

# Q&A コミュニティの質問文を用いた 観光情報の経年的変化の分析

吉見 憲二<sup>†1</sup> 谷本 和也<sup>†2</sup> 田中 康裕<sup>†3</sup>  
岩井 憲一<sup>†4</sup> 上田 祥二<sup>†5</sup> 針尾 大嗣<sup>†6</sup>

**概要**：本研究では、国立情報学研究所が提供する「Yahoo!知恵袋データ(第3版)」の2019年度提供版と2022年度提供版のデータを利用し、国内旅行カテゴリにおける質問文から観光情報の経年的な変化について分析を行った。計量テキスト分析の手法を用いて、先行研究にデータを追加し、2014年度から2019年度の6年間の動向について検討した。加えて、先行研究との比較を通して、アフターコロナを見据えた国内観光に関する有用な知見の取得を目指した。

**キーワード**：Q&A コミュニティ、観光情報、経年的変化、計量テキスト分析

## 1. はじめに

長引くコロナ禍は国内産業に深刻なダメージを与えている。特に「鉄道旅客運送業」「道路旅客運送業」「宿泊業」「旅行業」を含む観光関連産業は、2020年に入ってから世界中に拡大した新型コロナウイルス感染症(COVID-19)の影響により、旅行需要の消失に直面している。2020年の延べ宿泊者数は2019年同月比を大きく下回っており、観光関連産業を牽引してきた訪日外国人旅行者数はコロナ禍初期には9割以上も減少するに至った[1]。

幾度かの感染拡大期を経て、ワクチン接種の普及もあり徐々に入国規制等も緩和されてきたものの、2021年の訪日外国人旅行者数は25万人と2019年の3188万人には遠く及ばない。2023年になってからも感染拡大「第8波」の只中にあるように、コロナ禍前の観光需要を回復することは容易ではない。

他方で、2020年7月から12月にわたって実施されたGoToトラベルキャンペーンは感染拡大に寄与したと批判を浴びたものの、図1が示す通り観光関連活動の活性化には一定の成果を上げているように見える。今後アフターコロナの社会において観光関連産業を復活させるためには、訪日外国人旅行者への依存が難しい中で、日本人国内旅行需要の喚起を行っていくことが欠かせないだろう。

しかしながら、コロナ禍以前より国内旅行のスタイルには大きな変化が生じていた。図2に示す通り、国内旅行に占めるパック・団体旅行の比率が2010年から2020年にかけて右肩下がりに減少し、個人旅行の占める比率が伸びてきている。そのため、個人のニーズに合致した観光資源を提供する重

要性がますます高まっていると考えられる。

本研究では、こうした観光関連産業が抱える課題に対し、Q&A コミュニティの質問文を用いた分析から有用な知見を提供することを目的とする。

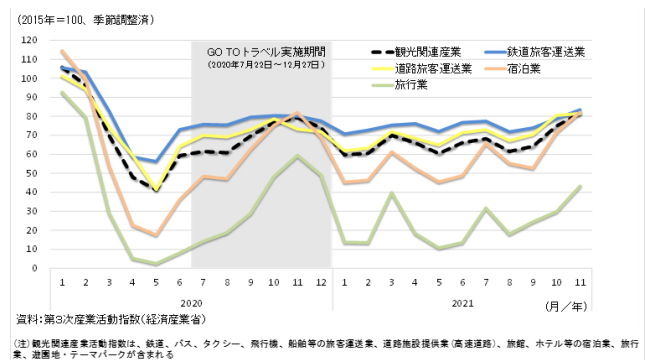


図1 観光関連産業活動指数の動向

(出典) 経済産業省

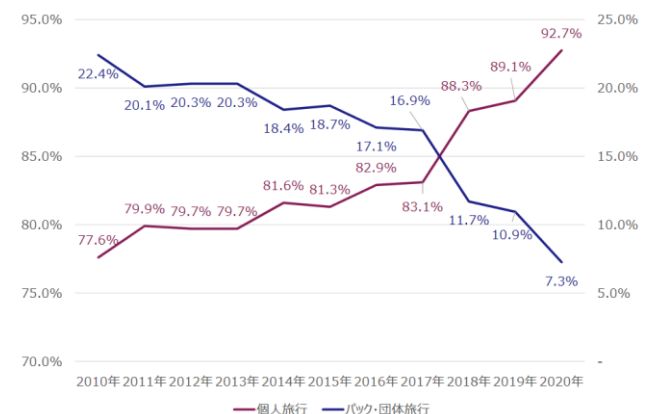


図2 国内旅行に占める個人旅行と  
パック・団体旅行の比率

(出典) 観光庁

†1 成蹊大学 Seikei University  
†2 佛敎大学 Bukkyo University  
†3 社会データ構造化センター Center for Social Data Structuring

