

東日本大震災について報じられた ウェブニュースコーパスの基礎的解析

Basic Analysis of the Web News Corpus Broadcasted the 2011 Great East Japan Earthquake Disaster

佐藤 翔輔¹, 今村 文彦¹, 林 春男²

Shosuke SATO¹, Fumihiko IMAMURA¹ and Haruo HAYASHI²

¹ 東北大学大学院工学研究科附属災害制御研究センター

Disaster Control Research Center, Graduate School of Engineering, Tohoku University

² 京都大学 防災研究所

Disaster Prevention Research Institute, Kyoto University

The web information, especially mass media and social media information, have a more increasing role in create operational picture about a crisis since the Great East Japan earthquake and tsunami disaster. In this paper, we have developed and basically analyze web news corpus of the earthquake and tsunami disaster with text mining method intended to clear the whole picture. The corpus consists of over 60,000 web news articles which tagged its topic name. Sub corpora about Nuclear power plant accident, Iwate, Miyagi, Fukushima, Aomori, Ibaraki, and Chiba from original corpus, and analyzed by comparison of extracted keywords and their chronological behaviors.

Keywords: *the Great East Japan earthquake disaster, social reality, corpus analysis, text mining, mass media, web news, common operational picture*

1. はじめに

2011年3月11日に発生した東北地方太平洋沖地震(M9.0)は最大震度7を観測し、東北地方の太平洋側を中心とした大規模な津波によって沿岸部に壊滅的な被害をもたらしただけでなく、福島第1原子力発電所の炉心融解等の事故や、関東地方での大規模な液状化現象を発生させ、国内外に大きな影響をもたらしている。死者・行方不明者は2万人を超え(警察庁¹⁾、2011年5月24日現在)、「東日本大震災」という独自の震災名称で呼ばれるようになり、複数の都県が同時に被災した超広域的な地震・津波災害となった。

東日本大震災に関する対応の特徴として、ウェブ情報(インターネット情報)が大量に発信され、震災に関する状況認識の情報源として大きな役割を担ったことが挙げられる。効果的な災害対応を行うためには、被害・対応に関する状況認識を的確に行えるかどうか、対応の質を大きく左右することは既に広く知られている²⁾。インターネット人口普及率が6割を超えた2003年³⁾以降には、2004年新潟県中越地震や2007年新潟県中越沖地震などでも、被災自治体による被害・対応状況の発信、ポータルサイトや新聞社サイトからの報道記事の発信が盛んに行われるようになっただけでなく、災害の現況把握に特化したウェブサイトが出現していた⁴⁾。東日本大震災と、

これまでの地震災害とでは、ウェブ情報を取り巻く状況は、量・質の両面で大きく異なる。例えば、新潟県中越地震で配信されたYahoo!ニュースの記事件数は、半年間で約2,600件であったのに対して、東日本大震災では最初の2ヶ月間で4万件を超えている。著者らの調査では、本災害に関連するツイートは1ヶ月間で170万件に到達している。野村総合研究所が関東域居住者を対象に行った調査によれば、信頼性が向上したメディアとして、1位のNHKの情報に次いで、第2位がポータルサイトの情報と第3位がソーシャルメディアの情報となっている⁵⁾。

本稿では、東日本大震災について報道されたウェブニュースを解析することにより、マスメディアが東日本大震災をどのように社会に伝えたかについて明らかにすることを目的にする。前述の内容を踏まえれば、個人や組織の対応活動の前提となる災害の状況認識の形成を行う上では、ウェブ上の情報が何らかの影響を与えていることは否定できない。ソーシャルメディアが普及した昨今においても、依然としてマスメディアが基本的な情報源の一つになっていることは言うまでもない。すべての社会的な現象がそうであるように、東日本大震災に関する対応には様々な要因が影響していることが考えられる。過去の災害においても、マスメディアが発信した情報を対象とした実態分析が行われており、新聞を分析対象として、被災地内外間の報道の温度差やその内容の時系列

的な変遷が明らかにされている^{6) 7)}。本震災に対するウェブ上のマスメディアの報道実態を科学的に記述することは、災害の全体像を捉える上の一助になると考え、東日本大震災に関連するウェブニュースの基礎的な解析を行い、その傾向について論ずる。

2. 研究方法

(1) 東日本大震災ウェブニュースコーパスの構築

解析対象となるデータを収集するために、ウェブニュース記事を自動クロールによって収集とアーカイブを行い、「東日本大震災ウェブニュースコーパス」を構築している。コーパスとは言語分析のための文字言語、あるいは音声言語資料の集合体として定義されるものである⁸⁾。現在は、Yahoo!ニュースサイト上の「東日本大震災」に配信・掲載されたウェブニュース記事をクロールの対象としている。本サイトのニュース記事を対象にしたのは、1) ニュースポータルサイトであり、全国紙や地方紙の複数の新聞社のニュース記事が掲載されており、網羅性が高いこと、2) すべてのニュース記事の掲載フォーマットが統一されているために、記事のタイトル、新聞社、配信日時、本文のそれぞれの同定が正確に行えることが挙げられる。ウェブの数あるポータルサイトから、Yahoo!を対象にしたのは、ポータルサイトや各新聞社のニュースサイトの中で、最も我が国で閲覧されているサイトであるためである⁹⁾。また、Yahoo!ニュースに掲載された記事のアーカイブ化とその解析は、既往の研究で継続しており¹⁰⁾、高い網羅性や正確性が確認されていることから、本稿での収集・解析対象とした。

クロールによって収集されたウェブニュース記事のデータについて、タイトル、配信社、配信日時、本文、トピック名、URL に区分し、データベースへの蓄積を行い、「東日本大震災ウェブニュースコーパス」としている。トピック名とは、Yahoo!の「東日本大震災」に関連するニュースを一定の内容ごとに分けた区分の名称である。1つの災害について複数のトピックに別れて掲載されることは、今回の震災が初めてであり、それだけ記事の件数が膨大であり、かつ記事の内容や対象が多様であること反映したものだと推察される。表1にトピックの一覧と、震災後2ヶ月間(2011年5月12日0:00まで)のそれぞれの記事件数を示した。林¹¹⁾によれば、災害を受けた人や社会は10のべき乗時間を節目として、失見当(0~10時間)、被災地社会の成立(10~100時間)、災害ユートピア(100~1,000時間)、現実への帰還(1,000時間以降)といった4つのステージをたどるという。本稿で震災発生から2ヶ月間のニュース記事を対象

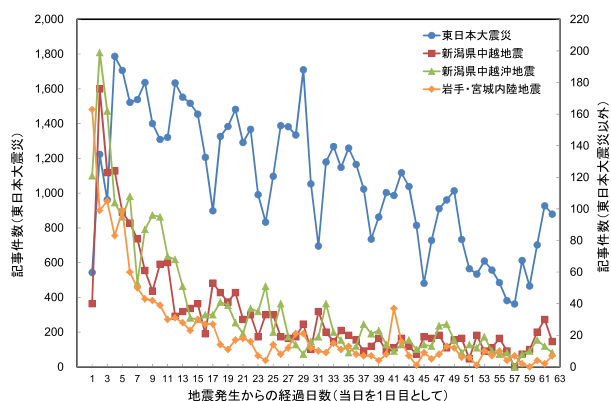


図1 東日本大震災と近年の地震災害のウェブニュース記事件数の推移 (Yahoo!ニュース)

表1 東日本大震災ウェブニュースコーパスの各トピックと記事件数 (Yahoo!ニュース)

