
Aleksei Gorgadze

National Research University - Higher School of Economics

Valery Gordin

National Research University - Higher School of Economics

Natalia Belyakova

National Research University - Higher School of Economics

Semantic Analysis of the Imperial Topic: Case of St. Petersburg¹

Tourism products' variety engenders a correspondingly large number of approaches to their promotion. One popular angle for promotion, applicable in a handful of cities, draws on the destination's imperial heritage. However, the patterns of using the imperial theme for tourism marketing vary in different capitals. This study is the first in a larger project aiming to study the former imperial capitals (St. Petersburg, Istanbul, Berlin, and Vienna). Our overall goal is to understand, first, how city representatives and travel agencies narrate the "imperial" past and turn it into a competitive advantage for the city, and second, how tourists react to such narratives. To collect and analyse the data, web scraping and text mining techniques are used. In this article, the first results of our study using the case of St. Petersburg are presented.

Key words: imperial capitals, semantic analysis, digital footprint analysis

Aleksei Gorgadze

Department of Management

St.Petersburg School of Economics and Management

National Research University - Higher School of Economics

3A Kantemirovskaya street, St. Petersburg, Russia

Phone: +7 (911) 834-88-02

Email: agorgadze@hse.ru

Valery Gordin

Department of Management

St.Petersburg School of Economics and Management

National Research University - Higher School of Economics

3A Kantemirovskaya street, St. Petersburg, Russia

Email: gordin@hse.ru

¹ The results of the project "Cultural and event activity as development factor of revitalized territories", carried out within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) in 2018, are presented in this work.

Natalia Belyakova
Department of Management
St.Petersburg School of Economics and Management
National Research University - Higher School of Economics
3A Kantemirovskaya street, St. Petersburg, Russia
Email: nubelyakova@hse.ru

Aleksei Gorgadze is a postgraduate student in the Department of Management at the National Research University - Higher School of Economics and the lecturer of the Text Mining course. His research interests include event management, tourism studies, text mining, and social network analysis.

Dr. Valery Gordin is a tenured Distinguished Professor of Management at the National Research University - Higher School of Economics and the academic supervisor of the Master's program "Cultural and Event Tourism Management". His current research focuses on the economics of cultural tourism, as well as on the field of management in cultural and creative industries.

Prof. Natalia Belyakova is an Associate Professor of Management at the National Research University - Higher School of Economics and the regional marketing and PR Director at Domina Holding.

Introduction

St. Petersburg is one of the world's few large cities whose appearance was entirely planned. From its founding in 1703, the city was preordained to become the capital of the Russian Empire – the status it enjoyed for three hundred years. But even a century after relinquishing it, the city's imperial spirit is still alive. Tour operators and the city authorities promote tourism products related to the city's imperial past. But how effective are these efforts?

Traditionally, the imperial theme has been prominently present in tourist-oriented English-language media materials representing the city. The destination's main connotative marker, which defines it on the international tourism market, is “the capital of the Russian Empire.” The focus of our attention in this paper, however, is not on media materials as such, but rather on user-generated content featuring the “imperial” marker, as well as on the analysis of this marker's representativeness and weight in the overall body of associations linked with Petersburg that were prompted by the commenters' own experience.

At the same time, we, as researchers, are conscious of the close links between the two datasets. It is impossible to neatly separate media-induced and experience-based images of Petersburg – if only because the commenters themselves do not always register the media's influence on them (and sometimes even deny it).

Some of them are consumers of “official” content, which represents Petersburg as a destination for cultural and historic tourism; others are forming their opinions in a tentatively independent manner. Even the latter group's judgement, however, is affected by the previously assimilated patterns in perceptions of Russia (stereotypes).

Without putting “imperial tourism” into a separate category, we analyse the extent to which the “imperial” discourse (a conglomerate of established perceptions of St. Petersburg in the media space) is assimilated (read, reiterated) by consumers. Using semantic analysis, we probe the following problem: how effective was the imperial rhetoric used in the city's positioning and promotion vis-à-vis the tourists who have visited the city? Our starting assumption is that the presence of “imperial” connotative terms in travellers' reviews testifies, at a minimum, to the users' familiarity with the concept, as well as to their effort at its validation (by comparing the induced image with one's own experience), the result of which may be either acceptance and assimilation or rejection.

