

わが国の成人のスポーツライフに関する テキストマイニングとヴィジュアル分析

～SSF「スポーツライフ・データ2022」における自由回答を対象として～

Text Mining and Visual Analysis of Japanese Adult Sports Life Based on the
Open-Ended Answers of the “Passion for Sports” and the “Promotion
and Development of Sports” in the SSF (Sasakawa Sports Foundation)
National Sports-Life Survey 2022.

横山 文人*

YOKOYAMA, Fumito

Abstract

This research paper describes text mining results of Japanese adult sports life based on the open-ended answers of the national sports-life survey 2022 conducted by the Sasakawa sports foundation in 2021. The primary survey data (n=3,000) was collected by using a multistage sampling method with a population of 3,000 men and women aged 18 or older residing in Japan, and is conducted by the door-to-door retention survey method, between June 10th and July 10th 2021. Text mining was applied to the open-ended answers of the “passion for sports” and the “promotion and development of sports” in the survey data. The results were as follows :

- 1) The relative importance of the exercise and health-related factor of “body,” “health,” and “exercise” was the greatest for men and women overall.
- 2) The human factors of “person,” “self,” and “children,” the “time” factor, and the behavioral factors of “want,” “can,” and “continue” were relatively important for both men and women.
- 3) The environmental elements of “facilities” and “environment” were also identified with a certain degree of importance.
- 4) Men’s keywords are characterized by the greatest relative importance of “want” and “health” and perceive the environmental factors to be more important than women.
- 5) The key word “body” is perceived to be outstandingly important as a characteristic of women’s keywords, and the environmental factors are perceived to be less important than for men.

Key Words

Text Mining, Visual Analysis, Adult Sports Life, National Sports-Life Survey 2022, Word Clouds

キーワード

テキストマイニング, ヴィジュアル分析, 成人のスポーツライフ, スポーツライフに関する全国調査2022年,
ワードクラウド

* 本学経営学部 ホスピタリティ・マネジメント学科 准教授

- 本成果は、令和5年度亜細亜大学特別研究奨励制度助成を受け進められた研究成果の一部である。ここに記して感謝の意を表明する。
- 本論文の作成にあたり、「スポーツライフデータ2022」の使用において特別の配慮を賜りました公益財団法人笹川スポーツ財団に深謝の意を表します。

目次

- I はじめに
 - II 研究方法
 - III 結果と考察
 - IV 結論
-

I はじめに

わが国のスポーツの振興や発展、スポーツに関する施策の総合的な推進を図ることを目的として設置されたスポーツ庁は、国民のスポーツライフ政策として、次のように提言している（スポーツ庁、2023）。

スポーツ庁では、第3期スポーツ基本計画において、成人の週1回以上のスポーツ実施率が70%程度となることを目標としています。

国民のスポーツ参加を促進する取組の充実を通じて、国民の誰もがいつでも、どこでも、いつまでもスポーツに親しむことができる社会の実現を目指します。

また、第3期スポーツ基本計画（文部科学省、2022）の中の「4. 第3期計画を支える様々な主体に期待される役割とそれに対する支援」において、以下のように記述している。

第2期計画でも示したように、基本計画は、国の施策を中心に国が定めるものであるが、飽くまでも「スポーツの主役は国民」であり、また、国民にその機会を提供する地方公共団体やスポーツ団体、民間事業者等が主役である。国としては、各主体が「主役」としてスポーツに参画し、そして最大限その価値を体感できるように、様々な取組・施策を図っていく必要がある。

すなわち、地方公共団体、スポーツ団体や民間事業者等には期待される役割と支援を示すとともに、計画を実現するための「主役は国民」であることを明言している。このことから、スポーツ庁の政策目標を実現し第3期スポーツ基本計画を達成していくためには、「主役である国民」からも生の声を聴く必要があると考えられる。

そこで、本稿では、公益財団法人笹川スポーツ財団（SSF）が実施した最新の全国調査「スポーツライフに関する調査2022」における『スポーツに対する思い』や『スポーツの普及や発展』に対する自由回答を分析対象として、成人のスポーツライフにおける重要なキーワードを探索的に抽出するためにテキストマイニングをおこなうとともに、その結果を視覚化するためヴィジュアル分析をおこなった。当該全国調査は、2021年6月10日から7月10日の調査期間において、日本国に居住する満18歳以上の男女を母集団として、多段標本抽出法により3,000人を調査対象とし訪問留置調査法により実施している。昨今はサンプル数を多くすることを目的の一つとし、簡便性と迅速性を優先するあまり、ウェブ調査によるデータが多い中、確立されたサンプリング理論に基づいた従来からの調査方法により収集されたデータは、質的にも貴重な調査データと認識されよう。なお、当該調査の自由回答データを対象とした分析は、公益財団法人笹川スポーツ財団の特別許可の下で実施したものであることを付記しておく。

