EMOTION DETECTION IN PUBLIC SPACE: A MULTILANGUAGE COMPARISON IN BARCELONA

Yang, Liya [1];* Marmolejo Duarte, Carlos [2]; Martí Ciriquián, Pablo [3]

Initial submission: 2019-06-07; **Definitive submission:** 2019-11-20; **Publication:** 2019-12-21

Citación: Yang, L. *et al.* (2019). Emotion Detection in Public Space: A Multilanguage Comparison in Barcelona. In *XIII CTV 2019 Proceedings: XIII International Conference on Virtual City and Territory: "Challenges and paradigms of the contemporary city":* UPC, Barcelona, October 2-4, 2019. Barcelona: CPSV, 2019, p. 8515. E-ISSN 2604-6512. DOI http://dx.doi.org/10.5821/ctv.8515

Abstract

Sentiment analysis via LBSN (Location-based social network) data has been a popular topic in urban studies since the booming of social media applications, such as work stress, the emotion of railway passengers, mapping sentiment, etc. Although it is difficult to measure variations of mass emotions on a precise level, there are some correlations between emotion and spatial environment. Therefore, understanding mass emotion is beneficial to improve the allocation of urban facilities and promote the urban environment. However, most researches are limited to English texts or single language due to the studied area or the technical problems of analyzing different languages. In fact, immigrants and visitors usually take an important portion in international metropolises. The analysis based on a single language is not sufficient to reveal perceptions about the same city from people who use other languages. Moreover, except for the cultural differences, the mass emotion is possibly different in different urban spaces, such as local and tourist spaces. As local language is usually distinct from visitors', the sentiment analysis based on multi-language could reflect the differences to some degree. Therefore, this study aims to detect the difference of mass emotion between people who use different languages in the same public space. Moreover, Previous studies mainly focus on a single type of land-use, such as tourist attractions or green parks. For filling the gap, the ultimate goal of the research is to explore the relationship between the urban environment and the mass emotion.

This study utilizes 30 months of Twitter data to analyze the mass emotions in Barcelona. Specifically, English, Spanish, and Catalan are involved in the comparison of emotions as the case study, because the number of tweets written by the three languages account for about 90% of our dataset. The analysis is composed of an analysis of high-frequency words and sentiment analysis on plazas. The sentiment analysis is implemented by two commonly used algorithms: Senti-strength that estimates sentiments in short informal texts and Svader that specifically focus on the social media texts. Based on the sentiment (positive, neutral, negative) given by the algorithm, a comprehensive score of sentiment is assigned to each tweet. In brief, the process includes: 1) cleaning data and removing non-individual tweets; 2) translating Spanish and Catalan tweets into English through Google Translate API; 3) calculating the sentiment score of each tweet via Senti-strength and Svader software; 4) comparing the sentiment classification from the two software; 5) a sample check of sentiment analysis via manual evaluation; 6) comparing the sentiment differences between the three groups of different language in twenty public spaces of Barcelona.

The result confirms the differences of high-frequency words between the three languages, though they have some words in common. The high-frequency Catalan tweets appeared more words which are names of local places. English tweets contained more words that are related to tourism. Spanish tweets seemed to be in between. In terms of sentiment variations, the proportion of positive emotion was higher than negative emotion in general.

Resumen

El análisis de sentimientos a través de los datos de LBSN (red social basada en la ubicación) ha sido un tema popular en los estudios urbanos desde el auge de las aplicaciones de redes sociales, como el estrés laboral, la emoción de los pasajeros del ferrocarril, el mapeo de sentimientos, etc. Aunque es difícil de medir variaciones de las emociones de masas en un nivel preciso, hay algunas correlaciones entre la emoción y el entorno espacial. Por lo tanto, comprender la

[1] PhD Candidate, Universitat Politècnica de Catalunya, Barcelona's School of Architecture (ETSAB), Centre for Land Policy and Valuations, Barcelona, Spain, https://orcid.org/0000-0002-0247-2701; [2] Associate Professor, Universitat Politècnica de Catalunya, Barcelona's School of Architecture (ETSAB), Centre for Land Policy and Valuations, Barcelona, Spain, https://orcid.org/0000-0001-7051-7337; [3] Full Professor, Department of Building and Urban Planning, Universidad de Alicante, https://orcid.org/0000-0002-6877-784X. * Correo de contacto: li.ya.yang@upc.edu

emoción masiva es beneficioso para mejorar la asignación de las instalaciones urbanas y promover el entorno urbano. Sin embargo, la mayoría de las investigaciones se limitan a textos en inglés o en un solo idioma debido al área estudiada o los problemas técnicos de analizar diferentes idiomas. De hecho, los inmigrantes y visitantes suelen ocupar una parte importante en las metrópolis internacionales. El análisis basado en un solo idioma no es suficiente para revelar percepciones sobre la misma ciudad de personas que usan otros idiomas. Además, a excepción de las diferencias culturales, la emoción de masas es posiblemente diferente en diferentes espacios urbanos, como los espacios locales y turísticos. Como el idioma local suele ser distinto del de los visitantes, el análisis de sentimientos basado en varios idiomas podría reflejar las diferencias hasta cierto punto. Por lo tanto, este estudio tiene como objetivo detectar la diferencia de emoción en masa entre las personas que usan diferentes idiomas en el mismo espacio público. Además, los estudios anteriores se centran principalmente en un solo tipo de uso de la tierra, como atracciones turísticas o parques verdes. Para llenar el vacío, el objetivo final de la investigación es explorar la relación entre el entorno urbano y la emoción masiva.