The historical legacy of a place can be used both in creating formalized brands and in forming less structured “association clouds,” as well as for rank-optimization with an eye

toward promoting pre-existing sub-brands (sites of cultural heritage, museums, etc.). Pioneering research by Ashworth and Page in the early 2010s (Kavaratzis et al., 2010 & Ashworth et al., 2011) lay the foundations for studying the “historic city” as an important concept in city tourism – including studying how residents, tourists, and businesses are influenced and informed by cities’ historical heritage. Currently, this topic is debated in research publications on destination branding and marketing (for a brief overview, see Scaramanga, 2012 and Evans, 2015). Nevertheless, the issue of perception of “historical” information by tourists – how in demand it is, whether actual experiences match the expectations, and how much it affects the final impression of a country – is typically left outside the research scope. Our research is meant to bridge this gap.

The purpose of this study is to assess the extent to which representatives of the city administration and tour operators call attention to the imperial past as a competitive advantage for the city and compare it with the response that these “imperial” - themed messages elicit in tourists.

The object of the study are the descriptions and reviews of sights, museums, and parks of St. Petersburg posted on official web city pages, as well as reviews posted on the TripAdvisor web site (with all data in the English language),

User-generated content (UGC), such as reviews on TripAdvisor, is an important source of information about consumer preferences. Digital footprint analysis is a relatively new, but very promising field in tourism research. Numerous studies have been devoted to the behaviour of tourists in the digital space, including the study of multifunctional tourist spaces (Salas-Olmedo et al., 2018), the analysis of Internet users’ search queries (Dergiades et al., 2018), the assessment of levels of satisfaction with hotel services by business travel representatives (Boo et al., 2018), the study of sentiment analysis (emotional colouring) and topic modelling of tourists’ reviews (Godnov et al., 2016), the study of digital footprint efficiency as an indicator of tourism demand (Önder et al., 2016), the study of how perceived hotel “value dimensions” are connected with the hotels’ categories (Kaspruk et al., 2017), etc.

Methodology

This study was set within the context of St. Petersburg – Russia’s second-largest city and the former capital of the Russian Empire. The design of the study involves comparing two corpora of text data: **the supply side** (descriptions of tourism products on the official city web sites and the web sites of travel agencies) and **the demand side** (tourists’ reviews on the

TripAdvisor web site). Thus, the methodology for data collection and analysis involves two basic components:

1. Analysis of the creation and promotion of the “imperial” tourist product.

This component analyses web pages of the city administration’s official representatives, as well as information about tourist products on travel web sites. In order to select the relevant web page data, we compiled a list of keywords to use in Google searches. For each search query, we selected the top 20 links to websites. Using scripts written in the R language, we collected the descriptions of tourist products (in the English language) from these sites. The resulting text data was processed and analysed using text mining techniques.

In order to assess the degree of use of imperial words in tour products, we generated a list of key-words/indicators, which were searched and counted. Additionally, we considered the average share of the imperial words use in each site’s description.

2. Analysis of the tourism products’ perception as “imperial” by consumers.

This component’s aim is to assess how much consumers of the tourist product really perceive and respond to the “imperial” branding when visiting the city’s tourist sights. For this purpose, we collected all English-language reviews of sights, parks, and museums of St. Petersburg posted on the TripAdvisor website. We then searched and analysed the imperial keywords that were used in the previous step of our analysis.

Web scraping was conducted in June, 2018. We collected the text data from 64 web site belonging to the city administration and the city’s travel agencies, and 8,354 reviews of St. Petersburg sights on the TripAdvisor website.

Data preparation involved several steps. First, we conducted lemmatization to transform the words into their infinitive forms. This method is superior to stemming because it doesn’t cut the ending and preserves the word’s sense. For this purpose, we used the R package “textstem”. We then removed all stopwords using the list of English stopwords contained in the R package “stopwords”. We also removed the punctuation, numbers, and extra white spaces and made all letters lowercase. After tokenization, we built the document-term matrix (documents indicate the reviews or texts of travel agent websites) for conducting our analysis.

We used several techniques from text mining and basic statistics in order to compare the supply side and the demand side. First, we used the t-test to identify significant differences in the average share of imperial words used. In other words, we first compiled a glossary of terms that identify imperial topics, then calculated these words’ share in each

Table 1: Description of the sub-corpora

Construct	Number of documents	Total word count	Unique word count
Supply side (official representatives and travel agencies)	64	139 115	11 802
Demand side (reviews)	8 354	267 891	10 342

Student’s t-test

We compared the average use of the “imperial” words in tour products (a channel for promoting the imperial theme) and in tourists’ reviews on the TripAdvisor website. The t-test showed a significant (p-value = 0.01411) difference in the mean (0.0047 & 0.0029) of these two (independent) samples. This indicates that the prevalence of the imperial theme in the tour products is, on average, higher than in the reviews.

