International Journal of Hospitality & Tourism Systems
Volume 15 Issue 1 June 2022
ISSN: 0974-6250 (Print)

Hospitality & Tourism Systems

How to Employ Zipf's Laws for Content Analysis in Tourism Studies?

Lucília Cardoso*, Noelia Araújo Vila, Mohammad Soliman***, Arthur Filipe Araújo****, Giovana Goretti Feijó de Almeida*****

Abstract *Although Zipf's laws were deployed in various contexts, no previous research has investigated their adoption in tourism academic publications. This study aims to fill this gap by employing Zipf's laws as a content analysis method for analyzing tourism literature, with emphasis on the gamification subject. Employing the DB Gnosis software, the findings revealed that the application of Zipf's laws on obtained tourism gamification articles not only broadens the academic understanding of these laws but also brings about relevant implications for academic literature in tourism. This research provides a new content analysis method in tourism and presents new insights into existing content analysis approaches in tourism. It also adds to the few studies that have addressed the topic of tourism gamification.*

Keywords: *Zipf's First Law, Zipf's Second Law, Content Analysis, Gamification, Tourism*

INTRODUCTION

Zipf proposes a set of empirical probability distribution power laws based on observation of text documents, which arguably apply to any given text on any given subject. Zipf's first law is inherently simple, to the point that it is difficult to explain how it can predict something as complex as human communication structures. The law simply states that the number of times a word appears in a text (its frequency) multiplied by its ranking in a list of most frequent words will be approximately a constant. Therefore, the second most frequent word should appear half as often as the most frequent; the third should appear a third as often, and so on (Zhu et al., 2018). As mentioned by Powers (1998), Zipf's law has been reappearing in many not entirely expected places.

Zipf's law was widely applied to all human languages (Montemurro & Zanette, 2002) and implemented in many contexts including statistics for firm sizes, city sizes, Internet file sizes, human behavior, family names, income, financial markets, the genome, etc. (Newman, 2005; Thurner et al., 2015). However, studies employing such methods to analyze scientific literature are extremely limited. Consequently, this study investigates whether the properties of Zipf's laws can be employed to detect key concepts of a certain topic within a complex set of qualitative information sources.

Further, it is argued that the increase of discipline literature determines its development. As a result, content analysis

* Centre for Tourism Research, Development and Innovation (CITUR), Portugal. Email: lucilia.a.cardoso@ipleiria.pt (Corresponding Author)
** Faculty of Business Sciences and Tourism, University of Vigo, Spain. Email: naraujo@uvigo.es
*** University of Technology and Applied Sciences-Salalah, Oman. Faculty of Tourism & Hotels, Fayoum University, Egypt. Email: msoliman.sal@cas.edu.om
**** TRIE - Transdisciplinary Research Center for Innovation and Entrepreneurship Ecosystems - Lusofona University, Porto, Portugal. Email: arthurfilipearaujo@gmail.com
***** Centre for Tourism Research, Development and Innovation (CITUR), Portugal. Email: goretti.giovana@gmail.com

has gained popularity and becomes a crucial approach to present a clear view of literature in different contexts and to follow-up with its evolution. In this regard, there is considerable growth in the number of studies that conducted content analysis within the tourism context in recent years (Domingo-Carrillo et al., 2019; Griffin, 2013). However, to the best of our knowledge, there is no study that has conducted a content analysis for gamification literature in tourism in general, applying Zipf's laws in particular. Moreover, numerous word frequency counting softwares and techniques were employed to identify the key themes of publications and to assess the performance of journals, authors, institutions, etc. (Tahamtan & Bornmann, 2019). It is evident that the starting point of these studies is the search for keywords in the online databases; however, the contents of the documents do not always match the keywords. At this end, Zipf's Laws are considered one of the crucial tools that can be used to solve such problems.

Considering these gaps, the following questions are developed: (1) How to apply Zipf's first law for content analysis in gamification articles published in tourism?; (2) How to apply Zipf's second law for content analysis of gamification articles published in tourism?; (3) Do the descriptive words in tourism gamification literature concentrate around the point of Zipf's second law?; and (4) Is there any similarity between the descriptive words concentrated around the point of Zipf's second law and the authors keywords announced in each paper?

For answering the above questions, this study provides empirical evidence on the application of Zipf's laws as a content analysis tool in the tourism gamification literature. In this research, the academic literature on the use of gamification in the tourism industry is adopted as settings for a study on the application of Zipf's and Goffman's laws (Goffman, 1964; Zipf, 1949) to content analysis purposes. In other words, Zipf's laws were applied to facilitate scientific literature analysis.

Considering the recent relevance associated with the attribute of gamification in enhancing tourism consumers' experiences and fostering their commitment toward the destination, the concept was selected to be analyzed in the present research. Employing Zipf's laws to analyze tourism literature, including gamification, can be an effective mechanism to facilitate the procedures of systematic review. In this context, a practical outcome of the present study could be an innovative process of systematic literature review, the "Zipf Analysis". Moreover, such a process will potentially be significantly useful in the context of papers' indexation and to generally facilitate scientific research in tourism by improving the identification of relevant documents for a literature review.

LITERATURE REVIEW

Zipf's Laws

According to Zipf's first law, if one ranks the words of a text in decreasing order, the ranking position of any given word multiplied by its frequency will be approximately a constant. This is expressed by the formula "R x F = K" (Zipf, 1949: 24), where "R" is the ranking position of a word, "F" its frequency, and "K" the constant (Bochkarev & Lerner, 2017; Huang et al., 2008; Tyagi et al., 2017). In other words, the frequency of any word is inversely proportional to its rank (Koplening, 2015; Newman, 2005). This means that the second most common word in a text appears half as often as the first, the third most common appears one third as often, and the fourth appears one fourth as often, and so on. In the English language, for instance, the most common word in any given text is most likely "the", followed by "of" (Bochkarev & Lerner, 2017). Together, these two words usually represent 10% of a text. Therefore, in a text in which "the" appears 270 times, for example, setting the constant at 270 (270*1), "of" is expected to appear 135 times (135*2 = 270), and the third most frequent word must appear around 90 times (90*3 = 270).

Zipf's first law applies to texts in any language, even extinct ones with very little in common with modern languages (Smith, 2007). The law even applies to realms beyond language, such as city population ranks, solar flare intensity, website traffic, etc. Amongst the least expected applications of Zipf's law, Odueke and Wier (2012) have employed it in the context of forensic accounting, while Ulubasogulu and Hazari (2004) used it to predict tourist arrivals. A probable reason why such law applies to texts and other instances is that it is in line with Pareto's principle, according to which 80% of the effects come from around 20% of the causes (the vital few). Zipf's first law uses this pattern, as it suggests that 18% of the words amount to 82% of the text.

Another probable explanation is the law of the 'least effort', according to which human beings will invariably follow the easiest or less risky path presented to them. In other words, as explained by Zhu et al. (2018: 1), "Each individual will adopt a course of action that will involve the expenditure of the probably least average of his work". This statement summarizes the 'least effort' law, which is often referred to a deterministic description of human behavior, and to which Zipf himself subscribed. The principle applies to Zipf's law as in the English language; frequently used words tend to be less than five letters long and are typically monosyllables. This is the result of a linguistic evolution process driven by speakers' desire to spend as little time and effort as possible while still conveying their message to the listener or reader.

There is also a third theory, according to which Zipf's law is related to preferential attachment processes, which occur when something (typically some wealth or credit) is distributed among individuals according to how much they already have. Such a process explains, for instance, how investors who have more money have access to better investment opportunities, and thus earn higher interests; or how bigger companies manage to expand further and reach greater profits through scale economies. In the context of languages, as they are developed, short words are the preferred choice, possibly due to the law of the least effort, and therefore, they become increasingly popular over time.

Still according to Zipf, words in a text can be divided into high-frequency words, low-frequency words and words whose frequency tend to one. In this context, Zipf himself concluded that his first law does not apply to low-frequency words, and consequently, proposed a second equation to deal with this anomaly (Zipf, 1949). The equation was refined by Goffman (1975) and gave origin to Zipf's second law, according to which, in any given text, many low-frequency words will have the same frequency (1).

