching pairs
道 青菜 4.99691898996	咖啡館 習慣 4.83146585229	水洩不通 no matching pairs
圍牆 道 4.86338759734	盤子 習慣 4.41862063688	軋 no matching pairs
原野 道路 5.55970435266	習慣 酒吧 4.21029217135	盈門 no matching pairs
○ 道路 5.55970435266	居民 習慣 4.20391595421	擠人 no matching pairs
土黃色 道路 5.40555367283	習慣 道路 4.13831867173	客棧 荒漠 3.11185145341
巡禮 道路 5.29734008819	買進 no matching pairs	客棧 景像 3.11185145341
農場 道路 5.22323211604	瞎拼 no matching pairs	客棧 市鎮 3.11185145341
大自然 植物 6.30641497614	選購 no matching pairs	動物園 客棧 3.11185145341
大自然 景像 6.06525291932	購 no matching pairs	客棧 旅行 3.02579907914
大自然 景象 5.80288865486	購入 no matching pairs	價位 旅店 3.81534712232
大自然 神父 5.7287806827	購得 no matching pairs	廚房 旅店 3.33568158217
大自然 頭頂 5.70857797538	購買 no matching pairs	旅店 老闆 3.2526846626
曲線 自然 5.11447601138	購置 no matching pairs	旅店 街 3.14677120249
動物 自然 5.06568584721	置 no matching pairs	旅店 治安 3.12869347107
植物 自然 5.01916583157	辦 no matching pairs	旅社 青年 4.41025238721
材料 自然 4.99669297572	收 no matching pairs	旅社 街道 4.2986613218
大自然 自然 4.98344774897	收買 no matching pairs	旅社 最愛 4.16079152758
保護區 no matching pairs	收購 no matching pairs	巷子 旅社 4.02053563607
世俗 no matching pairs	沽 no matching pairs	旅社 風 3.73285356362
傳承 風俗 7.44107598058	採買 no matching pairs	旅舍 電話 3.99905660439
民俗 風俗 7.33571546492	採購 藝術 4.54762255614	客棧 旅舍 2.19556072154

旅舍 旅行 2.15586496153	上 門市 1.15476756799	網誌版 高山 5.97955819822
旅舍 時 2.05314645534	個 門市 0.703318508698	粉絲團 高山 5.97955819822
中 旅舍 2.02497557837	我 門市 0.437631805632	山 荒漠 5.15954429678
老娘 飯店 3.90072070367	鋪子 no matching pairs	山 蒙 4.34861408056
原野 飯店 3.90072070367	鋪 no matching pairs	山 纜車站 4.31224643639
農舍 飯店 3.74657002384	莊 酒 4.66793916295	山 水泥 4.17871504377
田園 飯店 3.61303863122	莊 葡萄 4.53235508401	山 植物 4.1479433851
人聲 飯店 3.61303863122	中世紀 莊 3.98581137764	吊橋 no matching pairs
會館 no matching pairs	莊 鎮 3.83542718061	虹橋 no matching pairs
賓館 酒店 4.6966581353	景 莊 3.5825416601	橋 沙漠 5.37430112933
城市 賓館 2.66933259476	小攤 街頭 5.49071148117	橋 橋樑 5.37430112933
客棧 賓館 2.26455359303	客棧 小攤 2.60102582965	橋 河岸 4.90429750008
旅行 賓館 2.22485783302	小攤 旅行 2.56133006964	橋 水道 4.82604530318
天 賓館 1.40024273298	人 小攤 1.67536254819	橋 瑪麗亞 4.73999444879
賓館 酒店 4.6966581353	地攤 no matching pairs	景色 橋樑 5.71849541204
遊覽 酒店 3.49626315232	路 路邊攤 3.75774574912	橋樑 街道 5.65105413125
士 酒店 2.88136816866	客棧 路邊攤 2.60102582965	橋樑 速度 5.575704694
巷子 酒店 2.53580120214	旅行 路邊攤 2.56133006964	橋 橋樑 5.37430112933
左 酒店 2.36590216534	你 路邊攤 1.4251729182	橋樑 馬路 5.2407697367
754853090101 民宿	我 路邊攤 0.725313878084	原野 城 4.15831288826
3.99241206167	攤子 段 3.74441438602	城 美名 3.97599133146
楊子 民宿 3.79174136621	客棧 攤子 2.41870427285	城 巡禮 3.71362706699
民宿 農舍 3.65593982505	攤子 旅行 2.37900851285	城 店舖 3.57052622335
交流 民宿 3.58694695357	人 攤子 1.49304099139	城 擴建 3.50598770222
主 人民宿 3.45128108841	一 攤子 1.22596800711	沙漠 都市 5.74202590945
旅館 田園 3.40242032421	攤 酒吧 4.45669258483	植物 都市 5.24125062154
旅館 橋樑 3.36159832969	攤 生活 3.75611042578	河岸 都市 5.08970072341
旅館 景緻 3.1792767729	攤 音樂 3.72547345632	景象 都市 5.07419653688
原野 旅館 3.1792767729	攤 風景 3.51972164772	原野 都市 5.04887872889
旅館 泳池 3.04574538028	攤 街 3.35808029616	小道 小鎮 4.08728083781
合作社 no matching pairs	戒心 治安 3.38820466656	小鎮 野外 3.86413728649
百貨商店 no matching pairs	偷竊 治安 3.38820466656	小鎮 橋樑 3.86413728649
店 橋樑 4.50066557305	搶匪 治安 3.36985552789	原野 小鎮 3.86413728649
單價 店 4.50066557305	大衣 治安 3.36985552789	小鎮 鎮 3.73687550688
店 店舖 4.42370453192	治安 財 3.2828441509	屯子 no matching pairs
店 景像 4.31834401626	住民 no matching pairs	市鎮 路 3.69320722799
原野 店 4.31834401626	居民 田園 5.33761956269	客棧 市鎮 3.11185145341
客棧 店舖 2.55223566548	原野 居民 5.11447601138	市鎮 旅行 3.07215569341
店舖 旅行 2.51253990547	居民 村落 4.87152983277	原野 田園 7.1125719136
天 店舖 1.91106835674	居民 山谷 4.85211174691	田園 鄉村 6.7071068055
店舖 那 1.82071285712	居民 平原 4.83684427478	田園 高山 6.27966279067
店舖 這 1.26696440881	山地 no matching pairs	天色 田園 6.23710317625
店 店舖 4.42370453192	山陵 no matching pairs	田園 風光 5.85269147734
城 店舖 3.57052622335	嶺 no matching pairs	上 村子 1.33708912478
店舖 邊 3.37204922634	原野 高山 6.56734486312	你 村子 1.13749084575
店舖 餐廳 3.08276587252	田園 高山 6.27966279067	個 村子 0.885640065492
店舖 路 2.88227701177	老娘 高山 6.16187975501	巡禮 村莊 5.46391328802
人 門市 1.20535891894		

