

# Topic model analysis of attitudes of mothers with infants and toddlers toward travel before and after the COVID-19 pandemic

**Tadanari Taniguchi** (STEM Education Center, Tokai University, taniguchi@tokai-u.jp, Japan)

**Ryozo Kitajima** (Department of Engineering, Tokyo Polytechnic University, r.kitajima@eng.t-kougei.ac.jp, Japan)

**Shinji Sakamoto** (Department of Information and Computer Science, Kanazawa Institute of Technology, sskmt@neptune.kanazawa-it.ac.jp, Japan)

**Masato Koizumi** (Department of Cultural and Social Studies, Tokai University, koizumi@tsc.u-tokai.ac.jp, Japan)

## Abstract

*It is important for creating a travel-friendly environment to consider old people and young children, who have many issues when traveling. In this study, the attitudes of mothers with infants and toddlers toward travel before and after the COVID-19 pandemic were analyzed using data from mamari, which is a smartphone application for mothers to solve and share questions and concerns about pregnancy, childbirth, and childcare. Latent Dirichlet allocation topic models were constructed for the data, which were user question sentences written in Japanese. Thus, the attitudes of mothers toward travel were investigated using the topic model analysis.*

## Keywords

*travel, mother with infants and toddlers, natural language processing, topic modeling, COVID-19*

## 1. Introduction

Because of the outbreak of COVID-19, the Japanese government declared a state of emergency multiple times between 2020 and 2021 and asked to ensure people stayed at home. In the Tokyo metropolitan area, a state of emergency was declared four times [Cabinet Secretariat, 2022]. Therefore, travel demand was suppressed [JTB Foundation, 2021].

Kock et al. [2020] used evolutionary psychology to examine coronaviruses and their effect on the psychology of tourists. Sigala [2020] investigated the effect of COVID-19 tourism and their implications on tourism. Abbas et al. [2021] provided transformational potential and implications for a sustainable recovery of the travel and leisure industry. Yamamoto [2022] examined the impact of the COVID-19 on the transitions and characteristics of population in Nagoya City using mobile spatial statistics data [Suzuki et al., 2022] analyzed the impact of subsidy from the Japanese government to revitalize the tourism industry, which has been negatively affected by the COVID-19. Although studies have been conducted on COVID-19 and tourism, age groups and genders are yet to be considered in most studies.

It is important for creating a travel-friendly environment to consider old people and young children, who have many issues when traveling. In this study, the travel attitudes of mothers with infants and toddlers were analyzed. Data were provided by Connehito Inc. in the 2021 data analysis competition of the Joint Association Study Group of Management Science. These data were expressed in Japanese and consist of log data from mamari as a smartphone application. On the application sites [Connehito, 2022a; 2022b], it is an application for young mothers to solve and share questions and concerns related to pregnancy, childbirth, and childcare. The sites also express that

approximately 1.3 million questions are posted for each month and it is the No. 1 application for young mothers in Japan. It is readily available as a smartphone application and has numerous submissions. Thus, the log data is suitable for collecting information on the questions and concerns of young mothers in Japan.

Natural language processing was applied to the question data written in Japanese to analyze attitudes of mothers toward travel. The question data were categorized into 15 categories and the “outings” category was used in this study. The question sentences were tokenized and the bag of words model was created from the tokenized documents. Because the data included outings close to home as well as long-distance travels, a topic model was constructed to classify the outing category data. The latent Dirichlet allocation (LDA) model [Blei et al., 2003], which is a topic model, was created using the bag of words. The topic model analysis was conducted to determine a word cloud [Halvey and Keane, 2007] and extract most critical words for each topic. Furthermore, co-occurrence network diagrams [Edmonds, 1997] were obtained for travel word.

The rest of the paper is organized as follows. Section 2 describes the data used in the study, the users of the application, and the data analysis methods. Section 3 presents the results of the analysis, and Section 4 discusses the results of the analysis. Finally, Section 5 presents the conclusion to this paper.

## 2. Methods

### 2.1 Data and users

The data contains search words, question and answer sentences, and the information of female users and their children. We used question sentences, which were submitted between January 2019 and July 2021. The question data consists of approximately 5.4 million sentences. The data provided were stripped of personal information except for the age group and the name of the prefecture in which the user resides.