Este estudio utiliza 30 meses de datos de Twitter para analizar las emociones masivas en Barcelona. Específicamente, el inglés, el español y el catalán están involucrados en la comparación de las emociones como estudio de caso, porque el número de tweets escritos por los tres idiomas representa aproximadamente el 90% de nuestro conjunto de datos. El análisis se compone de un análisis de palabras de alta frecuencia y análisis de sentimientos en plazas. El análisis de sentimientos se implementa mediante dos algoritmos de uso común: Senti-strength que estima los sentimientos en textos informales cortos y Svader que se enfoca específicamente en los textos de las redes sociales. Según el sentimiento (positivo, neutral, negativo) dado por el algoritmo, se asigna una puntuación integral de sentimiento a cada tweet. En resumen, el proceso incluye: 1) limpiar datos y eliminar tweets no individuales; 2) traducir tweets en español y catalán al inglés a través de la API de Google Translate; 3) calcular el puntaje de opinión de cada tweet a través del software Senti-strength y Svader; 4) comparar la clasificación de sentimientos de los dos software; 5) una verificación de muestra del análisis de sentimientos mediante evaluación manual; 6) comparar las diferencias de sentimiento entre los tres grupos de diferentes idiomas en veinte espacios públicos de Barcelona.

El resultado confirma las diferencias de las palabras de alta frecuencia entre los tres idiomas, aunque tienen algunas palabras en común. Los tweets catalanes de alta frecuencia aparecieron más palabras que son nombres de lugares locales. Los tweets en inglés contenían más palabras relacionadas con el turismo. Los tweets en español parecían estar en el medio. En términos de variaciones de sentimientos, la proporción de emoción positiva fue mayor que la emoción negativa en general.

Palabras Clave: Emoción publica; Twitter emoción; Espacio público; Smart city

Key words: Mass emotion; Twitter sentiment; Public space; Smart city

1. Introduction

Sentiment analysis via Location-based social network (LBSN)data has been a popular topic in urban studies, such as work stress (Wang *et al.,*2016), the emotion of railway passengers (Collins *et al.*,2013), mapping sentiment (Li *et al.,*2016), etc. Although it is difficult to measure variations of mass emotions on a precise level, there are some correlations between emotion and spatial environment. For instance, the distance to landfill has a negative correlation with subjective well-being (Brereton *et al*. ,2008). Larger and greener parks delivered more happiness to people than small ones (Schwartz *et al.*,2018). Therefore, understanding mass emotion is beneficial to improve the allocation of urban facilities and promote the urban environment.

However, most researches are limited to English texts or single language due to the studied area or the technical problems of analyzing different languages. In fact, immigrants and visitors usually take an important portion in international metropolises. The analysis based on single language is not sufficient to reveal perceptions about the same city from people who use other languages. Moreover, except the culture differences, the mass emotion is possibly different in different urban spaces, such as local and tourist space. As local language is usually distinct from visitors', the sentiment analysis based on multi-language could reflect the differences in some degree. Therefore, this study aims to detect the difference of mass emotion between people who use

ISSN e: 2604-6512

different languages in the same public space. Moreover, Previous studies mainly focus on a single type of land-use, such as tourist attractions or green parks. For filling the gap, the ultimate goal of the research is to explore the relationship between the urban environment and the mass emotion. Since Twitter emotion is a fast access to observe people's perception of a city, it could be a social indicator in urban management in future.

Therefore, this study utilizes 30 months of Twitter data to analyze the mass emotions in Barcelona. Specifically, English, Spanish and Catalan are involved in the study of sentiment detection, because the number of tweets written by the three language account for about 90% of our dataset. The analysis is composed by analysis of high frequency words and the sentiment analysis on plazas. The sentiment classification is implemented by two commonly used algorithms which focus on short texts. The second step is to analyze the correlation between positive emotion and the urban environment. Due to the deficiency of the data itself, only Spanish and English tweets involve in this part of analysis.

The result confirms the differences of high frequency words between the three languages, though they have some words in common. The high frequency Catalan tweets appeared more words which are names of local places. English tweets contained more words which are related with tourism. Spanish tweets seemed to be in between. In terms of sentiment scores, the proportion of positive emotion was higher than negative emotion in general. The strength of emotions is correlated with the spatial proximity.

2. Literature review

2.1 Geographical studies of urban emotion

Mapping urban emotion could date back to the 1960s. Lynch (1960) brought up the concept of "mental map" to represent people's perception of their build environment. Based on the framework, Kuipers (1978) elaborated a theoretical model to state a person's cognitive map (i.e. how people store their spatial surroundings in their mind). However, the cognitive centered conception was challenged by an interactive theory in the 1990s. Space is a material that the body engages and works with (Luption,1998), rather than an objective existence solely. Therefore, "emotions can be conceptualized as the felt and sensed reaction that arise in the midst of the (inter) corporal exchange between self and world" (Davidson *et al.,* 2012). In other words, the spatial environment influences people's emotion to some degree. Although it is hard to measure the individual's emotional reaction to a spatial place, the common feeling is possible to be observed and aggregated. For example, Molz(2005) analyzed travelers' emotional responses when they were eating at McDonald's. It extracted contents from forty websites when these travelers were traveling around the world. The result showed that McDonald's evoked travelers' emotion of familiarity that mixed contentment and contempt.