No.	トピック名	件数	累積比率(%)
1	東北地方太平洋沖地震	23,639	23.5
2	電力	7,898	31.4
3	福島第1原発事故 ^{*1}	7,637	39.0
4	原子力	5,814	44.8
5	被災地、避難所の救援	5,498	50.2
6	宮城県震災被害	4,154	54.4
7	福島県震災被害	4,033	58.4
8	地震	3,625	62.0
9	被災地復興への取り組み	3,217	65.2
10	岩手県震災被害	2,720	67.9
11	福島原発周辺自治体の対応	2,285	70.2
12	福島原発による人的影響	2,124	72.3
13	東京電力の計画停電	1,599	73.9
14	募金、義援金	1,524	75.4
15	スポーツ選手による被災者支援	1,431	76.8
16	避難者の健康・生活	1,378	78.2
17	被災地の医療	1,262	79.4
18	被災住民の受け入れ	1,192	80.6
19	芸能人による被災者支援	1,181	81.8
20	災害ボランティア	1,059	82.8
21	茨城県の震災被害	1,043	83.9
22	被災地のライフライン	943	84.8
23	福島原発周辺からの住民避難	908	85.7
24	福島第1原発への関係機関の対応	841	86.6
25	宮城・石巻市の津波被害	781	87.3
26	交通情報	689	88.0
27	福島第1原発の安定化作業	614	88.6
28	放射性物質の拡散調査	593	89.2
29	岩手・陸前高田市の津波被害	574	89.8
30	海外からの災害支援	567	90.4
31	青森県の震災被害	553	90.9
32	福島原発事故の被害補償	518	91.4
33	宮城・南三陸町の津波被害	514	91.9
34	千葉県の震災被害	444	92.4
35	信越地方で強い地震相次ぐ	438	92.8
36	節電の取り組み	434	93.2
37	岩手・釜石市の津波被害	362	93.6
38	岩手・大船渡市の津波被害	345	94.0
39	岩手・宮古市の津波被害	342	94.3
40	食品への放射性物質の影響	319	94.6
41	津波	317	94.9
42	水への放射性物質の影響	317	95.2
43	宮城・気仙沼市の津波被害	316	95.6
44	福島・いわき市の震災被害	313	95.9
45	原発事故による第一次産業への風評被害	246	96.1
46	外国人、外資系企業の日本退避の動き	219	96.3
47	安否情報	206	96.5
48	被災者の心のケア	183	96.7
49	岩手・大槌町の津波被害	182	96.9
50	東北地方太平洋沖地震による住民孤立	180	97.1
51	東北地方太平洋沖地震の科学的分析	176	97.3
52	震災被災者の住居・仮設住宅	176	97.4
53	液状化現象による被害	169	97.6
54	宮城・東松島市の津波被害	168	97.8
55	東北電力の計画停電	167	97.9
56	宮城・名取市の津波被害	162	98.1
57	宮城・仙台市の津波被害	140	98.2
58	岩手・山田町の津波被害	139	98.4
59	福島・双葉町の集団避難	139	98.5
60	被災地のペットに関する情報	126	98.6
61	福島・南相馬市の津波被害	117	98.7
62	宮城・多賀城市の津波被害	111	98.9
63	被災地のがれき処理	110	99.0
64	災害時の在留外国人支援	103	99.1
65	震災便乗詐欺	101	99.2
66	宮城・山元町の津波被害	87	99.3
67	福島・相馬市の津波被害	85	99.3
68	震災犠牲者の埋葬	84	99.4
69	宮城・亘理町の津波被害	80	99.5
70	宮城・女川町の津波被害	78	99.6
71	原子力政策の見直し	77	99.7
72	海外の原子力政策見直し	76	99.7
73	震災犠牲者の身元確認	70	99.8
74	ガス業界	53	99.9
75	震災孤児	51	99.9
76	被災地の燃料不足	38	99.9
77	計画停電の在宅療養への影響	29	100.0
78	震災関連の迷惑メール	27	100.0

*1旧トピック名:大地震の福島原発への影響

にしたのは、速報性の観点から論文執筆時に収集・分析が可能な最長期間のデータであったことのほか、上記のすべてのステージを含むデータであるためである（震災後 2 ヶ月間は、約 1,500 時間）。なお、今後も継続して配信されるであろう震災後 2 ヶ月以降のニュース記事を含めた分析は今後の課題としたい。

図 1 に、東日本大震災のほか、新潟県中越地震、新潟県中越沖地震、岩手・宮城内陸地震の地震発生後 2 ヶ月間の記事件数の日変化を示した。なお、東日本大震災については、複数のトピックに掲載された記事を 1 件としてカウントしている。図 1 を見ると、過去の地震災害の記事件数は、地震発生から約 2 週間で急激に減少し、その後 0~30 件の間を推移している。それに対して、東日本大震災の記事件数は毎日の減少傾向が緩やかであり、2 ヶ月後現在でも 400~600 件程度に推移している。なお、東日本大震災の記事件数に周期的に減少する傾向が見られる日は日曜日であることを付記しておく。

(2) 解析サブコーパスの作成

本稿では、コーパス中の全記事に対して解析を行うと同時に、いくつかのサブコーパスを作成し、それらを対象にした解析も行う。コーパスの中から、任意の属性やフリーワードで抽出された記事によって作成される、元のコーパスよりも小さいコーパスをサブコーパスとして定義する。

東日本大震災は、福島第 1 原子力発電所の事故やその影響に関する記事が多いため、原発に関連するトピックの記事を対象にしたサブコーパスを作成する。また、今回の震災は複数の県域をまたがっており、各県に関連する報道に着目することは一定の価値がある。原発事故に関連するトピック、被災自治体に関連するトピックから、次の 7 つのサブコーパスを作成した：

- 1) 原子力発電所に関する被害・対応（原発：No. 3, 4, 11, 12, 23, 24, 27, 28, 32, 40, 45, 71, 72）
- 2) 岩手県の被害・対応（岩手県：No. 10, 29, 37, 38, 39, 49）
- 3) 宮城県の被害・対応（宮城県：No. 6, 25, 33, 44, 54, 56, 57, 62, 66, 69）
- 4) 福島県の被害・対応（福島県：No. 7, 44, 59, 61, 67）
- 5) 青森県の被害・対応（青森県：No. 31）
- 6) 茨城県の被害・対応（茨城県：No. 21）
- 7) 千葉県の被害・対応（千葉県：No. 34）

かつこの中は、トピック名を示す No. であり、表 1 のものに対応している。以上と併せて、元のコーパス（全体、No. 1~78）を実際の解析対象にする。

図 2 に元コーパスとサブコーパスについて、発災からの経過日数と毎日の記事件数の推移を示した。なお、元コーパス内、サブコーパス内で複数のトピックに重複して掲載された記事は 1 件としている。それぞれの合計の記事件数は、全体が 66,042 件、原発が 17,362 件、岩手が 3,298 件、宮城が 5,118 件、福島が 4,342 件、青森が 553 件、茨城が 1,043 件、千葉が 444 件であった。全体と原発は似たような増加・減少の傾向を示しており、原発が全体の記事件数に及ぼす影響が大きいことが想像される。なお、4/8 は前日に宮城県沖で発生した最大震度 6 強の余震(M7.1)による影響、5/11 はそれぞれ震災後 2 ヶ月を迎えたことによる影響によって記事件数が増加したものである。

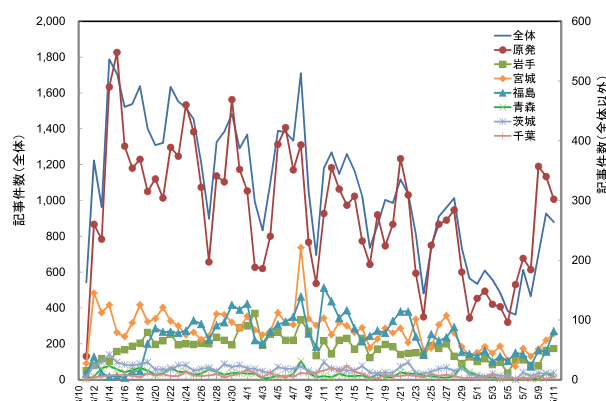


図 2 元コーパスとサブコーパスのウェブニュース記事件数の推移

3. 解析方法

(1) 解析アプローチ

本稿では、ウェブニュースコーパスをテキストマイニングのアプローチによって分析を進めていく。テキストデータを分析する方法には、KJ 法、コーディング、量的内容分析、テキストマイニングがある¹²⁾。前 3 つの分析方法は、災害報道の新聞記事进行分析した研究^{6) 7)}のほか、これまで各方面の研究で有用な知見を得てきた優れた方法であるものの、解析結果が分析者や評価者の経験や能力に大小の影響を受けることも指摘されている。一方、テキストマイニングは、解析の過程に人による介入がほぼない機械的な処理によって行われる再現性の高い分析手法である。解析結果を得るまで過程に、分析者の個体差による影響を可能な限り排除することを意図として、同アプローチを採用して分析することにする。

