

# ニュースにおける特定単語の文脈の変遷

寺田光里 高田悠希  
著者 3127 寺田光里  
群馬県立高崎高等学校

## 要旨

本研究は、ニュースの見出しや新聞記事における、「特定の単語の使われる文脈の変遷」を調べるため、複数の単語に対して「その単語の周辺に位置する単語の経年変化」を調査したものである。本研究の手法を用いることで、単語の出現回数を調べるだけではわからない、その単語にまつわる当時の社会情勢・出来事や、かつての人々の関心事などを明らかにできた。

## 1. はじめに

### 1.1 研究背景

数年で世の中が大きく変化する現代において、同じ単語であっても使われる文脈が変化しているのではないかと考えた。また、それを調べることで、その単語に対する人々の関心やイメージだけでなく忘れられてしまった社会課題などもわかるのではないかと考えた。それらを知ることは、社会研究をはじめ、企業の新製品開発などの多くの分野で重要であると思われる。

先行の研究でもメディア記事の内容をテキストマイニングで分析するものはあるが、時間経過で追っていないもの<sup>1</sup>や、「終活」など特定の単語のみに注目したもの<sup>2</sup>が多く、汎用的に特定の単語の周辺単語の変化をとれるようにしたものは存在しない。

## 2. 検証

### 2.1 仮説

特定単語の周辺単語の Dice 係数を調べることによって、その単語の使われる文脈を推測できるのではないかと。

### 2.2 方法

実験として、実験①と実験②を行った。

まず、実験①では分析対象の文書として、Yahoo! JAPAN ニュース(2018~2022)の見出し語を用いた。Yahoo! JAPAN ニュースの見出し語を取得するために、web ページの過去の履歴が保存されている Internet Archive の Way Back Machine を利用し、スクレイピングした。特定単語を「支援」という単語にし、「支援」を含む文章を KH Coder<sup>3</sup> を用いて対応分析した。

次に実験②では、分析対象の文書として、上毛新聞本文(2001~2023.1)を用いた。文章データを形態素解析し、各単語の Dice 係数を求め、グラフ化した。

---

<sup>1</sup> : 樋口, 「現代における全国紙の内容分析の有効性——社会意識の探索はどこまで可能か——」(2011)  
[https://www.jstage.jst.go.jp/article/jbhmk/38/1/38\\_1\\_1/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/jbhmk/38/1/38_1_1/_pdf/-char/ja)

<sup>2</sup> 木村, 安藤「マス・メディアにおける終活のとらえ方とその変遷 —テキストマイニングによる新聞記事の内容分析—」, 横浜国立大学技術マネジメント研究学会, 技術マネジメント研究第 17 号 1 - 19 (2018)