The question data were categorized into 15 categories and

Table 1: Number of questioners and outings' questions by age group

Age	2019 spring		2019 summer		2020 spring		2020 summer		2021 spring	
	Questioner	Question	Questioner	Question	Questioner	Question	Questioner	Question	Questioner	Question
-20	975	8,966	852	8,956	662	8,278	636	8,425	590	7,461
21-25	6,905	80,156	6,606	76,138	5,844	75,661	5,620	75,116	5,287	67,592
26-30	14,710	143,377	13,725	132,466	11,890	128,045	11,759	126,534	10,626	115,323
31-35	16,874	151,569	15,401	136,342	13,046	127,565	12,708	126,151	11,243	111,663
36-40	6,853	55,376	6,111	48,664	4,984	44,006	4,713	42,599	4,049	37,762
41-50	1,911	16,113	1,586	13,931	1,239	12,295	1,078	11,287	870	8,965
51-	22	248	27	350	19	404	20	338	19	298
N/A	46,070	321,700	42,964	300,086	38,287	303,027	38,405	316,408	37,928	325,945
Total	94,320	777,505	87,272	716,933	75,971	699,281	74,939	706,858	70,612	675,009

the “outings” category was used. In this study, the spring period from February to May and the summer period from June to September were analyzed because the number of Japanese people who engage in overnight tourism and recreation was higher in March, May, August, and September [JTB Foundation, 2021]. Table 1 details the number of questioners and the questions by the age group in outing category questions. The questions were written in Japanese and included emojis (pictograms). A question submitted in summer 2020 is as follows:

生後2か月の子どもですが、コロナウイルスの影響でお宮参りがまだ未です。実家は大阪と宮城なので、どこでいつなどなかなか決まりません。月齢近い方々もうお済みでしょうか?? まだな方はいつ頃予定してるかも教えてください。

## 2.2 Data analysis

Natural language processing was applied to the question data written in Japanese. Python was used to remove emojis (pictograms) as a preprocessing step for morphological analysis. The

question sentences were tokenized, and the bag of words model was created from the tokenized documents. Because the data included outings close to home as well as long-distance travels, a topic model was constructed to classify the outing category data. The LDA topic model, which was one of topic model, was created using the bag of words. Using topic model analysis, we revealed the word cloud and extracted most important words for each topic model. Co-occurrence network diagrams were displayed for travel. These natural language processes were executed using Text Analysis Toolbox of MATLAB [MathWorks, 2022], a numerical simulation software.

## 3. Results

Natural language processing was used to analyze attitudes of mothers toward travel for the question data in spring and summer 2019, spring and summer 2020, and spring 2021. To clarify word relationships, only nouns and proper-nouns were considered from the questions written in Japanese to create the bag of words. An LDA topic model was generated from the bag

