

日経記事の経済データを用いた コロナウイルス流行前後での トピックの変化

中央大学大学院 理工学研究科 経営システム工学専攻
修士1年 野中芽依

1. 背景・目的
2. データ概要
3. 分析
 1. 単語頻度解析
 2. 係り受け頻度解析
 3. ことばネットワーク
 4. 感情表現についての時系列単語頻度推移(割合)
 5. 感情表現の単語頻度割合とDI値の関係
4. まとめと今後の課題
5. 参考文献

現在コロナウイルスの流行は世界中に大きな影響を与え、世界経済は急速に悪化している。コロナウイルスの感染拡大による影響は、今までの災害や世界金融危機とは異なり、新しい経済ショックであると言われている[1].

経済の変化や状況について報じるメディアの中で、新聞記事が挙げられる。新聞記事の内容は、そのときの状況や変化をダイレクトに受けて報道される。

そのため、コロナウイルス前後での新聞記事内容は大きく変化し、**トピック内容やトレンドが大きく変化している**ことが想定できる。そして、そこには報道内容としての**感情の変化も投影されているのではないかと**考える。

本研究の目的

新聞記事を用いた

- コロナウイルス大流行の前後での**トピックの変化**
- **感情の変化と経済状況との関連**について分析を行う。

(株) NTTデータ数理システムより本学生研究奨励賞向けに提供いただいた、日本経済新聞データ
2019年9月1日～2020年9月30日の1年間の経済記事

大分類	経済	記事数
中分類	W301 金融政策	1857
	W302 経済活動	1774
	W303 マーケット	3394

なお、テキスト分析対象はbodysub_norm列（記事原文）のみとした。

コロナウイルス流行前後の記事を比較するために、経済活動として大きく変動したと考えられる、小中高の臨時休校を政府が発表し開始した3月からの期間を**コロナ流行前**、**コロナ流行後**としてデータを分割した。