Personal questionnaire is the traditional way to understand the relationship between spatial place and people's feeling. Matei *et al.* (2001) visualized a first digital map of emotion in Los Angeles, USA. They aggregated 215 participants' perception about the residential communities into a map of fear feeling. Obviously, such a method has great limit on the number of samples and places. Therefore, the LBSN data, such as Twitter or Facebook, naturally has been concerned by researchers in recent years. In addition to the geo-spatial information, the contents of LBSN data provide a fast access to understand people's opinion and emotion. It can provide valuable

ISSN e: 2604-6512

information about the work stress (Wang *et al.,*2016), elections (Wang *et al.,* 2012), social movement (LeFebvre & Armstrong ,2018), even the stock market (Pagolu *et al.,* 2016), etc. Along with the geographical information, researches also study the relationship between spatial place and mass emotion. Gallegos *et al.* (2016) utilized Foursquare data to study the happier places in Los Angeles. It concluded that the happier places tend to be observed in census tracts which have more Foursquare check-ins. Collins *et al.* (2013) studied the emotion of passengers of suburban trains near the city of Chicago. They found out that the dissatisfaction to incidents can be measured by social media data. With regard to the emotional reaction in specific places, Padilla *et al.* (2018) investigated tourists' emotions at tourist destinations via Twitter data in Chicago. Their result showed that seasonal temperature is positively correlated with the positive sentiment in general. Urban parks could also reduce people's negative feelings according to the investigation of Schwartz *et al.* (2018) in San Francisco because negation words such as 'no', 'not' decreased in frequency during visits to urban parks.

2.2 Sentiment analysis of Twitter

The sentiment classification is derived from psychological theories (Ekman & Cordaro, 2011; Vytal &Hamann,2010) which study the basic emotions like happiness, fear, sadness, etc. The sentiment classification of "positive - neutral-negative" is derived from basic emotions. Further, language as the most direct expression of emotions, people express their own feelings and evaluate other people's emotions through words. One of the most famous related studies is conducted by Shaver *et al*. (1987) which classified emotional words into six emotion categories: love, anger, joy, sadness, surprise, and fear. Therefore, it is reasonable to observe the mass emotion using Twitter texts. The simplest way to measure emotion automatically is to calculate the word-frequency (López-Ornelas & Zaragoza 2015; Lim, et al. 2016; Quercia, *et al.* 2012). However, single words usually cannot represent the completed emotional trend of a phase or paragraph. Therefore, various algorithms are developed to solve the problem automatically, such as Naïve Bayes classifier (Pak, 2010), Graph based Semi-Supervised Learning (Chapman et.al 2018), Latent Dirichlet Allocation(LDA) (Kovacs-Gyori, et al. 2018), etc. It has become an emerging academic field of natural language processing which is aim to let the computer "understand" human language.

In summary, Twitter and other LBSN data have provided a more economical way to access dynamic mass emotions. However, the majority urban emotion studies mainly focus on specific places or some groups of people. The relationship between the macro- built environment and mass emotions still have plenty room for investigation. Moreover, most researches are limited to English texts or single language, lack of comparison of different groups of people who come from different cultural backgrounds. Therefore, to fill this gap, this study aims to provide a triple-language comparison of sentiment analysis in Barcelona.

3. Methods

3.1 Study scope and description of Twitter datasets

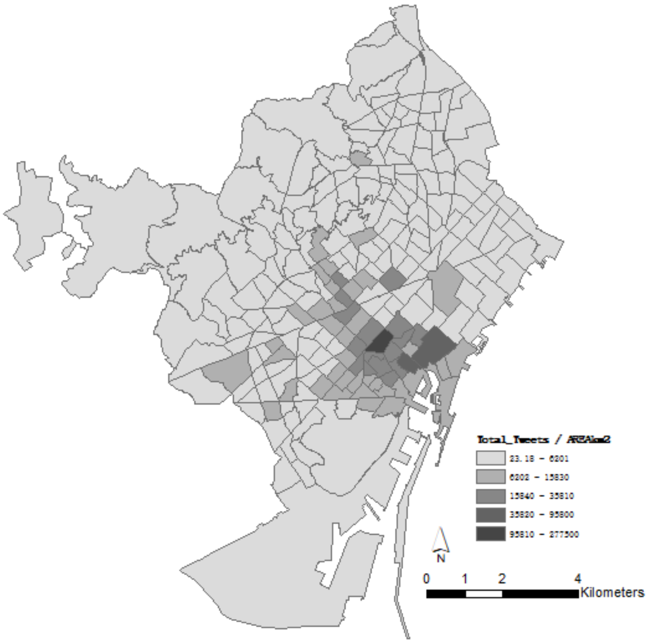
The study scope is restricted to Barcelona city, Spain. It is an international metropolis in which more than 160 million people live. 333,516 foreigners are registered as residents in Barcelona. Moreover, it is a famous tourism city all over the world which visitors were over 9 million in 2017.

ISSN e: 2604-6512

Apparently, a sentiment analysis based on single language cannot reflect the urban emotion comprehensively.

The cleaned Twitter dataset contains 707,549 tweets which were generated by 63,178 users, ranged from September of 2016 to April of 2019. For reducing noises, users who only appeared once in the studied area are excluded from the dataset. What's more, the non-individual accounts and their tweets, such as weather information, companies, websites, etc., were also removed as many as possible. Figure 1 shows the density of Tweets based on the basic statistical area (AEB) of Barcelona. It is clear that the densest area is located in the city center which is also the historical area of Barcelona.