村莊 鄉村 5.32081244438	城市 鄉鎮 3.0466268259	禮拜堂 阿拉伯 4.45371195669
村莊 鄉間 5.19564930142	城市 遊樂場 3.00580483138	皇家 禮拜堂 4.36915456866
天色 村莊 4.85080881513	田園 鄉村 6.7071068055	景 禮拜堂 4.34468171215
地勢 村莊 4.71104687275	原野 鄉村 6.48396325418	城堡 禮拜堂 4.32941424001
居民 村落 4.87152983277	綠地 鄉村 5.60849451683	教堂 禮拜堂 4.08160244228
城堡 村落 3.62443260181	天色 鄉村 5.60849451683	教堂 沙漠 3.93256686312
城市 村落 2.88064168843	斷崖 鄉村 5.54395599569	教堂 橋樑 3.93256686312
村落 遊客 2.86593554104	清清爽爽 no matching pairs	教堂 港灣 3.86802834199
村落 路線 2.73179606083	清新 no matching pairs	原野 教堂 3.7990354705
村鎮 路 3.69320722799	清潔 衛生 6.35880011123	古蹟 咖啡館 4.63393487683
天 村鎮 1.95985852091	消息 清潔 5.14896218745	古蹟 拱門 4.45161332004
城鄉 no matching pairs	客人 清潔 4.83530462859	古蹟 遺跡 4.37800150269
城鎮 農場 4.5719546568	清潔 生活 3.99905660439	古蹟 西元 4.08388853991
城鎮 田園 4.43842326418	店 清潔 3.80751839249	古蹟 宗教 3.96184110547
城鎮 巡禮 4.24059752085	淨 no matching pairs	花木 no matching pairs
城鎮 比基尼 4.21527971286	窗明几淨 no matching pairs	花樹 no matching pairs
原野 城鎮 4.21527971286	潔 no matching pairs	草木 no matching pairs
土黃色 鄉下 5.93237963794	潔淨 no matching pairs	葫蘆 no matching pairs
鄉下 高山 5.43594275163	清潔 衛生 6.35880011123	動物 植物 6.33986291021
農莊 鄉下 5.31334042954	旅遊 衛生 2.14371393131	大自然 植物 6.30641497614
斷崖 鄉下 5.28802262155	年 衛生 2.00517295107	植物 邊緣 6.14179299645
鄉下 鄉間 5.22739799974	交通 衛生 1.9523857885	天色 植物 5.91864944513
鄉野 no matching pairs	台灣 衛生 1.85124275411	宗教 植物 5.53423774622
巡禮 鄉間 6.09643584676	清爽 no matching pairs	教 no matching pairs
天色 鄉間 5.77101344632	明淨 no matching pairs	原野 宗教 5.62954792603
邊緣 鄉間 5.70647492519	純淨 no matching pairs	宗教 植物 5.53423774622
鄉間 高山 5.52589098829	乾乾淨淨 no matching pairs	宗教 金色 5.38268784809
賽車 鄉間 5.30100981708	政治 運動 5.48333137387	宗教 統治 5.34186585357
城市 鄉鎮 3.0466268259	政治 時期 5.29264156741	世人 宗教 5.34186585357
旅行 鄉鎮 2.0913264404	政治 概念 5.20302940872	山野 no matching pairs
各 鄉鎮 1.78508416976	宗教 政治 5.15954429678	郊野 no matching pairs
天 鄉鎮 1.77753696412	政治 道 4.64024404602	原野 田園 7.1125719136
西班牙 鄉鎮 1.05064890285	美酒 美食 4.6761429768	原野 高山 6.56734486312
原野 鎮 5.7325471655	主廚 美食 4.6581244713	原野 鄉村 6.48396325418
鎮 高山 4.89963804256	巡禮 美食 4.64262028476	原野 天色 6.30164169739
天色 鎮 4.85707842814	牛肉 美食 4.43498091999	原野 賽車 6.16811030476
耳 鎮 4.70292774832	美食 鄉間 4.37435629817	荒 no matching pairs
鄉間 鎮 4.69109329067	城堡 白雪公主 4.66588647664	荒地 no matching pairs
莊 酒 4.66793916295	城堡 農場 4.51173579681	荒郊 no matching pairs
莊 葡萄 4.53235508401	城堡 鄉間 4.47173046219	荒郊野外 no matching pairs
中世紀 莊 3.98581137764	城堡 迪士尼 4.44274292532	荒原 no matching pairs
莊 鎮 3.83542718061	城堡 時空 4.37820440418	荒野 no matching pairs
景 莊 3.5825416601	古跡 no matching pairs	山 荒漠 5.15954429678
鄉 no matching pairs	遺蹟 no matching pairs	段 荒漠 4.43756156658
城市 橋樑 3.29348690383	跡 no matching pairs	荒漠 路 3.98088930044
城市 民風 3.2289483827	天主堂 no matching pairs	交通 荒漠 3.39930477143
原野 城市 3.15995551121	世 禮拜堂 4.61277665132	客棧 荒漠 3.11185145341

