

形態素解析を用いた大規模停電に関するツイート分析 —北海道胆振東部地震におけるケーススタディー—

グループ演習 4 班
児島 怜, 段 煙煙, 新田 和樹, 楊 嵐
(アドバイザー教員: 庄司 学)

1 はじめに

1.1 研究背景

2018年9月6日に発生した北海道胆振東部地震では、最大震度7を記録し、北海道全域の約295万戸の大規模停電を引き起こした [1]。本地震以外に最大震度7を記録した例を挙げると、1995年の兵庫県南部地震では兵庫県および大阪府において約300万世帯、2011年の東北地方太平洋沖地震では東北地方および関東地方において約845万世帯の大規模停電を引き起こした。これらの地震による大規模停電は、交通機関やライフラインといった社会インフラだけでなく製造業をはじめとする産業に対しても打撃を与えた。特に、北海道胆振東部地震では国際的な観光都市であることから、観光産業への影響も見られた。図1は、2016年度から2018年度までの北海道への観光入込客数の推移を示しており、2018年度9月の観光入込客数は前年比の約15%減少している。北海道経済部観光局によると、観光消費の影響額は約292億円と推計された。また、北海道は日本全体の4分の1の農地面積を保有しているほどの代表的な農業地域であり、稲作、畑作、酪農など大規模な農業を展開している。酪農家や牛乳工場では、停電の影響により生乳の廃棄が相次ぐなど農林水産業に対しても影響を与えた。北海道庁によると、農林水産業の被害額は約397億円と推計された。以上のように、地震による大規模停電がもたらす影響は横断的に広がっている。そこで本研究では、停電による

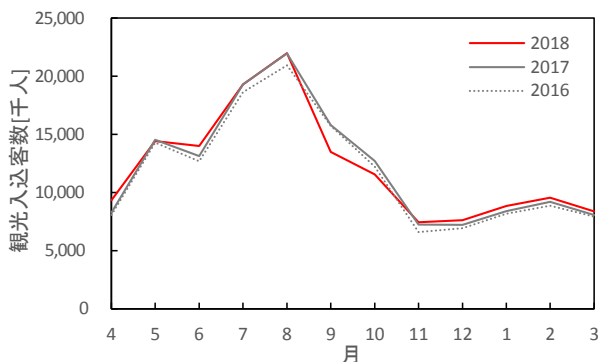


図1 観光入込客数の推移

影響を評価するために、北海道胆振東部地震における大規模停電に関するツイートの分析を行う。ツイート分析を行う手法の一つとして、本研究では形態素解析を用いる。形態素解析とは、辞書と呼ばれる単語の品詞等の情報にもとづき、テキストデータ(文章)を形態素(言語で意味を持つ最小単位)に分割し、それぞれの形態素の品詞等を判別する手法である。この処理によって、各単語の出現頻度または共起度(ある文字列とある文字列が同時に出現する度合い)を分析することが可能である。ツイート分析および形態素解析に関する既往研究については、次に記す。

1.2 既往研究

西村ら [2] は、観光庁の宿泊旅行統計データおよび東日本地域の宿泊業者に対し実施したアンケート調査結果を用いて、東日本大震災後を対象とした宿泊需要の減少に関する実証分析を実施した。その結果、震災による需要の減少や、復興需要による被害の軽減効果が時間・空間的に明らかとなった。また、震度が小さく観光資源や周辺のインフラ被害のない地域における需要減少は風評被害の影響を大きく受けているものと推察した。

庄司ら [3] は、東日本大震災において停電の被害率とその解消過程が異なる茨城県北茨城市、水戸市、日立市、および神奈川県横須賀市をそれぞれ取りあげ、Twitter 情報に基づいて停電の時空間分布の特徴を分析した。その結果、地震災害時における停電情報の時空間分布のセンシングにおいて Twitter 情報が極めて有用であり、町丁目程度の空間領域の停電の経時変化を適切に把握可能であることが明らかとなった。

湊ら [4] は、著者の心理情報を具体化するために茶筌 (Chasen) を用いた形態素解析により、(1) 事故や企業トラブルの概要に関する記述 (2) 事故または企業トラブルに関する感想文の2種類の文書を評価した。形態素解析の結果、感想文においては概要を記述した文章に比べ、形容詞、副詞、連体詞の出現頻度の比は約2倍の値となった。また、副詞における tf-idf 分析を求めることは、概要と感想文を分類する手段として使用できる可能性があることが示

唆された。

1.3 研究目的

そこで本研究では、北海道胆振東部地震における大規模停電の影響を評価することを目的とし、停電情報の可視化および形態素解析を用いたツイート分析を行う。停電情報の可視化では、北海道電力が公開している停電情報に基づいてデータを収集し、GISを用いて停電地域の遷移状況の可視化を行う。形態素解析を用いたツイート分析では、(1) 頻出単語 (2) 共起ネットワーク (3) 停電戸数と共起関係強度について3つの分析を行う。

2 停電情報の可視化

停電情報に関して、Twitter、Facebook、Google等の検索機能でデータを収集し、停電情報を時間的および空間的に明らかにする。その際に、各時間帯における停電解消地域および停電戸数のデータを収集し、(1) 停電戸数の遷移状況 (2) 停電情報の時空間分布センシングの2つの分析を行った。また、停電情報の時空間分布センシングではGIS（地理情報システム）を用いた。GISとは、地理情報および付加情報をコンピュータ上で作成、保存、利用、管理、表示、検索するシステムである。人工衛星、現地踏査などから得られたデータを、空間、時間の面から