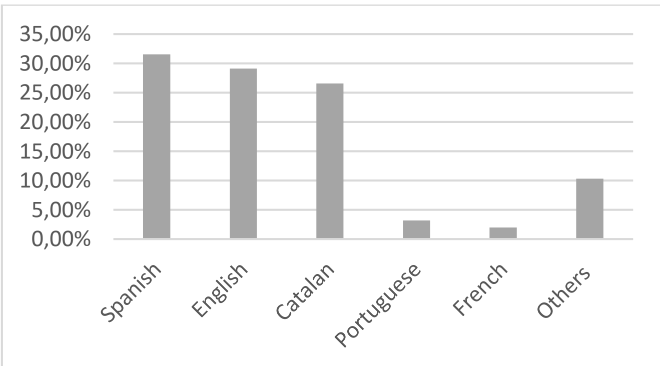
Figure 1. **Density of Tweets of Barcelona**



Source: self-elaboration

With regard to languages of tweets, 47 languages were detected by Twitter automatically. Spanish, English and Catalan account for nearly 80% of all tweets (Figure 2). Spanish and Catalan are the official languages of Catalonia autonomous community. Such language composition of Twitter is accord with the actual situation.

Figure 2. **Composition of Tweets' Languages**



Source: self-elaboration

ISSN e: 2604-6512

3.2 Sentiment classification

Firstly, Spanish and Catalan tweets were translated into English via Google Translate automatically. After the translation, two commonly used algorithms were introduced to generate the sentiment scores of each tweets: Svader that specifically focus on the social media texts (Hutto & Gilbert 2014), and Senti-strength that estimates sentiments in short informal texts (Thelwall, et al. 2010). Both of them are rule-based sentiment analysis which relies on lexical resources of emotionally charged words. Such dictionaries usually are a list of words which are labeled by the emotion orientation (positive or negative). Based on the dictionaries and rules of evaluation, the algorithm labels each word in the text and assign sentiment scores for them.

Svader rates the sentiment of lexicon in term of sentiment polarity (negative/ positive) and the intensity, using a range from -4(maximum negative valence) to 4(maximum positive valence) and normalizes the score to [-1,1]. It means that it considers the category of the sentiment as well as the intensity. For example, the word "great" has higher positive valence than "good", due to its intensity. According to its official guide, the sentiment of a given sentence can be measured by the compound score:

1. positive sentiment: compound score >= 0.05
2. neutral sentiment: (compound score > -0.05) and (compound score < 0.05)
3. negative sentiment: compound score <= -0.05

Similarly, SentiStrength contains an original collection of 298 positive terms and 465 negative terms classified with relative sentiment intensity. Ratings of 1 to 5 indicate a positive sentiment and -1 to -5 indicate a negative sentiment. The sum of positive and negative score (senti_compound) is the final sentiment measure of the sentence：

1. positive sentiment: senti_compound > 0
2. neutral sentiment: senti_compound = 0
3. negative sentiment: senti_compound < 0

For improving the preciseness of sentiment analysis, the final classification of emotion of each tweet is the intersection of the two algorithms. One positive tweet should be confirmed by both algorithms, otherwise, it is excluded from the sentiment dataset. Neutral and negative tweets are confirmed in the same manner. In addition, to measure whether the translation greatly affects emotion detection, a manual evaluation of Spanish and Catalan tweets is conducted by a sample test.

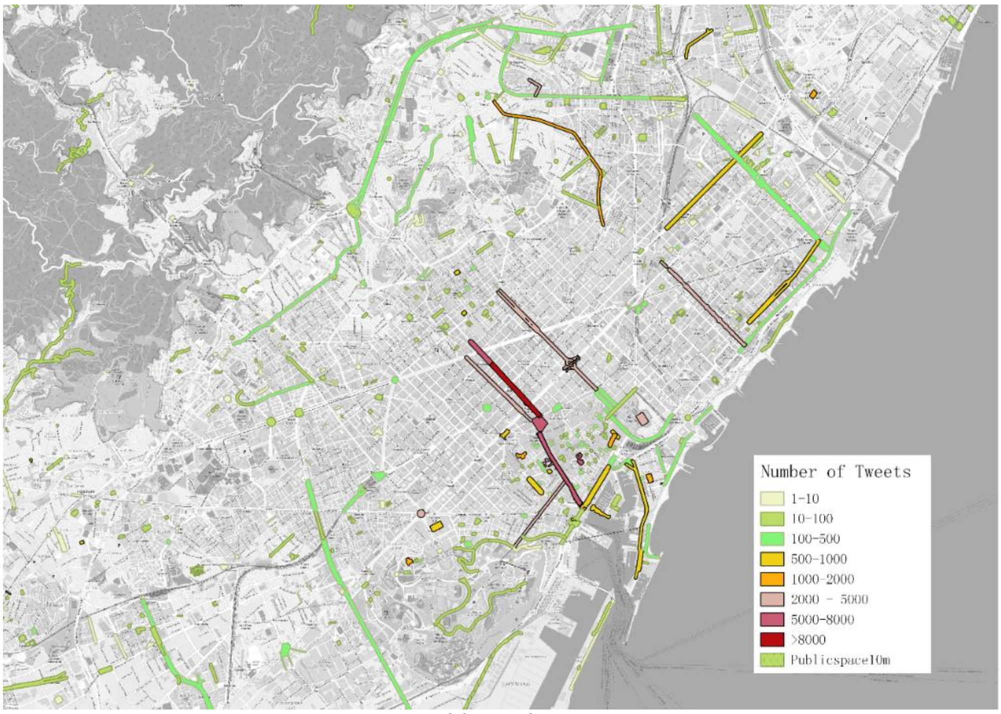
4. Results

4.1 Spatial variation of Twitter activities

According to London Assembly's 2011's report, public space refers to "all spaces including streets, squares and parks that everyone can use and access in principle." Therefore, based on OpenStreetMap, we choose plazas and major pedestrian avenues in Barcelona as the representatives of public space (Figure 3), because plazas and pedestrian avenues belong to the most common public space in European cities. What's more, all public areas were expanded by a 10-meter buffer for collecting tweets, as the boundary of these areas are not very precise.