データセット名	期間	記事数
コロナ流行前	2019年9月1日～2020年2月29日	3271
コロナ流行後	2020年3月1日～2020年9月30日	3748

Text Mining Studio (以降、TMS) を使用して、

- 単語頻度解析
- 係り受け頻度解析
- ことばネットワーク

の作成を**コロナ流行前**・**コロナ流行後**のデータセットそれぞれに対して行なった。

その際、新型コロナウイルスについての類似単語として類義語辞書に以下のように登録した。

代表語：新型コロナウイルス

類義語：新型コロナ，コロナウイルス，コロナ，新型ウイルス，新型肺炎

分析結果—単語頻度解析 上位20単語—

単語頻度解析では、名詞について上位20件を抽出した。

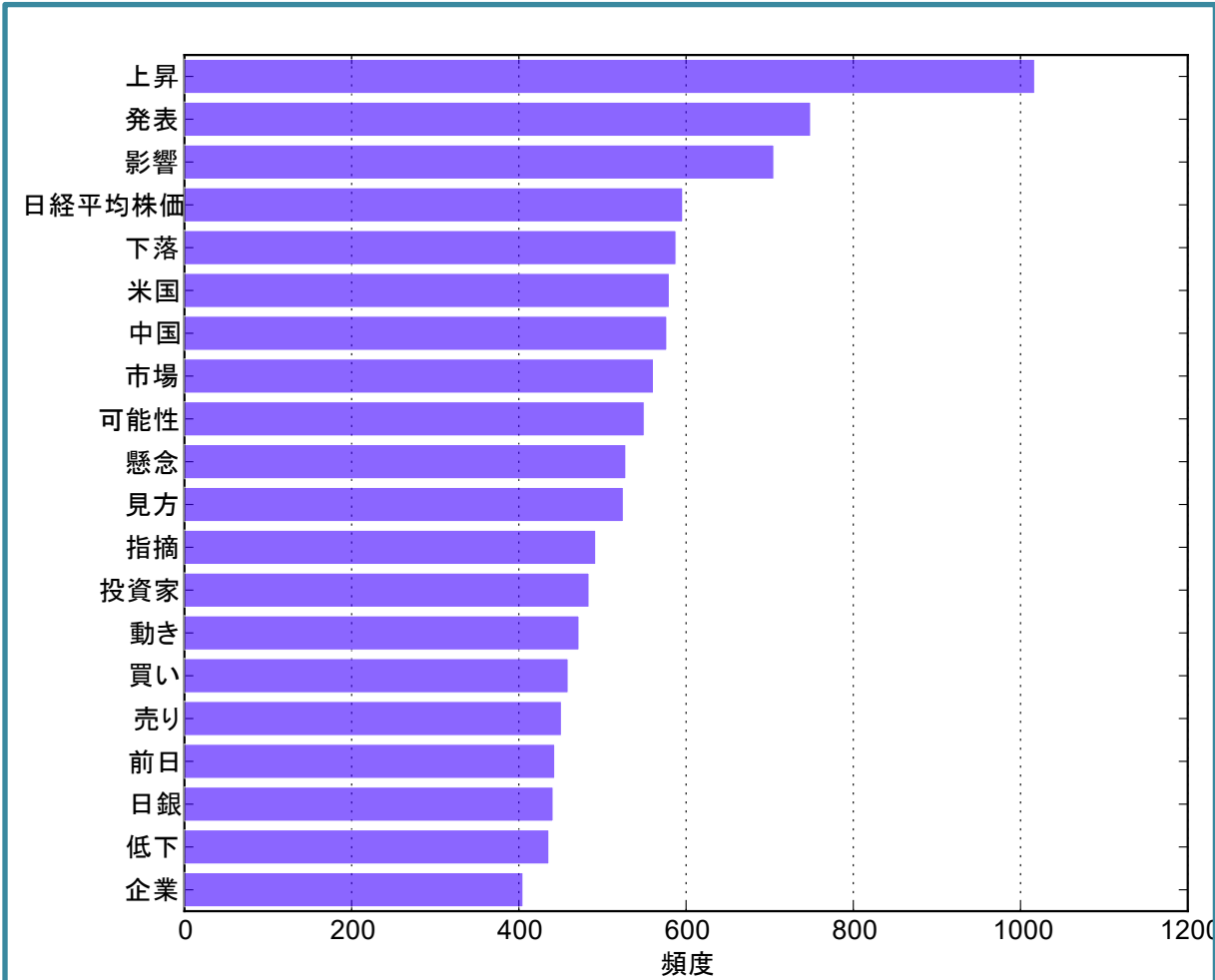
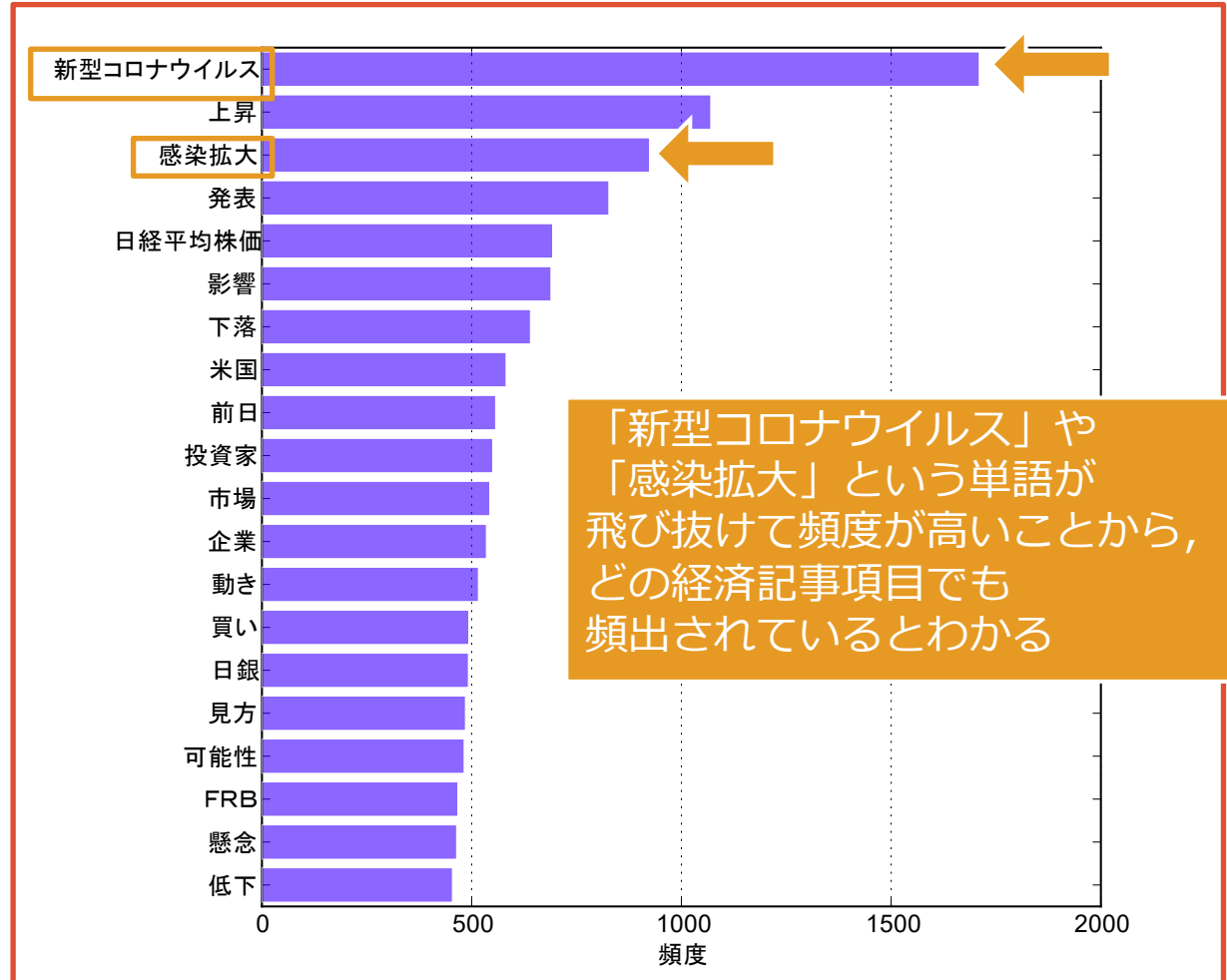