Log-Likelihood (G-squared)

To establish a more relevant distinction in the perception of sights by tourists and tour agents, it is useful to look at the likelihood of the appearance of each term. The log-likelihood compares the occurrence of a term in two corpora to determine if it shows up more often or less often than expected. This method is accurate even at low frequencies (Dunning, 1993).

Figure 3 shows the most relevant terms for the supply and demand side based on the log odds ratio. We observe, for example, that tourists used words like “wow”, “toilet”, “pocket”, “cloakroom”, “worthwhile”, “disappoint”, “glad”, etc. more often.

There are several types of narratives at play here. The first reflects the emotional aspect of the visit. It’s hardly surprising that tourists should be more concerned about the availability of basic infrastructure and that they may give advice on possible service improvements. In contrast, the official representatives and the travel agencies focus on the types and names of sights (“monastery”, “mariinsky”, “mikhail”, “portico”), and on those terms that are specific to the tour route (“literary”, “drawbridge”, “industrial”, “cemetery”).

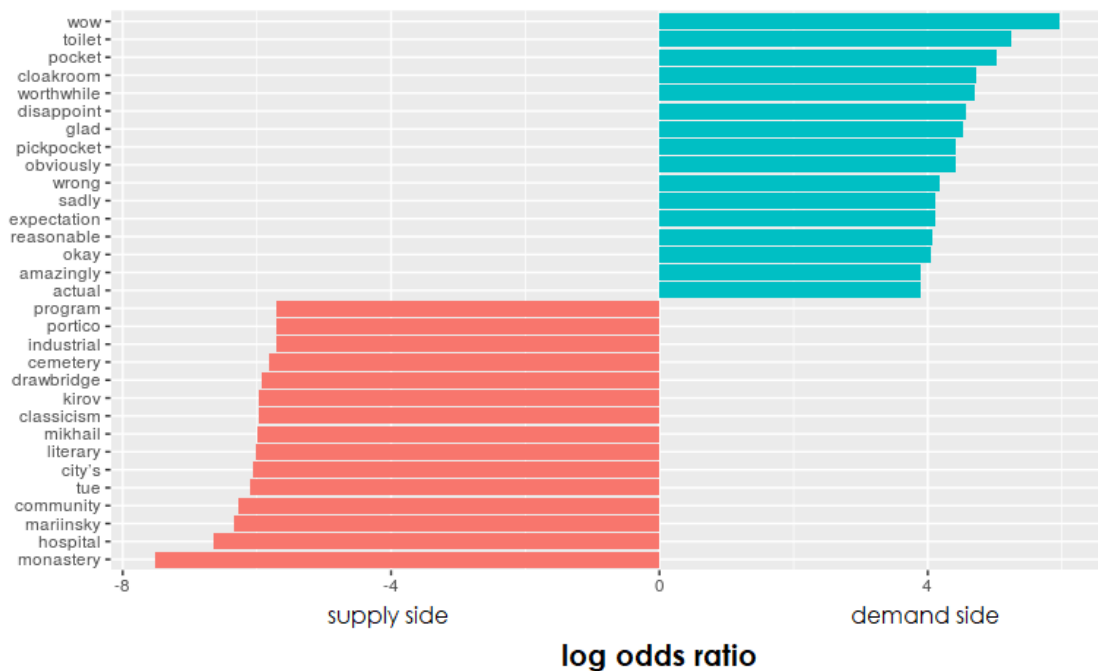


Figure 3. Log-likelihood ratios of frequencies by Supply side and Demand side.

Our research is aimed at studying the imperial factor. Therefore, we consider the use of such terms separately. First, we selected the terms related to the imperial topic. These are: “imperial”, “emperor”, “empire”, “tzar”, “czar”, and “tsar”. The term “tsar” is a title that the Russian monarchs often used in Russia. Etymologically, this title is derived from the Latin title for Roman emperors, “caesar” (Vodoff, 1978). However, there are different transcriptions of this term: “tsar” (the most common), as well as “tzar” and “czar”. For our purposes, we combined all these words into a single term and used it to calculate the log-likelihood. Despite the fact that the selected terms are used more in tourist reviews, they are more typical for travel agencies ($p < 0.0001$).