As explained in more detail by Molinos, Mesquita and Hoff (2016), any text presents three categories of words: (1) high-frequency words, also called "stopwords", which are operational words, such as articles, pronouns, conjunctions, prepositions, and some adjectives and adverbs; (2) average-frequency words, which carry more morphological and informative representations than those in the first zone, such as substantives, adjectives and verbs; and (3) unit-frequency words, including terms that happen in very specific contexts, and therefore, have frequencies of one or close to one.

High- and low-frequency words represent the extreme points in a text's frequency distribution list. In this regard, there is also a frequency transition zone (for high-frequency to low-frequency words), as well as a critical point, or Goffman's T point, which is the exact ranking position that separates high- and low-frequency words. High semantic content words (e.g., indexation terms, descriptors and keywords) are concentrated around the critical point (Tyagi et al., 2017). The properties of texts described by Zipf's law make it a useful tool that can be applied in many different contexts, as more minutely addressed in the following section.

Application of Zipf's Laws

Several authors (e.g., Moran & Bouchaud, 2018; Zhu et al., 2018) argue that Zipf's empirical observations not only follow the principles of power laws, but derive from the very laws of nature, and therefore are applicable to many contexts involving human behavior. According to Huang et al. (2008), Zipf's laws indicate the relationships between words' rank and frequency, so they can be employed to handle disaggregated account-level data. Moreover, they have the advantage of not requiring any pre-defined type of attributes to be employed. Thus, the law can potentially even be used as a tool for fraud detection.

Zipf's Laws have also been adopted in the area of Information Retrieval (IR), which typically involves problems inherent to the collection process for a corpus of documents. Such problems are frequently solved by providing functionalities that allow users to find a particular subset by making queries. The idea of IR implementation revolves around an attempt to systematically extract information from available document corpora, and then utilize it to determine whether each document is relevant to a particular request. IR is often facilitated by content analysis (Krippendorff, 2004), a quantitative tool that allows users to examine indicators in text documents.

According to Molinos, Mesquita and Hoff (2016), Zipf's laws are potentially powerful tools for content analysis to detected key-themes in articles. This study aims to verify that potential by investigating whether Zipf's power law can be used to facilitate scientific literature analysis. To this end, the principles were applied to literature on gamification in tourism.

Gamification in Tourism

It is evident that there is little accord associated with the definition of gamification; however, a widely accepted definition (Sigala, 2015a) was presented by Deterding, Dixon, Khaled, and Nacke (2011: 28), as follows: "the use of game-play mechanics (e.g. points, leaderboards, achievements/ badges, levels, story/theme, clear goals, feedback, and rewards) for non-game context". Deterding (2012) refers to the concept as the selective incorporation of game-play elements in an interactive system, without resulting in a complete game as a final product. Zichermann and Cunningham (2011) indicated that gamification might establish a brand, engage users and influence their behavior. Accordingly, Hamari (2013) defines gamification as the process of improving services with (motivational) affordances in order to invoke gameful experiences and further behavioral outcomes. According to Walz and Deterding (2015: 6), gamification is an instantiation of the interpretation of games in everyday life, the most visible and recent one. They argued that such penetration has a long history and is just a part of a wider trend referred to as "ludification of culture, ludification of society or the rise of a ludic society".

Regarding the tourism context, gamification is viewed as a marketing tool (Xu et al., 2016) and might be an effective way of inducing destination image (Xu et al., 2017). The concept is also employed in gamified tour guides (Luimul & Trygg, 2016; Weber, 2014), often employed to enhance tourists' experiences of cultural heritage. This use of gamification, especially through technologies such as virtual and augmented reality, has also attracted the interest of academia and the creative industries (Ioannides et al., 2017). This is something new, but there have already been other examples in the tourism sector that have benefited from new technologies to induce or facilitate consumption, such as apps of airlines (Qin et al., 2017) or hotels (Lei et al., 2019).

METHODS

Acquisition of Tourism Gamification Literature

To acquire literature on gamification in tourism, a database consisting of the most relevant studies on gamification in tourism was developed in March 2019, through a search in Scopus and Web of Science databases as in previous studies involving systematic reviews (de Mudarra-Fernández et al., 2019; Domingo-Carrillo et al., 2019). The Scopus and Web of Science databases were selected because they are regarded as the greatest databases of peer-reviewed publications in several fields.

The search was for articles containing "Gamification + Tourism" in the titles, abstracts and keywords, from 2010 to March 2019. The search on both databases resulted in 54 documents. This research focused on published articles only. Therefore, conference proceeding articles, books, book chapters and reviews were discarded. The final database consisted of 21 papers, which are described in Table 1.

Table 1: Database Description

Article	Authors/Year	Words	Different Words	W/D	I1	I1%
1	Xu, Tian, Buhalis, Weber, and Zhang (2016)	7484	1463	5.12	769	52.56
2	Moro, Ramos, Esmerado, and Jalali (2019)	4117	1018	4.04	558	54.81
3	Tan (2018)	6549	1260	5.20	676	53.65
4	Prasetyo and Suyoto (2018)	3154	790	3.99	456	57.72
5	Kolarand Čater (2018)	8217	1714	4.79	917	53.50
6	Skinner, Sarpong, and White (2018)	5906	1411	4.19	778	55.14
7	Goshevski, Velianoska, and Hatziapostolou (2017)	3594	1051	3.42	588	55.95
8	Huang, Chiao and Huang (2018)	3985	1017	3.92	562	55.26
9	Widawska-Stanisz (2018)	3062	887	3.45	506	57.05
10	Swacha and Ittermann (2017)	2277	709	3.21	412	58.11
11	Yoo, Kwon, Na, and Chang (2017)	7629	1481	5.15	784	52,94
12	Adukaite, Van Zyl, Er, and Cantoni (2017)	7845	1625	4.83	834	51.32
13	Xu, Buhalis, and Weber (2017)	5830	1284	4.54	637	49.61
14	Liang, Schuckert, Law, and Chen (2017)	7055	1272	5.55	623	48.98
15	Garcia, Linaza, Gutierrez, and Garcia (2019)	7770	1393	5.57	731	52.50
16	Thimmand Seepold (2016)	5053	1335	3.79	784	58.73
17	Mesáro et al. (2016)	2981	818	3.64	489	59.78
18	Sigala (2015a)	8319	1444	5.76	720	49.86
19	Negruşa, Toader, Sofică, Tutunea, and Rus (2015)	8537	1841	4.64	969	52.63
20	Sigala (2015b)	5928	1333	4.45	698	52.36
21	Cirulis, Paolisand Tutberidze (2015)	2927	983	2.98	609	61.95
å	-	117859	26129	4.52	14100	-
Mín	-	2277	709	2.98	412	48.98
Máx	-	8537	1841	5.76	969	61.95
Average	-	5629.47	1244.2	4.4	671.4	54.50

Processing of Data

The obtained articles were processed using the Destination Brand Gnosis (DB Gnosis) software to text analysis, this software is free available via http://favouritedestinations.com/en/dbgnosis/. The choice of this software was due to the big volume of text to be analyzed. Unlike other alternatives, DB Gnosis software allows for fast processing of such big volumes, which made it the most adequate tool for the analysis. In doing so, the 21 articles were prepared for the word frequency count. Firstly, the papers were converted from the PDF format to .txt. with UTF8 codification, a requirement of DB Gnosis. With the papers in .txt files, and therefore editable, the next step consisted of the actual preparation of their texts for the word count and application of Zipf's laws. It was considered that Zipf's work was initially applied to the analysis of books, which follow a certain writing pattern (Flores, 2016). To guarantee the efficiency of the application, titles, abstracts, key words, tables, figures and references were excluded from the word count of the obtained documents.

The following phase of the investigation consisted of the word frequency count within the retrieved scientific papers on gamification in the tourism industry. In this context, words were adopted as the registry unit for the content analysis, and the frequency was the enumeration rule for ranking. The adoption of frequency as the enumeration rule considers that the registry unit's (the word's) relevance increases with its frequency of appearance (Bardin, 2000), which makes it ideal for testing the application of Zip's laws. The .txt files were loaded to DB Gnosis software so that frequency queries could be carried out with their words. The analysis outputs were then converted to Excel to allow for the elaboration of the final database. First, word counts were carried out for each document, which resulted in a total of 117859 words, with an average of 5629.47 words per document. Within the 21 papers, 26129 different words were registered, from which 14100 had a frequency of 1. Such data was used for the application of Zipf's laws.