曠野 no matching pairs	值 天色 5.16380869557	我 運 0.725313878084
小鎮 野外 3.86413728649	起價 no matching pairs	運載 no matching pairs
客棧 野外 2.60102582965	價錢 餐具 3.0676802351	工具 運輸 2.92473700829
旅行 野外 2.56133006964	價錢 售價 2.88227701177	博物館 運輸 2.67463764699
天 野外 1.7367149696	價錢 巷口 2.86700953964	交通 運輸 2.45484316259
賭場 no matching pairs	價錢 溝通 2.80816903962	公車 運輸 1.91761706005
市價 no matching pairs	互動 價錢 2.77999816265	旅行 運輸 1.79122184794
售價 店 3.54515412803	商 資料 2.61338743867	輸送 no matching pairs
價錢 售價 2.88227701177	一 商 1.14592529944	駁 no matching pairs
元 售價 2.55462437099	商 這 1.05339030851	交通 荒漠 3.39930477143
售價 路 2.51455223164	上 商 1.00573198883	交通 民風 3.11162269898
售價 四 2.22818034608	個 商 0.687814322163	交通 沙漠 2.88847914767
單價 店 4.50066557305	經濟 行動電話 5.2484917828	交通 橋樑 2.88847914767
單價 客棧 2.60102582965	傳真 經濟 4.84302667469	交通 原野 2.88847914767
單價 旅行 2.56133006964	政治 經濟 4.55534460224	瑜伽 no matching pairs
單價 天 1.7367149696	危機 經濟 4.30403017395	體育 no matching pairs
牌價 no matching pairs	一年 經濟 3.91349071606	政治 運動 5.48333137387
價目 no matching pairs	汙染 no matching pairs	宗教 運動 4.81123760251
價位 牛肉 4.22081223043	文物館 no matching pairs	居民 運動 4.40152620352
價位 菜色 3.83596640953	博物院 no matching pairs	運動 道路 4.11278536972
價位 旅店 3.81534712232	美術館 運動場 3.76794488343	運動 鎮 3.99794611011
價位 牛排 3.73310902409	美術 美術館 3.65016184777	食堂 no matching pairs
價位 高山 3.72437534412	美術館 蘇菲 3.58562332664	菜館 no matching pairs
價碼 no matching pairs	美術館 聖保祿 3.54480133212	飯館 no matching pairs
標價 no matching pairs	收藏 美術館 3.54480133212	飯廳 no matching pairs
貴賤 no matching pairs	博物館 藝術品 4.06093200811	館子 no matching pairs
價 行情 5.454343837	博物館 遊樂園 3.77324993566	酒家 no matching pairs
值 價 3.5825416601	博物館 禮拜堂 3.77324993566	橋樑 餐廳 3.95823460987
價 道路 3.10296857984	博物館 湖泊 3.77324993566	原野 餐廳 3.82470321725
價 缺點 2.93664736439	博物館 情境 3.77324993566	植物 餐廳 3.57524235762
價 道 2.87665545376	音樂廳 音樂會 4.06590891138	巡禮 餐廳 3.56233895278
價 行情 5.454343837	音樂 音樂會 3.75966482107	道 餐廳 3.549599927
交通 行情 2.70615759087	街 音樂會 3.56662504806	咖啡廳 高山 5.53772544594
客棧 行情 2.41870427285	場 音樂會 3.3352238115	咖啡廳 大自然 5.03563350214
旅行 行情 2.37900851285	店 音樂會 3.26229134201	下午茶 咖啡廳 5.03563350214
天 行情 1.55439341281	演奏會 no matching pairs	咖啡廳 鄉下 4.72840683362
底價 no matching pairs	輸 no matching pairs	咖啡廳 都市 4.71240649227
物價 薪水 5.51318433702	捐 no matching pairs	咖啡館 海洋 4.89554470897
溝通 物價 4.15506085287	捐來 no matching pairs	咖啡館 習慣 4.83146585229
物價 礦泉水 4.08606798138	送來 no matching pairs	古蹟 咖啡館 4.63393487683
商家 物價 4.04684726823	搭載 no matching pairs	咖啡館 酒吧 4.39953417099
失業率 物價 3.93764797626	搬運 no matching pairs	咖啡館 景觀 4.34154691334
訂價 no matching pairs	載運 no matching pairs	天 荒漠 2.24754059337
值 原野 5.81613388161	行李 運 2.76010283117	天 小道 2.24754059337
值 田園 5.75159536047	天 運 1.55439341281	天 天色 2.20498097895
值 區塊 5.43314162935	上 運 1.44244964044	天 景緻 2.06521903657
值 景象 5.26608754469	一 運 1.22596800711	天 天候 2.06521903657

原野 天色 6.30164169739	天氣 柏油路 4.44534370702	度 氣象 4.0116353866
天色 田園 6.23710317625	天氣 天色 4.44534370702	氣象 生活 3.88127356874
天色 老娘 5.9369985838	原野 天氣 4.44534370702	意思 氣象 3.6103986146
天色 植物 5.91864944513	天氣 成就感 4.26302215023	陰晴 no matching pairs
天色 鄉間 5.77101344632	地中海型 氣候 5.3616344389	野獸 no matching pairs
天候 餐廳 3.48823098063	大雪 氣候 4.95616933079	動物 植物 6.33986291021
天候 客棧 2.70638634531	氣候 高山 4.71104687275	動物 曲線 5.55970435266
天候 旅行 2.6666905853	氣候 鄉下 4.4125538842	動物 牛奶 5.49516583152
天候 時 2.38165052231	氣候 鄉村 4.33998319136	動物 大自然 5.32331557459
天 天候 2.06521903657	天氣 氣象 4.62766526382	動物 天色 5.27202228021
天氣 氣象 4.62766526382	景色 氣象 4.61988312337	

## A. 6. 2. Qyer

病院 no matching pairs	沙漠 群島 4.89300118352	遊 風光 3.45488175454
衛生院 no matching pairs	阻塞 no matching pairs	遊山玩水 no matching pairs
保羅 醫院 5.99587787097	封死 no matching pairs	遊逛 no matching pairs
聖保羅 醫院 5.92176989882	堵 no matching pairs	遊園 no matching pairs
醫生 醫院 5.54389274723	堵住 no matching pairs	遊歷 no matching pairs
藝術品 醫院 5.0795871391	堵塞 no matching pairs	湖泊 遊覽 5.73746618082
維修 醫院 5.05141626213	淤 no matching pairs	遊覽 風貌 5.1088575214
工藝品 店鋪 6.57982575957	閉塞 no matching pairs	村落 遊覽 5.1088575214
居民 工藝品 6.06041639211	塞住 no matching pairs	景象 遊覽 5.08687861468
工藝品 港口 5.70220416885	壅 no matching pairs	港灣 遊覽 4.96427629259
工藝品 風 5.19353139845	梗 no matching pairs	遠足 no matching pairs
工藝品 街道 5.08730918519	含義 生活 5.30658425973	踏青 no matching pairs
手工 藝品 7.1337108728	世俗 生活 5.28638155241	遨遊 no matching pairs
動物 藝品 6.59721750228	生活 異教徒 4.94990931579	環遊 no matching pairs
港口 藝品 5.4145220964	災難 生活 4.94990931579	咖啡廳 觀光 6.04627188473
山 藝品 4.98777958985	夜巴薩羅 生活 4.94990931579	港灣 觀光 5.0907604397
生活 藝品 4.83212628014	生涯 no matching pairs	優惠券 觀光 5.07549296757
藝術品 銅像 6.84602880034	海邊 溫飽 4.96180410344	村落 觀光 4.94765959606
標價 藝術品 6.28641301241	旅遊 溫飽 4.04722423597	景象 觀光 4.80789765368
文物 藝術品 6.03509858413	很多 溫飽 3.21849297647	喜慶 旅遊 4.55804985974
大理石 藝術品 6.03509858413	問題 溫飽 3.19641006018	旅遊 湖泊 4.15258475163
回廊 藝術品 5.98103136286	天 溫飽 2.47689745366	旅遊 溫飽 4.04722423597
觀光局 no matching pairs	賭場 遊樂場 7.4438658011	旅遊 醫療單 3.86490267918
巷子 旅遊局 4.98194866954	海濱 遊樂場 5.59383742375	旅遊 鄉村 3.86490267918
位子 旅遊局 4.7358156	居民 遊樂場 5.5720636242	夜生活 賭場 6.15288161978
公路 旅遊局 4.59473700174	自然 遊樂場 5.0102524457	夜生活 夜色 5.43993181193
感受 旅遊局 3.96402529212	經濟 遊樂場 4.86364897151	咖啡館 夜生活 5.21678826061
旅遊局 酒吧 3.96262571176	遊樂園 no matching pairs	夜生活 酒吧 4.87891644364
道 酒館 5.5713645679	出遊 經驗 3.21181674757	夜生活 貓 4.8391579515
道路 酒館 5.39852175506	信息 出遊 2.93753200469	慶典 節日 5.95075743569
種類 酒館 5.3580807855	出遊 路線 2.57941301718	節慶 節日 5.60983084872
啤酒 酒館 4.83754634761	出遊 計劃 2.37499278602	售價 節日 5.32214877627
物價 酒館 4.8351338012	出遊 本人 2.28836959703	世俗 節日 5.21678826061
舞場 no matching pairs	巡禮 no matching pairs	禮拜堂 節日 4.95442399615
咖啡廳 酒吧 5.5720636242	郊遊 no matching pairs	周一 節假日 4.57596689906
酒吧 音樂家 5.23559138758	山水 旅行 4.62411161042	平時 節假日 4.14359433803
賭場 酒吧 4.87891644364	傳真 旅行 4.49894846746	周六 節假日 4.02286579214
夜生活 酒吧 4.87891644364	旅行 醫療單 4.44179005362	周日 節假日 3.86452223405
快餐 酒吧 4.69659488684	旅行 產權證 4.44179005362	周末 節假日 3.58947190851
bar no matching pairs	旅行 證明信 4.33642953797	喜慶 教堂 4.76346713133
舞廳 no matching pairs	野遊 no matching pairs	喜慶 旅遊 4.55804985974
沙漠 道 5.46600405224	泳池 遊 4.35282334774	喜慶 天 2.70004100497
沙漠 飯店 4.98777958985	值 遊 4.22498997623	慶典 節日 5.95075743569
景色 沙漠 4.93648629546	自助 遊 3.75856416833	宗教 慶典 5.74741651168
沙漠 港口 4.90369647264	同 遊 3.66374286656	慶典 風 5.31131443411