(2) 単語の重み付け指標

テキストマイニングでは、単語への重み付けによって絞り込みを行い、上位になった単語をキーワードとし、詳細な分析を行うことが一般的である。ここでは、テキストマイニングにおける単語への重み付け指標として、1) 出現頻度、2) 特異値 D^{13) 14)}を用いることにする。前者は、重み付け指標として最も単純かつ基礎的な指標である。一方で、言語情報学においては、重要な単語の使用頻度は中程度であると言われている¹⁵⁾のに対して、出現頻度による重み付けでは高頻度の単語に高い重みを与えることになる。このような問題を受けて、佐藤ら^{13) 14)}は時系列的に増加するコーパスに対して、時間断面ごとに特徴的な単語に高い重みを与える特異値 D を提案している。特異値 D は他の一般的な重み付け指標に比べて、中程度の頻度で使われる単語に高い重みを与える。本研究では、出現頻度という基礎的な指標によって高頻度の単語を抽出するほか、特異値 D を用いてキーワードらしい特徴的な単語を抽出することにした。

各コーパスの記事本文を形態素解析器¹⁶⁾で形態素への分かち書きを行い、各形態素に対して両指標での重み付けを行う。なお、便宜上、本稿では形態素を単語と呼ぶことにする。出現頻度は、コーパス内の各単語が出現した回数である。災害過程は 10 のべき乗時間ごとに状況が質的な変化をすることから^{17) 18)}、特異値 D の算定における解析時間幅を、発災から 1~10 時間における記事は 1 時間、10~100 時間は 10 時間、100~1,000 時間は 100 時間、1,000 時間以降を 100 時間とした。時間断面ご

表 2 各コーパスにおける出現頻度または累積特異値が高い単語

出現頻度					累積特異値SD				
	10時間	100時間	1,000時間	1,500時間		10時間	100時間	1,000時間	1,500時間
東日本大震災	1 地震(1,012)	地震(5,485)	避難(48,850)	避難(6,0293)	1 小学校(196)	募金(1,330)	選挙(7,534)	選挙(11,420)	
	2 午後(501)	午後(4,347)	福島(46,170)	福島(58,318)	2 施設(181)	グループ(837)	賠償(2,948)	賠償(7,169)	
	3 宮城(434)	福島(4,234)	原発(40,030)	原発(51,976)	3 原子力(170)	爆発(521)	ベクレル(5,150)	ベクレル(6,356)	
	4 津波(403)	津波(3,219)	被災(34,526)	被災(42,304)	4 同社(162)	運休(509)	候補(3,521)	候補(5,698)	
	5 東京(360)	原発(3,196)	東日本大震災(32,387)	東日本大震災(41,790)	5 発電(157)	水素(506)	容疑(2,487)	容疑(5,014)	
	6 発生(317)	午前(3,095)	支援(31,727)	支援(39,698)	6 協力(152)	ボランティア(468)	現職(2,813)	現職(4,867)	
	7 福島(316)	県(3,083)	県(30,015)	県(36,714)	7 冷却(142)	京都(464)	投票(2,832)	投票(4,824)	
	8 沖(312)	停電(3,012)	津波(25,410)	津波(32,336)	8 空港(131)	皆さん(437)	陛下(1,100)	陛下(4,109)	
	9 東北(279)	避難(2,943)	被災地(25,125)	被災地(31,673)	9 原発(128)	義援金(424)	議席(1,899)	議席(3,963)	
	10 被害(270)	被害(2,784)	地震(23,983)	被害(28,597)	10 委員(127)	被ばく(418)	新人(2,298)	新人(3,963)	
新潟県中越地震	1 地震(148)	地震(757)	避難(1,601)	避難(1,827)	1 震度(77)	同市(188)	ダム(424)	ダム(567)	
	2 新潟(121)	新潟(707)	地震(1,547)	地震(1,678)	2 震源(19)	断層(179)	ボランティア(330)	ボランティア(543)	
	3 震度(79)	長岡(466)	新潟(1,497)	新潟(1,632)	3 停電(17)	電話(147)	入居(267)	入居(542)	
	4 午後(63)	避難(427)	長岡(1,403)	長岡(1,625)	4 確認(14)	雨(146)	優太(258)	復興(319)	
	5 小千谷(45)	小千谷(425)	新潟県中越地震(1,279)	新潟県中越地震(1,540)	5 小千谷(14)	死亡(144)	学校(220)	児童(319)	
	6 発生(45)	震度(401)	震度(1,030)	小千谷(1,120)	6 市内(14)	自宅(140)	児童(220)	義援金(318)	
	7 長岡(42)	余震(357)	小千谷(992)	震度(1,093)	7 脱線(13)	レール(123)	優(210)	生徒(298)	
	8 地方(33)	午後(314)	県(980)	県(1,061)	8 観測(12)	ガル(116)	義援金(195)	募金(287)	
	9 確認(27)	新潟県中越地震(284)	余震(950)	被災(1,011)	9 気象庁(11)	IC(114)	仮設(183)	優太(283)	
	10 脱線(26)	発生(265)	被災(818)	余震(989)	10 新幹線(10)	ボランティア(111)	村長(181)	学校(282)	
新潟県中越沖地震	1 新潟(145)	柏崎(856)	柏崎(2,134)	柏崎(2,326)	1 道路(41)	電話(223)	相談(437)	相談(659)	
	2 柏崎(119)	地震(615)	地震(1,489)	地震(1,786)	2 市内(37)	耐震(142)	二塁打(335)	訓練(528)	
	3 地震(93)	新潟(572)	原発(1,056)	県(1,206)	3 ガス(33)	東電(128)	試合(325)	観光(444)	
	4 震度(78)	県(418)	県(1,029)	原発(1,194)	4 家屋(28)	ボランティア(125)	海水浴(272)	試合(409)	
	5 長野(59)	刈羽(296)	新潟(1,022)	原発(1,103)	5 倒壊(26)	物資(123)	観光(262)	入居(408)	
	6 発生(44)	避難(290)	被災(923)	被災(1,068)	6 女性(26)	判定(120)	選手(251)	選手(388)	
	7 午前(35)	原発(285)	被害(867)	被害(1,016)	7 運転(23)	仮設(112)	新(247)	生産(351)	
	8 午後(35)	被害(283)	避難(839)	避難(974)	8 今回(23)	知事(109)	生産(243)	海水浴(340)	
	9 上越(30)	発生(278)	刈羽(815)	刈羽(890)	9 役場(22)	ガル(108)	入居(219)	二塁打(335)	
	10 刈羽(30)	被災(277)	調査(673)	調査(769)	10 余震(19)	活断層(106)	花火(205)	花火(295)	
岩手・宮城内陸地震	1 地震(249)	地震(757)	地震(1,868)	地震(2,013)	1 断層(110)	温泉(223)	耐震(534)	耐震(823)	
	2 岩手(190)	宮城(574)	宮城(1,231)	栗原(1,346)	2 被害(86)	集落(203)	住宅(399)	住宅(600)	
	3 宮城(141)	岩手(518)	栗原(1,208)	宮城(1,286)	3 状況(72)	孤立(202)	断層(369)	選手(468)	
	4 震度(137)	栗原(447)	岩手・宮城内陸地震(1,160)	岩手・宮城内陸地震(1,279)	4 災害(62)	捜索(193)	選手(311)	募金(403)	
	5 午前(90)	岩手・宮城内陸地震(374)	被害(1,006)	被害(1,086)	5 話す(59)	旅館(180)	学校(299)	仮設(396)	
	6 揺れ(90)	被害(307)	岩手(996)	岩手(1,027)	6 起きる(57)	地区(175)	募金(281)	学校(382)	
	7 発生(80)	震度(298)	地区(837)	県(936)	7 揺れ(54)	泥(149)	相談(279)	断層(369)	
	8 震源(79)	発生(294)	県(825)	地区(905)	8 電話(52)	支援(145)	野球(239)	キャンセル(330)	
	9 被害(76)	災害(294)	避難(766)	避難(819)	9 運転(50)	断層(136)	仮設(239)	野球(318)	
	10 秋田(71)	温泉(262)	発生(748)	災害(799)	10 同市(47)	避難(129)	ボランティア(234)	大会(317)	