## 2.3 結果①

実験①の結果として、以下の図が得られた。

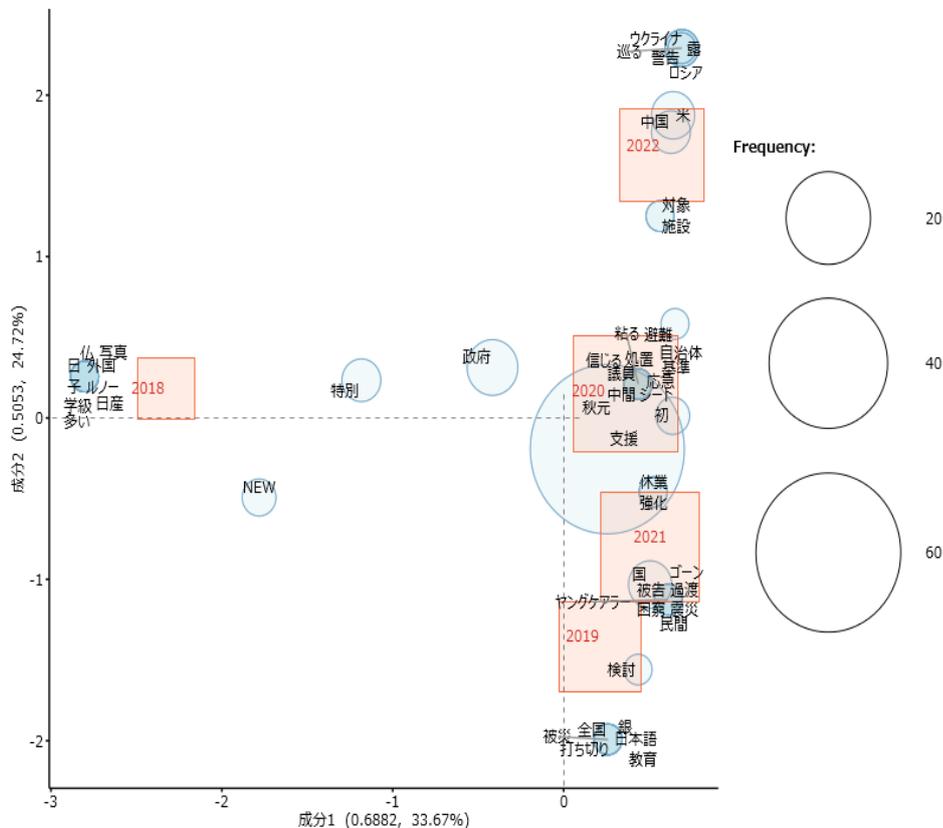


図 1 (支援という単語の周辺単語の対応分析)

対応分析の見方として、原点に近いものほど特徴的で、近い要素は関係が強いというものがある。また、軸は絶対的なものではない。

支援という単語は、例えば 2019 年では、ヤングケアラーという単語と関係が深く、2022 年ではウクライナやロシアといった単語と関係があった。

## 2.4 考察①

支援という単語の使われ方が 2022 年で大きく変わったのは、ロシアのウクライナ侵攻によるものだと考えられる。また、2022 年はヤングケアラーと支援が深く関係していないことから、ヤングケアラーという社会課題が忘れられてしまったものではないかと考察できる。

## 2.5 実験②

実験②の結果として以下、図 2～図 4 が得られた。図 2～図 4 の横軸は年で、縦軸は特定単語との Dice 係数の値である。ただしこのグラフに示された単語は Dice 係数の最大値が 0.1 以上のものである。

Dice 係数は、(単語 A, 単語 B をともに含む文章数) / (単語 A を含む文章数と単語 B を含む文章数の平均) で与えられる値で、この値が大きいほど 2 つの単語がより一緒に使われやすい、つまり関連が深いと分析できる。

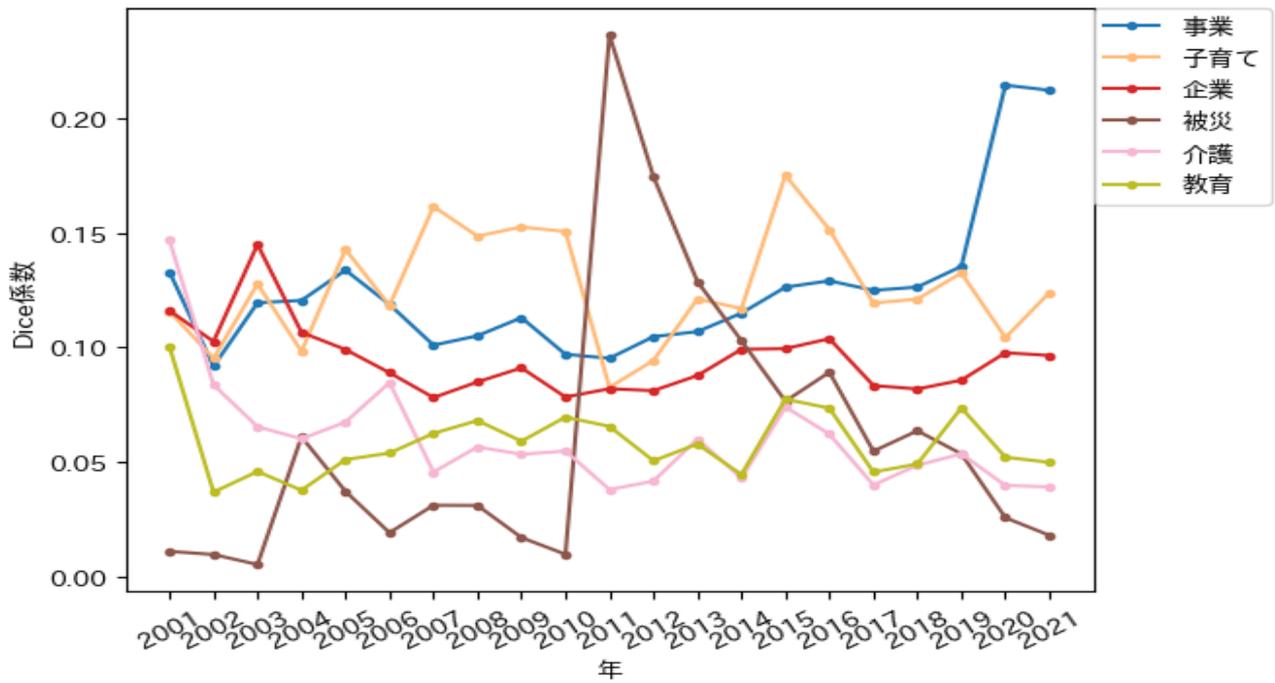


図 2(支援の周辺単語の Dice 係数の変化)