ISSN e: 2604-6512

Figure 3. **Spatial distribution of Tweets in public spaces of Barcelona city**



(a) overview

(b) central area of Barcelona

Source: self-elaboration

After aggregation, the total polygons of the public space are 1478 in Barcelona. There are 581 polygons contain Twitter messages. However, 75% of polygons contains less than 100 tweets. The majority of tweets were gathered in polygons of the central area of Barcelona (Figure 3 (b)). The densest public space is along with Passeig de Gracia- la Rambla.

ISSN e: 2604-6512

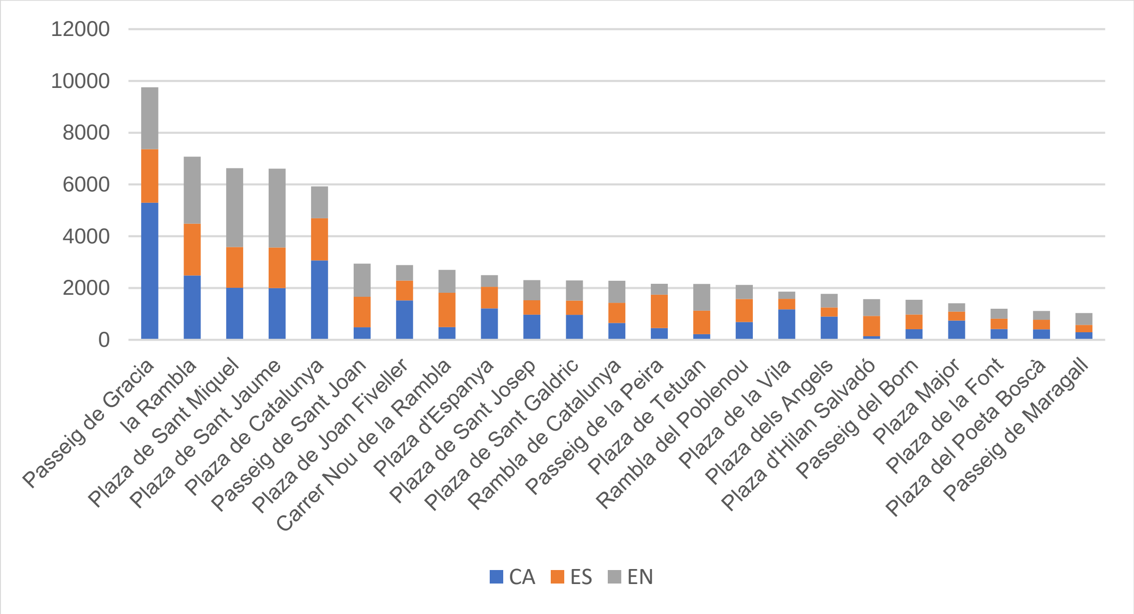
Table 7. **Spatial distribution of Tweets in public spaces**

N. of tweets	N. of polygon	%
10<	266	45.78%
10-100	182	31.33%
100-500	98	16.87%
500-1000	11	1.89%
1000-2000	8	1.38%
2000-5000	11	1.89%
5000-8000	6	1.03%
>8000	1	0.17%

Source: self-elaboration

Figure 4 lists the distribution of languages in 23 public spaces which contain more than 1000 tweets. In terms of language distributions, several areas show evident different components of languages. For example, Passeig de Gracia aggregated a lot of Catalan tweets, however, the proportion of Catalan reduced gradually when the space moves down to la Rambla and two plazas of the historical center (Plaza de Sant Miquel, Plaza de Sant Jaume). Spanish tweets take the largest proportion in Passeig de la Peira and Plaza de Tetuan.

Figure 4. **Distribution of the three languages in public spaces of Barcelona**



Source: self-elaboration

4.2 *Classification of emotions of tweets*

Although the percentage of classification is similar for both methods (Table 1), their consistency still shows differences. Positive and neutral categories have higher consistency than the negative category which only has the half consistency. The classification of English tweets presents the highest consistency which reaches to 78.75%. The following is the Spanish tweets which consistency reaches 73.15%. The worse one is Catalan tweets.

ISSN e: 2604-6512

Table 1**. Cross-validation of the consistency of two sentiment algorithms**

	English Tweets	Percentage	Spanish Tweets	Percentage	Catalan Tweets	Percentage
Svader_pos	64,429	31.28%	94,470	42.31%	34,766	18.49%
Senti_pos	64,963	31.54%	91,562	41.01%	38,584	20.52%
Both_pos	47,307	22.96%	68,356	30.62%	22,969	12.22%
Svader_neg	14,406	6.99%	21,218	9.50%	7,856	4.18%
Senti_neg	12,245	5.94%	18,604	8.33%	7,605	4.04%
Both_neg	6,756	3.28%	11,039	4.94%	4,186	2.23%
Svader_neu	127,162	61.73%	107,586	48.19%	114,335	60.81%
Senti_neu	128,789	62.52%	113,108	50.66%	110,768	58.91%
Both_neu	108,160	52.51%	83,923	37.59%	97,605	51.91%
Total consistency tweets	162,223	78.75%	163,318	73.15%	124,760	66.35%
Total Tweets	205,997	100.00%	223,274	100.00%	188,027	100.00%