慶典 生活 4.66222724334	消息 風 2.99102976401	湖泊 遊覽 5.73746618082
慶典 教堂 4.07031995077	涵義 no matching pairs	山 湖泊 5.39324469796
節 節慶 6.41489594549	意味 no matching pairs	教堂 湖泊 4.58114557453
節慶 節日 5.60983084872	意思 阿拉伯語 4.61141669094	旅遊 湖泊 4.15258475163
城 節慶 4.4581838634	意思 民風 4.45726601111	它 湖泊 3.99856454346
博物館 節慶 3.96823582466	意思 遊樂場 4.3619558313	港灣 湖 6.28355994343
教堂 節慶 3.66485484266	意思 沙漠 4.33948297545	村落 湖 6.14045909979
節 節慶 6.41489594549	意思 體育 4.27494445432	東部 湖 5.67045547054
民俗 節 6.15538475	意涵 no matching pairs	湖 美術館 5.53069352816
節 節日 4.57493437444	意義 階級 5.59653648072	制 湖 5.26499036243
節 車廂 4.27810310734	意義 破侖 5.46073493956	禮拜堂 美術 6.74618346537
節 美術館 4.27265350257	意義 歷史學家 5.3553744239	美術 美術館 5.7701734988
港口 飛機場 4.76393453026	思維 意義 5.2600642441	幅 美術 5.2857811321
醫院 飛機場 4.65214312427	勞動 意義 5.2600642441	畫 美術 5.18399843779
旅店 飛機場 4.60702268899	意義 概念 4.4729852437	主義 美術 5.15851935249
街道 飛機場 4.1490395466	大海 概念 4.28841396717	藝術 藝術品 5.51380166049
教堂 飛機場 4.11287956519	概念 設計 3.66025484698	含義 藝術 5.45176626958
16.3 航站 6.20945197127	概念 片 3.64715162198	禮拜堂 藝術 5.16919929779
單價 航站 5.80398686316	概念 選擇 3.53213374709	美術 藝術 5.15224973948
紫色 航站 5.51630479071	資料 鄉村 4.39072262033	藝術 赤足女 5.09509132564
照相 航站 5.51630479071	準假 資料 4.05425038371	灘 no matching pairs
專線 航站 5.33398323392	中英文 資料 3.90301941399	海邊 溫飽 4.96180410344
含意 no matching pairs	簽證費 資料 3.87704392758	海邊 賭場 4.82827271082
含義 政治 6.79723863617	3萬 資料 3.87192882692	工藝品 海邊 4.77948254665
含義 色彩 5.71204936784	旅行 資訊 3.8131813942	海邊 遊樂場 4.57881185119
含義 藝術 5.45176626958	博物館 資訊 3.45741020089	海邊 運動 4.50806099507
含義 經濟 5.31563409525	旅遊 資訊 3.23629401976	交流 街頭 4.04388127586
含義 生活 5.30658425973	它 資訊 3.08227381159	交流 生活 3.96908006278
旅館 見聞 3.80819535077	交通 資訊 2.94557243129	交流 餐廳 3.46096080061
交通 見聞 3.70771248333	義 no matching pairs	交流 旅行 3.46096080061
城市 見聞 3.59530553204	傳真 材料 4.88828419364	交流 材料 3.44483141868
次 見聞 2.37215251204	材料 醫療單 4.8311257798	你來我往 no matching pairs
天 見聞 2.14042521704	材料 簽證表 4.8311257798	相通 no matching pairs
定義 no matching pairs	材料 產權證 4.8311257798	居民 通訊 5.85974569665
信息 傳真 3.97551967154	信頭紙 材料 4.8311257798	經濟 通訊 5.15133104396
信息 湖泊 3.81300074205	異教徒 風 5.59899650656	藝術 通訊 4.48895552207
信息 簽字 3.65885006222	慶典 風 5.31131443411	店 通訊 4.22099795496
信息 白底 3.52531866959	工藝品 風 5.19353139845	博物館 通訊 3.76756512919
信息 紙片 3.47652850543	世俗 風 5.08817088279	溝通 電話 4.00629265842
界定 no matching pairs	風 黑夜 5.01120984166	他 溝通 3.00432096883
音訊 no matching pairs	訊息 no matching pairs	國家 溝通 2.98854646158
風信 no matching pairs	動物園 米羅 5.12040913362	溝通 點 2.51171166397
風聲 no matching pairs	凱旋門 動物園 4.77663759452	溝通 第一 2.47541706849
小道 消息 3.71926826438	動物園 街頭 4.61924542077	大哥大 no matching pairs
消息 習慣 3.17767098195	動物園 鐵票 4.3296234835	外線 no matching pairs
消息 道路 3.04186944079	動物園 山 4.23009388815	資訊網 no matching pairs
消息 缺點 3.00550179662	水潭 no matching pairs	網際 no matching pairs