2つ以上の災害で上位になった単語

とに求まる特異値 D と、コーパスにおける各単語の特異値 D を総和して求める累積特異値 $\Sigma D^{19)}$ も計算する。

4. 東日本大震災と過去の地震災害のウェブニュースコーパスとの比較分析

出現頻度と累積特異値 ΣD を計算し、10 時間、100 時間、1,000 時間、1,500 時間のそれぞれ上位 10 位になった単語（キーワード）を災害別に分けて表 2 のように示した。同一の指標で 2 つ以上の災害で上位になった単語については、網かけをしている。各単語のかっこ付の右側の数値は、それぞれの重み付け指標の値を示している。

4 つの地震災害において、出現頻度が上位の単語についての明確な差は認められない。一部、単一の災害でしか上位になっていない単語もあるが、多くは被災地域の名称であり、現象や内容に関する差異ではない。一方、累積特異値が高い単語については、東日本大震災と、それ以外の過去の地震災害とでは、その傾向に差が見られる。1,000 時間、1,500 時間では、過去の地震災害において「仮設」「入居」「ボランティア」「相談」「募金」などが共通しているのに対して、東日本大震災では共通する単語が一つもない。過去の地震災害では、以上のよ

うに復旧・復興に関連する単語が特徴的になっているにも関わらず、東日本大震災では「選挙」「候補」「現職」「投票」「議席」「新人」などの選挙に関連する単語が特徴的になっている。また、10 時間では、過去の 3 つの地震災害で共通する単語はないが、中越では「震度」「震源」、中越沖では「余震」、岩手・宮城内陸では「断層」「揺れ」などのハザードに関連する単語が上位になっているのに対して、同時間区分における東日本大震災の単語にハザードに関連すると思われる単語は上位に位置していない。

ここまでに見てきたように、東日本大震災に関するウェブニュースを通した報道の質は、これまでの地震災害と大きく異なることがわかった。次章では、東日本大震災のウェブニュースコーパスのみに着目した解析を行い、過去の地震災害と異なっていた傾向の内容について明らかにする。

5. 東日本大震災のサブコーパスに基づいた比較分析

(1) コーパス別の単語の比較

表 3 各コーパスにおける出現頻度上位 30 位の単語とその比較

	全体	原発	福島県	岩手県	宮城県	青森県	茨城県	千葉県
複数のコーパスで上位になったキーワード	1.避難(60,293)	5.避難(13,983)	3.避難(6,658)	1.避難(5,316)	1.避難(7,416)	15.避難(289)	6.避難(746)	7.避難(347)
	2.福島(58,318)	2.福島(30,886)	1.福島(9,480)	9.福島(1,834)	9.福島(2,761)	4.福島(392)	2.福島(991)	15.福島(232)
	7.県(36,714)	13.県(7,938)	8.県(2,057)	7.県(2,180)	12.県(2,612)	3.県(402)	4.県(975)	4.県(430)
	10.被害(28,597)	29.被害(5,675)	11.被害(1,947)	6.被害(2,246)	5.被害(3,705)	8.被害(337)	3.被害(980)	2.被害(543)
	14.受ける(26,253)	16.受ける(7,226)	15.受ける(1,710)	13.受ける(1,528)	13.受ける(2,455)	17.受ける(226)	12.受ける(544)	16.受ける(225)
	9.被災地(31,673)			14.被災地(1,491)	16.被災地(2,275)	20.被災地(223)	24.被災地(330)	20.被災地(192)
	4.被災(42,304)		9.被災(1,989)	4.被災(3,044)	4.被災(4,124)	12.被災(308)	9.被災(628)	11.被災(304)
	5.東日本大震災(41,790)		6.東日本大震災(2,348)	5.東日本大震災(2,369)	6.東日本大震災(3,619)	2.東日本大震災(403)	5.東日本大震災(772)	6.東日本大震災(368)
	8.津波(32,336)		5.津波(2,410)	2.津波(4,709)	3.津波(6,439)	13.津波(304)	14.津波(454)	5.津波(371)
	12.地震(27,136)		10.地震(1,958)	10.地震(1,582)	8.地震(3,018)	24.地震(208)	7.地震(638)	10.地震(307)
	13.震災(26,338)		17.震災(1,623)	12.震災(1,548)	11.震災(2,616)	10.震災(317)	10.震災(585)	13.震災(272)
	15.話す(23,189)		25.話す(1,439)	11.話す(1,553)	17.話す(2,249)		16.話す(426)	17.話す(210)
	17.宮城(21,939)		23.宮城(1,484)	8.宮城(1,939)	2.宮城(7,104)	7.宮城(362)	23.宮城(355)	
	23.月(19,695)	22.月(6,374)	18.月(1,621)	30.月(928)	22.月(1,597)			
	3.原発(51,976)	1.原発(34,085)	2.原発(6,797)				13.原発(498)	30.原発(148)
	6.支援(39,698)			15.支援(1,426)	18.支援(2,112)	26.支援(191)	18.支援(396)	
	11.事故(28,302)	3.事故(18,227)	4.事故(3,585)				28.事故(301)	
	29.生活(16,981)		22.生活(1,491)	19.生活(1,273)	21.生活(1,773)			
	18.復興(21,906)			23.復興(1,141)	20.復興(1,810)	18.復興(224)		
	19.行う(21,467)	25.行う(5,844)					22.行う(356)	18.行う(206)
	21.午後(19,849)	24.午後(5,912)					15.午後(430)	29.午後(158)
	27.影響(17,704)	21.影響(6,488)				30.影響(180)	21.影響(362)	
		28.住民(5,737)	7.住民(2,257)	27.住民(1,065)	26.住民(1,469)			
	22.東京電力(19,814)	4.東京電力(14,187)	14.東京電力(1,724)					
	25.作業(19,144)	9.作業(10,192)	16.作業(1,643)					
	28.原子力(17,270)	6.原子力(12,385)	12.原子力(1,944)					
	16.人(22,450)		19.人(1,617)		14.人(2,422)			
	30.県内(16,494)						11.県内(548)	28.県内(159)
				3.岩手(4,228)	15.岩手(2,282)	9.岩手(329)		
	26.対策(18,133)	20.対策(6,505)						
		7.東電(10,722)	21.東電(1,615)					
		11.安全(8,694)	26.安全(1,425)					
		14.放射線(7,705)	27.放射線(1,422)					
		15.政府(7,260)	20.政府(1,616)					
		23.対応(6,068)	24.対応(1,474)					
		27.首相(5,756)	13.首相(1,874)					
				21.不明(1,247)	19.不明(1,825)			
				26.行方(1,122)	23.行方(1,586)			
					27.再開(1,396)	5.再開(390)	17.再開(423)	
					7.仙台(3,090)	14.仙台(294)		
						16.東京(253)		19.東京(197)
						21.復旧(219)	20.復旧(374)	
						23.3月(209)	26.3月(306)	
						25.予定(194)	25.予定(321)	
						29.4月(181)	19.4月(377)	
					28.発生(1,379)		27.発生(301)	
			30.出る(1,297)					25.出る(163)
				17.住宅(1,278)				26.住宅(161)
				28.自宅(1,012)	25.自宅(1,515)			
単一のコーパスで上位になったキーワード	20.日本(20,428)	8.放射(10,235)	28.相馬(1,395)	16.宮古(1,333)	10.石巻(2,650)	1.青森(775)	1.茨城(1,271)	1.千葉(665)
	24.活動(19,329)	10.発電(9,000)	29.圏内(1,341)	18.陸前高田(1,277)	24.気仙沼(1,566)	6.八戸(388)	8.水戸(631)	3.選挙(475)
		12.物質(7,986)		20.釜石(1,258)	29.流す(1,375)	11.運輸(309)	29.午前(287)	8.浦安(343)
		17.原子(7,040)		22.電話(1,162)	30.南三陸(1,367)	19.東北(223)	30.対策(284)	9.液状化(310)
		18.水(7,012)		24.流す(1,125)		22.運行(210)		12.旭(290)
		19.電力(6,911)		25.大船渡(1,125)		27.盛岡(187)		14.市(247)
		26.可能(5,761)		29.大槌(937)		28.東北新幹線(187)		21.投票(189)
		30.発表(5,678)						22.実施(171)
								23.県議(166)
								24.市長(164)
								27.運営(160)