よって、本研究の目的は、日本における成人男女から発せられるスポーツの振興や思いに対する「生のメッセージ（自由回答）」を対象として分析することにより、スポーツライフにおける重要なキーワードについて探索的に明らかにすることとする。

II 研究方法

ここでは、本稿で用いられた研究方法について説明していく。

1. 分析データ：分析対象となるデータは、公益財団法人笹川スポーツ財団（SSF）が実施した最新の全国調査「スポーツライフに関する調査2022」における『スポーツに対する思い』や『スポーツの普及や発展』に対する自由回答データである。回収数をみると3,000票で、その内訳は男性1,503票、女性1,497票であった。この3,000票の中で自由回答欄に回答を記入した票において、「無回答」や「とくになし」といった分析対象とはならない回答を除いた分析対象とする有効回答数は1,180票で（有効回答率39.3%）、その内訳は男性543票（36.1%）、女性637票（42.6%）であった。
2. 分析ツール：当該有効回答1,180票を分析対象とし、統計解析プログラミング言語の一つである R 言語をベースとした UI（User Interface）ツールの Exploratory を用いて分析をおこなった。分析手順は、対象データについてデータラングリングを（生データのクレンジング、リストラ、および強化を行うこと（Any Connector, 2021））実施し、クリーンなデータにした後、テキストマイニングを適用した。
3. 分析方法：テキストマイニングの初期手順である形態素解析（日本語文章を最小単位である品詞に分解し、その意味を分析する手法）を実施した。次に、抽出された単語の出現頻度について分析した。その際には、ストップワードと呼ばれる非語（句読点などの区切り記号類）、非内容語（助詞や助動詞などの品詞）、および低内容語（内容推測に貢献しない語）は分析から除去した（國府ら、2013）。さらに、自由回

答の設問に含まれる単語の「スポーツ」と「思い」については出現頻度は高いが、設問文に含まれる単語であり自明であることから分析から除外した。

次に、この形態素解析の結果に基づいて頻度分析をおこない、抽出されたキーワードの出現頻度を男女別に、その内容を分析した。

さらに、頻度分析の結果を視覚化するためにヴィジュアル分析を実施し、グラフ化およびワードクラウドを適用した。ワードクラウドとは、文章やテキストから単語の出現頻度にあわせて文字の大きさを変えて視覚化したグラフのことである（AI Academy Media, 2023）。例えば、出現頻度の多いキーワードは大きなフォントで表示され、少ないキーワードは小さなフォントで表示される。フォントの大きさで、キーワードの相対的な重要性を視覚的に把握することが可能となる。

III 結果と考察

ここでは、「スポーツライフに関する調査2022」における『スポーツに対する思い』や『スポーツの普及や発展』に対する自由回答を分析対象データとしてテキストマイニングをおこなった分析結果および考察について記述する。

1) 男女全体からみた頻度分析

分析対象データに形態素解析をおこない、その解析結果を男女別に頻度分析したものを表1の男女別頻度分析結果（上位25以上）に示す。

表1において、男性と女性の出現頻度を合計した頻度合計に着目する。最も出現頻度が多かった単語は「身体」の252であった。次に「健康」の234、「運動」の211となり、これら3つの単語だけが出現頻度200を超えている。第4位の「欲しい」は167で、第3位の「運動」とは頻出頻度に

表1 男女別頻度分析結果
(上位25以上；頻度合計でソート)

順位	出現単語	男性	女性	頻度合計
1	身体	73	179	252
2	健康	86	148	234
3	運動	56	155	211
4	欲しい	87	80	167
5	人	58	95	153
6	良い	47	99	146
7	自分	36	103	139
8	時間	50	81	131
9	できる	47	71	118
10	続	46	70	116
11	もっと	53	55	108
12	動かす	26	70	96
13	子供	34	61	95
14	好き	31	61	92
15	コロナ	27	60	87
16	行	36	50	86
17	体力	24	62	86
18	見	31	49	80
19	施設	37	36	73
20	今	23	49	72
21	環境	33	35	68
22	必要	31	36	67
23	楽	20	46	66
24	参加	20	45	65
25	観戦	21	42	63