電訊 no matching pairs	民風 背景 6.03509858413	小道 風光 4.97730828998
傳真 證明信 6.82622617305	商人 背景 6.00692770716	公路 小道 4.92324106871
傳真 公章 6.60308262173	田園 風光 5.95813754299	小道 氣候 4.77663759452
傳真 職位 6.42076106494	民風 風光 5.85277702733	小道 醫院 4.60958350985
傳真 公民 6.29559792199	大自然 風光 5.39852175506	公路 鎮 5.41571755381
傳真 申請人 6.16944663666	村莊 風光 5.1108396826	公路 高速路 5.23339599701
網路 no matching pairs	小道 風光 4.97730828998	公路 小道 4.92324106871
wifi no matching pairs	植物 風貌 5.69862634751	140 公路 4.85870254757
行動電話 no matching pairs	橋 風貌 5.14583652464	停車位 公路 4.76909038888
內褲 手機 4.75169864617	宗教 風貌 5.11880785225	出路 no matching pairs
小包 手機 4.33048518109	遊覽 風貌 5.1088575214	巷子 音樂會 5.02277066406
手機 職務 4.21270214544	街道 風貌 4.86416563388	巷子 旅遊局 4.98194866954
傳真 手機 4.15207752362	景 景象 5.13765094006	巷子 店鋪 4.6103851131
手機 軟件 4.00790773279	景 音樂會 5.09509132564	巷子 海鷗 4.55450465471
傳真 電話 4.99079164032	景 造型 4.65683639471	巷子 景象 4.55450465471
醫療單 電話 4.4918004742	夜生活 景 4.58426570187	奧迪 段 3.70256641477
產權證 電話 4.4918004742	山 景 4.41241544494	段 藝術品 3.59275154876
證明信 電話 4.38643995854	景色 村落 5.34195140357	段 風土 3.55353083561
總額 電話 4.3094789174	山地 景色 5.27741288243	旅遊局 段 3.45922015613
電信 no matching pairs	景色 湖 5.15962984677	段 酒館 3.45125198649
自由港 no matching pairs	景色 港灣 5.0795871391	省道 no matching pairs
民風 港 6.62288524903	景色 樹林 4.97730828998	16.3 馬路 5.46978477508
港 都市 5.8809479043	城市 景物 3.2386305881	音樂家 馬路 5.13331253846
制 港 5.52427296036	景致 道 5.32624210986	3多 馬路 5.13331253846
港 港口 5.49148313754	景致 植物 5.27118233268	單價 馬路 5.06431966697
港 背後 5.21678826061	居民 景致 4.83446468107	主樓 馬路 5.06431966697
村落 港灣 6.87134660833	景致 港口 4.47625245781	路途 3.53147922541
港灣 湖 6.28355994343	景致 路邊 4.38387913768	東西 途 3.45649922355
港口 港灣 5.89409517666	景象 賭場 6.70626685797	城市 途 3.08447990827
港灣 遊人 5.22268798274	景象 震撼 5.56843385615	後 途 3.03471049904
出口 港灣 5.17756754746	景象 植物 5.27118233268	中 途 2.80973716957
碼頭區 no matching pairs	景 景象 5.13765094006	陸路 no matching pairs
避風港 no matching pairs	景象 遊覽 5.08687861468	街 邊小 4.73218583195
港口 港灣 5.89409517666	景像 no matching pairs	街 賭場 4.73218583195
工藝品 港口 5.70220416885	景緻 no matching pairs	習俗 街 4.73218583195
港 港口 5.49148313754	景觀 運輸 6.50955656372	蘭布 街 4.61440279629
港口 都市 5.4145220964	典範 景觀 5.50055643374	景象 街 4.59242388957
港口 藝品 5.4145220964	景觀 自然 5.45386931977	街路 no matching pairs
山水 風景 5.42896278056	15世紀 景觀 5.43993181193	民風 街道 5.26963074199
山 山水 5.2878841823	人文 景觀 5.38277339809	山地 街道 5.20509222085
城堡 山水 5.02551991783	觀瞻 no matching pairs	工藝品 街道 5.08730918519
山水 旅行 4.62411161042	曠野 風景 5.65210633187	山丘 街道 4.95377779257
山水 教堂 4.25264150756	山水 風景 5.42896278056	街道 馬賽克 4.9049876284
地景 no matching pairs	賭場 風景 4.7358156	聖女 街頭 5.24785408019
常人 背景 6.20945197127	村落 風景 4.7358156	藝人 街頭 5.17313053399
背景 階層 6.18924926396	港灣 風景 4.69659488684	街頭 風貌 4.80156697756
背景 鄉村 6.03509858413	小路 小道 5.21092314116	人行道 街頭 4.76234626441

美術街頭 4.67640383461	習俗街 4.73218583195	頂下 no matching pairs
賭場路 4.27341657014	旅遊習俗 3.86490267918	買入 no matching pairs
路鄉村 4.15563353448	城市習俗 3.2386305881	買下 no matching pairs
攤子路 4.15563353448	天習俗 2.29457589686	買下來 no matching pairs
湖泊路 3.97331197769	上習俗 1.96778269429	買去 no matching pairs
景象路 3.88234019948	俗 no matching pairs	買來 no matching pairs
後路徑 2.56470686979	民俗節 6.15538475	16.3購物 5.45474689771
天路徑 2.00689382441	民俗附近 3.42856444622	折扣卡購物 5.27242534092
裏路徑 1.44855411826	城市民俗 3.08447990827	單價購物 5.2316033464
路徑這 1.40579529164	天民俗 2.36356876835	存包處購物 5.11827466109
你路徑 1.32693869063	上民俗 1.81363201447	主樓購物 5.04928178961
收銀臺路線 3.78338582151	民風都市 6.79723863617	購 no matching pairs
路線高山 3.74256382699	民風港 6.62288524903	代購 no matching pairs
紙片路線 3.62923514168	民風背景 6.03509858413	代買 no matching pairs
升降梯路線 3.62923514168	民風風光 5.85277702733	買 no matching pairs
勝利門路線 3.62923514168	民風港口 5.30916158075	摩肩接踵 no matching pairs
貴賤道 7.07544196467	異教徒風 5.59899650656	摩肩擦踵 no matching pairs
世俗道 5.87146916035	慶典風 5.31131443411	擁擠 no matching pairs
道酒館 5.5713645679	工藝品風 5.19353139845	擠 no matching pairs
沙漠道 5.46600405224	世俗風 5.08817088279	擠滿 no matching pairs
政治道 5.46600405224	風黑夜 5.01120984166	人滿為患 no matching pairs
成果道路 6.20945197127	人情風土 7.71352936805	水泄不通 no matching pairs
變革道路 5.87297973465	自然風土 5.13088043349	水洩不通 no matching pairs
傳奇道路 5.62166530637	天氣風土 3.87228654468	軌 no matching pairs
沖擊道路 5.51630479071	間風土 3.70166102765	盈門 no matching pairs
政治道路 5.44731191923	機會風土 3.63502386986	擠人 no matching pairs
大自然植物 6.05530129145	店鋪習慣 4.76971715167	交通客棧 3.57418109071
元素大自然 5.97306319321	建築物習慣 4.5534940432	客棧路 3.39794783278
人類大自然 5.57346320455	前者習慣 4.48895552207	天客棧 2.00689382441
大自然象征 5.55020634239	習慣色 4.34377351222	人客棧 1.90446801372
大自然懸崖 5.51630479071	習慣風光 4.1480289351	一客棧 1.68467728887
保護區自然 6.12721387304	買進 no matching pairs	旅店酒館 4.62900159571
自然花木 6.02185335738	瞎拼 no matching pairs	旅店飛機場 4.60702268899
自然運輸 5.46223756944	選購 no matching pairs	旅店被子 4.34131952326
大自然自然 5.46223756944	購 no matching pairs	旅店旅社 4.18716884343
景觀自然 5.45386931977	購入 no matching pairs	文筆旅店 4.00484728664
保護區自然 6.12721387304	購得 no matching pairs	旅社青年 4.98350026023
保護區公園 4.43047222869	購買 no matching pairs	床位旅社 4.55957308928
保護區教堂 4.35800202322	購置 no matching pairs	旅社物價 4.21097949213
保護區路 3.68562990524	置 no matching pairs	旅店旅社 4.18716884343
保護區天 2.00689382441	辦 no matching pairs	旅社街道 4.10647993218
世俗道 5.87146916035	收 no matching pairs	旅舍青年 5.16296874804
世俗宗教 5.52427296036	收買 no matching pairs	旅舍清真寺 4.56876151533
世俗生活 5.28638155241	收購 no matching pairs	旅舍飯店 4.43773325293
世俗經濟 5.24664122376	沽 no matching pairs	60旅舍 3.93324344761
世俗節日 5.21678826061	採買 no matching pairs	一般旅舍 3.85974098562
城市風俗 3.64409569621	採購 no matching pairs	賭場飯店 5.39324469796