Table 2: The statistics of united Imperial words

Term	Supply side	Demand side	G-squared	Log Ratio
imperial words	599	692	109.96***	-0.8563

*** - $p < 0.0001$; critical value = 15.13

We also considered these terms separately. Table 3 shows the distribution of these terms, along with some statistics. All the words except “tsar” demonstrate a significant difference ($p < 0.0001$ and $p < 0.05$ for the term “czar”) in the supply and demand sides. The terms “imperial”, “emperor”, “empire” are more typical of texts by official representatives

and travel agencies. Tourists use the words “tzar” and “czar” more frequently. This leads us to the hypothesis that international tourists see the “imperialness” of Russia in Eastern/Asian terms.

Table 3: The distribution of Imperial words

Term	Supply side	Demand side	G-squared	Log Ratio
imperial	185	100	126.42***	-1.9522
emperor	157	77	118.11***	-2.0925
empire	46	19	40.14***	-2.3403
tzar	7	56	15.61***	1.9354
czar	36	116	5.51*	0.6234
tsar	168	324	0.72	-0.1171

*** - $p < 0.0001$; critical value = 15.13

* - $p < 0.05$; critical value = 3.84

The methodology and the results of this study can be applied to place brand-making using UGC. Our results can be used by city authorities and travel agencies for the creation of tourist products in order to help them to better represent the city’s brand on their websites.

Future plans

In parallel to this study, we are conducting a comparative analysis of the use of imperial themes in other former imperial capitals (Istanbul, Berlin, Vienna). Comparison of the results between these cases will allow us to test the hypothesis which states that the active use of the imperial theme by travel agents does not result in the commensurate amount of relevant feedback on tourists’ part. Moreover, we plan to apply modern accurate text mining methods such as LDA and LSA. This will allow us to find hidden phenomena and dependencies in our database.

References

- Boo, S., & Busser, J. A. (2018). Meeting planners' online reviews of destination hotels: A twofold content analysis approach. *Tourism Management*, 66, 287-301.
- Bradley, James V. (1968). *Distribution-Free Statistical Tests*. Prentice Hall.
- Cheltenham: Edward Elgar; Ashworth, G. and Kavaratzis, M. (2011) *Why Brand the Futures with the Past? The Roles of Heritage in the Construction of Place Brand Reputations*. *International Place Branding Yearbook*. Springer: 25-38

- Dergiades, T., Mavragani, E., & Pan, B. (2018). Google Trends and tourists' arrivals: Emerging biases and proposed corrections. *Tourism Management*, 66, 108-120.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational linguistics*, 19(1), 61-74.
- Evans, G.L. (2015) The role of culture, sport and heritage in place-shaping: A Literature Review, Department for Culture Media & Sport, CASE Evidence Programme
- Godnov, U., & Redek, T. (2016). Application of text mining in tourism: case of Croatia. *Annals of Tourism Research*, 58, 162-166.
- Hardie, A. (2014). Log Ratio: An informal introduction. Retrieved from <http://cass.lancs.ac.uk/?p=1133> (last accessed October 2018)
- Kaspruk, N., Silyutina, O., & Karepin, V. (2017, June). Hotel Value Dimensions and Tourists' Perception of the City. The Case of St. Petersburg. In *International Conference on Digital Transformation and Global Society* (pp. 341-346). Springer, Cham.
- Kavaratzis, M. and Ashworth, G. (Eds) (2010) Place branding: where do we stand? In: Ashworth, G. and Kavaratzis, M. (eds) *Towards Effective Place Brand Management. Branding European Cities and Regions* (1-14).
- Mood, A. M.; Graybill, E A.; and Boes, D. C. (1974). *Introduction to the Theory of Statistics*. McGraw Hill.
- Önder, I., Koerbitz, W., & Hubmann-Haidvogel, A. (2016). Tracing tourists by their digital footprints: The case of Austria. *Journal of Travel Research*, 55(5), 566-573.
- Scaramanga, M. (2012). Talking about art (s) A theoretical framework clarifying the association between culture and place branding. *Journal of Place Management and Development*, 5(1), 70-80.
- Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*, 66, 13-25.
- Vodoff, V. (1978). Remarques sur la valeur du terme "czar" appliqué aux princes russes avant le milieu du 15e siècle, in "Oxford Slavonic Series", new series, vol. XI. Oxford University Press

Acknowledgements

We would like to express our gratitude to Ilya Musabirov and Anastasia Kuznetsova for their assistance in data collection, Slava Borisov for his help in preparing this article for publication, and all our reviewers for their helpful comments.