において44の開きがあることから、上位3位までの出現単語は、それ以下のものと比較すると突出して出現回数が多いと考えられる。換言すると、「身体」「健康」「運動」という運動・健康関連要素の相対的重要性が際立っているとも言えよう。

以下、第5位「人」の153、第6位「良い」の146、第7位「自分」の139、第8位「時間」の131、第9位「できる」の118、第10位「続」116、第11位「もっと」の108、までが出現頻度100以上であった。「人」や「自分」という人的要素、「時間」要素、「できる」や「続」という行動的要素が、全体として相対的に重要であると認識されていると考えられる。

第12位以下を概観すると次のように要約される。

- ・第12位「動かす(96)」, 第16位「行(86)」, 第18位「見(80)」, 第24位「参加(65)」, 第25位「観戦(63)」という行動的要素がみられた。
- ・第13位「子供(95)」という人的要素がみられた。
- ・第17位「体力(86)」という運動・健康関連要素

男女全体の単語出現頻度 上位25

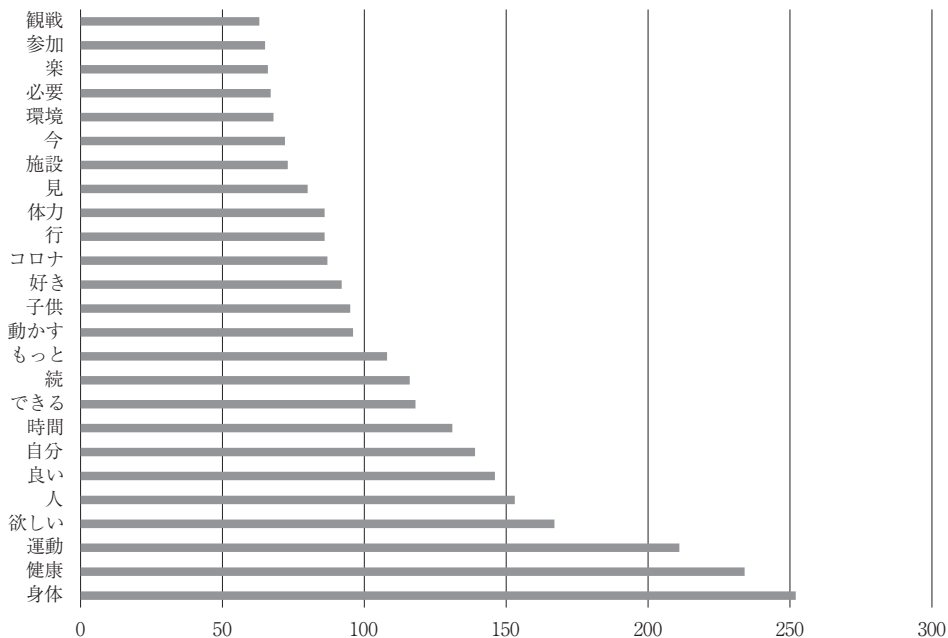


図1 男女全体の頻度分析結果
(上位25以上；頻度合計でソート)

がみられた。

・第19位「施設(73)」や第21位「環境(68)」という環境要素がみられた。

次に、表1の結果を視覚化したものが図1のグラフである。

この図1をみると、表1の結果に基づいて上述したことが視覚化されることで、より鮮明に理解しやすくなると考えられる。例えば、上位3位までの運動・健康関連要素を他の出現単語と比較すると、頻出数が顕著であることが容易に読み取れる。

また、図1全体を概観すると、単語の出現頻度の分布は最大値から直線的に減少するのではなく、指数関数状に減少することがわかる。これは、多くの自然現象や社会現象においてもみられる冪乗則に従っているように推察されよう。

さらに、ワードクラウドを適用してヴィジュアル分析をおこなったものを図2に示す。

この図2をみると、図1と同様に、表1の結果に基づいて上述したことが視覚化されることで、より明解に理解しやすくなることがわかる。例えば、中央に位置し同色で最も大きなフォントの「身体」と「健康」の出現頻度が多く、次に、ほぼ同じ大きさのフォントであるが濃度が異なる「運動」が中央付近に位置することから、これら3つのキーワードの重要性が相対的に高いことが容易に読み取れる。