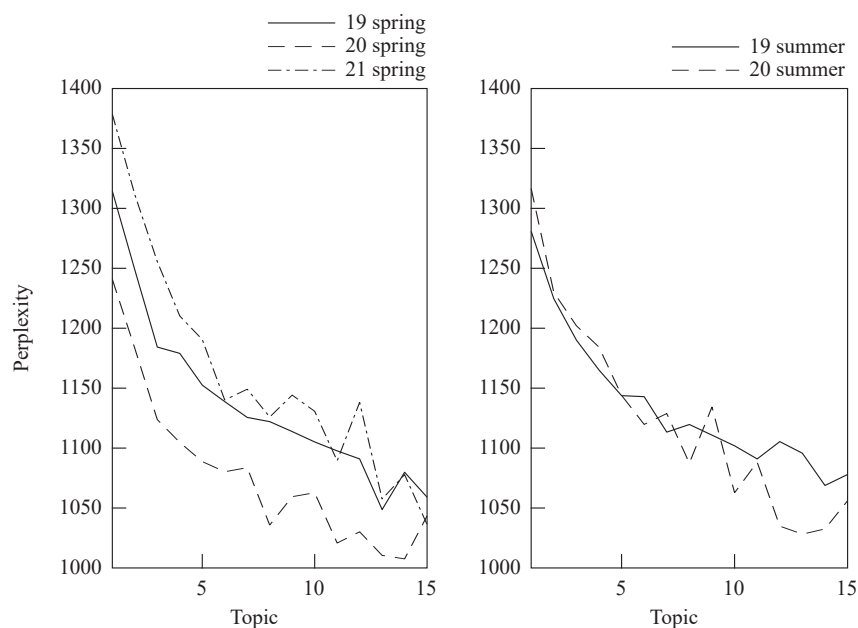


Figure 1: Relationships between perplexities and topics for each season



Table 4: Most important words for each LDA topic model in 2021

Topic	Spring
1	lunch, child, Costco, suggestion, location
2	location, railroad station, suggestion, parking, please
3	child, mama, daughter, son, friend
4	<i>Coronavirus</i> , <b>zoo</b> , <i>statement</i> , <i>emergency</i>
5	park, child, playground equipment, today, laugh
6	park, Center, support, hall, location
7	recently, weekday, status, weekend, reservation
8	<i>Coronavirus</i> , <b>travel</b> , <b>hotel</b> , plan
9	<b>car</b> , <b>parents' home</b> , <b>homecoming</b> , <b>train</b> , <b>bus</b>
10	suggestion, <b>hot spring</b> , <b>recommendation</b> , meal, meal
11	today, rain, walking, short sleeve, clothes
12	wedding, bicycle, sheet, child, car
13	<b>ticket</b> , <b>Disney</b> , <b>purchase</b> , <b>Disneyland</b> , tent
14	baby buggy, baby, sling, shopping, car
15	photo, photographing, shrine, studio, rental

not possible. Only the most important five words for each topic model are listed in Tables 2 to 4. Five words were extracted as Japanese words, but some topics had a reduced number of words in English notation.

