

# Q&A コミュニティにおける応答に着目した 子育ての悩みへの対応

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> 小舘 亮之<sup>†7</sup>

**概要**：本研究では、Q&A コミュニティにおける応答に着目し、子育ての悩みに対して推奨すべき対応について検討する。具体的には、Yahoo!知恵袋の子育てカテゴリにおける悩みの中から、特に母乳・ミルクに関する応答を取り上げ、ベストアンサーとなった回答の特徴についてテキストマイニングの手法を用いて明らかにすることを試みた。結果より、少なくない質問者が感情的な同意を求めていることが示唆された。

**キーワード**：Yahoo!知恵袋，育児の悩み，ベストアンサー，計量テキスト分析

## 1. はじめに

著者ら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って提供されている「Yahoo!知恵袋データ(第3版)」を用いた研究をこれまで行ってきており、質問文の投稿内容から様々な知見が得られることを明らかにしてきた[1][2][3][4][5]。他方で、研究に利用したデータセットでは質問と回答がセットで提供されているため、この両者の対応関係を分析することでより深い洞察を得ることが期待できる。

本研究では、Yahoo!知恵袋の「子育ての悩み」カテゴリにおける質問文から育児の悩みの分析を行った先行研究を踏まえて、そこで行われる応答に着目し、子育ての悩みに対して推奨すべき対応について検討する[5]。

## 2. 先行研究

### 2.1 IDR データセット提供サービス

国立情報学研究所 (NII) のデータセット共同利用研究開発センター (DSC) はデータセットの共同利用事業として情報学研究データリポジトリ

(Informatics Research Data Repository: 以下, IDR) を運営しており、本研究の対象である Yahoo!知恵袋の投稿データだけでなく、楽天やニコニコ、クックパッドといった多くの企業がデータセットの提供に協力している。こうした公開データの研究成果については、各々の学会や論文誌だけでなく、IDR データフォーラム等でも公開されている[6]。

### 2.2 Yahoo!知恵袋の質問文を用いた先行研究

前述の通り、著者らは国立情報学研究所の IDR データセット提供サービスで提供される「Yahoo!知恵袋データ(第3版)」を用いた研究を行ってきた。近年では、ソーシャルリスニングとしてソーシャルメディアの投稿データの分析が盛んに行われているが、ノイズの除去といったデータクリーニングへの対応などの面で課題がある。他方で、Yahoo!知恵袋に代表される Q&A コミュニティでは質問がカテゴリ化されていることによってより精度の高い分析を行うことが可能になる。実際に、著者らの研究においてもこうした Q&A コミュニティの投稿データを計量テキスト分析の手法で分析することが「大学受験」や「国内観光」、「スマートフォンの製品評価」といった幅広い分野に活用できることを明らかにしている[2][3][4]。

ただし、これらの研究はいずれも質問文のみを対象としたものであり、データセットで提供されている回答データは扱っていない。ソーシャルメディアの分析においては質問と回答が明示的に区分されていないため、分析者の側で何らかの対応を行う必要がある。その点、Q&A コミュニティの分析では、最初から質問と回答が分離しているため、一連の応答をより精緻に扱うことが可能となる。

### 2.3 育児に関する先行研究

Yahoo!知恵袋に投稿された育児の悩みを対象とした先行研究では、月齢別の特徴語や質問者の悩みの実態について質問文の分析から明らかにし、育児に必要なサポートの在り方について検討している[7]。

†1 成蹊大学 Seikei University  
†2 佛教大学 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  
†7 津田塾大学 Tsuda University

さらに、当該の先行研究が手動で 1 年分のデータ取得しかできなかったことを課題に挙げていたため、著者らの先行研究では IDR データセット提供サービスを用いて 4 年分の Yahoo!知恵袋における「育児カテゴリ」の質問文について分析を行った。「半年未満」「半年以上 1 歳未満」「1 歳以上 2 歳未満」の 3 つの時期に分けて行なった分析からは、各時期において表出する悩みが異なること、時間の経過によって自然と解消していく傾向があることが確認された[5].