2) 男性からみた頻度分析

表1の男女別頻度分析結果（上位25以上）を男性の出現頻度の高い順に並べ替えたものを表2に示す。

男性の上位5位までの出現単語は、順位は異なるが、表1の男女合計の上位5位までと全く同じであることは興味深い。男性の第1位「欲しい(87)」は男女合計では第4位であったが、第2位

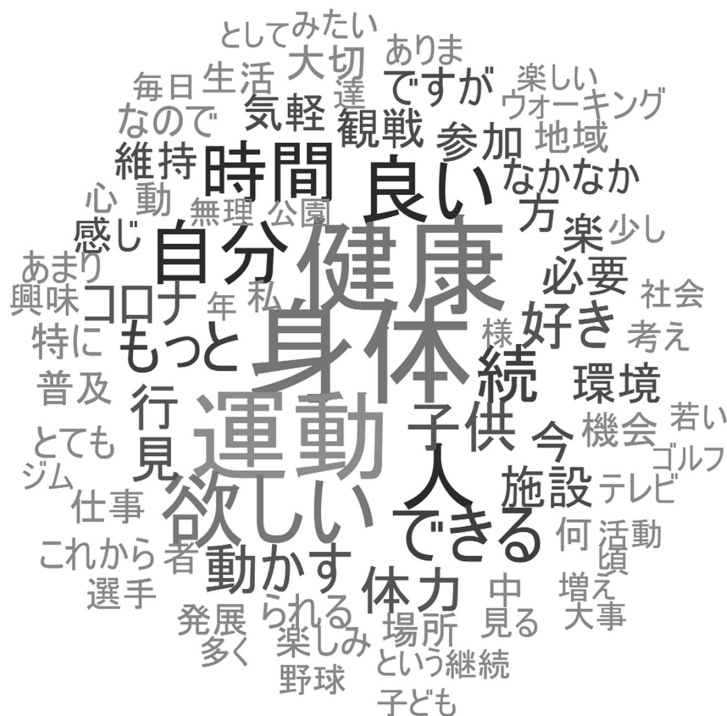


図2 ワードクラウド：「スポーツへの思いや意見」に関する自由回答
(回答者数 男女1180；「スポーツ」「思い」を除く回答頻度30以上)

の「健康(86)」は男女合計と同じ第2位であった。第3位の「身体(73)」は男女合計では第1位であった。第4位の「人(58)」は男女合計では第5位であった。第5位の「運動(56)」は男女合計では第3位であった。このことから、男性にとって重要と認識されている単語の第5位までは、男女合計のものと同じで、運動・健康関連要素、人的要素、行動的要素となっている。

第6位以降について、男女全体の順位と顕著に異なる点は、第11位の「施設(37)」(男女全体では第19位)と第15位の「環境(33)」(男女全体では第21位)という環境要素の順位が上位に位置するということである。すなわち、男性は環境要素について、より重要性が高いと認識しているように推察される。

次に、表2の結果を視覚化したものが図3のグラフである。

この図3をみると、表2の結果に基づいて上述

表2 男女別頻度分析結果

(上位25以上；男性の頻度でソート)

合計順位	男性順位	出現単語	男性	女性	頻度合計
4	1	欲しい	87	80	167
2	2	健康	86	148	234
1	3	身体	73	179	252
5	4	人	58	95	153
3	5	運動	56	155	211
11	6	もっと	53	55	108
8	7	時間	50	81	131
6	8	良い	47	99	146
9	9	できる	47	71	118
10	10	続	46	70	116
19	11	施設	37	36	73
7	12	自分	36	103	139
16	13	行	36	50	86
13	14	子供	34	61	95
21	15	環境	33	35	68
14	16	好き	31	61	92
18	17	見	31	49	80
22	18	必要	31	36	67
15	19	コロナ	27	60	87
12	20	動かす	26	70	96
17	21	体力	24	62	86
20	22	今	23	49	72
25	23	観戦	21	42	63
23	24	楽	20	46	66
24	25	参加	20	45	65