Analysis Technique

Firstly, content analysis can be applied to identify and understand patterns (subjects) within the analyzed texts (Braun & Clarke, 2006). Content analysis is an objective procedure to describe and quantify phenomena through the analysis of texts (Schreier, 2012), which adequately applies to the present study's objectives. The specific type of content analysis employed in the present work was categorical content analysis, which consists of dismembering the texts into units, or categories, according to pre-established criteria (Bardin, 2000).

To conduct a content analysis, Zipf's laws were employed to identify the relevance of specific themes, in order to verify their applicability to this purpose. As described in Fig. 1, Zip's Laws application was carried out in three phases: (1) application of Zipf's first law; (2) application of Zipf's second law; (3) verification of the theme's relevance through Zipf's laws. It should be considered that we applied the aforementioned steps of Zipf's law on the first paper as an example that was systematically selected. Finally, an analysis was conducted to verify whether the most frequent semantically significant words in a text corresponded to those that expressed the study's main objectives and contributions. In this context, semantically significant words refer to those that express meaning on their own, as opposed to those of purely syntactical value, such as prepositions and definite articles. Therefore, for the purposes of this final analysis, articles and prepositions, as well as formal elements, such as references, were excluded from the word count. Fig. 1 summarized the steps of the research design.

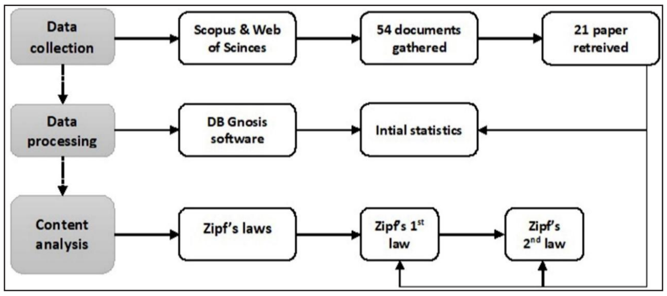

Fig. 1: The Research Design of the Current Study

RESULTS

Application of Zipf's First Law

To apply Zipf's first law, and to answer the first research question, the following formula was employed:

$$r \cdot f = c$$

where f = frequency of occurrence; r = word's ranking position; and c = Zipf's constant. The latter is supposedly the same within all the words in a given text and varies from text to text depending on the format and language. As addressed in the literature overview, the formula is explained by assuming that the frequency of any word in a text is inversely proportional to its rank, which is explained by the following function:

$$\text{f (r)} \; \alpha \, r^{-\alpha} \Leftrightarrow \text{f (r)} \; \alpha \, r^{-1}$$

Considering this assumption, a Zipf plot was performed on each paper. This aimed to verify whether its words' frequencies followed a Zipfian distribution (i.e., if they were inversely proportional to their rankings). Fig. 2, which displays the Zipf plot of the first paper, shows that the distribution of the papers' words frequencies indeed follows a Zipfian distribution, that is, each word's frequency is inversely proportional to its ranking. Therefore, the figure shows a parabolic distribution, which is the graphical representation of a Zipfian word frequency configuration.

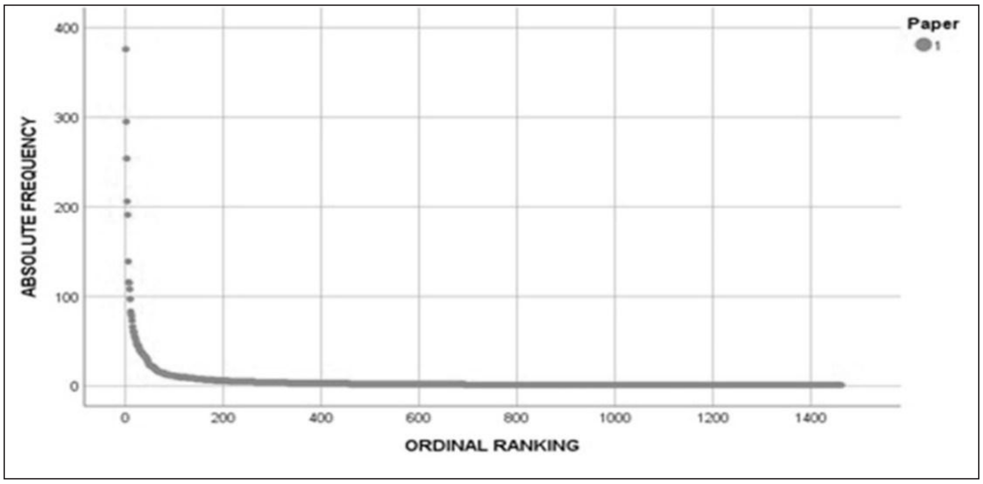

Fig. 2: Zipf Plot of Paper 1

In face of this result, the next phase of the investigation was based on a specific property of Zipfian distribution, according to which the distribution becomes a linear function when taking the logarithm of both sides. Against this backdrop, following the example of previous studies on the subject (Odueke & Weir, 2012; Powers, 1998), the next step consisted of transforming the first paper's word frequencies and ranking in logarithms and then applying the formula $r \cdot f = c$. (Zipf, 1949: 24) to find Zipf's constant for the document.

Table 2 shows the top and bottom 20 words for the first paper, that is, the 20 words with highest frequencies and the 20 with lowest (all equal to 1). The table also shows the Zipf C value for each word, that is, Zipf's constant, as per the mentioned formula, and the results show that this value is indeed relatively constant. Moreover, the logarithm values of absolute frequencies and ranking positions are also shown.

In order to analyze the relationship between variables, the representative line for the data presented in Table 3 was calculated by employing the minimum squares statistic technique, through which a straight line is obtained by minimizing the sum of the squared differences between the ordinates of the data and the values predicted by the equation of the line (Jovell, 2006). According to Zhu et al. (2018), when the rank positions and frequencies are plotted on a double logarithmic chart, it can be argued that the frequencies follow a Zipfian distribution if the points fall on a straight line from the left to the right, at an angle of 45º. The application of the described method to all points (X*, Y*) resulted in the blue line shown in Fig. 3, to which the points align between 1 and 3, as predicted by X*, the position logarithm. Therefore, considering Zhu et al.'s (2018) description of the graphic representation of a Zipfian distribution in a double logarithmic chart, this result suggests that the distribution of words' frequencies on the first paper indeed follows the principles of Zipf's first law.

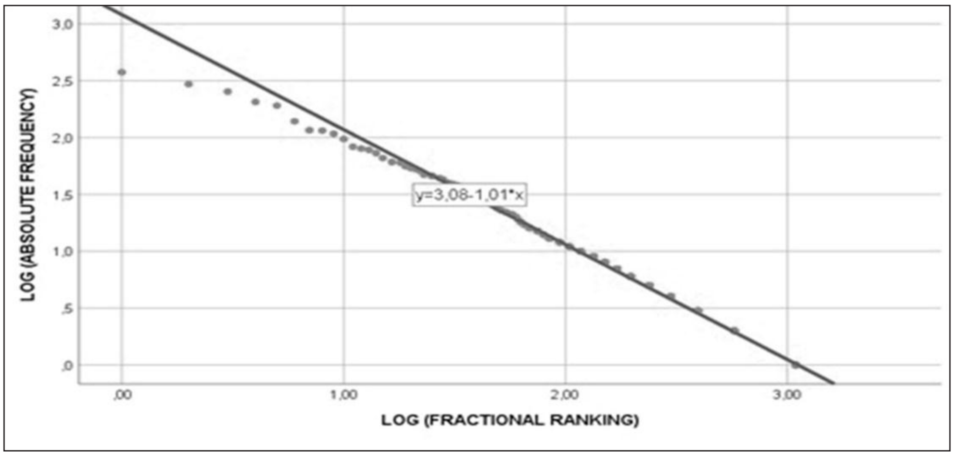

Fig. 3: Zipf Plot for Paper 1 (log rank - log absolute frequency plot of all tokens)

Table 2: First Zipf's Law Analysis of Paper 1 (1.848 Words)