## 2.4 問題意識

これらの先行研究からは Yahoo!知恵袋の質問文がさまざまな分野に応用できること、育児カテゴリにおいては子育てに悩む親のリアルな声を可視化できることが成果として挙げられている。ただし、質問と回答をセットで分析している例は見られていない。

特に、育児カテゴリの悩みにおいては回答が悩める親の貴重な情報源となっていることから、質問と回答をセットにした応答に着目することで、子育ての悩みの別の側面を捉えることが期待できる。そのため、本研究では IDR データセットを用いて Yahoo!知恵袋における「育児カテゴリ」の質問文とそれに対応する回答について計量テキスト分析の手法を用いて分析し、有用な知見を獲得することを目的とする。

## 3. 分析

### 3.1 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 2020 年度提供版における質問文と回答を利用した[1]。収録データは収録期間(2015 年 4 月 1 日-2018 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り、全体の質問数は約 270 万件、回答数は約 838 万件となっている。

分析に当たっては、提供されたデータから「>子育てと学校>子育て、出産>子育ての悩み」カテゴリの質問投稿 9,113 件を抽出した。加えて、今回は回答も含めたため、各質問に対するベストアンサーを同様に抽出した。ベストアンサーが設定されていなかった 4 件を除き、最終的に質問と回答(ベストアンサー)それぞれ 9109 件ずつを分析に用いるデータとした。

### 3.2 分析対象

「子育ての悩み」カテゴリでは幅広いテーマが扱われているため、カテゴリ全体を対象とした場合には個々の応答における特徴的な単語が一般的な単語に埋もれてしまう懸念が生じてしまう。そのため、著者

らの先行研究において主要なテーマとして登場した「母乳・ミルク」「寝る・泣く」「離乳食」の中から、件数の多さや抽出のしやすさを踏まえて「母乳・ミルク」を分析対象として設定した。具体的には、「ミルク」「母乳」「授乳」「完ミ」「混合」「断乳」を対象語とし、これらを含む質問投稿を「母乳・ミルク」に関連するものとした。最終的に、1859 件が対象となり、全体に占める割合は 20.4%となった。

### 3.3 分析方法

分析に当たってはテキストマイニングのフリーソフトウェアである KH Coder (<https://kncoder.net/>) を使用した。形態素解析は付属の ChaSen (茶筌) を用いているが、「母乳・ミルク」に関連する単語として「断乳」「完ミ」「完母」「哺乳瓶」を強制抽出の対象語とした。

分析の手順として、まず質問と回答(ベストアンサー)における頻出上位単語の差異について概観した。次に、質問と回答(ベストアンサー)の共起ネットワークをそれぞれ描写し、特徴について比較した。最後に、回答(ベストアンサー)における特徴的な単語を取り上げ、どのような単語との共起関係が目立ったのかについて検討した。

なお、頻出上位語のカウントに際しては、単語の登場数ではなく、1 投稿当たりの登場の有無を基準とした。これは特定の投稿における言及の多さの影響を避けるためである。加えて、共起ネットワークの描写にあたっては、先行研究と同様に、全体の投稿の 1 割以上の投稿に登場している単語を対象に Jaccard 係数 0.2 以上の共起関係を基準とした。共起ネットワーク分析においては明示的な基準は存在しないため、先行研究と同様の基準を用いることで結果の比較が可能となることを重視した。

## 4. 分析結果

### 4.1 頻出上位単語

質問と回答(ベストアンサー)における頻出上位 30 語の抽出結果を表 1 と表 2 に示している。質問への応答となることから多くの単語が重複していたが、質問では「生後」「お願い」「育てる」「教える」「最近」「現在」の 6 語が回答(ベストアンサー)にない単語となった。時期の詳細や質問文特有の表現に用いられる単語だと解釈できる。

一方で、回答(ベストアンサー)に含まれて質問にない単語としては、「大丈夫」「気」「量」「増える」「多い」「心配」の 6 語が挙げられた。こちらは質問者の心情に寄り添う表現や回答に際して具体性を付与し

た表現だと理解できる。

表1 質問文の頻出上位30語 (1投稿当たり)

抽出語	登場数	比率	抽出語	登場数	比率
ミルク	1043	56.1%	お願い	478	25.7%
飲む	912	49.1%	前	449	24.2%
思う	886	47.7%	子	435	23.4%
母乳	859	46.2%	育てる	432	23.2%
授乳	801	43.1%	離乳食	431	23.2%
寝る	774	41.6%	いい	422	22.7%
生後	726	39.1%	食べる	395	21.2%
赤ちゃん	675	36.3%	教える	384	20.7%
時間	663	35.7%	子供	356	19.2%
泣く	627	33.7%	最近	355	19.1%
今	615	33.1%	少し	347	18.7%
言う	535	28.8%	体重	324	17.4%
出る	529	28.5%	良い	322	17.3%
夜	527	28.3%	現在	318	17.1%
起きる	479	25.8%	聞く	318	17.1%