Note: Senti is Senti-strength. Pos, neg, neu means positive, negative, and neutral respectively.
Source: self-elaboration

There are several automatic methods to evaluate the quality of translation, however, they still require a human translation for reference. Human manual evaluation is the most reliable method but extremely expensive and time consumption. What's more, unlike the edited texts, the original texts of tweets actually contain many wrong spelling or grammatical errors. Such natural defect brings difficulty to adopt automatic methods of evaluation. Therefore, considering the economic costs and the aim of the research, we measure the impact of translation. Firstly, under 95% confidence level and 4.6 confidence interval, 453 tweet samples were extracted from our dataset of Spanish and Catalan tweets. Secondly, a native Spanish and Catalan speaker with advanced English level was invited to classify the original Spanish tweets into positive, negative, and neutral. The advertising tweets are considering as neutral emotion. Finally, we compared the two results and analyzed their differences. Those tweets that have different classifications of emotion would be further investigated.

Table 2. **Result of sentiment evaluation of Spanish**

Spanish	Number of samples	Positive tweets	Negative tweets	Neutral tweets
Manual classification *(original Spanish tweets)*	453	204(45.25%)	17(3.75%)	230(50.99%)
Algorithm classification *(translated English tweets)*	453	187(41.28%)	38(8.38%)	228(50.33%)
Consistency of both methods		127	8	166

Source: self-elaboration

In general, the neutral tweets accounts for the largest portion of all samples (Table 2), the following is the positive group. The consistency rate of the positive samples is not very high, only half of samples. Negative tweets are very few; and only eight of them are confirmed by both classifications. The number of the difference of each emotional category is listed in Table 3.

ISSN e: 2604-6512

Table 3. **Number of difference between manual evaluation and machine evaluation**

	Positive(A)	Negative(A)	Neutral(A)
Positive (M)	127	18	58
Negative (M)	5	8	4
Neutral (M)	55	12	166

Note: M: manual, A: Algorithm.
Source: self-elaboration

When we investigated the concrete text of these tweets, we concluded three reasons which led to the unmatched classification (Table 4). Regard to the problem of translation, as we mentioned before, it is a complex issue. Therefore, the error of translation is specifically defined as the wrong translation of word-to-word and untranslated words. In the group of Neutral (A)-Positive(S), nearly 60% of these tweets belong to commercial advertisements which only can be detected by manual examination. The understanding of the contents leads to the algorithm cannot recognize positive/ negative emotion very well, which includes idioms, lack of emotional words, metaphor, satire and etc.

Table 4. **Reasons of different evaluation**

Difference	Reason of differences
Neutral(A)-Positive (S)	Commercial advertisement: 58.18% Problem of understanding the meaning of text: 34.54% Word-translation error:5.45%
Neutral(S) –Positive(A)	Problem of understanding the meaning of text: 72.41% Word-translation error: 27.58
Positive(M) – Negative(A)	Word-translation error: 16.67% Problem of understanding the meaning of text: 83.33%
Positive(A) – Negative(M)	Problem of understanding the meaning of text: 40% Word-translation error: 60%
Negative(A) – Neutral (S)	Word-translation error: 25% Problem of understanding the meaning of text: 75%
Negative(S) – Neutral (A)	Problem of understanding the meaning of text: 100%

Note: M: manual, A: Algorithm.
Source: self-elaboration

It is necessary to emphasize that the aim is to measure whether the translation greatly affects emotion detection, rather than to measure the quality of translation. In brief, word-translation does not greatly affect the sentiment analysis. In fact, understanding the meaning of texts is the core issue of sentiment analysis, no matter manual or machine detection. It is undeniable that both methods of classification exist inherent bias. However, such an issue is beyond the discussion of our research.

With regard to the result of manual evaluation of Catalan translation (Table 5), the major problem of Catalan tweets is that over 50% of all samples do not contain any useful information, except the location information. For example, many messages are: finished to upload a picture at a place. Therefore, the percentage of neutral tweets is very high, reaches 72.62% of all samples. Therefore, the Catalan tweets only involved in the analysis of word-frequency and spatial density. The further sentiment analysis only includes Spanish and English.

ISSN e: 2604-6512

Table 5. **Result of sentiment evaluation of Catalan**

Catalan	Number of samples	Positive tweets	Negative tweets	Neutral tweets
Manual classification *(original Spanish tweets)*	453	96(21.12%)	28(6.18%)	329(72.62%)
Algorithm classification *(translated English tweets)*	453	74(16.33%)	7(1.5%)	372(82.11%)
Consistency of both methods		41	5	301

Source: self-elaboration

4.3 Word-Frequency analysis

In terms of high-frequency words (Table 6), English and Spanish appear more emotional words than Catalan tweets, such as good, love, beautiful, etc. Catalan high-frequency words contain more names of places of Barcelona, such as 'sant','cugat', 'plaça' (square in English), which are accord with the finding of previous sample test. English tweets contained more words which are related to tourism, such as photo, drinking, hotel, etc. Spanish tweets seem to be in between.