震災から 2 ヶ月間のコーパスに対して、出現頻度と累積特異値 ΣD を計算し、上位 30 位になった単語（キーワード）をそれぞれ、表 3、表 4 のように整理した。表の上部は、複数のコーパスで上位になった単語で、上位になったコーパスの種類数が多い単語を上から順に同一行上に記載している。それ以外の単語については表の下部に出現頻度順に列挙した。各単語のかっこ付の右側の数値は、それぞれの重み付け指標の値を示している。

表 3 と表 4 を見ると、すべてのコーパスや多くのコーパスに共通して見られる単語のほか、2～3 つのコーパスだけに共通して見られる単語が存在することが分かる。後者として、1) 全体と原発と福島県、2) 岩手県と宮城県、3) 青森県と茨城県と千葉県、のそれぞれが似たような単語が上位になっている例が挙げられる。1) 全体と原発と福島県では、「東京電力」「作業」「原子力」、原発と福島県では「東電」「安全」「放射線」「政府」「対応」「首相」が共通で高い頻度を示している。累積特異値でも、全体と原発が「トレンチ」「放水」「ロボット」、原発と福島県が「損害」「観光」「計画的避難区域」「捜索」「土壌」が共通して高い値をしている。2) 岩手県と宮城県では、「不明」「行方」が共通して高

い出現頻度を示しており、「支所」「雇用」「診療」「放送」「交付」が共通して高い累積特異値を示している。3) 青森県と茨城県では、「再開」「復旧」「3月」「予定」「4月」が共通して高い出現頻度を示しており、青森県と千葉県では「住宅」「県議」、茨城県と千葉県では「漁協」「規制」「水揚げ」「営業」「農家」が共通して高い累積特異値を示している。以上をまとめると、東日本震災を県域に分けて報道したウェブニュースは、それぞれ 1) 原発事故影響を強く受けた地域（福島県）、2) 津波被害が甚大だった地域（岩手県、宮城県）、3) その他の被害を受けた地域（青森県、茨城県、千葉県）という 3 つのグループの存在が浮かび上がってくる。

(2) 単語の時系列なふるまい

本節では、単語の時系列なふるまいについての解析を行う。また、紙面の関係から記事件数が少なかった青森県、茨城県、千葉県を除く、全体、原発、岩手県、宮城県、福島県についての解析と考察を行う。

各コーパスについて、震災から 10 時間、100 時間、1,000 時間、1,500 時間（2 ヶ月後）のそれぞれの段階で累積特異値を算出し、上位 10 位になった単語についてそ

表 4 各コーパスにおける累積特異値上位 30 位の単語とその比較

	全体	原発	福島県	岩手県	宮城県	青森県	茨城県	千葉県
複数のコーパスで上位になったキーワード	1. 選挙(11,420)	2. 選挙(4,167)	3. 選挙(990)	1. 選挙(729)	7. 選挙(660)	1. 選挙(681)	5. 選挙(395)	1. 選挙(704)
	8. 陸下(4,110)		8. 陸下(719)	3. 陸下(690)	2. 陸下(1,040)		11. 陸下(275)	4. 陸下(239)
	3. ベクレル(6,356)	3. ベクレル(2,753)	7. ベクレル(759)				1. ベクレル(644)	13. ベクレル(145)
	7. 投票(4,825)			15. 投票(364)		3. 投票(199)	24. 投票(216)	5. 投票(206)
	18. ホウレンソウ(2,999)	13. ホウレンソウ(1,510)					2. ホウレンソウ(597)	21. ホウレンソウ(106)
		14. 出荷(1,485)	26. 出荷(400)				7. 出荷(346)	6. 出荷(191)
	2. 賠償(7,169)	1. 賠償(4,710)	2. 賠償(1,317)					
	14. 村(3,465)	4. 村(2,593)	1. 村(1,343)					
	21. 牛(2,596)	10. 牛(1,711)	5. 牛(948)					
	4. 候補(5,698)	6. 候補(2,049)				2. 候補(269)		
	15. 勝俣(3,251)	5. 勝俣(2,222)						18. 候補(118)
			29. 首相(391)	6. 首相(501)	1. 首相(1,168)			
						30. 震度(101)	27. 震度(210)	7. 震度(177)
					9. 選手(526)		12. 選手(269)	8. 選手(177)
	22. トレンチ(2,561)	7. トレンチ(1,966)						
	28. 放水(2,402)	22. 放水(1,303)						
	30. ロボット(2,366)	16. ロボット(1,471)						
	5. 容疑(5,015)		4. 容疑(957)					
	24. ペット(2,537)		21. ペット(421)					
	25. 当選(2,477)					16. 当選(134)		
	16. 選管(3,201)							2. 選管(288)
	29. 小沢(2,397)		22. 小沢(418)					
	10. 新人(3,963)							25. 新人(101)
		9. 損害(1,737)	6. 損害(762)					
		12. 観光(1,520)	14. 観光(462)					
		21. 計画的避難区域(1,305)	11. 計画的避難区域(568)					
		25. 捜索(1,263)	12. 捜索(507)					
		28. 土壌(1,162)	30. 土壌(387)					
		18. 漁協(1,445)			26. 漁協(383)			
		20. コウナゴ(1,325)					6. コウナゴ(391)	
		27. 野菜(1,163)					3. 野菜(440)	
				2. 支所(691)	5. 支所(706)			
				5. 雇用(585)	8. 雇用(605)			
				11. 診療(395)	12. 診療(462)			
				18. 放送(356)	30. 放送(378)			
				28. 交付(315)	27. 交付(381)			
			19. 児童(423)	12. 児童(392)				
			27. 授業(396)	19. 授業(353)				
			28. 処理(395)		23. 処理(391)			
					25. 往復(387)	9. 往復(156)		
					24. 募金(388)		17. 募金(246)	
			24. 客(414)			12. 客(148)		
						28. ボランティア(104)	18. ボランティア(245)	
						27. 住宅(108)		15. 住宅(138)
						29. 県議(104)		23. 県議(103)
							8. 漁協(326)	30. 漁協(90)
							10. 規制(307)	16. 規制(128)
							19. 水揚げ(243)	10. 水揚げ(158)
							23. 営業(216)	20. 営業(107)
							29. 農家(197)	26. 農家(99)
							4. 風評(429)	3. 発電(248)
							9. 検出(310)	9. 自衛(162)
単一のテーマで上位になったキーワード	6. 現職(4,867)	8. 乳児(1,865)	9. 清水(575)	4. 障害(627)	3. ランドセル(746)	4. 自民(177)		
	9. 議席(3,964)	11. 資金(1,525)	10. 警戒(572)	7. 防潮堤(501)	4. カキ(730)	5. 運行(172)		
	11. 定数(3,903)	15. ビット(1,478)	13. 作付け(479)	8. 船(437)	6. 養殖(676)	6. 民主(163)	13. 黨(267)	11. 電力(148)
	12. 連立(3,899)	17. 窒素(1,471)	15. 釈放(457)	9. 火葬(420)	10. 集団(518)	7. 全職(160)	14. 店舗(264)	12. 市長(148)
	13. 現(3,649)	19. 流出(1,390)	16. 指針(441)	10. 社員(397)	11. 回答(464)	8. 応援(159)	15. 検査(259)	14. 高い(145)
	17. 動物(3,034)	23. 浄水(1,285)	17. 支部(439)	13. 廃棄(374)	13. 七ヶ浜(458)	10. 新幹線(152)	16. 市場(253)	17. 基準(121)
	19. 国債(2,980)	24. JCO(1,281)	18. 保護(424)	14. 入所(367)	14. 校長(450)	11. 水産(150)	20. 漁業(239)	19. 教授(115)
	20. 陣営(2,618)	26. 伊方(1,184)	20. 村長(421)	16. 及川(359)	15. 労働(448)	13. 原発(140)	21. 液状化(225)	22. 需要(104)
	23. 増税(2,560)	29. 自主避難(1,137)	23. 延期(416)	17. 撤去(358)	16. こいのぼり(440)	14. 局長(139)	22. 相談(218)	24. 香取(102)
	26. 無所属(2,465)	30. 浜岡原発(1,13.7)	25. 米(410)	20. 田老(346)	17. 神戸(437)	15. 弘前(135)	25. 公園(213)	27. 東電(98)
	27. 犬(2,456)			21. 事業(344)	18. 卒業生(415)	17. バス(133)	26. 生産(212)	28. 鎌子(97)
				22. 介護(342)	19. 義援金(404)	18. 立候補(130)	28. 魚(201)	29. 調査(90)
				23. 配分(340)	20. 島(400)	19. 電源(128)	30. 採取(197)	
				24. 視察(339)	21. 組合(398)	20. 大会(126)		
				25. 店(331)	22. 桜(393)	21. 余震(118)		
				26. 内定(327)	28. 本吉(379)	22. 政策(117)		
				27. 漁港(317)	29. 小野寺(378)	23. まつり(115)		
				29. 松(312)		24. 盛岡(111)		
				30. 教諭(304)		25. 試合(110)		
						26. 入居(110)		