†4 滋賀大学 Siga University  
†5 株式会社セールスフォースドットコム salesforce.com, Inc  
†6 摂南大学 Setsunan University

## 2. 分析内容

### 2.1 分析に用いるデータ

本研究では、国立情報学研究所が提供する「Yahoo!知恵袋データ(第3版)」の2019年度提供版および2022年度提供版における質問文を利用した[2]。Yahoo!知恵袋は日本最大のQ&Aコミュニティであり、多くのユーザーから質問と回答が寄せられている。

2019年度提供版における質問数は約264万件、回答数は約611万件となっている。収録データは収録期間(2014年4月1日-2017年3月31日)に投稿され解決した質問の10%がランダムサンプリングされたものである。

2022年度提供版における質問数は約247万件、回答数は約649万件となっている。収録データは収録期間(2017年4月1日-2020年3月31日)に投稿され解決した質問の10%がランダムサンプリングされたものである。

上記より、2019年度提供版および2022年度提供版を用いることで、2014年度から2019年度にかけて6年分の分析が可能となる。

### 2.2 Q&A コミュニティの質問文を用いる意義

前述の通り、日本人旅行者の観光需要を調べるためには、日本語で発信された情報にアプローチすることが効果的だと考えられる。ソーシャルメディアの投稿は大量のデータにアクセスできる点で優位性があるが、多くのノイズが含まれるため観光情報に限定して取得することは困難を伴ってしまう。

その一方で、Q&A コミュニティにおける質問文は投稿件数の制約はあるものの、カテゴリ化されている点で精度の高いデータを取得することが可能である。さらに、質問文の形式を取っていることから、質問者にとって関心の高いテーマを扱っていることが予想される。そのため、より潜在的な観光需要に関連した情報にアクセスすることが期待できる。

### 2.3 先行研究

著者らは本研究と同様に「Yahoo!知恵袋データ(第3版)」を用いた研究を実施し、国内旅行カテゴリにおける質問文の分析から「北海道」「京都」「沖縄」それぞれの地域に特有の観光客の関心や懸念事項を明らかにした[3]。さらに、2014年度から2017年度までの経年的な変化についても分析を行い、「神社」のような用語において特徴的な登場傾向が見られた[4]。

しかしながら、前者の研究では経年的な変化については検討できておらず、後者の研究では年度とい

う枠組みのみであったため、例えばシーズンにおける傾向等については十分に議論することができなかった。

図2では2017年から2018年にかけて個人旅行の比率が急激に高まっていることから、2018年以降のデータを追加することで新たな知見が得られるかもしれない。そのため、本研究では先行研究を踏まえた分析手法を採用し、以降のデータを追加することで、結果の比較を行うこととする。

### 2.4 分析対象

本研究では、先行研究と同様にYahoo!知恵袋における国内旅行カテゴリの質問文を分析対象とし、2014年4月から2020年3月までの28,822件を取得した。これらの質問文を各年度の上半期と下半期でそれぞれグルーピングし、分析に用いた。

なお、分析にあたっては閾値として1,200回以上の登場回数を設定した。これは全体の4%強に該当し、各質問文に対しての登場の有無でカウントしている。

### 2.5 分析手法

分析に当たってはフリーのテキストマイニングソフトウェアであるKH Coder (<https://khcoder.net/>)を使用した。形態素解析は付属のChaSen(茶筌)を用い、先行研究と同様に一部観光施設名等で辞書に登録されていないものについて強制抽出の対象とした。