男性の単語出現頻度 上位25

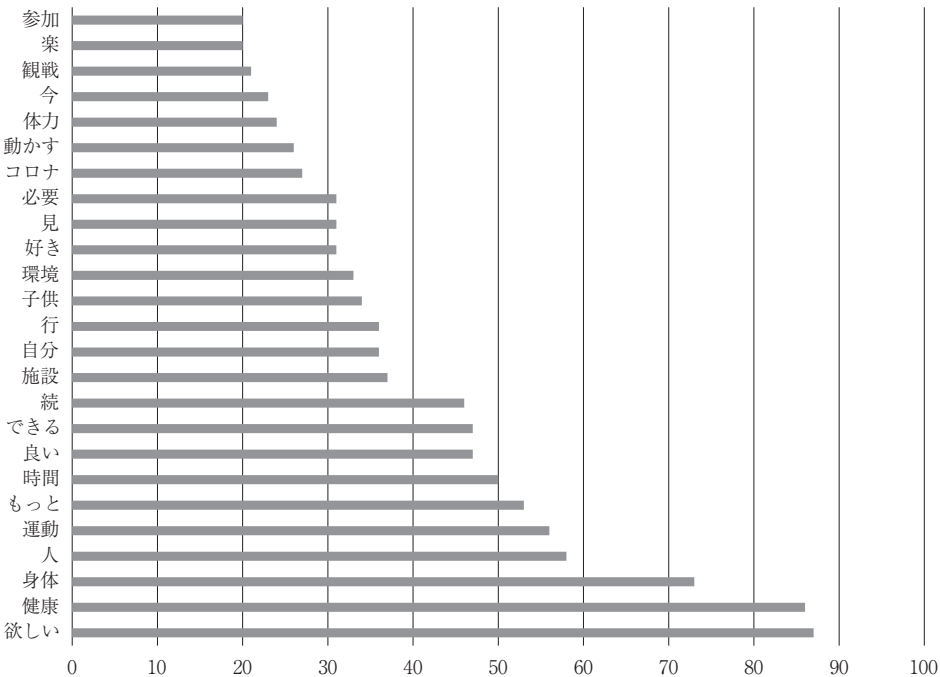


図3 男性の頻度分析結果
(上位25以上；頻度合計でソート)

したことが視覚化されることで、より明確に理解しやすくなっている。例えば、上位3位までの運動・健康関連要素と行動的要素が、他の出現単語と比較すると、頻出数が顕著であることが容易に読み取れる。さらに、図1（男女全体）と単語出現頻度の順位を比較すると、環境要素が図1では下位（第19・21位）に位置しているが、図3では中位（第11・15位）に位置することも読み取りやすくなっている。

また、図3全体を概観すると、単語の出現頻度の分布が、図1と同様に指数関数的に減少することが認識でき、冪乗則に従っているのではと推察されよう。

さらに、ワードクラウドを適用してヴィジュアル分析をおこなったものを図4に示す。

この図4をみると、図3と同様に、表2の結果に基づいて上述したことが視覚化されることにより、より鮮明でインパクトのある形で理解しやすい。

くなることが認識できる。例えば、中央に位置し最も大きなフォントである「欲しい」と「健康」の出現頻度が多く、次に、ほぼ同じ大きさのフォントであるが、濃度が異なる「身体」が中央付近に位置することから、これら3つのキーワードの重要性が相対的に高いことが容易に理解できる。

3) 女性からみた頻度分析

表1の男女別頻度分析結果（上位25以上）を女性の出現頻度の高い順に並べ替えたものを表3に表す。

女性の上位3位までの出現単語は、順位が若干異なるが、表1の男女合計の上位3位までと全く同じということは特徴的であると考えられる。女性の第1位「身体(179)」は男女合計でも同位であったが、第2位の「運動(155)」は男女合計では第3位で、第3位の「健康(148)」は男女合計では第2位であった。上位3位の中でも、第1位