表2 回答の頻出上位30語 (1投稿当たり)

抽出語	登場数	比率	抽出語	登場数	比率
思う	1120	60.2%	良い	327	17.6%
飲む	631	33.9%	食べる	321	17.3%
ミルク	587	31.6%	離乳食	294	15.8%
赤ちゃん	564	30.3%	夜	285	15.3%
子	562	30.2%	少し	270	14.5%
母乳	545	29.3%	気	268	14.4%
寝る	465	25.0%	体重	251	13.5%
言う	446	24.0%	前	240	12.9%
授乳	430	23.1%	子供	233	12.5%
出る	429	23.1%	量	233	12.5%
今	405	21.8%	起きる	231	12.4%
いい	388	20.9%	増える	229	12.3%

時間	377	20.3%	多い	228	12.3%
大丈夫	344	18.5%	心配	217	11.7%
泣く	335	18.0%	見る	212	11.4%

#### 4.2 共起ネットワーク分析

続いて、質問と回答(ベストアンサー)それぞれにおける、特に強い単語間の共起関係について、共起ネットワーク分析を用いて可視化した。結果は図2と図3に示している。図2と図3を比較すると、図2の方が単語間の共起関係が強く見られている。このことは、質問が全体的に似通っている一方で、回答(ベストアンサー)の一致度が必ずしも高くないことを示していると考えられる。

図2の特徴としては、「混合」「完母(完全母乳)」が頻出上位30語にはない単語として出現していることが挙げられる。必ずしも全体の投稿に占める割合は多くはないものの、母乳に関する主要な話題として登場していることが読み取れる。

加えて、図2では「体重」に関する共起が「母乳」となっているのに対し、図3では「体重」に関する共起が「ミルク」となっているのも両者の差異であった。また、「ミルク」の共起語に「量」があるのは図3のみの特徴となっている。この辺りも質問と回答(ベストアンサー)の傾向の違いを反映しているものと考えられる。

さらに興味深いのは、回答(ベストアンサー)特有の単語として頻出上位(14位)に登場した「大丈夫」という単語が共起ネットワーク分析では出てこなかったことである。言い換えると、「大丈夫」という単語はさまざまなベストアンサーとなった回答に用いられているものの、特定の単語とは必ずしも共起していないということになる。

そのため、次節では特に「大丈夫」という単語に注目し、どのような単語と強い共起関係にあるのかについて確認した。

#### 4.3 「大丈夫」の共起関係

「大丈夫」という単語に限定した共起関係を示したのが表3である。もっとも共起関係が強かった単語は「心配」であり、Jaccard係数は0.15となっている。今回基準とした0.2には達していないものの、両単語が質問者の不安な心情に寄り添うものとして同時に使われる傾向にあったことが示唆されている。ただし、他の共起語からは特に強い傾向を読み取ることはできなかった。