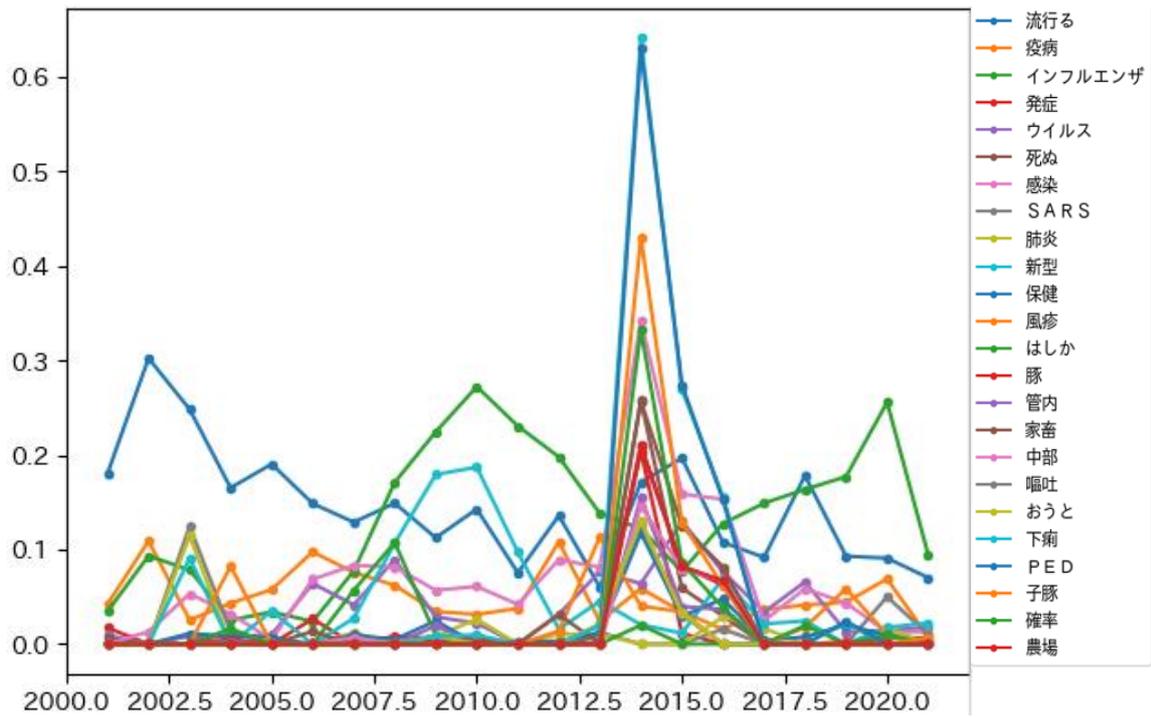


図 3 (流行の周辺単語の Dice 係数の変化)