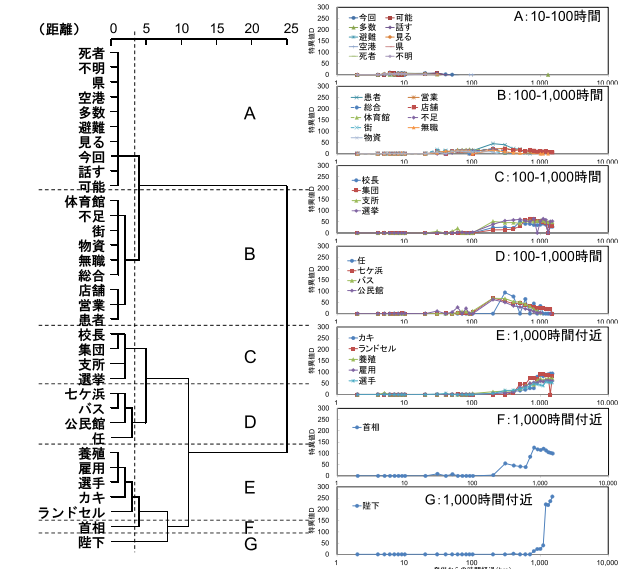
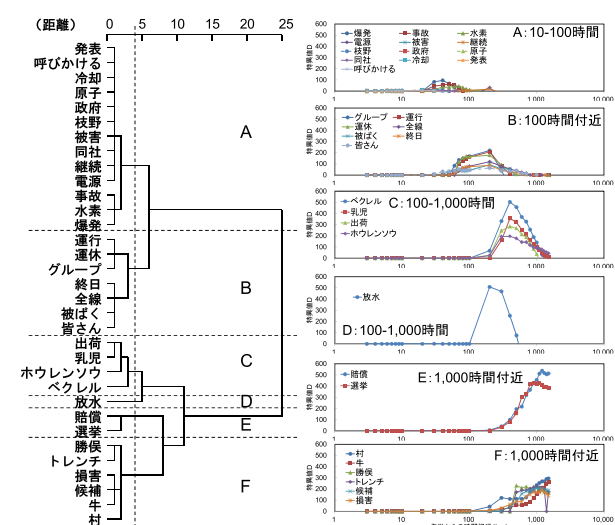
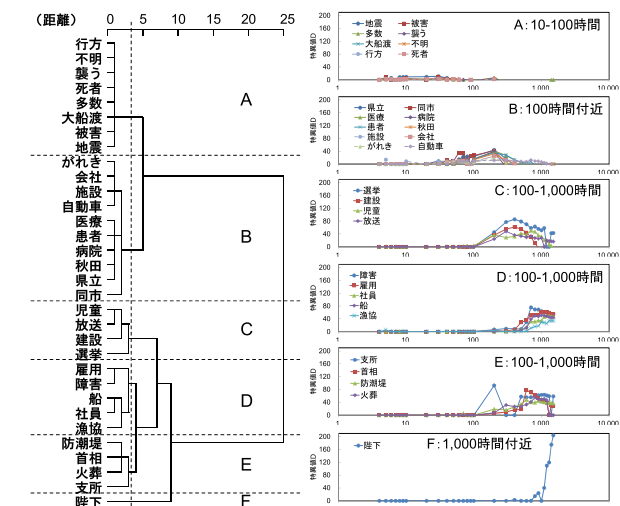
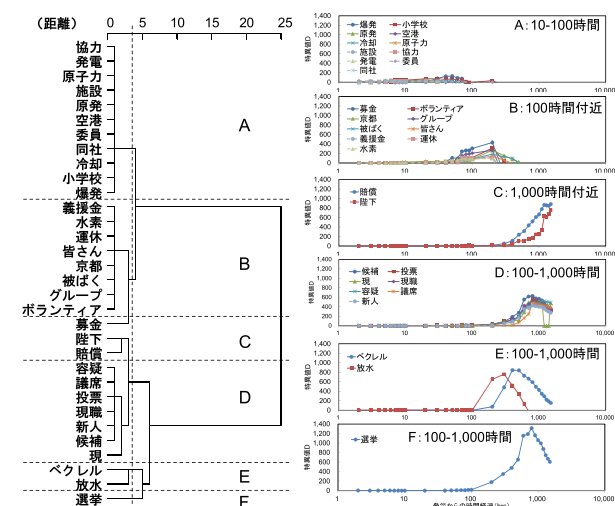
の時系列なふるまいを観察する。発災から 2 ヶ月後の段階で重みが高くなった単語を対象にすれば、発生から間もない頃や、その半ばで高い重みを示した単語のふるまいを適切に捉えることができない。このような問題を解消するために、それぞれの時間断面ごとの重み付け結果を元に、各時間フェーズで特徴的だった単語を選ぶ。複数の時間断面で高い累積特異値を示す単語もあることから、重複を取り除いてそれぞれのコーパスで 32～34 の単語が観察の対象になった。

また、各単語をケース、各時間断面の特異値を変数としたデータセットを作成し、クラスター分析を行うことによって、その時系列的なふるまいが類似した単語のグループ化を行い、時間フェーズごとに報道の特徴を捉えるを試みる。なお、距離計算には平方ユークリッド

距離を、クラスター化には Ward 法を採用してクラスター分析を行った。

図 3～図 7 に各コーパスで抽出された単語のクラスター分析の結果としてデンドログラムと、クラスター分析の結果のグループごとに単語の特異値の時系列的なふるまいの図を併記した。デンドログラム中に示した縦方向の点線は、グループ化を行った距離の値を示し、横方向の点線は各グループの区分し、A～E または A～F の記号を付した。デンドログラム右側のふるまい図中のアルファベット記号はそれに対応している。また、各グループに含まれる単語の特異値のピークの概ねの位置を、10～100 時間、100 時間付近、100～1,000 時間、1,000 時間付近というラベルを付けている。

図 3～図 7 の解析結果を表 5 のように整理した。表 5



は縦方向にコーパス、横方向はラベルとして付した時間フェーズをとり、そこにグループ化された単語を整理している。単語が示している意味を理解するために、単語をキーとした KWIC 検索を行い、その時間フェーズにおける単語の用いられ方を参考にラベル化した代表的な文脈情報をもとに示している。