沙漠 飯店 4.98777958985	小攤 山 4.18927189363	橋樑 no matching pairs
港灣 飯店 4.84319836104	地攤 no matching pairs	城 山地 4.86364897151
分店 飯店 4.82793088891	路邊攤 no matching pairs	城 景象 4.81919720894
酒館 飯店 4.80545803305	攤子 餐廳 5.02957671853	城 際 4.70949829168
會館 no matching pairs	博物館 攤子 4.55602248956	城 鄉間 4.70949829168
藝術品 賓館 4.27867848135	攤子 教堂 4.54032358001	城 王朝 4.68132741471
四星 賓館 4.21413996021	攤子 路 4.15563353448	民風 都市 6.79723863617
招牌 賓館 4.0831116978	後來 攤子 3.96402529212	港 都市 5.8809479043
管理 賓館 4.03181840341	天 攤 2.70004100497	16世紀 都市 5.78200795645
賓館 走廊 3.99099640889	扒手 治安 4.00782791478	宗教 都市 5.47548279619
賭場 酒店 3.36602023826	治安 脖子 3.76834794414	地位 都市 5.42099461091
特價房 酒店 3.36602023826	治安 胸 3.76834794414	小鎮 鎮 4.26581197075
東站 酒店 3.36602023826	失業率 治安 3.74999880548	小鎮 運動 4.19506111463
家堂巴特洛 酒店	治安 警惕 3.62789610867	夜巴薩羅 小鎮 4.17880059376
3.36602023826	住民 no matching pairs	1085年 小鎮 4.17880059376
女裝 酒店 3.36602023826	居民 拜占庭 6.17819942777	5435 小鎮 4.14106026578
民宿 華人 3.9299892957	居民 工藝品 6.06041639211	屯子 no matching pairs
旅館 民宿 3.71103160232	居民 通訊 5.85974569665	市鎮 no matching pairs
天氣 民宿 3.54686414425	3多 居民 5.84172719115	田園 風光 5.95813754299
交通 民宿 3.45639805505	主樓 居民 5.77273431966	宗教 田園 5.62963347602
民宿 間 3.19391707042	山地 氣候 5.87524988319	田園 自然 5.21092314116
旅館 醫療單 3.9623460306	山地 自然 5.32870617682	城堡 田園 4.72541532538
旅館 產權證 3.9623460306	山 山地 5.32870617682	田園 風景 4.5534940432
旅館 見聞 3.80819535077	山地 景色 5.27741288243	村子 村落 6.951389316
旅館 老板人 3.80819535077	山地 街道 5.20509222085	村子 風 4.905849326
旅館 正本 3.80819535077	山陵 no matching pairs	村子 經濟 4.77663759452
合作社 no matching pairs	嶺 no matching pairs	村子 藝術 4.40194414508
百貨商店 no matching pairs	港口 高山 5.30916158075	村子 海邊 4.08633536609
咖啡廳 店 5.00945531533	山 高山 5.1055626255	政治 村莊 5.69862634751
店 廚師 4.70935072288	城 高山 4.35282334774	村莊 風光 5.1108396826
工藝品 店 4.70935072288	教堂 高山 4.07031995077	居民 村莊 4.92543645927
店 熱量 4.64481220174	海邊 高山 3.98097485043	宗教 村莊 4.78233561563
店 店鋪 4.63789175889	山 湖泊 5.39324469796	村莊 白色 4.68339566778
工藝品 店鋪 6.57982575957	山 山地 5.32870617682	村子 村落 6.951389316
店鋪 習慣 4.76971715167	山 山水 5.2878841823	村落 港灣 6.87134660833
巧克力 店鋪 4.64991595186	山 高山 5.1055626255	村落 湖 6.14045909979
店 店鋪 4.63789175889	山 藝品 4.98777958985	村 村落 6.11848019307
店鋪 街道 4.6103851131	吊橋 no matching pairs	景色 村落 5.34195140357
店鋪 no matching pairs	2 虹橋 2.54670465744	村鎮 no matching pairs
門市 no matching pairs	天 虹橋 2.1892153812	城鄉 no matching pairs
鋪子 no matching pairs	上 虹橋 2.15010425109	城鎮 政治 5.15962984677
鋪 no matching pairs	時間 虹橋 2.1169609777	城鎮 景觀 4.87194777432
旅遊 莊 3.86490267918	光芒 橋 5.36898007596	城鎮 統治 4.79498673319
小攤 路邊 4.81132315251	橋 風貌 5.14583652464	城鎮 背景 4.75416473867
小攤 食物 4.74311490248	大橋 橋 4.96351496785	動物 城鎮 4.63555899626
傳說 小攤 4.44995336426	對岸 橋 4.64074157558	鄉下 風景 4.67127707886
城 小攤 4.23504031209	植物 橋 4.62704273123	城 鄉下 4.57596689906