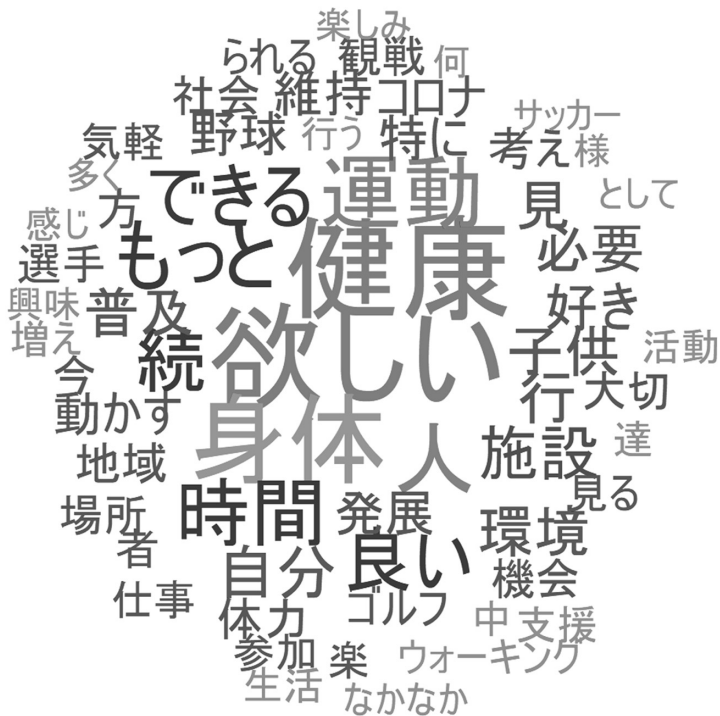


図4 ワードクラウド：「スポーツへの思いや意見」に関する自由回答
(回答者数 男性543；「スポーツ」「思い」を除く回答頻度15以上)

の「身体」の出現頻度を第2位・第3位の出現頻度と比較すると、突出して大きいことがわかる。すなわち、女性にとっては、「身体」というキーワードが際立って重要であることが推察される。以上のことから、女性にとって重要と認識されている単語の第3位までは、男女合計のものと同じであり、運動・健康関連要素となっている。

第4位以降について、男女全体の順位と相対的に大きく異なる点は、第23位の「施設(36)」(男女全体では第19位)と第25位の「環境(35)」(男女全体では第21位)という環境要素の順位が最下位に位置するということが挙げられる。すなわち、女性は環境要素について、相対的に重要性が低いと認識していることが示唆される。

次に、表3の結果を視覚化したものが図5のグラフである。

この図5をみると、表3の結果に基づいて上述したことが視覚化されることで、より明白に把握

表3 男女別頻度分析結

(上位25以上；女性の頻度でソート)

合計順位	女性順位	出現単語	男性	女性	頻度合計
1	1	身体	73	179	252
3	2	運動	56	155	211
2	3	健康	86	148	234
7	4	自分	36	103	139
6	5	良い	47	99	146
5	6	人	58	95	153
8	7	時間	50	81	131
4	8	欲しい	87	80	167
9	9	できる	47	71	118
10	10	続	46	70	116
12	11	動かす	26	70	96
17	12	体力	24	62	86
13	13	子供	34	61	95
14	14	好き	31	61	92
15	15	コロナ	27	60	87
11	16	もっと	53	55	108
16	17	行	36	50	86
18	18	見	31	49	80
20	19	今	23	49	72
23	20	楽	20	46	66
24	21	参加	20	45	65
25	22	観戦	21	42	63
19	23	施設	37	36	73
22	24	必要	31	36	67
21	25	環境	33	35	68

しやすくなっている。例えば、上位3位までの運動・健康関連要素が、他の出現単語と比較すると、頻出数が突出していることが一目で読み取れる。さらに、図1(男女全体)と単語出現頻度の順位を比較すると、環境要素が図1では下位(第19・21位)に位置しているが、図2ではさらに下位(第23・25位)に位置することも判別しやすくなっている。

また、図5全体を概観すると、単語の出現頻度の分布が、図1や図3と同様に指数関数的に減少することが認識でき、冪乗則に従っているように推察されよう。

さらに、ワードクラウドを適用してヴィジュアル分析をおこなったものを図6に示す。

この図6をみると、表3の結果に基づいて上述したことが視覚化されることにより、より明瞭でインパクトのある形で理解しやすくなることが明らかである。例えば、中央に位置し最も大きなフォントである「身体」の出現頻度が最も多く、次に、ほぼ同じ大きさのフォントであるが濃度が異なる「運動」と「健康」が中央付近に位置している。このことから、これら3つのキーワードの重要性が、顕著に高いことが認識できよう。

4) 考察

以上のように、笹川スポーツ財団による最新の全国調査「スポーツライフに関する調査2022」における『スポーツに対する思い』や『スポーツの普及や発展』に対する自由回答を分析対象として、成人のスポーツライフにおける重要なキーワードを探索的に抽出するためにテキストマイニングをおこなうとともに、その結果を視覚化するためヴィジュアル分析をおこなった。その結果として、「身体」「健康」「運動」という運動・健康関連要素の相対的重要性が最も大きく、続いて、人的要素、行動的要素、および環境要素の重要性が明らかになった。