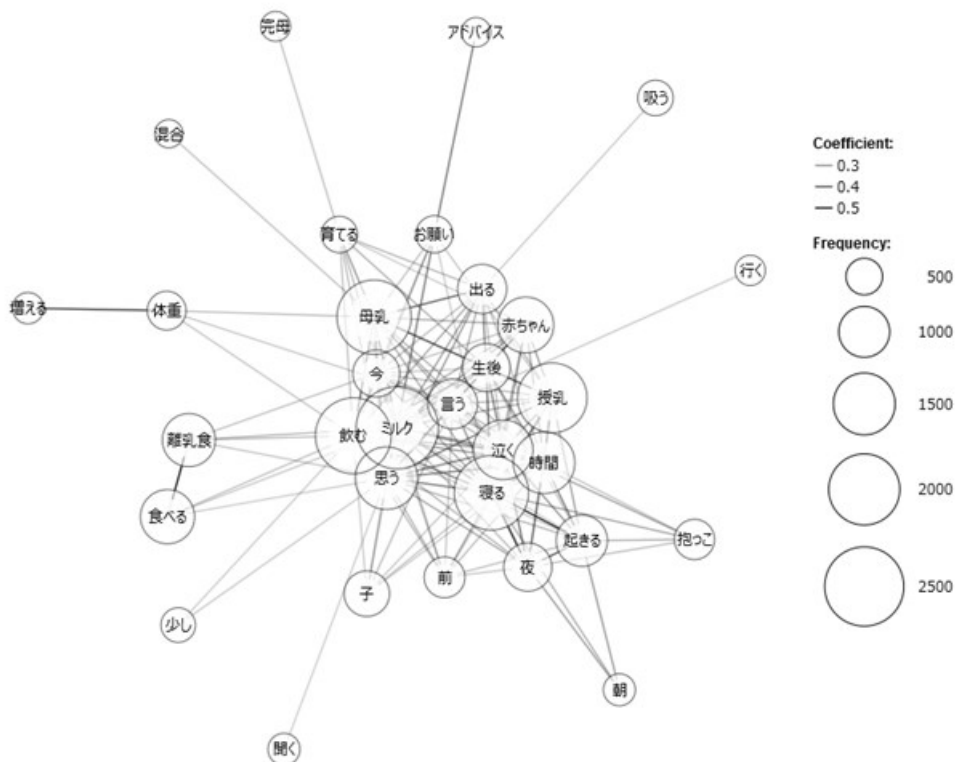


図1 共起ネットワーク分析の結果（質問）

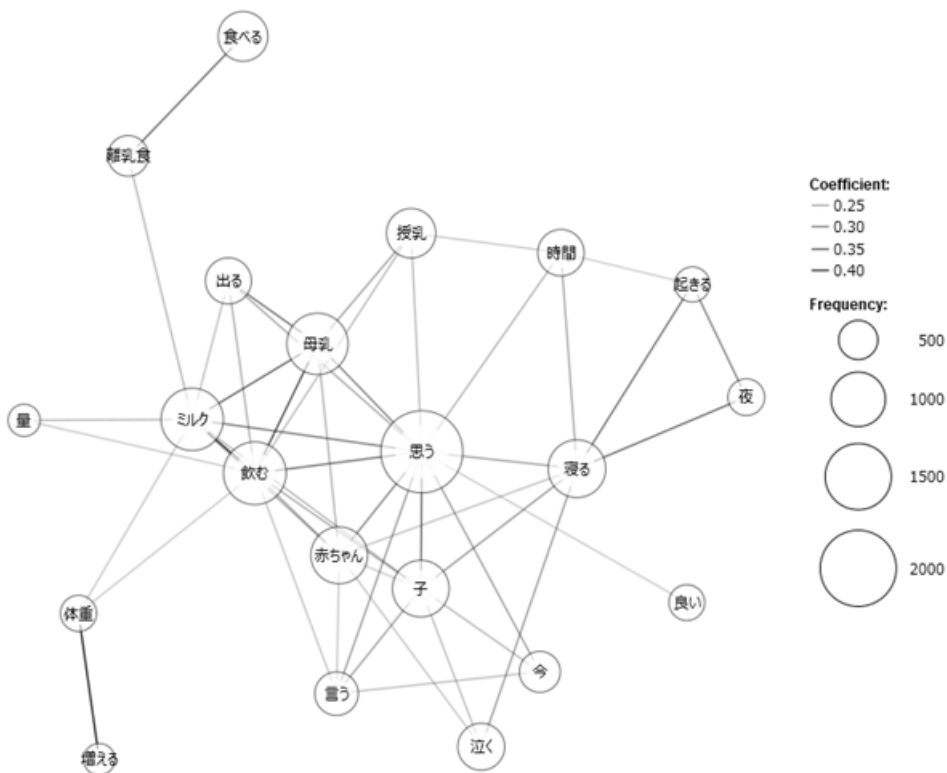


図2 共起ネットワーク分析の結果（回答）

表3 「大丈夫」との共起語

	単語	全体	共起	Jaccard
1	心配	287 (0.154)	73 (0.272)	0.1515
2	思う	886 (0.477)	140 (0.522)	0.1381
3	赤ちゃん	675 (0.363)	114 (0.425)	0.1375
4	ミルク	1043 (0.561)	154 (0.575)	0.1331
5	飲む	912 (0.491)	137 (0.511)	0.1314
6	授乳	801 (0.431)	120 (0.448)	0.1264
7	言う	535 (0.288)	90 (0.336)	0.1262
8	生後	726 (0.391)	110 (0.410)	0.1244
9	時間	663 (0.357)	103 (0.384)	0.1244
10	お願い	478 (0.257)	76 (0.284)	0.1134

表2に示している通り、「大丈夫」という単語の投稿に占める出現割合は18.5%と必ずしも突出して多いわけではない。しかしながら、特定の単語と強い共起関係を示さないかたちで「大丈夫」という単語がベストアンサーの頻出上位語として登場している背景には、感情的な同意を求める質問者の姿があることが予想される。