全体と原発の 2 つのコーパスは、報道内容の時間的な展開が酷似している。10-100 時間では福島第 1 原発の爆発（爆発，発電），原子炉・格納容器の冷却措置（冷却），100 時間付近ではグループ別の計画停電（グループ），原爆作業員の被ばく（被ばく），公共交通機関の運休（運休），100-1,000 時間では放射性物質の検出（ベクレル），1,000 時間付近では東電原発事故の損害賠償の資金確保対応（賠償）が内容・フェーズともに一致している。また、全体のコーパスの特徴的な単語の中に、福島第 1 原発事故による被害と対応に関するもの以外がいくつ也存在するが、その数は多くない。したがって、東日本大震災に関するウェブ上のニュース報道は、福島第 1 原発事故による被害と対応に関する報道が代表していることが読み取れる。

岩手県と宮城県のコーパスの解析結果において見られた特徴的な単語は、全体と原発のコーパスではあまり見

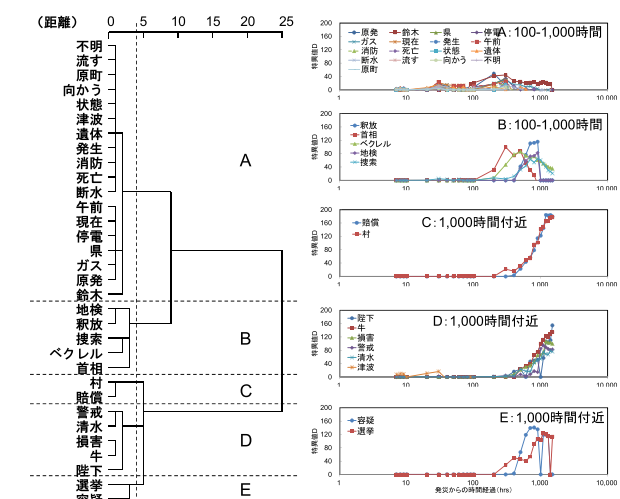


図 7 福島県コーパスにおける単語の時系列的なふるまいのクラスタ分析

られず、報道内容の時間的な変化の傾向は大きく異なる（表 5）。岩手県と宮城県は、10-100 時間に死者・行方不明者の多数の発生（死者、不明、多数）という報道が

表 5 時間と各コーパスの特徴的な報道内容の関係(単語のクラスター分析の整理)

	10-100時間	100時間付近	100-1,000時間	1,000時間付近
全体	- 福島第1原発の爆発 (原発, 原子力, 発電, 爆発) - 原子炉・格納容器の冷却措置の決定 (原発, 原子力, 発電, 冷却) ・小学校への避難(小学校) ・仙台空港の冠水, 各空港の閉鎖(空港) ・野党からの協力申し出, 与党からの協力要請(協力, 委員) ・その他(同社, 施設)	・義援金・募金回収活動(義援金, 募金) ・ボランティアの受付・活動開始(ボランティア) ・グループ別の計画停電(グループ) ・原発作業員の被ばく(被ばく) ・福島第1原発の水素爆発(水素) ・公共交通機関の運休(運休) ・首相・官房長官の被災者・国民への呼びかけ(皆さん) ・京都府警・京都消防の被災地派遣(京都)	・統一地方選選挙 ・選挙, 候補, 投票, 現, 現職, 議員, 新人) ・福島第1原発への上空・地上からの放水冷却(放水) ・放射性物質の測定・検出(ベクレル) ・計画停電中の交通事故/募金箱盗難/便乗詐欺による容疑者逮捕(容疑)	・東電原発事故賠償の資金確保対応(賠償) ・天皇皇后両陛下の被災地訪問(陛下)
原発	・福島第1原発の爆発/水素爆発の可能性 (原子, 発電, 爆発, 事故, 水素) ・原子炉・格納容器の冷却措置の決定 (原子, 発電, 爆発, 事故, 冷却) ・官房長官による政府記者会見(政府, 枝野)	・グループ別の計画停電(グループ) ・原発作業員の被ばく(被ばく) ・福島第1原発の水素爆発(水素) ・公共交通機関の運休(全線, 運休, 運行, 終日) ・首相・官房長官の被災者・国民への呼びかけ(皆さん)	・放水口付近からの高濃度の放射性物質の検出(ベクレル, 放水) ・水道水からの放射性物質の検出/乳児摂取控えの要請(ベクレル, 乳児) ・ホレンソウ等野菜からの放射性物質の検出/出荷制限(ベクレル, 出荷, ホレンソウ)	・東電原発事故賠償の資金確保対応(損害, 賠償, 勝俣) ・統一地方選選挙(選挙, 候補) ・避難区域の家畜対応(牛, 村) ・トレンチにたまった汚染水への(トレンチ)
岩手県	・地震とそれによる被害の発生(地震, 襲う, 被害) ・死者・行方不明者が多数発生(死者, 行方, 不明, 多数) ・大船渡市内行方不明者の生存確認(大船渡, 行方, 不明)		・統一地方選選挙(選挙) ・がれき化した家屋と自動車の処分(がれき, 自動車) ・患者受入可能病院の情報(県立, 医療, 病院, 患者) ・仮設住宅の入居募集(支所, 建設) ・失業者の雇用支援(雇用, 会社) ・学校の再開(児童) ・地域FM放送開始/地デジ放送完全移行の延期(放送) ・首相の被災地視察(首相) ・遺体の火葬(火葬) ・住民の防潮堤に対する当時の思い(防潮堤) ・漁業の復旧・復興対応(船, 漁船) ・その他(同市, 秋田, 施設, 社員)	・天皇皇后両陛下の被災地訪問(陛下)
宮城県	・死者・行方不明者が多数発生(死者, 不明, 多数) ・津波からの避難(避難) ・仙台空港の冠水, 各空港の閉鎖(空港) ・その他(今回, 可能, 話す, 見る, 県)		・統一地方選選挙(選挙) ・救援物資の受付・輸送・受取り(物資, 体育館, 公民館) ・救援物資の不足(物資, 不足, 体育館, 公民館) ・学校の再開対応(体育館, 校長) ・商業の再開対応(体育館, 営業, 街) ・住民の集団避難(集団) ・人工透析患者送迎用のバス提供(バス, 患者) ・沿岸部の被害のふりかえり(七ヶ浜) ・津波倒壊家屋からの9日ぶりの救出(任) ・仮設住宅の入居募集(支所)	・カキ養殖業の再開対応(カキ, 養殖) ・行政による各種雇用支援事業(雇用) ・スポーツ選手の被災地訪問(選手) ・復興構想会議(首相) ・ランドセルの希贈(ランドセル) ・天皇皇后両陛下の被災地訪問(陛下)
福島県			・放射性物質の測定・検出(原発, ベクレル) ・首相の被災地視察(首相) ・避難指示区域の遺体確認継続(死亡, 遺体, 不明, 捜索, 発生) ・ライフラインの緊急停止・復旧状況(停電, 断水, ガス) ・南相馬市原町区での対応(原町) ・福島地検による容疑者の釈放(地検, 釈放) ・その他(県, 現在, 午前, 状態, 向かう, 鈴木, 流す)	・統一地方選選挙(選挙) ・東電原発事故賠償の資金確保対応(損害, 賠償, 清水) ・避難区域の家畜対応(牛, 村) ・避難地域での窃盗容疑者逮捕(容疑) ・20km圏内の「警戒区域」として指定(警戒) ・その他(津波)

両コーパスで共通して特徴的であり、100-1,000 時間ではそれぞれに特徴的な復旧・復興の対応に関する内容が特徴的になっている。両者で共通しているものとして、仮設住宅の入居募集（支所）、学校の再開がある。岩手県では失業者の雇用支援（雇用、会社）や漁業の復旧・復興（船、漁協）が特徴的であり、宮城県では救援物資の状況（物資、体育館）や住民の集団避難（集団）に関する内容が特徴的である。また、両コーパス（両地域）では、住民の防潮堤に対する当時の思い（防潮堤）、沿岸部の被害のふりかえり（七ヶ浜）といったように沿岸部を津波が襲った当時の状況についての回想が特徴的な報道内容になっていた。1,000 時間付近では、今回の解析対象になったコーパスの中で岩手県が最も記事件数が少ない状況を反映しているせいか、宮城県に比べて岩手県の特徴的な記事が見られない。今後の報道活動の沈静化が懸念される。