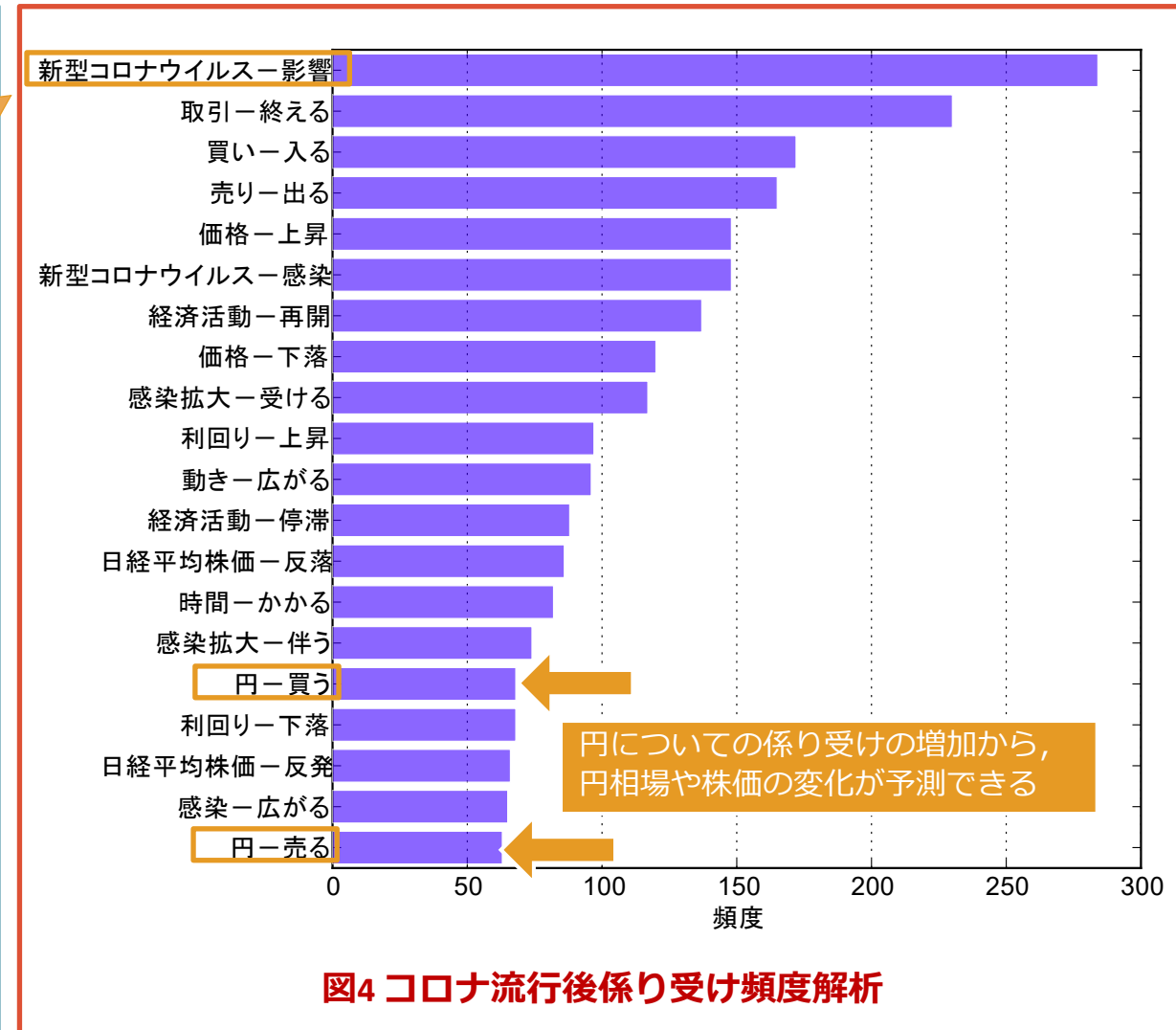
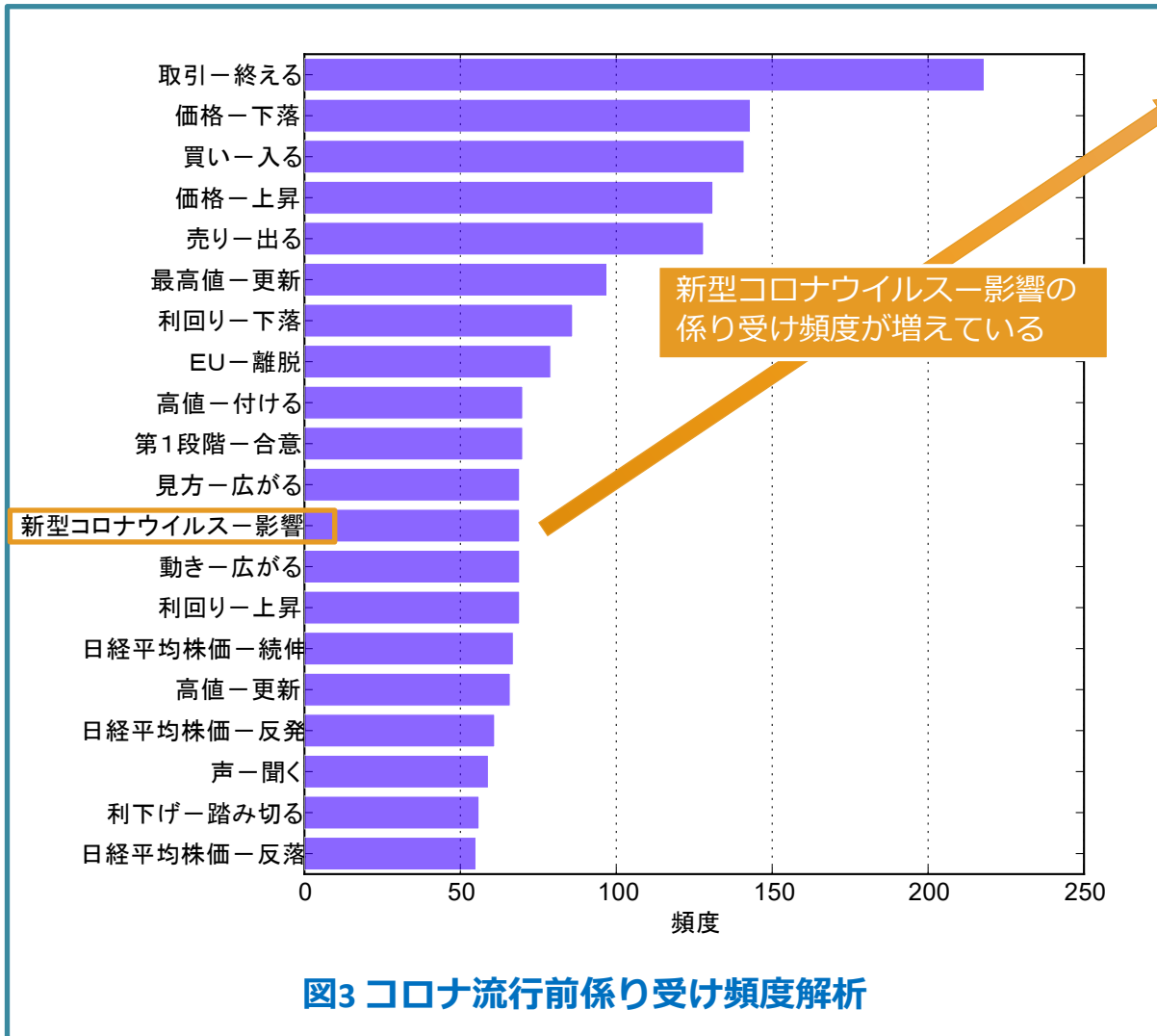
図1 コロナ流行前単語頻度解析