TOP 20 Ranked words | Bottom 20 ranked words

Ranking (X)	Words	Absolute Frequency (Y)	Zipf C	LOG(f) (Y*)	LOG (Ranq) (X*)	Ranking (X)	Words	Absolute Frequency (Y)	Zipf C	LOG(f) (Y*)	LOG (Ranq) (X*)
1	The	376	376	2.575188	0	64	Voluntary	1	64	0	1.80618
2	And	295	590	2.469822	0.30103	64	Volunteered	1	64	0	1.80618
3	To	254	762	2.404834	0.477121	64	We	1	64	0	1.80618
4	Of	206	824	2.313867	0.60206	64	Week	1	64	0	1.80618
5	A	191	955	2.281033	0.69897	64	Weeks	1	64	0	1.80618
6	In	139	834	2.143015	0.778151	64	Welldesigned	1	64	0	1.80618
7	Game	116	812	2.064458	0.845098	64	Whose	1	64	0	1.80618
8	Games	115	920	2.060698	0.90309	64	Why	1	64	0	1.80618
9	Is	108	972	2.033424	0.954243	64	Widely	1	64	0	1.80618
10	As	97	970	1.986772	1	64	Without	1	64	0	1.80618
11	That	83	913	1.919078	1.041393	64	Wonder	1	64	0	1.80618
12	Are	80	960	1.90309	1.079181	64	Wonderful	1	64	0	1.80618
13	For	78	1014	1.892095	1.113943	64	Word-of-mouth	1	64	0	1.80618
14	Players	73	1022	1.863323	1.146128	64	Works	1	64	0	1.80618
15	Destination	66	990	1.819544	1.176091	64	World's	1	64	0	1.80618
16	Tourism	61	976	1.78533	1.20412	64	Worthwhile	1	64	0	1.80618
16	With	61	976	1.78533	1.20412	64	youtube.com	1	64	0	1.80618
17	Gaming	60	1020	1.778151	1.230449	64	Year-old	1	64	0	1.80618
18	Can	56	1008	1.748188	1.255273	64	Yes	1	64	0	1.80618
19	On	54	1026	1.732394	1.278754	64	Younger	1	64	0	1.80618
20	This	53	1060	1.724276	1.30103	64	YouTube	1	64	0	1.80618

Following the methods employed by Odueke and Weir (2012) and Cleophas and Zwinderman (2018), from a simple linear regression (Y = a + bx), the equation "*Y* = 3.08 – 1.01 X**" (with data turned into logarithms) was obtained. Within this equation, -1.01 is the regression coefficient (b) that is, for each unit increased by variable Y (absolute frequency), variable X (average of the ranking positions) decreases 1.01 units, as also shown in Fig. 3. In addition, when solving the equation, as Y* = log (Y) and X* = log (X), by substituting these values, the following equations are obtained:

$$\log(Y) = \log(10^{3.08}) - 1.01 \log(X) \Leftrightarrow \log(Y)$$

$$\log(Y) = \log(1202.26) + \log(X^{-1.01}) \Leftrightarrow \log(Y) = \log(1202.26\, X^{-1.01})$$

$$Y = f(X) = 1202.26\, X^{-1.01}$$

This reinforces that the relationship between Y (absolute frequencies) and X (words' ranking positions) follow the principles of Zipf's law, whose application to paper no. 1 is therefore confirmed.

Next, the same procedure was applied to the other 20 papers. As shown in Fig. 4, the points of all papers align with the regression line between 0 and 3, as predicted by X*, the position logarithm, which suggests that the word distribution in all papers indeed follows the principles of Zipf's first law. Moreover, the simple regression line shown in Fig. 4 is translated by the following equation:

Y* = 2.69 – 0.89 X* (with data turned into logarithms)

By solving said equation, and considering that Y* = log (Y) and X* = log (X), the following equations are obtained:

$$\log(Y) = \log(10^{2.69}) - 0.89 \log(X)$$

$$\log(Y) = \log(10^{2.69}) - 0.89 \log(X) \Leftrightarrow$$

$$\log(Y) = \log(489{,}778) + \log(X^{-0.891}) \Leftrightarrow \log(Y) = \log(489{,}778\, X^{-0.89})$$

$$Y = f(X) = 489{,}778\, X^{-0.89}$$

This reinforces that the relationship between Y (absolute frequencies) and X words' ranking positions) for the 21 papers follows the principles of Zipf's first law, whose application to all 21 papers is, therefore, confirmed.

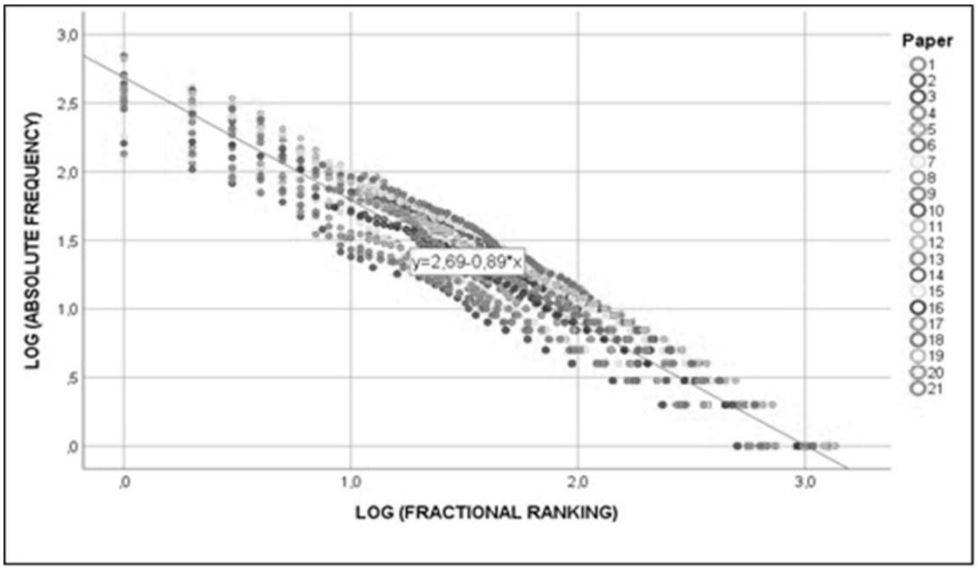

Fig. 4: Zipf Plot for the 21 Papers (log rank - log absolute frequency of all tokens)

Application of Zipf's Second Law

In order to answer the second research question and to apply the Zipf's second law, enounced by Goffman (1975) and becoming known as Goffman's T point (Zhu et al., 2018), and attempt to find Goffman's T point, that is, the point in which the distribution of frequencies changes its behavior, first, the following equation was considered:

$$\frac{I1}{In} = \frac{n(n-1)}{2}$$

Where,

*I*1: number of words with an absolute frequency equal to 1.

In: number of words with an absolute frequency equal to n.

2: Zipf's constant for texts in English.

Next, considering Zipf's second law, as in the study carried out by Tyagi et al. (2017), $InIn$ was substituted by 1. Therefore, only the positive root was considered, which resulted in the following equation:

$$n = \frac{-1 + \sqrt{1 + 8.I1}}{2}$$

By applying this procedure to each of the papers, which had already been done as part of the analysis described in the previous sections and presented in Table 2, the transition point (Goffman's T point), for each paper is obtained. As shown in Table 3, for paper 1, for example, the transition point is located among words with an absolute frequency of 38.20.

Table 3: Analysis of the Second Law of Zipf

Paper	Goffman's T point	N=words	Z1	%Z1	Z2	%Z2	Z3	%Z3
1	38.20	7484	3140	42	3575	48	769	10
2	31.90	4117	1441	35	2218	51	558	14
3	35.50	6549	3037	46	2836	43	676	10
4	28.10	3154	1170	37	1528	48	456	15
5	41.40	8217	3014	37	4286	52	917	11
6	37.60	5906	2207	37	2921	50	778	13
7	32.40	3594	1084	30	1922	54	588	16
8	31.90	3985	1235	31	2188	55	562	14
9	29.80	3062	948	31	1608	52	506	17
10	26.60	2277	626	28	1239	54	412	18
11	38.50	7629	2755	36	4090	54	784	10
12	40.30	7485	2855	36	4156	53	834	11
13	35.80	5830	1997	34	3196	55	637	11
14	35.70	7055	3007	43	3425	49	623	8
15	37.30	7770	3408	44	3631	46	731	10
16	36.50	5053	1713	34	2556	51	784	15
17	28.60	2981	934	31	1558	52	489	17
18	38.00	8319	3751	45	3848	46	720	9
19	42.90	8537	3014	35	4554	54	969	11
20	36.50	5928	2105	35	3125	53	698	12
21	31.40	2927	788	27	1530	52	609	21
å	-	117859	44229	-	59990	-	14100	-
Min	-	2277	788	27	1239	43	412	8
Máx	-	8537	3751	46	4554	54	969	21
Média	-	5357.22	2106	35.9	2857	51.04	671.4	13

In this context, Z1 is the zone before the transition point, which encompasses 3140 words; Z2 is the zone between the transition point and the zone of unitary frequency, which encompass 3575 words. It should be noted that Z1 includes 42% of a paper's words, on average, while Z2 encompasses 48% and Z3 comprehends 10%. These results are graphically represented in Fig. 5.