Table 6. **The top forty high-frequency words**

English		**Spanish(translated)**		**Catalan(translated)**	
Word	Frequency	Word	Frequency	Word	Frequency
barcelona	7599	barcelona	6066	barcelona	15977
posted	1611	spain	1521	catalonia	4909
photo	1606	day	1426	casa	4614
gothic	1494	today	1277	plaza	3245
quarter	1117	good	1240	batllo	2919
spain	1032	new	1035	sant	2821
day	738	plaza	918	gaudi	2266
one	711	catalunya	875	square	1853
love	661	beach	837	spain	1626
new	633	barceloneta	810	photo	1625
city	589	photo	783	rambla	1620
ES	553	love	780	pedrera	1594
time	529	one	780	mila	1495
night	529	happy	695	ciutadella	1378
good	526	time	674	park	1343
like	519	see	672	gracia	1279
last	509	like	619	cugat	1160
city	475	best	618	vallès	1091
drinking	470	great	583	posted	1046
beautiful	441	rambla	556	market	1006
happy	437	city	549	paseo	969
best	430	want	548	today	918
barceloneta	417	life	533	museum	909

ISSN e: 2604-6512

beach	415	boqueria	531	art	902
back	412	night	514	boqueria	901
mercat	400	always	490	poble	821
boqueria	397	thanks	482	plaça	817
playa	393	go	474	day	801
hotel	391	ramblas	471	macba	794
barcelona.	370	apolo	461	passeig	747
great	368	bcn	458	cataluña	702
amazing	365	sala	456	espanyol	688
rambla	364	know	441	gràcia	679
catalunya	357	us	441	contemporary	677
see	352	casa	439	new	669
running	349	year	437	jaume	622
finished	333	morning	435	barceloneta	622
get	332	club	426	published	585

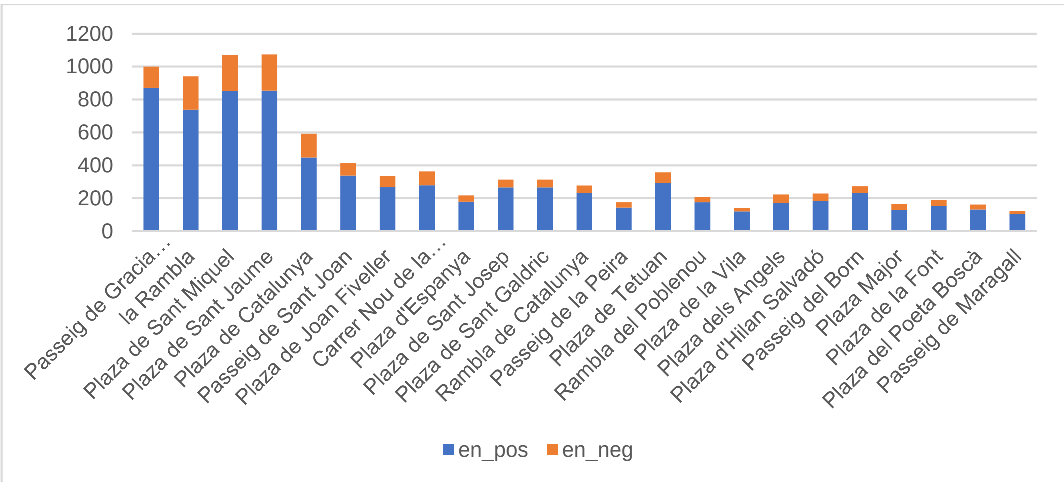
Source: self-elaboration

4.4 Spatial distribution of mass emotion

Figure 5 displays the distribution of positive and negative sentiments of Spanish and English tweets in the public spaces separately. The negative tweets only account for a small portion of all these places. Positive emotion takes the dominant role. In general, negative tweets decrease as the total number of tweets decrease.

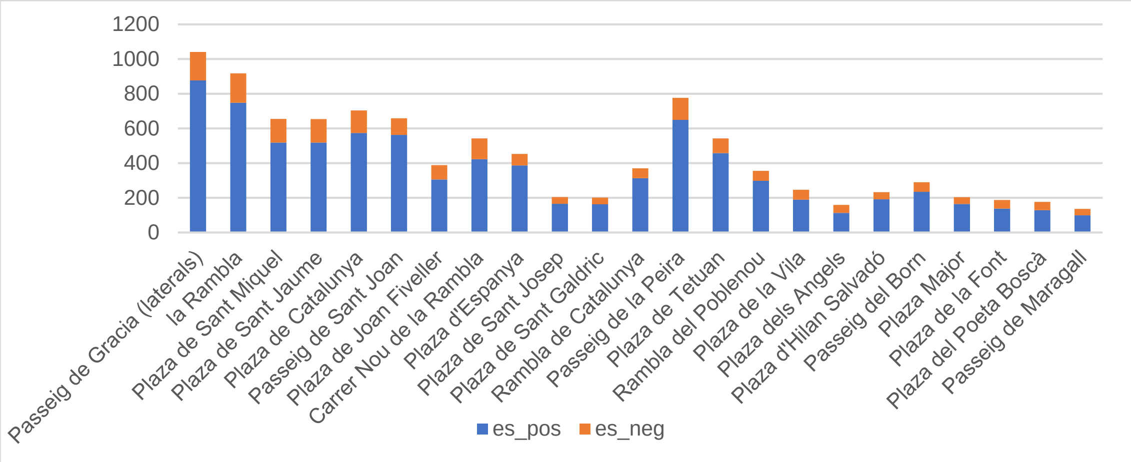
However, the proportion of English negative tweets is higher in Plaza de Sant Miquel and Plaza de Sant Jaume. One possible explanation is that the government of Catalonia and Barcelona are located in these two places, thus there are more political events that happened and could evoke negative emotion in these two places.