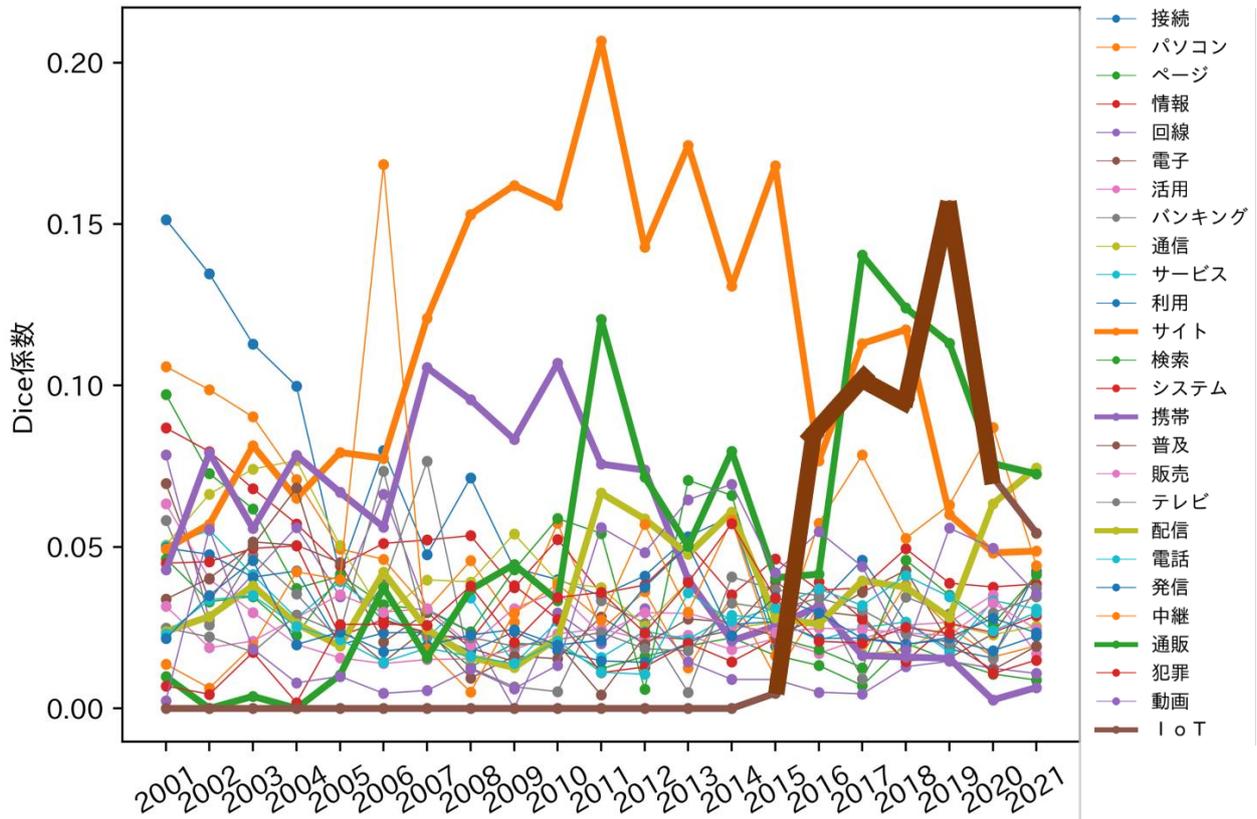


図 4 (インターネットの周辺単語の Dice 係数の変化)

## 2.6 考察②

まず, 図 2 をみると, 2011 年に被災という言葉と支援という単語の関連が強いことが分かり, これは 2011 年に発生した東日本大震災の支援という文脈で支援が使われていることが分かる。また, 事業という単語の Dice 係数が 2020 年に急激に大きくなっているのは, COVID-19 と緊急事態宣言に伴う事業支援という文脈で支援が使われていると分析できる。

次に, 図 3 を見ると, 2014 年に PED や子豚という単語の Dice 係数が急激に大きくなっている。このことから, 豚流行性下痢 (PED) の世界的パンデミックという文脈で流行が使われていることが分かる。また, 2010 年にはインフルエンザという単語の Dice 係数が大きくなっており, このことから H1N1 インフル (豚インフルエンザ) の世界的流行という意味の中で流行が使われたことが分かる。

最後に, 図 4 を見ると, 2016 年に IOT という単語の Dice 係数が急増していることが分かる。このことから, IOT は, IOT 元年と呼ばれている 2017 年の 1 年前からすでに話題になっていたことが読み取れる。

## 3. まとめ

### 3.1 結論

本研究の手法を用いることで, 単語の使われる文脈の変遷を追うことができた。

### 3.2 今後の展望

分析対象のデータをニコニコ動画のコメントデータやニコニコ大百科のコメントデータに広げ, メディア間での単語の文脈の変遷を比較できるようにしたい。

また, 文脈が変化した原因を Chat GPT などの AI に推測させるといったことも実現したい。

そして、これらの機能を入力だけで利用できるようなツール開発まで進め、多くの人が手軽に本研究の手法を使って研究、調査が行えるようにしたい。

## 謝辞

分析対象とした上毛新聞の記事データを提供して下さった上毛新聞社様、ご指導いただいた群馬大学の数理データ科学教育研究センター教授の青木悠樹様、北海道大学大学院情報科学研究院准教授の林克彦様、指導教員の岡田直之先生をはじめとする本研究に協力して下さった皆様に深く感謝申し上げます。