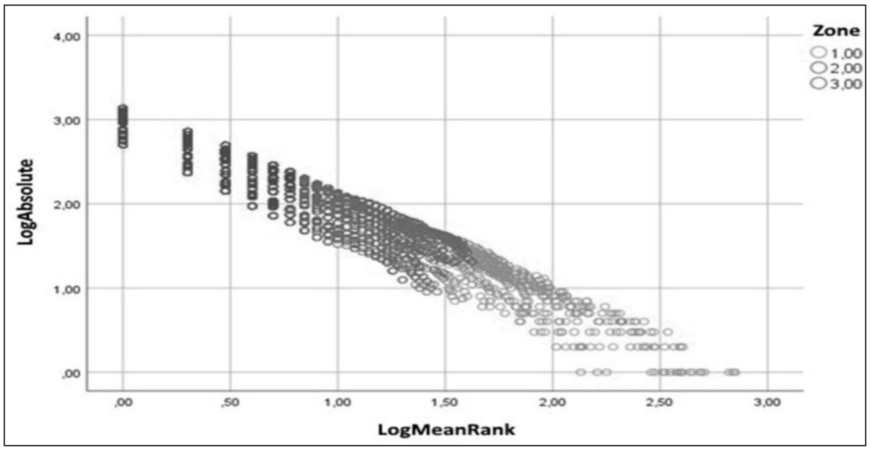

Fig. 5: Zipf Plot for the Z1, Z2 and Z3 of the 21 Papers (log rank - log absolute frequency of all tokens)

Confirmation of Zipf's Laws

The last step aims to answer the third and fourth questions of this research. To achieve this, the analysis consisted of a mixed method analysis (qualitative and quantitative) from a comparative perspective. In this framework, a first qualitative-quantitative analysis was done, the retrieved papers were systematically summarized in two mind maps (Fig. 6, Fig. 7) containing two variables (author keywords and two Zipf's words zones). Regarding the quantitative method, the papers' word frequencies were summarized in Fig. 6 (papers 1 to 10) and Fig. 7 (papers 11 to 21), which displayed the words consisting of the first and second zones of word frequencies.

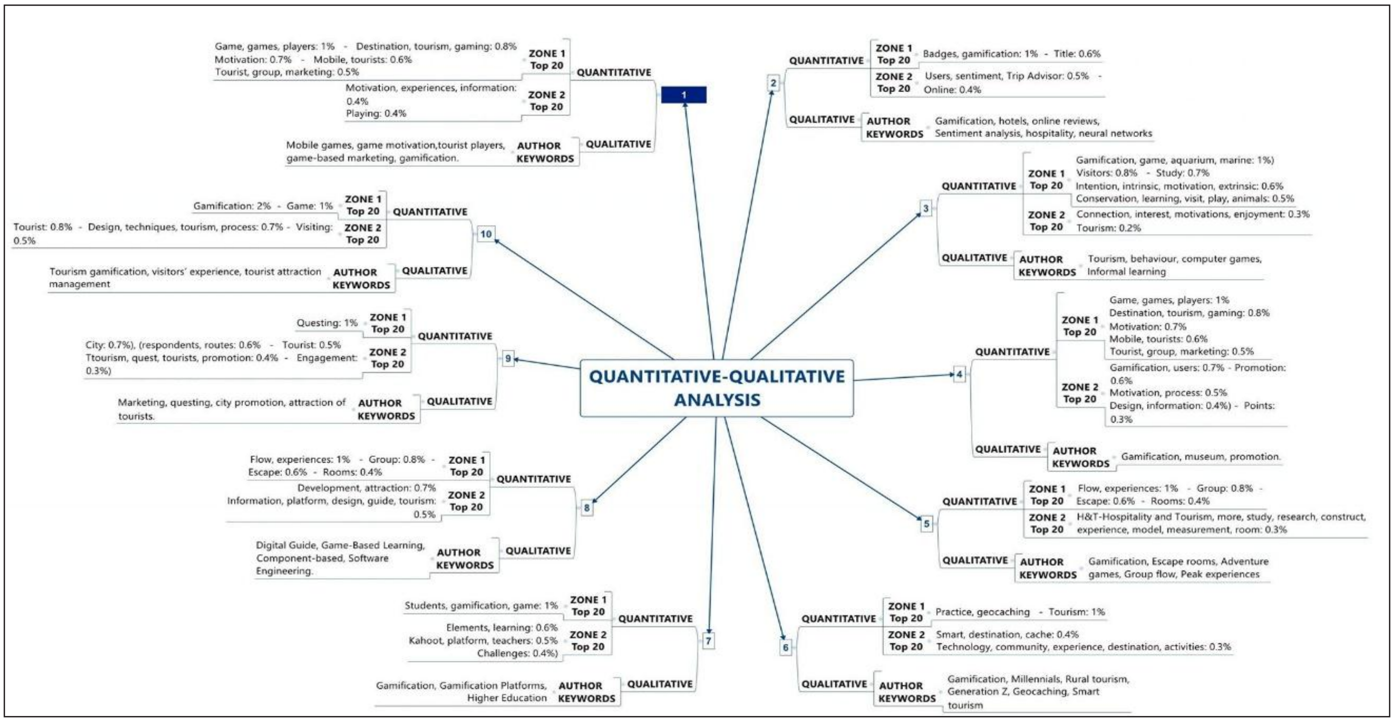

Fig. 6: Quantitative-Qualitative Analysis from Papers 1 to 10

That is, in zone 1 those words that are before Goffman's T (transition) point, and in zone 2 those words that are after Goffman's point, and which have high frequencies and are supposedly intrinsically related to the text's main subject. For this analysis, stop words, such as articles and prepositions, as well as formal elements such as citations, were excluded; otherwise, they would show up in zone 1.

To understand the relationship between the words of Zipf's two transition zones and the study focus of the papers, two more variables were added (study focus and applications to tourism) and summarized in Table 4.

The results presented in Table 4, Fig. 6 and Fig. 7 compared the keywords of the two transition zones with the qualitative variables (identified by reading the papers) study focus, key-theme and applications to tourism, they are related. We point out as an example "experiences" (e.g. paper 15) is related with "mobile experiences" of the study focus analysis, "motivations" (paper 3) is related with "behavior and tourism" of the key-theme, "flow experiences" (paper 5) is related with key-theme of "group flow and adventure games". This evidence shows that by applying Zipf's second law, next to the transition zones of Goffman's point are the words that identify the key-theme of the paper and that is not always evident through the author keywords. This confirms that descriptive words in tourism gamification literature are concentrated around the point of Zipf's second law.

The results reveal that there is some similarity between the descriptive words concentrated around the point of Zipf's second law and the author keywords announced in each paper. However, the results also revealed that there are some papers whose author keywords are not present in zones 1 and 2 of Goffman's point. Some examples are "neural networks" (paper 2), "peak experiences" (paper 5), "gamification" (paper 6), "marketing" (paper 9), "hedonic goods" (paper 11) and "Airbnb" (paper 14).

Focusing the analysis on the transition zones, the results revealed that, in some cases, words that are not present in the author keywords appear in Z1 and Z2 zones. It is the case of "experiences + gamified" (paper 15), "gamified + quality" (paper 11), "gamified + funware + experiences" (paper 18) "crowdsourcing + experience + gamified" (paper 20), "intrinsic + motivation + extrinsic (Zone 1) + connection + interest + motivations + enjoyment (Zone 2)" (paper 3), "flow + experiences " (paper 5), "development + attraction" (paper 8), "questing + routes + tourist + tourism + quest" (paper 9), "sustainability + hedonic" (paper 11), " perceived + playfulness + gamification" (paper 12) and finally " experiences + information" (paper 15).