Figure 5. **Distribution of positive and negative emotion**



(a) English

ISSN e: 2604-6512



(b) Spanish

Source: self-elaboration

5. Conclusion and Discussion

The study compares the spatial distribution differences between English, Spanish and Catalan tweets in Barcelona. Further, it reveals the sentiment differences between the tweets of the three language in terms of the word-frequency and the proportion of different emotion categories. In general, most of tweets either are positive or neutral. Negative tweets only account for less than 5%. However, the Spanish tweets has higher percentage of positive tweets than English tweets. Moreover, the result confirms the differences of high frequency words between the three languages, though they have some words in common. Spanish and Catalan tweets present more words which are related with local people.

Acknowledgements: Is very appreciated all the valuable and vital advice given by the supervisor, Professor Carlos Marmolejo, during the process of the research. Meanwhile, the dataset provided by Professor Martí Pablo.

Author's contributions: The first author has been in charge of data collection, analysis, and writing paper; the second author of the guidance and revision of the paper, and the third author providing the *Twitterdata*, support the analysis of it, and guidance.

Conflict of Interest: The authors declare no conflict of interests.

Bibliografía

Brereton, F., Clinch, J. P., & Ferreira, S. (2008). Happiness, geography and the environment. *Ecological economics*, *65*(2), 386-396.

Chapman, L., Resch, B., Sadler, J., Zimmer, S., Roberts, H., & Petutschnig, A. (2018). Investigating the emotional responses of individuals to urban green space using twitter data: A critical comparison of three different methods of sentiment analysis. *Urban Planning*, *3*(1), 21-33.

ISSN e: 2604-6512

Collins, C., Hasan, S., & Ukkusuri, S. V. (2013). A novel transit rider satisfaction metric: Rider sentiments measured from online social media data. *Journal of Public Transportation*, *16*(2), 2.
Davidson, J., Smith, M. M., & Bondi, L. (Eds.). (2012). *Emotional geographies*. Ashgate Publishing, Ltd.

Do, H. J., Lim, C. G., Kim, Y. J., & Choi, H. J. (2016, January). Analyzing emotions in twitter during a crisis: A case study of the 2015 Middle East Respiratory Syndrome outbreak in Korea. In *2016 international conference on big data and smart computing (BigComp)* (pp. 415-418). IEEE.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion review*, *3*(4), 364-370.

Gallegos, L., Lerman, K., Huang, A., & Garcia, D. (2016, April). Geography of Emotion: Where in a City are People Happier? In *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 569-574).

Kovacs-Gyori, A., Ristea, A., Havas, C., Resch, B., & Cabrera-Barona, P. (2018). # London2012: Towards Citizen-Contributed Urban Planning Through Sentiment Analysis of Twitter Data. *Urban Planning*, *3*(1), 75-99.

Kuipers, B. (1978). Modeling spatial knowledge. *Cognitive science*, *2*(2), 129-153.

LeFebvre, R. K., & Armstrong, C. (2018). Grievance-based social movement mobilization in the# Ferguson Twitter storm. *New Media & Society*, *20*(1), 8-28.

Li, M., Ch'ng, E., Chong, A., & See, S. (2016, July). The new eye of smart city: novel citizen sentiment analysis in twitter. In *2016 International Conference on Audio, Language and Image Processing (ICALIP)* (pp. 557-562). IEEE.

López-Ornelas, E., & Zaragoza, N. M. (2015, August). Social media participation: A narrative way to help urban planners. In *International Conference on Social Computing and Social Media* (pp. 48-54). Springer, Cham.

Lupton, D. (1998). *The emotional self: A sociocultural exploration*. Sage.
Lynch, K. (1960). *The image of the city* (Vol. 11). MIT press.

Matei, S., Ball-Rokeach, S. J., & Qiu, J. L. (2001). Fear and misperception of Los Angeles urban space: A spatial-statistical study of communication-shaped mental maps. *Communication Research*, *28*(4), 429-463.

Molz, J. G. (2005). Guilty pleasures of the Golden Arches: mapping McDonald's in narratives of round-the-world travel. *Emotional geographies*, 63-76.
Padilla, J. J., Kavak, H., Lynch, C. J., Gore, R. J., & Diallo, S. Y. (2018). Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. *PloS one*, *13*(6).

Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016, October). Sentiment analysis of Twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system (SCOPES)* (pp. 1345-1350). IEEE.

ISSN e: 2604-6512

Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, No. 2010, pp. 1320-1326).

Quercia, D., Ellis, J., Capra, L., & Crowcroft, J. (2012, February). Tracking" gross community happiness" from tweets. In *Proceedings of the ACM 2012 conference on computer supported cooperative work* (pp. 965-968).

Schwartz, A. J., Dodds, P. S., O'Neil-Dunne, J. P., Danforth, C. M., & Ricketts, T. H. (2018). Exposure to urban parks improves affect and reduces negativity on Twitter. *arXiv preprint arXiv:1807.07982*.

Shaver, P., Schwartz, J., Kirson, D., & O'connor, C. (1987). Emotion knowledge: further exploration of a prototype approach. *Journal of personality and social psychology*, *52*(6), 1061.

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *Journal of cognitive neuroscience*, *22*(12), 2864-2885.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012, July). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 system demonstrations* (pp. 115-120). Association for Computational Linguistics.

Wang, W., Hernandez, I., Newman, D. A., He, J., & Bian, J. (2016). Twitter analysis: Studying US weekly trends in work stress and emotion. *Applied Psychology*, *65*(2), 355-378.

ISSN e: 2604-6512