「新型コロナウイルス」や「感染拡大」という単語が飛び抜けて頻度が高いことから、どの経済記事項目でも頻出されているとわかる

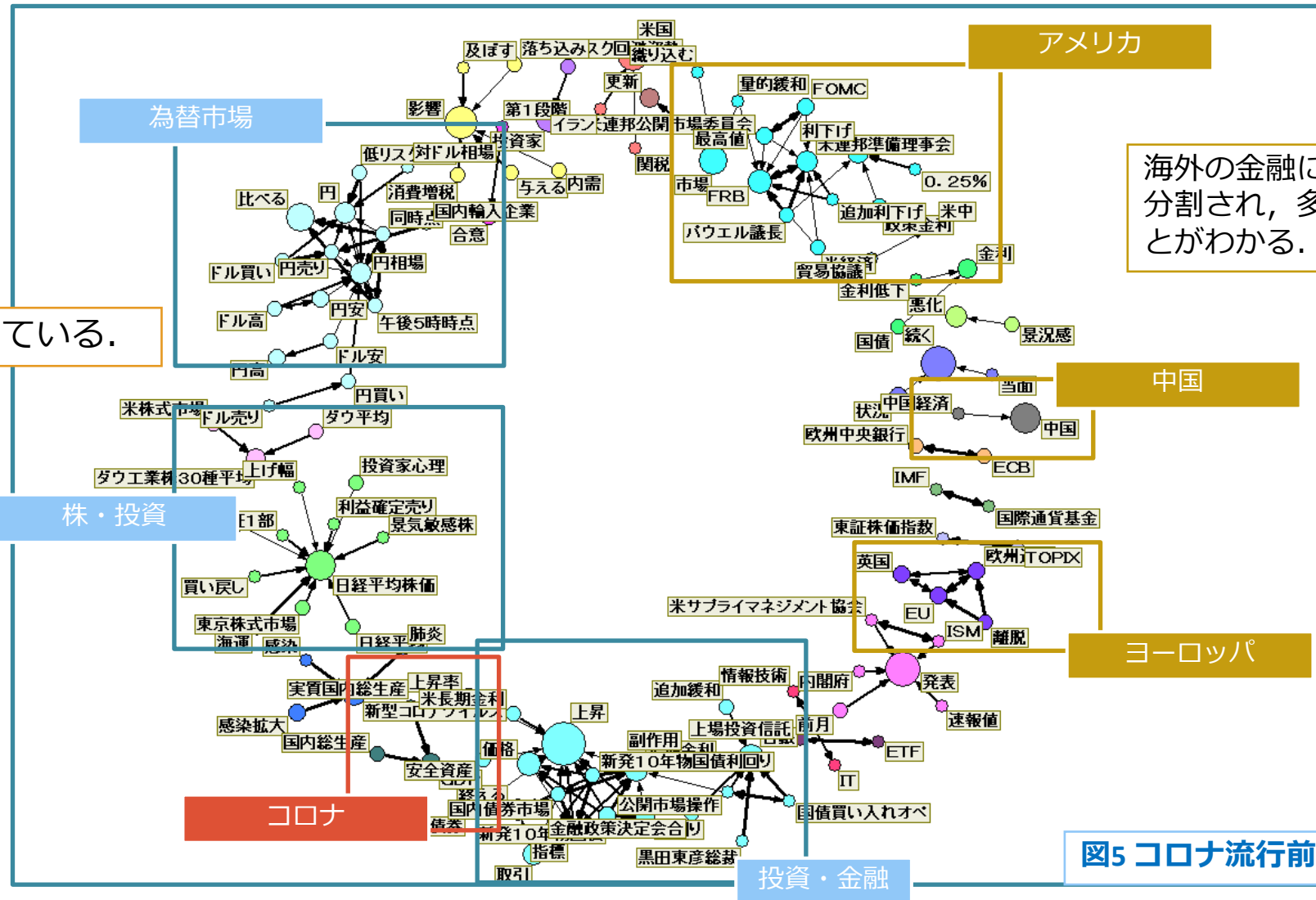
図2 コロナ流行後単語頻度解析

係り受け頻度解析では、名詞ー動詞・サ変名詞の係り受け上位20単語について抽出した。



ことばネットワークでは、共起関係を抽出し、名詞-動詞・サ変名詞の品詞設定を行った。

コロナ流行前



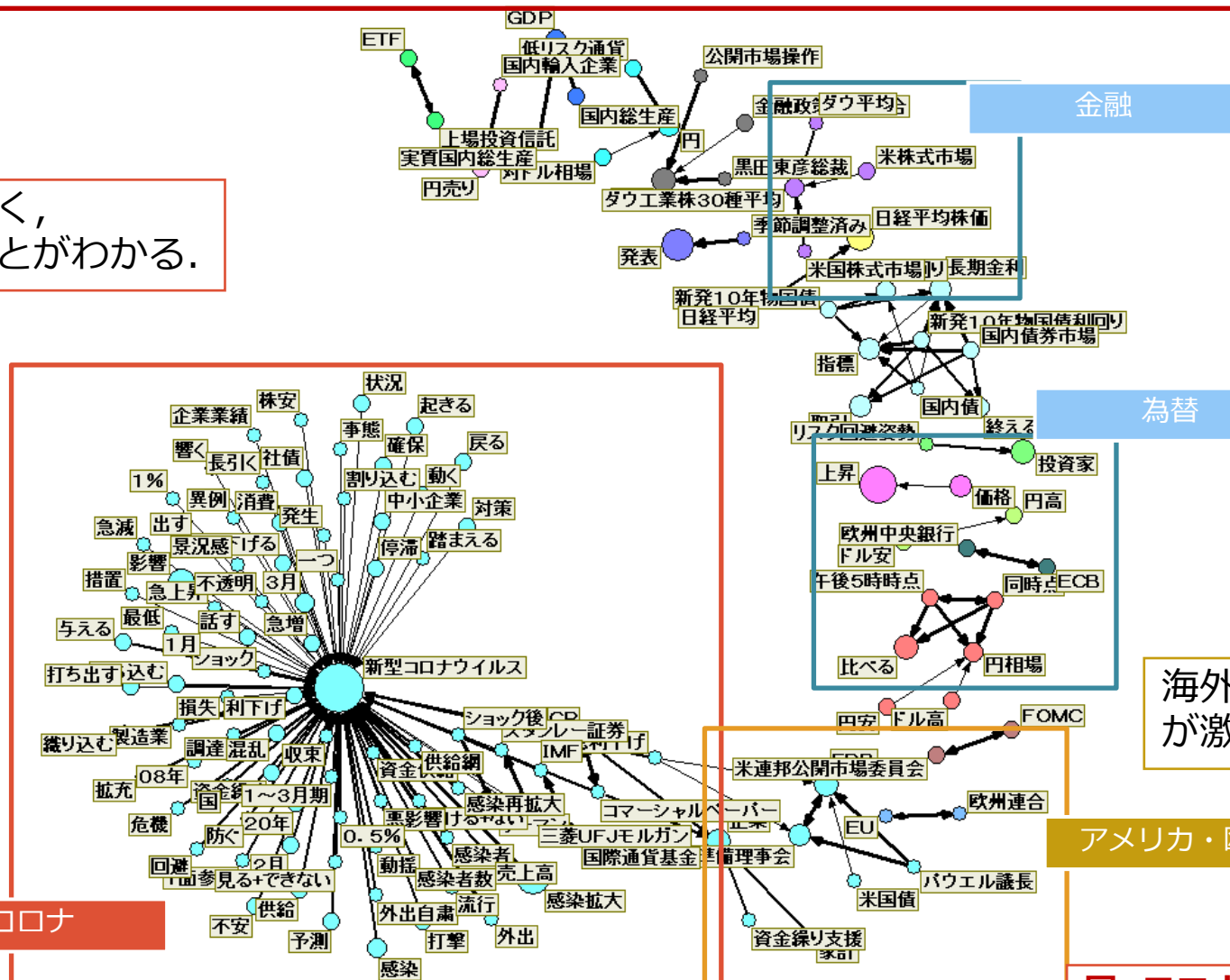
トピックが複数存在している。

海外の金融についてのトピックが分割され、多く報道されていることがわかる。

図5 コロナ流行前ことばネットワーク

コロナ流行後

コロナの話題がかなり大きく、
話題の中心になっていることがわかる。



海外情勢に関する記事
が激減している。

図6 コロナ流行後ことばネットワーク

1年間のデータ期間の中で、記事の中での**感情表現に変化が生じている**と考えた。

そのため、『感情表現辞典』[2]を使用して、1年間のデータ期間の中での感情表現についての時系列推移で分析を行う。

『感情表現辞典』には10感情で分類されている感情辞書である。例を以下に示す。

表1 感情表現例

感情	感情語
喜	喜び, 悦び, 歓び, 嬉しい, 祝意
哀	悲しさ, 哀しさ, 感傷, 傷つく, 泣く
安	麗か, 緩む, 晴れやか, 吐息, 落ち着く
厭	不快, 嫌気, むかむか, 苦い, 自己嫌悪
驚	驚く, 衝撃, ショック, 動揺, 驚愕
好	情熱, 情, 愛情, 相思, 愛しい
恥	恥ずかしい, 面映い, 眩しい, 羞恥, 不面目
怒	怒り, 腹立たしさ, 業腹, 激怒, 憤怒
怖	不気味, 無気味, 不気味さ, 気味が悪い, 怖さ
昂	焦る, 焦り, 焦燥, やきもき, 緊張