分析の手順として、最初に経年的な傾向を対応(コレスポネンス)分析で可視化し、特徴的な単語を明らかにした。

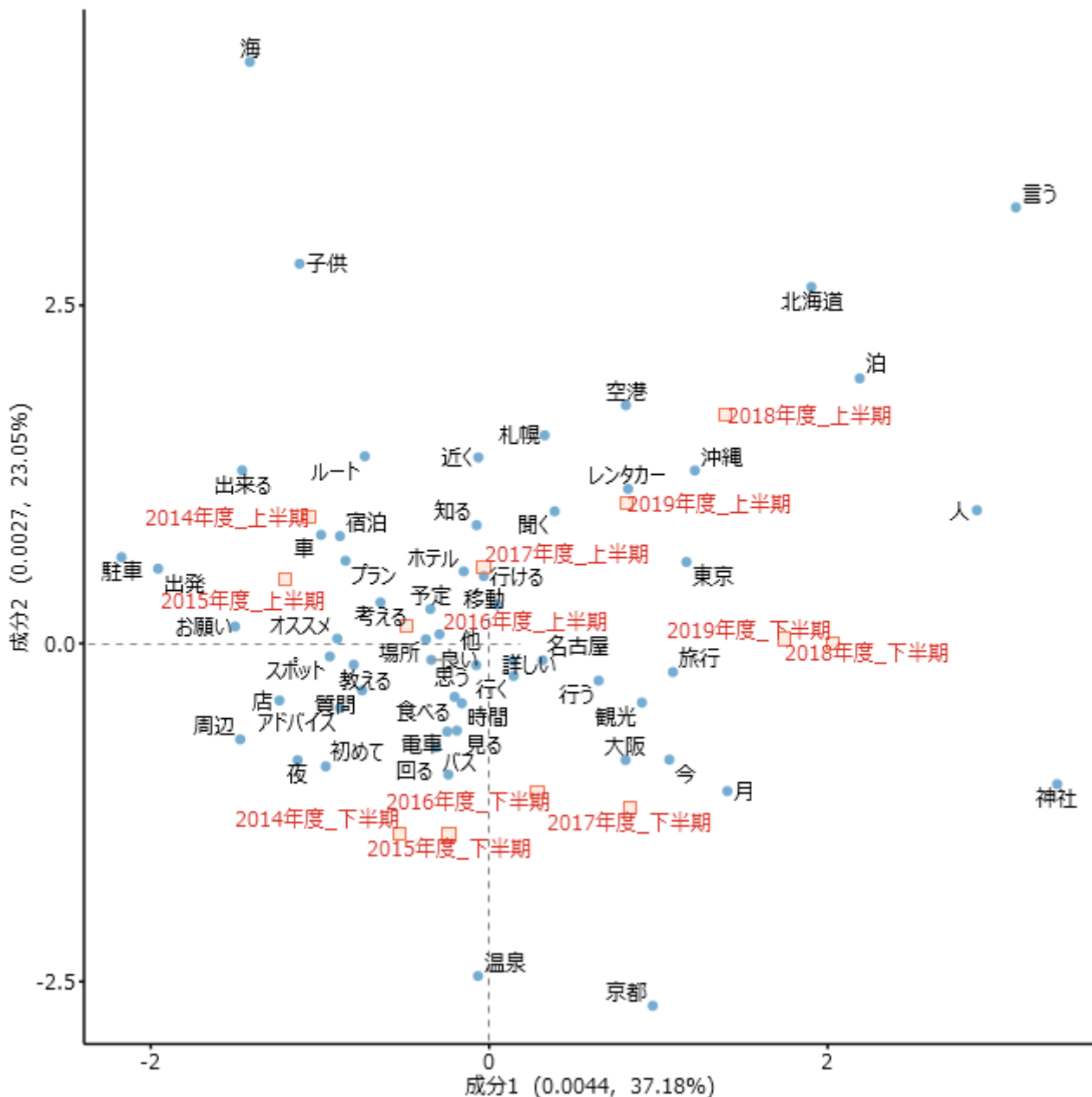
次に、先行研究の対象である「北海道」「京都」「沖縄」と特徴的な単語について経年的な登場傾向を検討した。

最後に、「北海道」「京都」「沖縄」と共起頻度が高い単語に対して、先行研究との結果の比較を行った。なお、参考までにJaccard係数0.1以上の共起関係を基準とした共起ネットワーク図についても描写する。

## 3. 分析結果

### 3.1 対応(コレスポネンス)分析

対応分析の結果は図3の通りである。上半期と下半期に各年度を分割したことから、先行研究と比べてよりシーズンを反映した結果となっている。縦軸は上半期と下半期の差異を示しており、「海」や「温泉」といった季節を反映した単語が目立っていた。横軸は時間的な経過を示しており、概ね2014年から2020年にかけて使用する単語の傾向が変化してきたことがうかがえる。