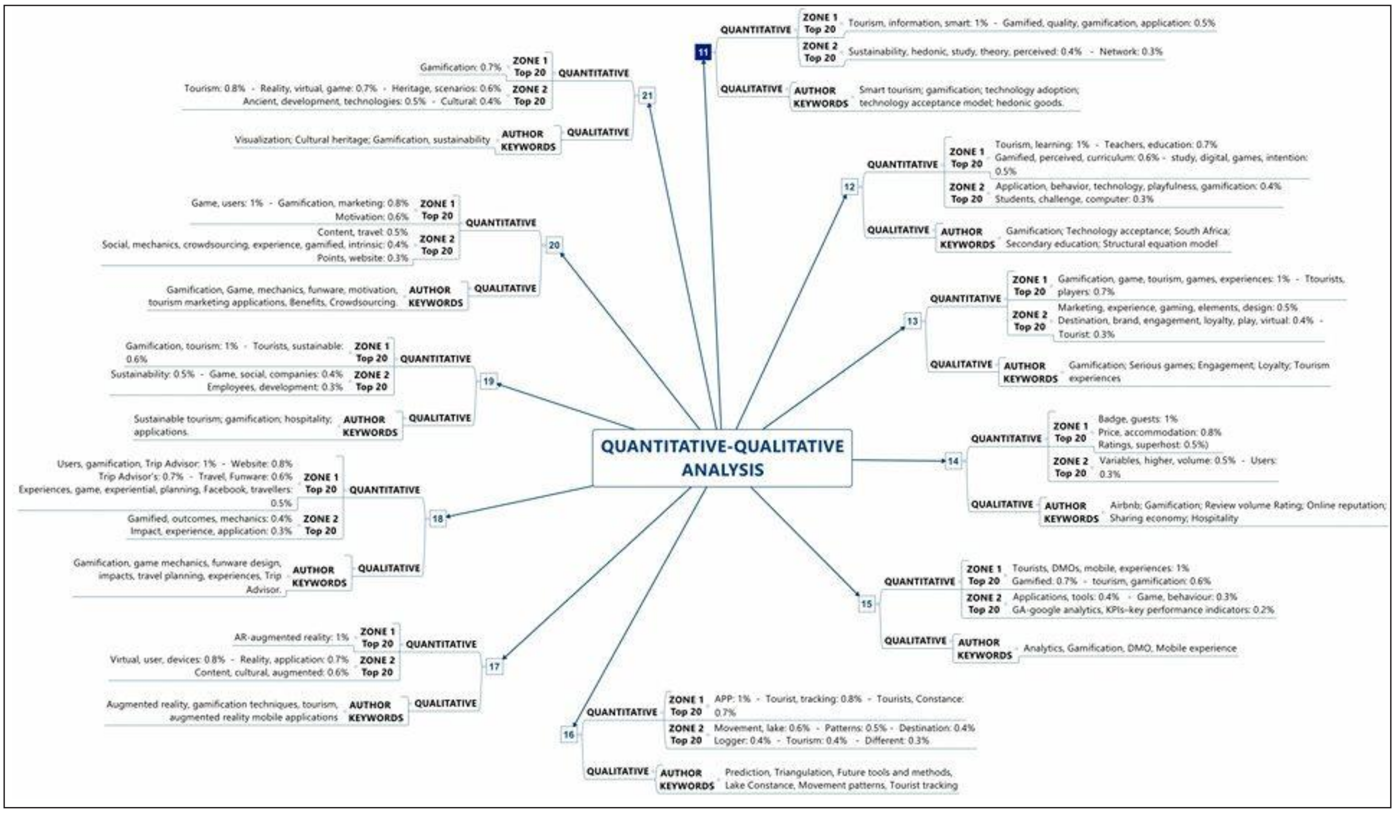

Fig. 7: Quantitative-Qualitative Analysis from Papers 11 to 21

Table 4: Papers Qualitative Analysis

Sr. No.	Study Focus	Applications to Tourism	Sr. No.	Study Focus	Applications to Tourism
1	Tourists players and game-based marketing.	Gamification for tourism experience.	11	Smart tourism and gamification.	Perceived enjoyment–hedonic experience–influences Gamified Smart Tourism Applications.
2	Online reviews and gamification.	Gamification and travelers' behavior.	12	Gamification and technology acceptance.	Gamified tourism learning.
3	Computer games, behavior and tourism.	Play a game imparts knowledge and promotes marine animal conservation.	13	Gamification and serious games in tourism experiences	Games provide tourists hedonistic experiences and increased brand awareness and brand loyalty.
4	Museum promotion and gamification.	Mobile applications used to promote museums.	14	Online reviews and gamification designs.	Gamification design influences accommodations' review volume and ratings.
5	Gamification, group flow and adventure games.	Flow experience theory. Group flow experiences with "escape rooms".	15	Analytics, gamification and mobile experience.	Gamified mobile experiences potential to enrich tourists' destination experiences.
6	Gamification, tourism, geocaching, millennial and generation Z.	Millennial and Generation Z gamified experiences to satisfaction of destination needs.	16	Future tools and methods of tourist tracking.	An algorithm-based prediction tourism movement.
7	Gamification platforms and higher education.	Gamification platforms used for tourism education.	17	Augmented reality mobileapplications in tourism.	Outstanding examples of Augmented Reality applications on tourism and cultural heritage.
8	Game-based learning and digital guides.	Game-based digital guides increase the entertainment effect in cultural attractions.	18	Impacts of gamification and gaming mechanics in tourism.	Gamification to help travelers (co)-learn and (co)-discover a destination, influences on their perceptions, images, and behaviors.
9	Promotion, gamification, questing and storytelling.	Questing employed to promote cultural and natural sites.	19	Gamification applied to sustainable tourism.	Gamification techniques used as an interface between tourists, organizations and local communities to leverage sustainable behavior.
10	Gamification and visitors' experiences.	Gamification techniques to improve tourists' experiences.	20	Gamification applications and Benefits in Tourism.	Gamification influencing tourist behavior at any stage of the consumption process.
			21	Gamification and virtual platforms and courses.	Incorporating Vokis recordings in virtual language courses.

DISCUSSION

Considering tourism contribution, first, this paper analyzes the content of almost 10 years of scientific research in tourism gamification. Table 4 summarizes study focus, and applications to tourism from 2010 to March 2019. Second, this study identifies some words through the application of the two laws of Zipf. The study identified some words, not present in the author keywords, "fuzzy words" of the topic "tourism gamification".

The first fuzzy word identified was "gamified". In paper 15, Garcia et al. (2019) apply "gamified" to the "mobile" "experiences" by stating that games played on-site should be simple, relaxed and not very challenging and should provide useful information about the destination as well as allow interaction with other people. Yoo et al. (2017), in paper 11, apply "smart", "gamified" and "quality", and argue that perceived gamified quality benefits the hedonic motivation, hedonic gaming and hedonic tourism. Also, Kim and Hall's (2009) research applies hedonic motivation to virtual reality.

On the other hand, in paper 18, Sigala (2015a: 190) applies "gamified experience", "gamified outcomes" and "gamified mechanics", and also "gamified and funware experiences", defending that gamification is the use of game-play mechanics for non-game applications and making the "distinction between 'serious game' (the design of full-fledged game for non-entertainment purposes) from 'gamified' applications that incorporate some (and not all) of the game design elements". Funware is the action to "transforma desired user action into a 'play' task and motivate the users' engagement with this activity". In paper 20, Sigala (2015b) applies the word "gamified" to "gamified marketing practices".

In the case of paper 5, there is an interesting situation, the "peak experiences" concept, present in author keywords, and pointed out as one of the originalities of this paper, are not around Goffman's point. However, a similar concept emerges that we identified as a new fuzzy word, “flow experiences”. Kolar and Cater (2018: 2637) argue that the sharing of customer experiences is exemplified by a group of flows and that “shared fun is doubled fun yet knowledge of how to manage shared experiences is lacking”.

In zone 2, at higher word frequencies of paper 9, we found a word linked to “questing”, namely “questing routes”. Widawska-Stanisz (2018) claims that questing and questing routes are a new tourism solution.

The present study aimed to verify whether Zipf’s laws apply to academic texts. The state-of-the-art research on gamification in the tourism industry was adopted as a topic of analysis. First, a detailed search for scientific articles in the subject was made on Elsevier’s Scopus database, through which the authors obtained and analyzed 21 articles. The retrieved publications were all subjected to categorical content analysis, performed by applying the DB Gnosis software, a purpose-made tool for the analysis of big volumes of text. Next, more specific quantitative analysis techniques, mainly linear regression analysis, were performed on SPSS.