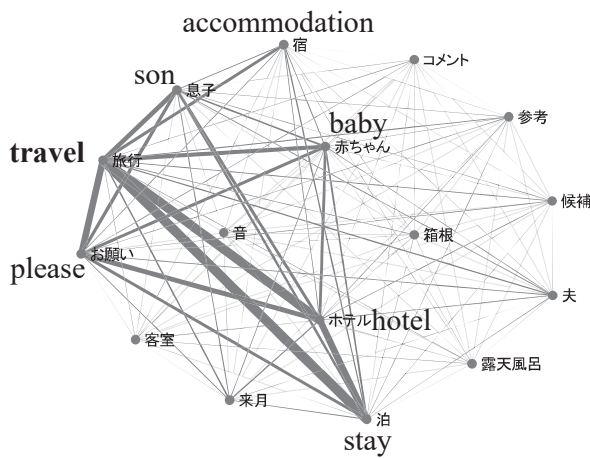


Figure 3: Co-occurrence network in Spring 2019

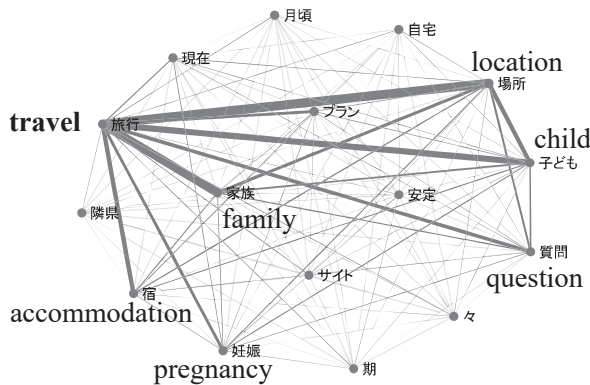


Figure 4: Co-occurrence network in Summer 2019

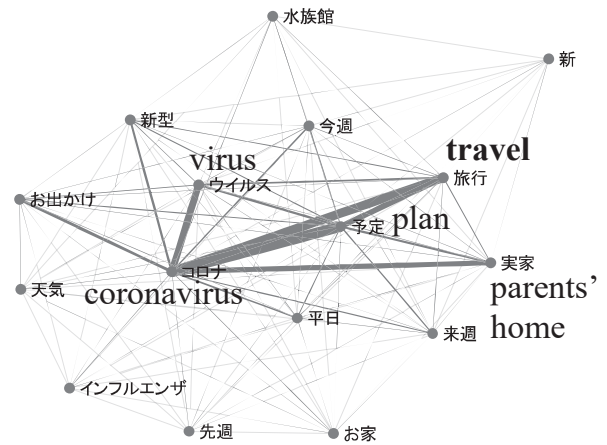


Figure 5: Co-occurrence network in Spring 2020

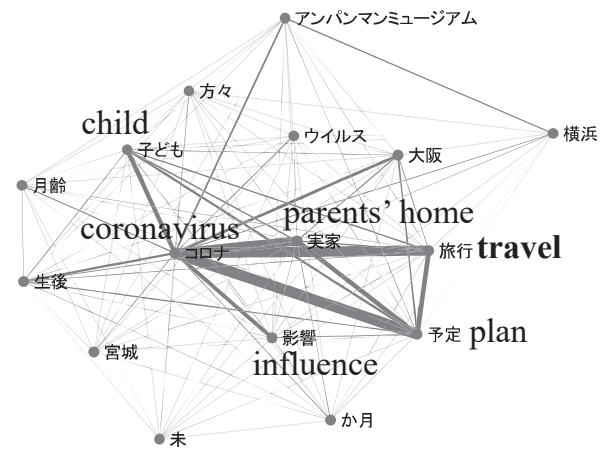


Figure 6: Co-occurrence network in Summer 2020

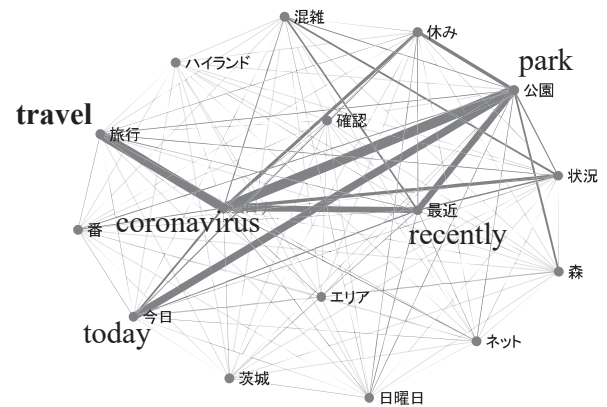


Figure 7: Co-occurrence network in Spring 2021

The co-occurrence networks were created for the bag of words to visualize the relationship between travel and the other words. In Figures 3 to 7, the thicker the line connecting the words is, the stronger the relationships are.

#### 4. Discussion

The question data for the outing category used in this study included outings close to home as well as long-distance travels. To classify the contents, we extracted the most important five words for each topic model in Tables 2 to 4. In these tables,

regular and bold font words mean outings close to home and long-distance travels, respectively, were detailed. Words with negative meaning for travel were italicized and underlined.

Table 2 details the most important five words for each LDA topic model in spring and summer 2019. In 2019, no negative words were considered for travel and mothers queried about the stay for the travel to Okinawa, parking at zoo, and reservation and hotel at Disneyland. Table 3 reveals the words in spring and summer 2020. Many negative words were present for travel because coronavirus first emerged in February 2020 in Japan. The table indicates mothers were concerned about outings during coronavirus, especially in spring, because limited information was available on coronavirus. Although few, questions regarding booking tickets to Disneyland were observed in summer. Table 4 reveals the words in spring 2021. Fewer words indicating negative sentiments for travel were observed compared to last year. Users enquired about travel cancellations because of coronavirus. This year, however, only a few mothers queried about travel sites where they could travel even with coronavirus.

Travel demand was suppressed considerably because of the Japanese government declaring COVID-19 state of emergency multiple times between 2020 and 2021, asking people to stay at home. However, tables reveal interest in travel is slowly recovering

Figures 3 to 7 detail the co-occurrence network of the relationship between travel and the other words using the same data of the previous analysis.

Figure 3 reveals that the lines from travel to hotel, and from travel to stay are thicker, which indicated strong relationships in spring 2019. Figure 4 reveals a strong relationship between travel and location, and travel and family in summer 2019 because many people returned to their parents' home in summer vacation. Figure 5 reveals a strong relationship between travel and coronavirus as well as travel and plan in spring 2020 because mothers were re-examining their travel plans because of the coronavirus outbreak. Figure 6 reveals strong relationships between travel and coronavirus as well as plan and coronavirus in summer 2020. This figure shows that mothers were concerned about coronavirus; however, their concerns were specific to their children and its effects. Figure 7 details strong relationships between travel and coronavirus as well as park and coronavirus. The increased interest in neighborhood parks is thus revealed.