この辞書を類義語辞書に単語をカテゴリとして登録し、TMSを用いて感情表現についての時系列単語頻度推移(割合)を月単位で行った。

分析—感情表現についての時系列単語頻度推移(割合)—

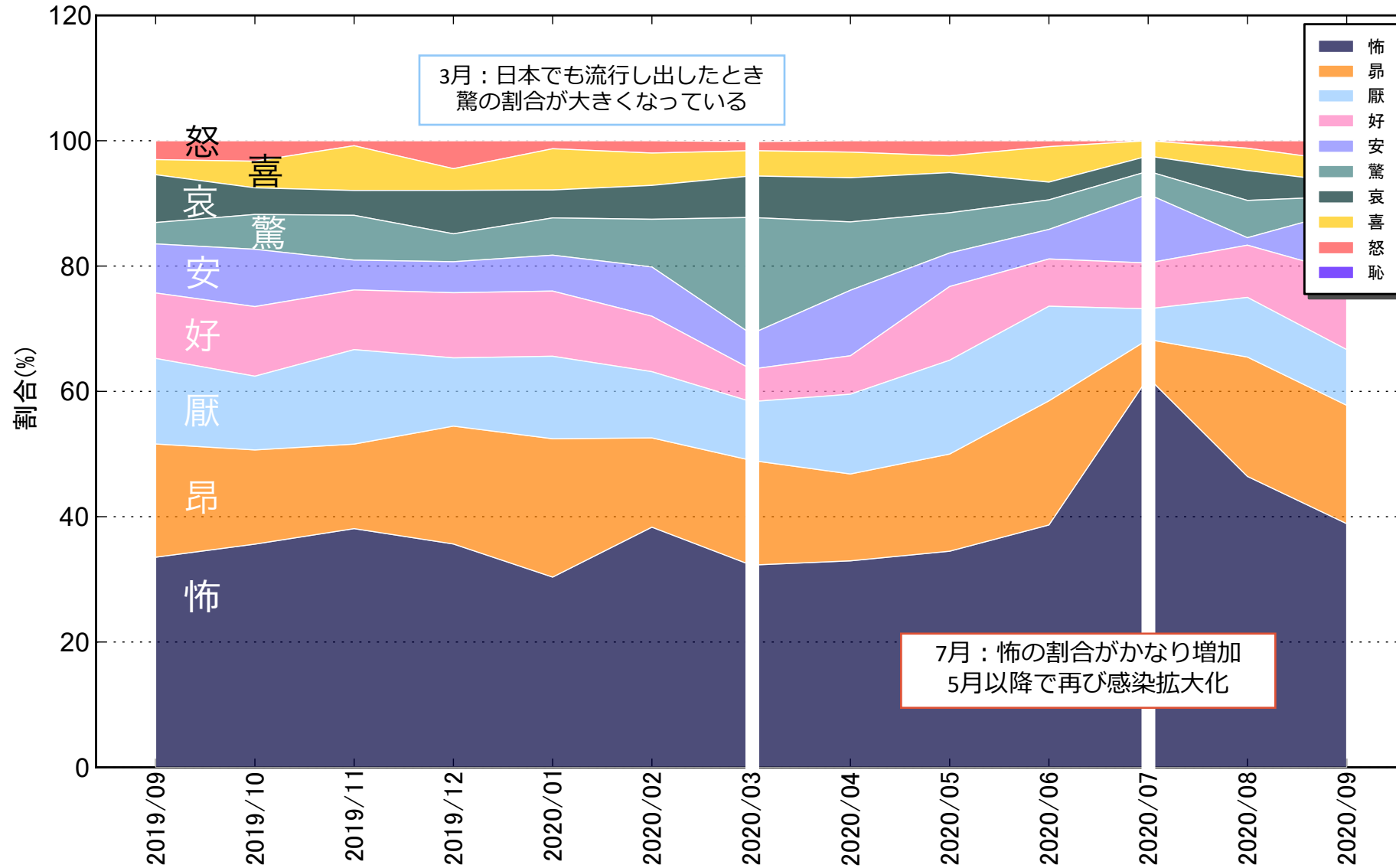


図7 感情表現についての時系列単語頻度推移(割合)

経済の変化についての指標として、**DI値（景気動向指数）**を導入する。

DI値とは、景気が「上向き」か「下向き」かという、**景気**の方向性を判定するための指標である。景気の影響を受けやすい28の指標のデータの中で、景気を先取りして動く「先行指数」、景気とほぼ一致して動く「一致指数」、景気に遅れて動く「遅行指数」の3つの系列に分類される。