The results clearly show that Zipf’s laws are indeed applicable to scientific articles. The findings are in line with Zhu et al.’s (2018) theoretical description of Zipf’s law. More specifically, the findings not only show that academic papers on gamification in the tourism industry follow a Zipfian word frequency distribution, but they also corroborate and extend the findings obtained by Molinos et al.(2016) through their analysis of Zipf’s laws applied to text in English. In line with the mentioned study’s findings, in the present analysis, zone 1, which encompasses the most frequent words in a text, was mostly composed by the words that are the basis of the English language’s syntax, namely, definite articles and prepositions.

The results regarding the application of Zipf’s first law also corroborate the methods employed by Odueke and Weir (2012) in the context of forensic accounting. Accordingly, results regarding the application of Zipf’s second law corroborate the methods suggested by Tyagi et al. (2017) and applied in the context of a documentary collection. These methods were here applied to the analysis of scientific articles with similar results, expanding the academic knowledge on the application of Zipf’s power laws.In sum, results are in line with the general knowledge and recognized theory on Zipfs’ law and expand its application to the realm of scientific texts.

CONCLUSIONS

Content analysis involves several tools applied to explore, organize and analyze scientific literature. It presents solid views into the intellectual structure, present status, research subjects, and essential factors among a scientific field. As a result, content analysis has been conducted in various studies within different contexts. The current study employed Zipf’s laws for content purposes.

Concerning the theoretical aspect, the results of the present study corroborate the assumption that Zipf’s laws hold true for every text, in every human language (Molinos et al., 2015). Although the properties of Zipf’s laws to texts in general, and even to realms beyond written and spoken language, have been acknowledged for several decades, their application to academic literature had not yet been tested by any previous study. Confirming that Zipf’s laws hold true also for scientific texts is an original result of the present research, which therefore provides a significant theoretical contribution. Furthermore, the current study contributes to knowledge in various fields by providing a new method of content analysis, namely through the application of Zipf’s laws. It also adds to the few studies that have tested gamification within the tourism context.

The study also provides a methodological contribution. The authors employed DB Gnosis software to perform a detailed and comprehensive categorical content analysis on the papers. DB Gnosis, as the name suggests, was originally conceived to help concerned scholars analyzing destination brand perceptions through qualitative inquiries. The software was purpose-made for a specific research project, in which researchers analyzed mega data of over 20,000 observations (Cardoso et al., 2019) consisting of descriptions of dream and favorite destinations through three words, hence its capabilities to quickly process big amounts of text data. The mechanisms behind the analysis performed by software, however, are simply text analysis mechanisms, only perfected to the desired purpose. In the present study, DB Gnosis was applied to the analysis of structured, second-hand texts, namely scientific articles, rather than unstructured first-hand text data on destination perceptions. Moreover, it was used to quantify qualitative data. Therefore, it could be used in a further purely quantitative analysis on SPSS (logistic linear regression). This unveiled a potential offered by software and until then unexplored. Future studies can capitalize on this contribution by applying DB Gnosis, alone or paired with SPSS (or other quantitative analysis software), for a diverse array of research projects including content analysis of big volumes of text.

From a practical perspective, the present study confirms that Zipf’s laws apply to scientific texts. In this context, they can indeed be employed to detect key concepts of a certain topic within complex qualitative information sources. In other words, the adoption of Zipf’s laws provides useful practical implications for academic literature in tourism. In addition, the current study provides clear insights for scholars to identify better views and develop the directions of future research in several fields, especially within the tourism context. Researchers could use Zipf’s power laws in order to identify the main sub-topics within the more comprehensive subject they are researching, as well as which papers or documents of other types are relevant for that particular subject. In this vein, applying Zipf's power laws to literature reviews has the potential of making the whole process more efficient, avoiding loss of time in scanning texts just to realize they are not relevant for the topic after all.

Naturally, when simply applied by the researcher, the potential of Zipf's laws is quite limited, since carrying out the whole process manually, especially for the first time, would imply a more significant volume of work than actually scanning through a bigger set of papers. In addition, it would presuppose an intermediate knowledge by the researcher regarding the properties of such laws. However, if incorporated into search engines (such as Scopus) through algorithms, these laws could have a higher impact, turning the whole process of searching and selecting documents for a literature review smarter and more efficient.

In sum, the present research presents original results regarding the adoption of Zipf's laws for content analysis purposes. However, as with any study, some limitations should be considered. First, being an original initiative in testing Zipf's power laws' application to academic texts, the authors did not have much to rely on in terms of references, beyond descriptive material on Zipf's laws per se and studies applying them to other areas. In order to focus on a feasible set of qualitative data, the authors selected a specific topic within the literature in tourism, namely gamification in the tourism industry, which was selected due to its current importance within the context of destination loyalty, destination engagement, smart tourism and the experience economy. Based on these limitations, future studies should expand the present findings by testing Zipf's power laws' application to other tourism subjects, as well as to studies of other phenomena and different disciplines.

REFERENCES

Adukaite, A., vanZyl, I., Er, Ş., & Cantoni, L. (2017). Teacher perceptions on the use of digital gamified learning in tourism education: The case of South African secondary schools. *Computers & Education, 111,* 172-190. doi:doi.org/10.1016/j.compedu.2017.04.008

Bardin, L. (2000). *Análise de conteúdo.* Lisbon: Edições 70. ISBN: 972-44-0898-1.

Bochkarev, V., & Lerner, E. (2017). Calculation of precise constants in a probability model of Zipf's law generation and asymptotics of sums of multinomial coefficients. *International Journal of Mathematics and Mathematical Sciences.* doi:org/10.1155/2017/9143747

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Cardoso, L., Dias, F., de Araújo, A. F., & Marques, M. I. A. (2019). A destination imagery processing model: Structural differences between dream and favourite destinations. *Annals of Tourism Research, 74,* 81-94. doi:org/10.1016/j.annals.2018.11.001

Cirulis, A., De Paolis, L. T., & Tutberidze, M. (2015). Virtualization of digitalized cultural heritage and use case scenario modeling for sustainability promotion of national identity. *Procedia Computer Science, 77,* 199-206. doi:org/10.1016/j.procs.2015.12.384

Cleophas, T. J., & Zwinderman, A. H. (2018). Continuous outcome regressions. In *Regression Analysis in Medical Research* (pp. 1-39). Cham: Springer,

De Mudarra-Fernández, A., Carrillo-Hidalgo, I., & Pulido-Fernández, J. (2019). Factors influencing tourist expenditure by tourism typologies: A systematic review. *An International Journal of Tourism and Hospitality Research, Anatolia 30*(1), 18-34. doi.org/10.1080/13032917.2018.1495086

Deterding, S. (2012). Gamification: Designing for motivation. *Interactions, 19*(4), 14-17.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: Defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9-15). ACM.

Domingo-Carrillo, M., Chávez-Miranda, E., & Escobar-Pérez, B. (2019). Scientific production on revenue management in tourism on Web of Science and SCOPUS. *Current Issues in Tourism,* 1-26.

Flores, J. (2016). *Master in linguistics: The study of the language faculty.* Netherlands: Utrecht Institute of Linguistics OTS.

Garcia, A., Linaza, M. T., Gutierrez, A., & Garcia, E. (2019). Gamified mobile experiences: Smart technologies for tourism destinations. *Tourism Review, 74*(1), 30-49. doi.org/10.1108/TR-08-2017-0131

Goffman, E. (1964). The neglected situation. *American Anthropologist, 66*(6), 133-136.

Goffman, E. (1975). Characteristics of total institutions. In W. Derczyński, A. Jasińska-Kania & Szacki, J. (Ed.). *Elements of Sociological Theories*. PWN, Warsaw.

Goshevski, D., Veljanoska, J., & Hatziapostolou, T. (2017). A review of gamification platforms for higher education. In *Proceedings of the 8th Balkan Conference in Informatics* (p. 28). ACM.

Griffin, T. (2013). Research note: A content analysis of articles on visiting friends and relatives tourism, 1990–2010. *Journal of Hospitality Marketing & Management, 22*(7), 781-802. doi.org/10.1080/19368623.2012.708960

Hamari, J. (2013). Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications, 12*(4), 236-245. doi:org/10.1016/j.elerap.2013.01.004

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*(1), 77-89. doi:org/10.1080/19312450709336664

Huang, S. M., Yen, D. C., Yang, L. W., & Hua, J. S. (2008). An investigation of Zipf's Law for fraud detection (DSS# 06-10-1826R (2)). *Decision Support Systems, 46*(1), 70-83. doi:org/10.1016/j.dss.2008.05.003

Ioannides, M., Magnenat-Thalmann, N., & Papagiannakis, G. (Eds.) (2017). *Mixed reality and gamification for cultural heritage* (vol. 2). Switzerland: Springer.