#### 4.4 考察

こうした質問者の不安な心情に寄り添うベストアンサーの在り方には2つの側面があると考えられる。肯定的な側面としては、「ソーシャルメディアを利用して寝かしつけに関連した悩みを発言し、他の育児経験者の意見をとおして問題解決を図っていこうとする現代の育児中の母親の生の声」「子どもを育てながらひとつひとつ意思決定していかなければならない現代の母親の、専門職には言えない、または相談するチャンスがない発言」[7]と先行研究でも言及されているような相談相手のいない不安に対して、主にリアルでの関係性が薄い経験者の立場からケアできることが挙げられる。「大丈夫」という単語に強い共起語が見られなかったことも、具体的な解決策というよりも感情的な同意を求める層が少なからず存在することを示唆している。

それに対して、否定的な側面として、ベストアンサーとなった回答者の多くが、経験者ではあったとしても医療や育児の専門職ではないことが挙げられる。特に母乳やミルクの話題は子どもの成長に直結するものであり、科学的な背景のない助言が場合によっては問題を引き起こしてしまうことも考えられる。実際に、図2において「混合」「完母(完全母乳)」が登場していたことも、こうした育児における考え方の違いが表出していることを示している。さらに、ベストアンサーは質問者が決められることから、アドバイスの有用性よりも当初の自分の考えを支持する回答をより強く評価してしまう、選択的接触を強化してしまうことが懸念される。

もちろん、専門家に相談できない内容を気軽に投稿できることがYahoo!知恵袋の魅力でもあるため、科学的な根拠との兼ね合いについては慎重に検討する必要があるだろう。

#### 5. 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2020年度提供版データを利用し、「子育ての悩み」カテゴリにおける質問文と回答から育児の悩みの分析を行った。特に母乳・ミルクに関する応答に着目し、ベストアンサーとなった回答の特徴についてテキストマイニングの手法を用いて明らかにすることを試みた。

一連の分析では、質問への応答として多くの単語の利用が重複しているものの、質問では時期の詳細や質問文特有の表現が頻出上位に登場していること、回答(ベストアンサー)では質問者の心情に寄り添う表現や回答に際して具体性を付与した表現が頻出上位に登場していることが明らかとなった。また、共起ネットワーク分析からは質問者の投稿内容が似通っている一方で、ベストアンサーとなった回答では類似度が下がっていることが確認された。加えて、「大丈夫」や「心配」などの単語がベストアンサーに特有の表現として特定された。こうした結果は、少なくとも質問者が感情的な同意を求めている傾向を示したものと解釈できる。

ただし、こうした感情的な同意の希求が強くなりすぎることは、科学的な背景に基づかない選択的接触を強化してしまう懸念にもつながってしまう。今後の研究ではこのような否定的側面にも着目し、より良いQ&Aコミュニティの応答の在り方について引き続き研究していきたい。

## 謝辞

本研究は ROIS-DS-JOINT(課題番号:00032,研究代表者:小舘亮之)の助成を受けた。また,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の「Yahoo!知恵袋データ(第 3 版)」を利用した。

## 参考文献

- [1] 国立情報学研究所：“情報学研究データリポジトリ” (2021/1/27 閲覧), [https://www.nii.ac.jp/dsc/idr/yahoo/chiebkr3/Y\\_chiebukuro.html](https://www.nii.ac.jp/dsc/idr/yahoo/chiebkr3/Y_chiebukuro.html)
- [2] 吉見憲二 (2020)「計量テキスト分析を用いた Q&A コミュニティからの評判情報の推定」『情報処理学会全国大会講演論文集』
- [3] 吉見憲二・田中康裕・針尾大嗣・谷本和也・源城かほり・岩井憲一・小舘亮之 (2020)「Q&A コミュニティ

- における質問文からの製品情報の分析」『情報コミュニケーション学会第 17 回全国大会予稿集』
- [4] 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.
  - [5] 吉見憲二・谷本和也・田中康裕・岩井憲一・上田祥二・針尾大嗣・小舘亮之 (2020)「Q&A コミュニティの質問文を用いた育児の悩みの可視化」『情報処理学会研究報告電子化知的財産・社会基盤 (EIP) ,2020-EIP-90(12)』 pp. 1-4.
  - [6] 国立情報学研究所：“IDR ユーザフォーラム” (2021/1/27 閲覧), <https://www.nii.ac.jp/dsc/idr/userforum/>
  - [7] 佐々木裕子, 高橋眞理 (2015)「インターネットの Q&A コミュニティサイトにみる 0~4 ヶ月児の母親の育児における寝かしつけの悩み テキストマイニングによる分析」『医療看護研究』11 巻 2 号, pp.28-35.