データは、e-Statで公表されている令和2年10月分速報[3]の2019年9月から2020年9月の1年分のDI値を用いる。

表2 2019年9月から2020年9月のDI値

DI値	19-Sep	19-Oct	19-Nov	19-Dec	20-Jan	20-Feb	20-Mar	20-Apr	20-May	20-Jun	20-Jul	20-Aug	20-Sep
先行指数	27.3	18.2	36.4	54.5	36.4	45.5	18.2	18.2	9.1	18.2	72.7	100	100
一致指数	40	0	0	10	40	60	0	0	0	10	80	80	70
遅行指数	33.3	22.2	38.9	50	66.7	44.4	22.2	11.1	5.6	22.2	22.2	50	33.3

感情の変化と景気の動向の関係について調べるために、感情表現についての時系列単語頻度推移(割合)の表データを用いて、ピアソンの積率相関係数と無相関検定のp値を算出する。

結果は以下ようになった。

表3 感情推移割合とDI値の相関係数

DI値相関係数	先行指数	一致指数	遅行指数
怖	0.579	0.640	-0.072
昂	-0.001	-0.106	0.483
厭	-0.656	-0.658	-0.075
好	0.164	0.087	0.229
安	-0.044	0.113	-0.403
驚	-0.430	-0.452	-0.244
哀	-0.472	-0.430	-0.087
喜	-0.230	-0.283	0.424
怒	-0.011	-0.213	0.058
恥	-0.238	-0.274	-0.179

表4 感情推移割合とDI値の無相関検定p値

p値	先行指数	一致指数	遅行指数
怖	0.04*	0.02*	0.81
昂	1.00	0.73	0.09
厭	0.01*	0.01*	0.81
好	0.59	0.78	0.45
安	0.89	0.71	0.17
驚	0.14	0.12	0.42
哀	0.10	0.14	0.78
喜	0.45	0.35	0.15
怒	0.97	0.48	0.85
恥	0.43	0.37	0.56

「怖」の先行指数・一致指数と「厭」の先行指数・一致指数について相関があるとわかり、その2感情について、p値が有意水準5%を満たす結果となった。

「**怖**」について該当する単語が入る文章を参照した結果を示す。

「金融市場の動揺が止まらないのは、政策ツールの不十分さを**不安**視したのもでもある。」

「市場では一段の下振れを**懸念**する声もある。」

「**厭**」について該当する単語が入る記事を参照結果を示す。

「物価高騰への**不満**が高まっていることに配慮し、公務員給与を平均で15%引き上げることを盛り込む一方、税収を13%引き上げる方針を示した。」

「トランプ米大統領は自国のエネルギー産業の振興に熱心だが、有権者が**嫌う**ガソリン高を避けるため選挙前には原油価格の抑え込みに動く可能性が高い。」

ポジティブな感情については、「**好**」のDI値の先行指数と一致指数、「**安**」のDI値の一致指数、「**喜**」のDI値の遅行指数の相関係数が正の値を取っていた。

しかし、無相関の検定として有意にはならなかった。したがって、**データ期間の拡張や他の期間での検証を行うことで、ポジティブな感情表現が含まれる記事が増えると景気動向もプラスに働くという可能性がある**と考えられる。

まとめと今後の課題

まとめ

本研究では、新聞記事を用いてコロナウイルス大流行の前後でのトピックの変化と、記事内容の感情の変化と経済状況との関連について分析を行った。

- コロナ流行前と後では**大きくトピックが変化し**、コロナウイルスの流行による**経済への大きなダメージ**が反映されていることがわかった。
- ことばネットワークの結果から、コロナ流行前では複数のトピックについて散らばっている傾向が見られたが、コロナ流行後では、**コロナ関連のトピックが大きくなり吸収されている**ということがわかった。
- 感情表現とDI値の相関から、「怖」と「厭」という感情について、**経済の動向と感情表現に関連がある**ということがわかった。

今後の課題

感情表現とDI値の相関について、**他の期間で比較を行うことや、データ期間を拡張して行った際の変化**について調査すること、**ポジティブな感情表現が含まれる記事と景気動向の関連**についての調査を行いたい。

- [1] 経済産業省, 令和2年版通商白書,
<https://www.meti.go.jp/report/tsuhaku2020/2020honbun/i1100000.html>
2020年12月17日閲覧
- [2] 中村明, 感情表現辞典, 東京堂出版, 1993
- [3] e-Stat 政府統計の総合窓口, 景気動向指数 / 令和2年10月分 速報,
https://www.e-stat.go.jp/stat-search/files?page=1&layout=dataset&toukei=00100406&stat_infid=000032030532
2020年12月17日閲覧