また、ヴィジュアル分析による視覚化は、単にテキストマイニングによる頻度分析表を提示するよりも、結果の特徴を一目で明確にインパクトのある形で表現することの有効性を示すこととなった。ただし、ワードクラウドには、キャッチーでわかりやすいという利点がある一方で、次のような懸念があることも視野に入れておかなければならない (e-TEA, 2021)。

- ・読み取れる情報が少ない
- ・文脈がわからない
- ・希少意見が無視される
- ・結局「So what?」となる

IV 結論

本稿の研究目的は、日本における成人男女から発せられるスポーツの振興や思いに対する「生のメッセージ (自由回答)」を対象として分析することにより、成人のスポーツライフにおける重要なキーワードについて探索的に明らかにすることであった。そのために、公益財団法人笹川スポーツ財団 (SSF) が実施した最新の全国調査「スポーツライフに関する調査2022」における『スポーツに対する思い』や『スポーツの普及や発展』に対する自由回答を分析対象として、成人のスポーツライフにおける重要なキーワードを探索的に抽出するためにテキストマイニングによる頻度分析を実施し、その結果を視覚化するためヴィジュアル分析をおこなった。その結果は以下の通りである。

- 1) 男女全体として「身体」「健康」「運動」という運動・健康関連要素の相対的重要性が最も大きかった。
- 2) 続いて、「人」や「自分」「子供」という人的要素、「時間」要素、「欲しい」「できる」および「続」などの行動的要素が、男女とも

に相対的に重要であった。

- 3) 「施設」や「環境」という環境要素も一定の重要度で認識された。
- 4) 男性の特徴としては、「欲しい」と「健康」の相対的重要性が最も大きく、環境要素について女性よりも重要性が高いと認識していた。
- 5) 女性の特徴としては、「身体」というキーワードが際立って重要で、環境要素は男性よりも重要性が低いと認識していた。

この結果はテキストマイニングの初期段階である頻度分析やワードクラウドにより、相対的に多く出現するキーワードを抽出することはできたが、キーワード間の関係性については分析されていない。今後は、こうしたキーワード間の関係性を明らかにするような追加的な研究が望まれる。さらに、頻度分析結果の分布関数を特定することも興味深い研究の一つになり得ると考えられる。

参考文献

- e-TEA (2021) 「なぜ「ワードクラウドで分析しました」は良くないのか」
<https://note.com/sanskruithiya/n/n46280fb7f656>
- 上田太一郎 (監修)・村田真樹・小木しのぶ・高山泰博・末吉正成・今村誠、淵上美喜 (2008) 『事例で学ぶテキストマイニング』 共立出版
- Any Connector (2021) 「データラングリングとは何ですか? 6つの主要なステップ」
<https://anyconnector.com/ja/data-transformation/what-is-data-wrangling.html>
- AI Academy Media (2023) 「【テキストマイニング】WordCloud (ワードクラウド) とは」
<https://aiacademy.jp/media/?p=3659#:~:text=%E3%83%AF%E3%83%BC%E3%83%89%E3%82%AF%E3%83%A9%E3%82%A6%E3%83%89%E3%83%88%EF%BC%88%EF%BC%89%E3%81%A8%E3%81%AF,%E3%81%AF%E3%80%81%E5%B0%8F%E3%81%95%E3%81%8F%E8%A1%A8%E7%A4%BA%E3%81%95%E8C%E3%81%BE%E3%81%99%E3%80%82>
- Exploratory (n.a.) URL <https://ja.exploratory.io/>
- 國府久嗣, 山崎治子, 野坂政司 (2013) 「内容推測に適したキーワード抽出のための日本語ストップワード」日本感性工学会論文誌12巻4号
- 笹川スポーツ財団 (2023) 「スポーツライフ・データ2022—スポーツライフに関する調査報告書」笹川スポ

スポーツ財団
スポーツ庁（2023）「国民のスポーツライフ」[https://
www.mext.go.jp/sports/b_menu/sports/mcatetop05/
1371876.htm](https://www.mext.go.jp/sports/b_menu/sports/mcatetop05/1371876.htm)

文部科学省（2022）「スポーツ基本計画」
[https://www.mext.go.jp/sports/content/000021299_
20220316_3.pdf](https://www.mext.go.jp/sports/content/000021299_20220316_3.pdf)