The results of these analyses indicate that mothers were interested in parks near their homes so that they could visit these places easily. Because of the COVID-19 state of emergency, people were asked to stay at home, and leisure facilities were temporarily closed. However, the analysis results also revealed that interest in Disneyland and the other facilities remained strong even among COVID-19 pandemic. In the future, if a similar pathogen resulted in stay-at-home orders, online tours should be provided for infants and young children.

## 5. Conclusion

In this study, the attitudes of mothers with infants and toddlers toward travel before and after the COVID-19 pandemic were analyzed using data from mamari, a smartphone application for mothers to solve and share questions and concerns about pregnancy, childbirth, and childcare. Natural language processing was applied to the question data written in Japanese. We constructed an LDA topic model of "outings" category question data. Using the models, we extracted the most important words in the constructed topic models and conducted classification of the question data. Furthermore, we visualized attitudes of mothers toward travel by the co-occurrence network diagrams.

This analysis focused on the question data. In the future, we plan to analyze the attitudes of mothers using answers to the questions, word searches, and data on children's ages that were not used in this study.

## 6. Contribution

To analyze the attitudes of mothers toward travel, an LDA topic model was constructed for approximately seven hundred thousand Japanese questions posted between 2019 and 2021. Using the topic model, we categorized the question data, which included travel to neighborhoods and long-distance travels, into 13-15 topics. The topic model was visualized in a word cloud. The relationship between travel and other words was visualized using a co-occurrence network.

## Acknowledgements

We would like to thank Connehito Inc. for providing us with notable data and the Joint Association Study Group of Management Science for running the 2021 Data Analysis Competition. We would also like to thank Ms. Teruko Miyagawa and Ms. Ayumi Mori, who are administrative staff members of Tokai University, for valuable comments as mothers with infants and toddlers.

## References

- Abbas, J., Mubeen, R., Iorember, P. T., Raza, S., Mamirkulova, G. (2021). Exploring the impact of COVID-19 on tourism: Transformational potential and implications for a sustainable recovery of the travel and leisure industry, *Current Research in Behavioral Sciences*, Vol. 2, 100033.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, Vol. 3, 993-1022.
- Cabinet Secretariat (2022). *COVID-19 information and resources* (Retrieved September 1, 2022 from <https://corona.go.jp/en>).
- Connehito (2022a), mamari, App Store (Retrieved October 21, 2022 from <https://apps.apple.com/jp/app/ママリ-妊娠-出産アプリ育児や授乳のお悩みや情報を共有/id867414015>. (in Japanese)
- Connehito (2022b). mamari, Google Play (Retrieved Oc-



- tober 21, 2022 from <https://play.google.com/store/apps/details?id=com.connehito.mamariq&hl=ja&gl=US>). (in Japanese)
- Edmonds, P. (1997). Choosing the word most typical in context using a lexical co-occurrence network. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, 507-509.
- Halvey, M. and Keane, M. (2007). An assessment of tag presentation techniques. *Proceedings of the 16th international conference on World Wide Web*, 1313-1314.
- JTB Foundation (2021). *Annual report on the tourism trends survey 2021*. Japan Travel Bureau Foundation. (in Japanese)
- Kock, F., Nørfelt, A., Josiassen, A., Assaf, A. G., and Tsonas, M. G. (2020). Understanding the COVID-19 tourist psyche: The evolutionary tourism paradigm. *Annals of Tourism Research*, Vol. 85, 103053.
- MathWorks (2022). Text analysis toolbox user guide (R2022a) (Retrieved August 24, 2022 from [https://www.mathworks.com/help/pdf\\_doc/textanalytics/textanalytics\\_ug.pdf](https://www.mathworks.com/help/pdf_doc/textanalytics/textanalytics_ug.pdf)).
- Sigala, M. (2020). Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. *Journal of Business Research*, Vol. 117, 312-321.
- Suzuki, S., Okano, Y., and Takahashi, K. (2022), The impact of travel subsidy on individual accommodations under COVID-19 pandemic: Analyzing changes of plans listed on an online travel service. *Journal of Global Tourism Research*, Vol. 7, No. 1, 61-68.
- Yamamoto, M. (2022). Examining the changes in the transitions and characteristics of populations in Nagoya City after COVID-19. *Journal of Global Tourism Research*, Vol. 7, No. 1, 53-60.

(Received September 26, 2022; accepted October 26, 2022)