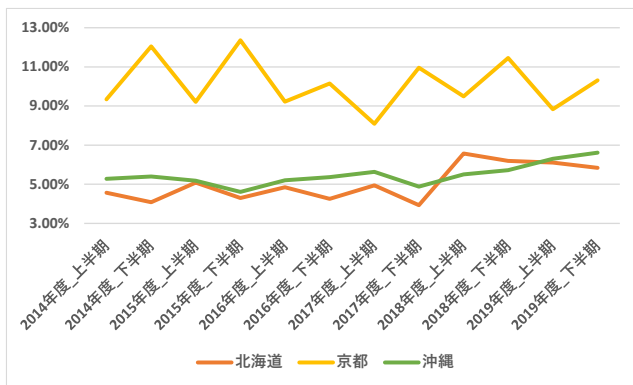


図4 地域名ごとの登場傾向

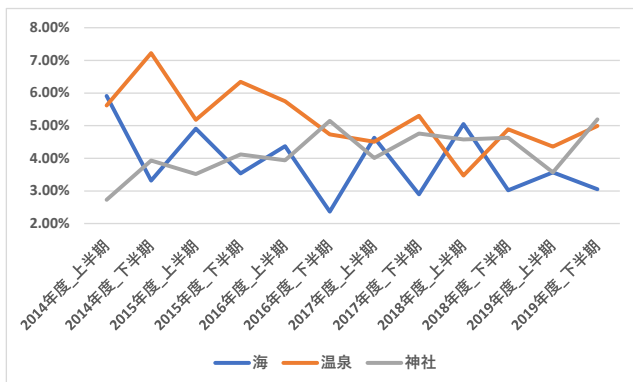


図5 特徴的な単語ごとの登場傾向

### 3.3 各地域の経年比較

最後に先行研究で検討した各地域(「北海道」「京都」「沖縄)の傾向について2017年度以降のデータを用いた分析と先行研究との比較を行った。

結果については、「北海道」「京都」「沖縄」の順番に以下に示す。

#### (1) 北海道の観光情報

2017年度以降のデータにおける「北海道」を含む関連語の共起ネットワーク分析の結果は図6の通りとなった。「札幌」「小樽」「富良野」「旭川」「美瑛」といった地名や「新千歳空港」「旭山動物園」といった施設名が登場している点は先行研究と同様であった。しかしながら、「帯広」「釧路」「知床」「洞爺湖」といった地名は今回登場しなかった。

「レンタカー」「移動」の共起関係や「美味しい」「海鮮」「食べる」という共起関係も先行研究と一致しており、観光における興味は継続しているようであった。

Jaccard係数上位の観光用語は表1に示しており、表2が先行研究の結果である。順位の変動こそあるものの、大きな変化はないようであった。ただし、全体的に共起の強さ(Jaccard係数)は落ちていた。

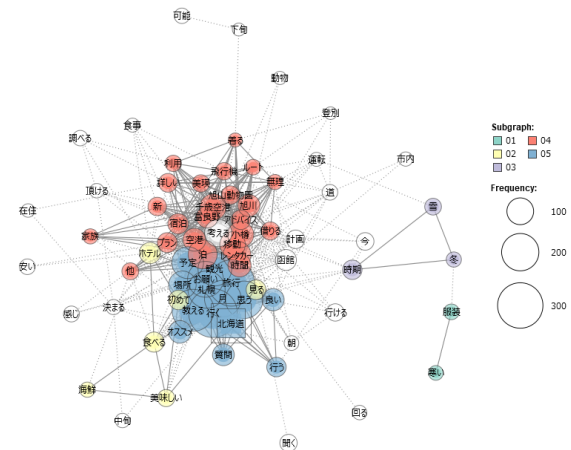


図6 「北海道」の共起ネットワーク(2017-2019)

表1 Jaccard係数上位の用語(北海道2017-2019)

	観光用語	Jaccard係数
1	札幌	0.2383
2	千歳空港	0.1373
3	小樽	0.1342
4	旅行	0.1082
5	旭川	0.0954
6	富良野	0.0923
7	レンタカー	0.0919
8	函館	0.0815
9	空港	0.0748
10	旭山動物園	0.0667

表2 Jaccard係数上位の用語(北海道2014-2016)

	観光用語	Jaccard係数
1	札幌	0.2605
2	小樽	0.1782
3	千歳空港	0.1522
4	レンタカー	0.1234
5	富良野	0.1137
6	旭川	0.1134
7	函館	0.1046
8	旅行	0.0912
9	旭山動物園	0.0853
10	飛行機	0.0801