Jovell, A. J. (2006). Análisis de regresión logística. *Cuadernos Metodológicos, 15.* Madrid: Centro de Investigaciones Sociológicas.

Kim, M. J., & Hall, M. (2019). A hedonic motivation model in virtual reality tourism: Comparing visitors and non-visitors. *International Journal of Information Management, 46*, 236-249. doi:org/10.1016/j.ijinfomgt.2018.11.016

Kolar, T., & Čater, B. (2018). Managing group flow experiences in escape rooms. *International Journal of Contemporary Hospitality Management, 30*(7), 2637-2661. doi:org/10.1108/IJCHM-03-2017-0152

Koplening, A. (2015). Using the parameters of the Zipf-Mandelbrot law to measure diachronic lexical, syntactical and stylistic changes - A large-scale corpus analysis. Degruytermouton: Corpus Linguistics and Ling. Theory.

Krippendorff, K. (2004). Reliability in content analysis. *Human Communication Research, 30*(3), 411-433.

Lei, S. I., Wang, D., & Law, R. (2019). Perceived technology affordance and value of hotel mobile apps: A comparison of hoteliers and customers. *Journal of Hospitality and Tourism Management, 39,* 201-211. doi:org/10.1016/j.jhtm.2019.02.006

Liang, S., Schuckert, M., Law, R., & Chen, C. (2017). Be a "Superhost": The importance of badge systems for peer-to-peer rental accommodations. *Tourism Management, 60*, 454-465. doi.org/10.1016/j.tourman.2017.01.007

Luimula, M., & Trygg, M. N. B. (2016, October). *Cultural heritage in a pocket: Case study "Turku castle in your hand".* In 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 000055-000058). IEEE.

Mesáro, P., Mandičák, T., Hernandez, M. F., Sido, C., Molokáč, M., Hvizdák, L., & Delina, R. (2016). Use of augmented reality and gamification techniques in tourism. *E-review of Tourism Research, 2.*

Molinos, D., Mesquita, D., & Hoff, D. (2016). ZipfTool: Uma ferramenta bibliométrica para auxílio na pesquisa teórica. *Revista de Informática Teórica e Aplicada, 23*(1), 293-317.

Montemurro, M., & Zanette, D. (2002). New perspectives on Zipf'slaw in linguistics: From single texts to large corpora. *Glottometrics, 4,* 87-99.

Moran, J., & Bouchaud, J. P. (2018). Greedy algorithms and Zipf laws. *Journal of Statistical Mechanics: Theory and Experiment, 2018*(4), 043402.

Moro, S., Ramos, P., Esmerado, J., & Jalali, S. (2019). Can we trace back hotel online reviews' characteristics using gamification features? *International Journal of Information Management, 44*, 88-95. doi:org/10.1016/j.ijinfomgt.2018.09.015

Negruşa, A., Toader, V., Sofică, A., Tutunea, M., & Rus, R. (2015). Exploring gamification techniques and applications for sustainable tourism. *Sustainability, 7*(8), 11160-11189. doi.org/10.3390/su70811160

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics, 46*(5), 323-351. doi:org/10.1080/00107510500052444

Odueke, A., & Weir, G. (2012). Triage in forensic accounting using Zipf's law.

Powers, D. (1998, January). Applications and explanations of Zipf's law. In *Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning* (pp. 151-160). Association for Computational Linguistics.

Prasetyo, N. (2018). Design mobile app for increase the visitor museum using gamification method. *Telkomnika, 16*(6).

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation, 25*(4), 348-349.

Qin, M., Tang, C., Jang, S., & Lehto, X. (2017). Mobile app introduction and shareholder returns. *Journal of Hospitality and Tourism Management, 31,* 173-180. doi:doi.org/10.1016/j.jhtm.2016.11.006

Schreier, M. (2012). *Qualitative content analysis in practice*. Sage Publications.

Sigala, M. (2015a). The application and impact of gamification funware on trip planning and experiences: The case of TripAdvisor's funware. *Electronic Markets, 25*(3), 189-209.

Sigala, M. (2015b). Gamification for crowdsourcing marketing practices: Applications and benefits in tourism. In *Advances in Crowdsourcing* (pp. 129-145). Cham: Springer.

Skinner, H., Sarpong, D., & White, G. R. (2018). Meeting the needs of the Millennials and Generation Z: Gamification in tourism through geocaching. *Journal of Tourism Futures, 4*(1), 93-104. doi:org/10.1108/JTF-12-2017-0060

Smith. (2007) Investigation of the Zipf-plot of the extinct Meroitic language. *Glottometrics, 15,* 53-61.

Swacha, J., & Ittermann, R. (2017) Enhancing the tourist attraction visiting process with gamification: Key concepts. *Engineering Management in Production and Services, 9*(4), 59-66. doi:doi.org/10.1515/emj-2017-0031

Tahamtan, I., & Bornmann, L. (2019). What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. *Scientometrics.* doi.10.1007/s11192-019-03243-4

Tan, W. K. (2018). Gamification in aquarium context: Intention to play game that imparts knowledge and promotes marine animal conservation. *Information Technology & People, 31*(6), 1070-1090.

Thimm, T., & Seepold, R. (2016). Past, present and future of tourist tracking. *Journal of Tourism Futures, 2*(1), 43-55. doi.org/10.1108/JTF-10-2015-0045

Thurner, S., Hanel, R., Liu, B., & Corominas-Murtra, B. (2015). Understanding Zipf's law of word frequencies through sample-space collapse in sentence formation. *Journal of the Royal Society Interface, 12*(108), 20150330. doi:doi.org/10.1098/rsif.2015.0330

Tyagi, M. S., Yadav, M. D., Khandelwal, M. S., Agal, S., Gits, U., & Aits, U. (2017). A new research methodology for preprocessing of documents: Information recovery systems. *IJCSIS, 15*(4).

Ulubasogulu, M., & Hazari, B. (2004). Zipf's law strikes again: The case of tourism. Journal of Economic Geography, 4, 459-472. doi:doi.org/10.1093/jnlecg/lbh030

Walz, S. P., & Deterding, S. (Eds.) (2015). *The gameful world: Approaches, issues, applications.* MIT Press.

Weber, J. (2014). Gaming and gamification in tourism. Retrieved January 13, 2015, from http://thinkdigital.travel/wp-content/uploads/2014/05/Gamification-in-Tourism-BestPractice.pdf

Widawska-Stanisz, A. (2018). Questing in city promotion on the example of the city of Częstochowa.

Xu, F., Buhalis, D., & Weber, J. (2017). Serious games and the gamification of tourism. *Tourism Management, 60,* 244-256. doi.org/10.1016/j.tourman.2016.11.020

Xu, F., Tian, F., Buhalis, D., Weber, J., & Zhang, H. (2016). Tourists as mobile gamers: Gamification for tourism marketing. *Journal of Travel & Tourism Marketing, 33*(8), 1124-1142. doi.org/10.1080/10548408.2015.1093999

Yoo, C., Kwon, S., Na, H., & Chang, B. (2017). Factors affecting the adoption of gamified smart tourism applications: An integrative approach. *Sustainability 9*(12), 2162. doi.org/10.3390/su9122162

Zanjirchi, S. M., Abrishami M. R., & Jalilian, N. (2019). Four decades of fuzzy sets theory in operations management: Application of life-cycle, bibliometrics and content analysis. *Scientometrics,* 1-21. doi:doi.org/10.1007/s11192-019-03077-0

Zhu, Y., Zhang, B., Wang, Q., Li, W., & Xu, C. (2018). The principle of least effort and Zipf distribution. *Journal of Physics: Conference Series, 1113.*

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. O'Reilly Media, Inc.

Zipf, G. K. (1949). Human behavior and the principle of least effort.

Funding: "This research is financed by national funds through FCT - Foundation for Science and Technology, I.P., within the scope of the UIDB/04470/2020 project".

Conflicts of Interest: "The authors declare no conflict of interest."