海邊 鄉下 4.20411840175	後世 政治 6.83360628035	宗教 異教徒 6.03509858413
城市 鄉下 3.2386305881	年駕齡 美食 5.01504861796	宗教 災難 6.03509858413
天 鄉下 2.23003737572	美酒 美食 4.95789020412	宗教 法制 6.03509858413
鄉野 no matching pairs	20000 美食 4.83272706117	山野 no matching pairs
城 鄉間 4.70949829168	山地 美食 4.72736654551	郊野 no matching pairs
旅行 鄉間 4.2876393738	美食 那不勒斯 4.60958350985	原野 no matching pairs
教堂 鄉間 4.20385134339	城堡 花木 5.24866346914	荒 no matching pairs
路 鄉間 3.53147922541	城堡 山水 5.02551991783	教堂 荒地 4.47578505888
城市 鄉間 3.08447990827	城堡 顛峰 4.84319836104	城市 荒地 3.64409569621
鄉鎮 no matching pairs	城堡 聖喬治 4.84319836104	荒郊 no matching pairs
公路 鎮 5.41571755381	城堡 特茹河 4.84319836104	荒郊野外 no matching pairs
植物 鎮 5.31563409525	古跡 名勝 5.64692497313	荒原 no matching pairs
白色 鎮 4.41373210083	古跡 政治 5.4927742933	荒野 no matching pairs
鎮 首都 4.33838266359	公元 古跡 5.20509222085	荒漠 no matching pairs
山 鎮 4.31710526514	古跡 統治 5.08730918519	曠野 風景 5.65210633187
旅遊 莊 3.86490267918	古跡 王國 4.95377779257	教堂 曠野 4.76346713133
鄉 no matching pairs	遺蹟 no matching pairs	天 曠野 2.70004100497
城市 山水 3.70863421734	跡 no matching pairs	野外 no matching pairs
咖啡廳 城市 3.70863421734	天主堂 no matching pairs	賭場 遊樂場 7.4438658011
城市 港灣 3.66941350419	禮拜堂 美術 6.74618346537	景象 賭場 6.70626685797
城市 風俗 3.64409569621	油畫 禮拜堂 6.02977942265	夜生活 賭場 6.15288161978
城市 荒地 3.64409569621	世紀 禮拜堂 5.52587424173	午餐 賭場 5.50055643374
20世紀 鄉村 6.71500053794	國王 禮拜堂 5.36409963679	賭場 飯店 5.39324469796
背景 鄉村 6.03509858413	王宮 禮拜堂 5.33909833459	市價 no matching pairs
居民 鄉村 5.667373804	教堂 曠野 4.76346713133	售價 節日 5.32214877627
它們 鄉村 5.63408782637	喜慶 教堂 4.76346713133	售價 註 4.51252660472
宗教 鄉村 5.52427296036	教堂 禮拜堂 4.68342442365	售價 線路 4.34869963056
清清爽爽 no matching pairs	教堂 湖泊 4.58114557453	人員 售價 4.20091839902
清新 no matching pairs	攤子 教堂 4.54032358001	售價 市區 4.08633536609
清潔 no matching pairs	古蹟 no matching pairs	單價 航站 5.80398686316
淨 no matching pairs	自然 花木 6.02185335738	單價 從 5.67045547054
窗明几淨 no matching pairs	城堡 花木 5.24866346914	單價 居民 5.54959076835
潔 no matching pairs	花樹 no matching pairs	單價 購物 5.2316033464
潔淨 no matching pairs	草木 no matching pairs	單價 馬路 5.06431966697
條件 衛生 4.532874756	價錢 葫蘆 4.79382999506	牌價 no matching pairs
早飯 衛生 4.31233198639	旅遊 葫蘆 3.74711964352	價目 no matching pairs
衛生 衛生間 4.2849330122	教堂 葫蘆 3.66485484266	價位 飯店 3.98714770954
衛生 風 4.12308998675	感覺 葫蘆 3.52387462602	價位 餐 3.9193251132
組織 衛生 4.12308998675	英語 葫蘆 3.5148247905	價位 意義 3.90845544096
清爽 no matching pairs	大自然 植物 6.05530129145	價位 酒吧 3.76050152767
明淨 no matching pairs	植物 植被 5.92176989882	價位 店 3.66789684805
純淨 no matching pairs	植物 風貌 5.69862634751	價碼 no matching pairs
乾乾淨淨 no matching pairs	動物 植物 5.61638824927	標價 藝術品 6.28641301241
政治 異教徒 6.951389316	植物 熱帶 5.47993714654	價 標價 5.1232622026
政治 法制 6.951389316	教 no matching pairs	標價 藝術 4.24779346525
成果 政治 6.951389316	宗教 長袍 6.03509858413	店 標價 3.97983589815
政治 民主 6.83360628035	宗教 軍費 6.03509858413	標價 街 3.88488797156

貴賤道 7.07544196467	博物館 美術 4.79858412673	種 食堂 3.01408850339
天 貴賤 2.70004100497	博物館 禮拜堂 4.6991233332	菜館 no matching pairs
價 標價 5.1232622026	博物館 藝術品 4.68118563251	天色 飯館 5.31455242938
9歐 價 3.99647903697	博物館 工藝品 4.66138300522	403 飯館 4.91711949297
價 海景 3.96908006278	博物館 攤子 4.55602248956	飯店 飯館 4.64452766624
價 景 3.89111852131	背景 音樂會 5.15962984677	沙拉 飯館 4.63753463075
7歐 價 3.84232835714	景 音樂會 5.09509132564	街 飯館 4.5870038221
山 行情 4.95141194568	巷子 音樂會 5.02277066406	飯廳 no matching pairs
城市 行情 3.08447990827	醫院 音樂會 4.89726558231	景色 館子 4.60958350985
歐 行情 2.6420080484	特區 音樂會 4.89215048164	街道 館子 4.31411929696
天 行情 2.36356876835	演奏會 no matching pairs	面包 館子 4.27311127323
2 行情 2.21023242081	輸 no matching pairs	城堡 館子 4.22415915263
底價 no matching pairs	捎 no matching pairs	飯店 館子 4.15005118048
港灣 物價 5.16055620163	捎來 no matching pairs	酒家 no matching pairs
村落 物價 5.01745535799	送來 no matching pairs	賭場 餐廳 5.02957671853
值 物價 5.01745535799	搭載 no matching pairs	攤子 餐廳 5.02957671853
中餐館 物價 4.97823464484	搬運 no matching pairs	拜占庭 餐廳 4.84725516173
物價 酒館 4.8351338012	載運 no matching pairs	女裝 餐廳 4.84725516173
訂價 no matching pairs	天運 2.70004100497	酒館 餐廳 4.72947212608
值 物 6.34525351243	運載 no matching pairs	咖啡廳 觀光 6.04627188473
值 物價 5.01745535799	景觀 運輸 6.50955656372	咖啡廳 酒吧 5.5720636242
值 意義 4.62140524882	自然 運輸 5.46223756944	咖啡廳 店 5.00945531533
位子 值 4.5534940432	地區 運輸 4.60894268958	咖啡廳 城市 3.70863421734
值 自然 4.47695396608	海邊 運輸 4.33764979437	咖啡廳 歐 2.79615872822
起價 no matching pairs	交通 運輸 4.2185381071	咖啡館 天色 5.68679188986
價錢 葫蘆 4.79382999506	輸送 no matching pairs	去處 咖啡館 5.52694318892
價錢 港灣 4.42610521493	駁 no matching pairs	咖啡館 老店 5.39910981741
價錢 展品 4.16967568599	交通 運輸 4.2185381071	咖啡館 夜生活 5.21678826061
值 價錢 4.05986081998	交通 湖泊 3.86186316316	咖啡館 景觀 4.92910618816
價錢 飯店 3.98945717939	交通 汙染 3.86186316316	天運 2.70004100497
商天 2.1892153812	交通 樞紐 3.86186316316	天 貴賤 2.70004100497
政治 經濟 5.59494791803	交通 東站 3.86186316316	天 曠野 2.70004100497
經濟 醫療單 5.46978477508	瑜伽 no matching pairs	天 攤子 2.70004100497
經濟 軍費 5.46978477508	購物 體育 4.76159971715	天 攤 2.70004100497
產權證 經濟 5.46978477508	山 體育 4.41241544494	咖啡館 天色 5.68679188986
災難 經濟 5.46978477508	意思 體育 4.27494445432	天色 飯館 5.31455242938
交通 汙染 3.86186316316	店 體育 4.13398657797	天色 小朋友 5.14128070811
天 汙染 2.29457589686	海鮮 體育 3.85671709459	天色 部 4.77285687168
一 汙染 1.79003780453	夜巴薩羅 運動 6.24171283349	午飯 天色 4.76658725866
個 汙染 1.24454486731	5435 運動 6.20397250551	天候 no matching pairs
文物館 no matching pairs	弗朗明哥舞 運動	天氣 預報 4.58990658176
博物院 no matching pairs	6.19292266932	天氣 晴空 4.24001132481
格爾尼卡 美術館	華僑 運動 6.17025386951	天氣 暴雨 4.08586064498
5.83079812062	沖浪 運動 6.13635231783	天氣 時節 4.05768976801
名畫 美術館 5.83079812062	分鐘 食堂 3.70412041663	天氣 陰雨 3.95232925236
美術 美術館 5.7701734988	城市 食堂 3.2386305881	山地 氣候 5.87524988319
湖 美術館 5.53069352816	每 食堂 3.20660222995	植被 氣候 5.65210633187
繪畫 美術館 4.94840894042	各 食堂 3.0818563622	