#### (2) 京都の観光情報

2017年度以降のデータにおける「京都」を含む関連語の共起ネットワーク分析の結果は図7の通りとなった。「金閣寺」「清水寺」「伏見稲荷」「嵐山」といった定番の観光地が他の多くの単語と共起している点は先行研究と同様であった。さらに、「ルート」「出

発「効率」や「日帰り」「時間」といった時間の制約に関する共起関係も類似した結果となった。ただし、「銀閣寺」「二条城」は今回には登場しなかった。

Jaccard 係数上位の観光用語は表 3 に示しており、表 4 が先行研究の結果である。こちらは大きな変化は見られず、やや順位の変動がある程度であった。

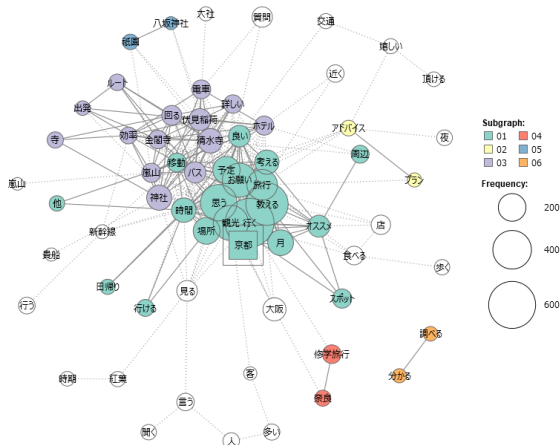


図 7 「京都」の共起ネットワーク (2017-2019)

表 3 Jaccard 係数上位の用語 (京都 2017-2019)

	観光用語	Jaccard 係数
1	清水寺	0.1256
2	神社	0.1014
3	伏見稲荷	0.1007
4	観光	0.0974
5	嵐山	0.0761
6	大阪	0.0718
7	旅行	0.0713
8	金閣寺	0.0707
9	バス	0.0656
10	修学旅行	0.0627

表 4 Jaccard 係数上位の用語 (京都 2014-2016)

	観光用語	Jaccard 係数
1	清水寺	0.1853
2	嵐山	0.107
3	伏見稲荷	0.1046
4	観光	0.0985
5	金閣寺	0.0942
6	バス	0.0916
7	大阪	0.0892
8	神社	0.0831
9	祇園	0.0667
10	修学旅行	0.0667

### (3) 沖縄の観光情報

2017 年度以降のデータにおける「沖縄」を含む関連語の共起ネットワーク分析の結果は図 8 の通りとなった。「美ら海水族館」「国際通り」「首里城」という観光名所や「本島」「離島」「石垣島」「古宇利島」といった島に関する用語が見られる点は先行研究と一致していた。ただし、「シュノーケリング」や「ダイビング」といったアクティビティは今回の結果では見られなかった。

Jaccard 係数上位の観光用語は表 5 に示しており、表 6 が先行研究の結果である。「首里城」と「ビーチ」が共起語の上位から落ちている点は特徴的であった。ただし、「首里城」は 2019 年 10 月に火災で焼失したため、その影響があったことが考えられる。

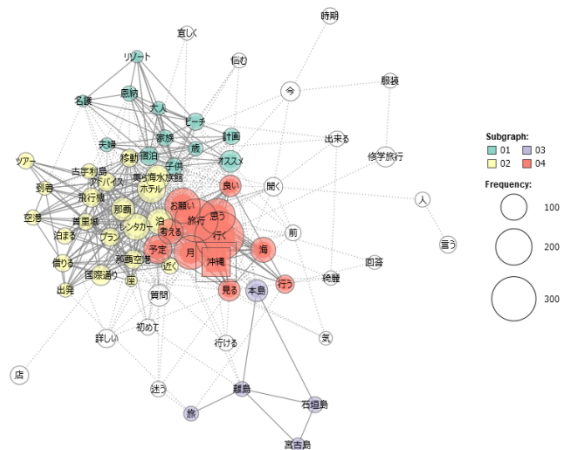


図 8 「沖縄」の共起ネットワーク (2017-2019)

表 5 Jaccard 係数上位の用語 (沖縄 2017-2019)