福島県は 10-100 時間、100 時間とも特徴的な報道が見られない。当時は福島県に関する情報は、ほぼ福島第 1 原子力発電所の事故の被害と対応であり、福島県の震動や津波による被害や対応の状況が適切に報道されなかった状況がうかがえる。また、100 時間以降も原子力発電所の事故に伴う県内の対応状況であり、津波の被害や復旧の対応に関する情報は特徴だっで見ることができない。すべてのコーパスで特徴的だった報道内容として、統一地方選挙（選挙）、天皇皇后陛下の被災地訪問（陛下）がある。東日本大震災が発生した半月後の 4 月は統一地方選挙の時期であり、被災地外の各候補者が防災の設備や体制の充実化や原発の稼働反対が演説で謳われたというニュース記事が多数配信され、全体や原発のコーパスでも特徴的になっていた。特に全体コーパスでは、これに関連する多数の単語（選挙、候補、投票、現、現職、議席、新人）があり、被害や復旧・復興の対応に関する情報を席卷している様子がうかがえる。被災地内である岩手県、宮城県、福島県では、統一地方選挙に関する報道として、首長並びに議会議員選挙の延期を可能とする特例法の成立に関する内容が見られた。

6. おわりに

本稿では、災害の被害や対応に関する状況認識のための情報源として役割が増大しているウェブニュースに着目し、東日本大震災に関するウェブ上の報道実態を科学的に記述することをねらいとし、データアーカイビングによるコーパスの構築と、テキストマイニングを用いたコーパスの基礎的な解析を行った。東日本大震災の 2 ヶ月間の中で、東日本大震災がウェブのマスメディアを通してどのように報道されたのか、ここでの解析を通して得られた傾向は次の通りである：

- 1) 東日本大震災のウェブニュースから得られた特徴的な単語は、新潟県中越地震、新潟県中越沖地震、岩手・宮城内陸地震のウェブニュースから得られた単語と一致する傾向は見られず、東日本大震災に関するウェブ報道は量だけでなく、質的な側面からもこれまでの地震災害のものとは大きく異なることが明らかになった。
- 2) 東日本大震災について、被災県域に分けて報道したウェブニュースは、①原発事故影響を強く受けた地域（福島県）、②津波被害が甚大だった地域（岩手県、宮城県）、③その他の被害を受けた地域（青森県、茨城県、千葉県）という 3 種類に報道内容の質が異なっている傾向が見られた。
- 3) 東日本大震災に関するすべてのウェブニュース記事を見た場合、福島第 1 原子力発電所の事故による被害とその対応に関する情報と統一地方選挙に関する情報が主だった報道内容になっていた傾向が示された。後者は、発災後半月後に被災地外の全国で実施された統一地方選挙は、防災や原発の対応が争点となった影響を受けていた。
- 4) 福島県は、岩手県や宮城県と報道内容とは大きく異なる。原発事故の影響が大きいためか、発災後間もない頃の被害速報が乏しく、その後の教育や仕事の再開に関する具体的な復旧・復興に関する報道も主だっで見られない。

後者 2 点は、被災地の被害や対応の実態が適切に発信されることを妨げる要因やその結果になっており、適切な

状況認識を阻害する要因になっており、解決の必要性の高い課題であると言える。

本稿での検討は、基礎的な解析に過ぎないため、内容分析や質問紙調査等を用いた補完をする必要があることを今後の課題として注記しておく。また、Yahoo!ニュースサイトが整理した7県のトピックではなく、記事の本文に着目し、その他の都道府県や市町村・字レベルでの地域差の分析も今後の検討事項としたい。

謝辞

本研究の一部は、日本学術振興会特別研究奨励費「災害・危機に関する言語資料解析にもとづく社会現象としての災害・危機の将来展開予測（21・1950，研究代表者：佐藤翔輔）」によるものである。また、東日本大震災に関するYahoo!ニュース記事のクロール収集と解析においては株式会社サイエンスクラフト、有限会社アールツー・メディアソリューションにご協力いただいた。

参考文献

- 1) 警察庁緊急災害対策本部：平成23年(2011年)東北地方太平洋沖地震の被害状況と警察措置，p.1, 2011.
- 2) 林春男：「日本社会に適した危機管理システム基盤構築」が目指したもの—ICSに準拠した一元的危機管理システム確率に向けた検討—，文部科学省科学技術振興調整費「日本社会に適した危機管理システム基盤構築」研究成果ワークショップアブストラクト集，pp.9-23, 2006.
- 3) 総務省：情報通信の現況，平成22年度情報通信白書，pp.160-174, 2010.
- 4) たとえば，澤田雅浩，八木英夫，林春男：震災発生時における関連情報集約とその提供手法に関する研究—新潟県中越地震復旧・復興GISプロジェクトの取組みを通じて—，地域安全学会論文集，No.7, pp.97-102, 2005.
- 5) 野村総合研究所：ツイッターが震災時に果たした役割，NRIニュースレター，Vol.103, 2011.
- 6) 中林一樹，村上大和：阪神・淡路大震災に関する新聞報道の比較分析—阪神版と東京版の情報の相違について—，地域安全学会論文報告集，No.8, pp.226-231, 1998.
- 7) 松本誠，鶴田和治：震災報道の“温度差”を見る—全国4紙の東西版紙面比較調査から—，兵庫地域研究，No.16, pp.67-80, 1999.
- 8) 伊藤雅光：計量言語学入門，大修館書店，285pp., 2002.
- 9) IT media ニュース：よく見るニュースサイトは「Yahoo ニュース」86%，<http://www.itmedia.co.jp/news/articles/0707/09/news103.html>（最終閲覧 2011.5.26）
- 10) 林春男：いのちを守る地震防災学，岩波書店，pp.55-65, 2003.
- 11) 京都大学防災研究所巨大災害研究センター：Yahoo!ニュース中の災害・危機に関するニュース記事からのTR解析<http://www.drs.dpri.kyoto-u.ac.jp/hayashi/tr/trendreader.html>（最終閲覧 2011.5.26）
- 12) 佐藤郁哉：質的データ分析法，新曜社，pp.53-58, 2008
- 13) 佐藤翔輔，林春男，牧紀男，井ノ口宗成：TFIDF/TFを用いた危機管理分野における言語資料体からのキーワード自動検出手法の開発—2004年新潟県中越地震を取り上げたウェブニュースへの適用事例—，地域安全学会論文集，No.8, pp.367-376, 2006.
- 14) 佐藤翔輔，林春男，井上和治，西野隆博：ウェブニュースに見る災害・危機における社会的側面の時系列展開の可視化—TRENDREADER(TR)によるキーワード自動抽出結果の公開を通じた危機管理関係者の状況認識の支援—，Vol.29, No.7, pp.17-26, 2009.
- 15) Luhn H.P.：A stoical approach to mechanized encoding and searching of literary information, IBM Journal of Research and Development, Vol.1, No.4, pp.390-317, 1957.
- 16) 工藤拓，松本裕治：チャンキングの段階適用による係り受け解析，情報処理学会論文誌，Vol.43, No.6, pp.1834-1842, 2002.
- 17) 青野文江，田中聡，林春男，重川希志依，宮野道雄：阪神・淡路大震災における被災者の対応行動に関する研究—西宮市を事例として—，地域安全学会論文報告集，No.8 pp.36-39, 1998.
- 18) 田中聡，林春男，重川希志依：被災者の対応行動にもとづく災害過程の時系列展開に関する考察，自然災害科学，Vol.18, No.1, pp.21-29, 1999.
- 19) 佐藤翔輔，林春男：TRENDREADER(TR)を用いた災害・危機に関する言語資料体の解析にもとづく災害・危機事象の将来展開予測手法の開発—新潟県を襲った2つの地震災害を例にして—，地域安全学会論文集，No.10, pp.281-291, 2008.

(原稿受付 2011.6.5)

(登載決定 2011.9.10)