氣候 領地 5.33625338245  
景觀 氣候 5.24664122376  
平方 氣候 5.12147808081  
氣象 no matching pairs

陰晴 no matching pairs  
野獸 no matching pairs  
動物 藝品 6.59721750228  
動物 植物 5.61638824927

動物 大自然 5.46223756944  
動物 藝術品 5.2878841823  
動物 手工 5.2878841823

## A. 7. Lago\_ruta (湖泊\_路線 2.txt)

網誌/0.0 圖文版/0.0

網誌圖文版

\*\*\*SCORE=0.0

巴塞隆納/0.0 的/0.0 凱旋門/0.0 往/0.0 海邊/0.0 走/0.0

巴塞隆納的凱旋門往海邊走

\*\*\*SCORE=-0.3125

會/0.0 到達/0.2945837 城堡/0.0 公園/0.0 PARQUE/0.0 DE/0.0 LA/0.0 CIUDADELA/0.0 , /0.0

會到達城堡公園 PARQUEDELA CIUDADELA,

\*\*\*SCORE=0.2945837

距離/0.0 市中心/0.0 出發/0.0361295999999996 約/0.0 10/0.0 幾/0.0 分鐘/0.0 車程/0.0 , /0.0

距離市中心出發約 10 幾分鐘車程,

\*\*\*SCORE=-0.3888704

城堡/0.0 公園/0.0 佔地/0.0 31/0.0 公頃/0.0 , /0.0 包括/0.05 市議會/0.0 大廳/0.0 、 /0.0 動物園/0.0 和 /0.0 幾/0.0 個/0.0 博物館/0.0

城堡公園佔地 31 公頃, 包括市議會大廳、動物園和幾個博物館

\*\*\*SCORE=-0.27832275000000006

城堡/0.0 公園/0.0 有/0.0 寬廣/0.3 的/0.0 綠地/0.0 和/0.0 湖泊/0.0

城堡公園有寬廣的綠地和湖泊

\*\*\*SCORE=0.3

靠近/0.10533890000000001 凱旋門/0.0 這/0.0 一/0.0 側/0.0

靠近凱旋門這一側

\*\*\*SCORE=0.10533890000000001

還/0.0 有/0.0 寬廣/0.3 的/0.0 綠/0.0 廊道/0.0

還有寬廣的綠廊道

\*\*\*SCORE=0.3

草地/0.0 上/0.0 常/0.0 有/0.0 人/0.0 悠閒/0.0 的/0.0 躺/0.0 著/0.0 曬太陽/0.0

草地上常有人悠閒的躺著曬太陽

\*\*\*SCORE=0.0

高/0.0 第/0.0 ANTONI/0.0 GAUDI/0.0 I/0.0 CORNET/0.0 在/0.0 巴塞隆納市/0.0 省立/0.0 建築/0.05 學校/0.0 時/0.0

高第 ANTONI GAUDI CORNET 在巴塞隆納市省立建築學校時

\*\*\*SCORE=0.392468

到/0.0 建築師/0.0 事務所/0.0 實習/0.0

到建築師事務所實習

\*\*\*SCORE=0.41419105000000006

當時/0.0 就/0.0 開始/0.11623320000000004 協助/0.525 事務所/0.0 的/0.0 工作/0.0

當時就開始協助事務所的工作

\*\*\*SCORE=0.96669245

他/0.0 曾經/0.0 參與/0.05 從事/0.2433249 景觀/0.0 的/0.0 相關/0.3623905999999999 裝飾/0.0 設計/0.013457100000000007 , /0.0

他曾經參與從事景觀的相關裝飾設計,

\*\*\*SCORE=0.3172583999999999

翠綠/0.0 的/0.0 湖泊/0.0 , /0.0 成群/0.0 的/0.0 雁鴨/0.0

翠綠的湖泊, 成群的雁鴨

\*\*\*SCORE=0.4369127

城堡/0.0 公園/0.0 以前/0.0 是/0.0 城堡/0.0

城堡公園以前是城堡

\*\*\*SCORE=0.0

後來/0.0 拆除/0.04289055000000001 之後/0.0 只/0.0 剩下/0.0 教堂/0.0 和/0.0 軍械庫/0.0

後來拆除之後只剩下教堂和軍械庫

\*\*\*SCORE=-0.51789055

原文/0.0 載/0.0 於/0.0 :/0.0 背包/0.0 客棧/0.0 自助/0.4579737 旅行/0.0 論壇/0.0

原文載於:背包客棧自助旅行論壇

\*\*\*SCORE=0.1079737

軍械庫/0.0 後來/0.0 做為/0.0 市議會/0.0 之/0.0 用/0.0

軍械庫後來做為市議會之用

\*\*\*SCORE=0.0

城堡/0.0 公園/0.0 裡/0.0 的/0.0 溫室/0.0

城堡公園裡的溫室

\*\*\*SCORE=0.375

\*\*\*SCORE=0.012424000000000102