	観光用語	Jaccard 係数
1	美ら海水族館	0.1211
2	那覇	0.1114
3	本島	0.0982
4	旅行	0.0968
5	海	0.0925
6	レンタカー	0.0874
7	国際通り	0.0853
8	ホテル	0.0795
9	那覇空港	0.0721
10	修学旅行	0.0613

表 6 Jaccard 係数上位の用語 (沖縄 2014-2016)

	観光用語	Jaccard 係数
1	那覇	0.1689
2	美ら海水族館	0.1296

3	国際通り	0.1039
4	本島	0.103
5	ビーチ	0.0951
6	旅行	0.0947
7	海	0.0922
8	レンタカー	0.091
9	ホテル	0.0847
10	首里城	0.0797

### 3.4 考察

これまでに見てきた分析結果では、2017 年度以降のデータを追加したことによる大きな変化は見られなかった。これは頻出上位に登場するような観光資源については堅調な観光需要があると捉えることができるかもしれない。

他方で、大きな流れとして「海」に関する情報の登場量の低下が示唆されている。図3では「海」が左上に位置しており、経年的な変化において出現傾向が過去に偏っていることを意味している。その証左として、沖縄における共起語から「シュノーケリング」や「ダイビング」、「ビーチ」といった語句が消えてしまっている。これが直ちに需要の減退を示すわけではないが、長期的なデータにおける変化の兆しとして注意する必要があるだろう。

また、先行研究に加えて上半期と下半期の区分を行ったことにより、対応（コレスポネンス）分析の軸の解釈をより明確にすることができた。こうした結果を踏まえて、シーズンによる変動が大きい観光地や観光用語においては、そのギャップを埋めることで新たな観光ニーズにつなげることが期待される。例えば、京都では冬のシーズン（下半期）のニーズが高いことが示されているが、貴船の川床や保津川下りといった夏場でも楽しめる観光情報を強調することで、新たな観光客の需要が喚起できるかもしれない。そうした観点では、季節の変動が少ない「神社」やパワースポットのような打ち出し方が参考になるだろう。

なお、分析対象の各地域（「北海道」「京都」「沖縄」）では共起上位語の Jaccard 係数が先行研究と比べて軒並み低下することとなった。これは個人旅行へのシフトによって、より個人的な興味・関心を旅行動機にしている可能性として解釈することができるかもしれない。本研究の結果からは十分に検討することはできないが、そうした興味・関心の分散を前提とすれば、むしろ頻出下位後の経年的変化の方が重要になるものと考えられる。

## 4. おわりに

本研究では、国立情報学研究所が提供する「Yahoo! 知恵袋データ(第3版)」の2019年度提供版と2022年度提供版とデータを利用し、国内旅行カテゴリにおける質問文から観光情報の経年的な変化について分析を行った。先行研究との比較を通して、大きな変化は見られなかったものの、「海」に関連した単語のニーズの提言が示唆される結果となった。加えて、シーズンにおける質問量の変動に着目することで、観光需要の喚起の可能性が示された。

ただし、当初の目的であるアフターコロナを見据えた有用な知見の取得に関しては、成果は限定的であった。考察でも触れたように、本研究の結果からはむしろ頻出下位語の経年的変化に着目することが有効かもしれない。こうした課題については、今後の研究で引き続き検討していきたい。

## 謝辞

本研究は ROIS-DS-JOINT 2022(課題番号: 028RP2022, 研究代表者: 針尾大嗣)の助成を受けた。また、本研究では国立情報学研究所の IDR データセット提供サービスによりヤフー株式会社から提供を受けた「Yahoo! 知恵袋データ(第3版)」を利用した。

## 参考文献

- [1] 観光庁 (2021) 「観光を取り巻く現状及び課題等について」(アフターコロナ時代における地域活性化と観光産業に関する検討会資料)。
- [2] ヤフー株式会社 (2018) Yahoo! 知恵袋データ (第3版)。国立情報学研究所情報学研究データリポジトリ。(データセット)。 <https://doi.org/10.32130/idr.1.3> (2022/1/20 閲覧)。
- [3] 吉見憲二 (2020) 「Q&A コミュニティにおける質問文からの観光情報の分析」『情報処理学会研究報告電子化知的財産・社会基盤 (EIP) , 2020-EIP-87(11)』 pp. 1-4。
- [4] 吉見憲二 (2020) 「Q&A コミュニティにおける投稿内容の経年的変化」『第19回情報科学技術フォーラム講演論文集 (FIT)』 第4分冊, pp. 157-158。