分析・編集することができる。今回は、北海道の停電解消地域を時間的および空間的に示すために分析に用いる。

2.1 停電戸数の遷移状況

北海道胆振東部地震発生後、離島などを除いた道内全域約295万戸で停電が発生した。道内全域停電は1951年の北海道電力創設以来初の出来事である。北海道電力の公式Twitterアカウントでは停電戸数および停電解消戸数を公開していた [5]。図2に停電戸数の遷移状況を示す。図2より、9月7日6時に停電地域の約半数は解消し、9月8日6時に停電地域の約99%は解消した。今回の停電は、2日余りにわたって北海道の人々に深刻な影響を与えた。

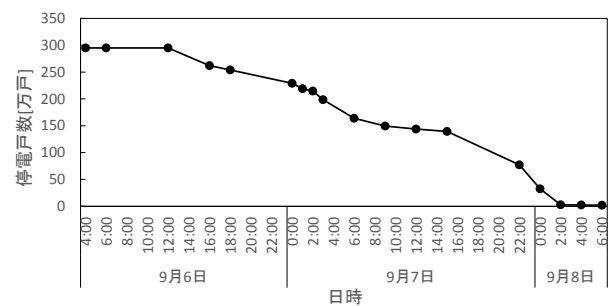


図2 停電戸数の遷移状況

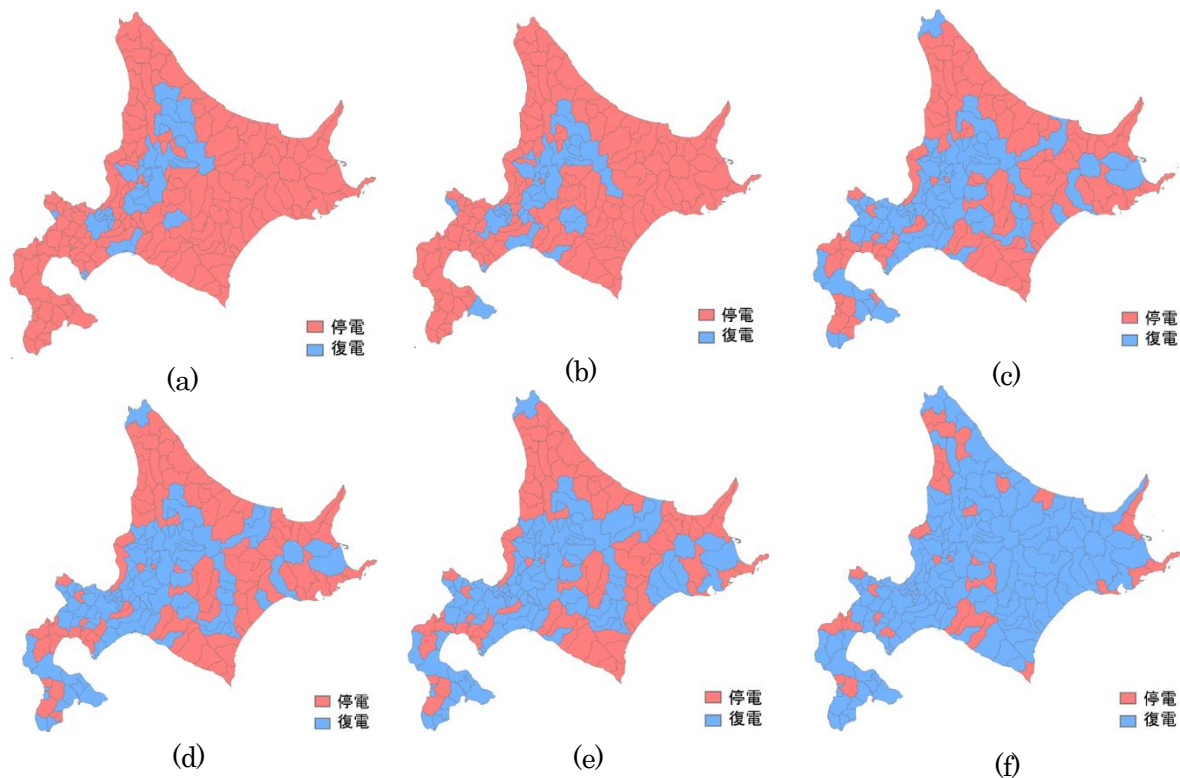


図3 停電情報の時空間分布センシング
(a)9月6日16時, (b)9月6日18時, (c)9月7日1時,
(d)9月7日2時, (e)9月7日3時, (f)9月7日6時

2.2 停電情報の時空間分布センシング

停電情報の時空間分布センシングを行うにあたり、2種類のデータを集めた。1つは国土交通省が公開している国土数値情報から北海道の行政区域データ [6]、もう1つは北海道電力のホームページから各時間における停電解消地域の市町村名である。これらのデータをGISに入力することで、時系列の停電解消地域を示す地図を作成する。しかし、具体的に停電解消地域の記述がある資料を不足しているため、資料がある6つの時間帯における図を作成した。図3より、停電解消の地域順序は北海道中央部から東西方向へ拡大して、9月7日3時以降に南北方向へ徐々に回復したことが分かる。

3 形態素解析を用いたツイート分析

3.1 形態素解析による頻出単語の集計

3.1.1 分析手法と結果

形態素解析とは、文法的な情報の注記の無い自然言語のテキストデータから、対象言語の文法や、辞書と呼ばれる単語の品詞等の情報にもとづき、形態素の列に分割し、それぞれの形態素の品詞等を判別する作業である。今回、私たちは北海道胆振東部地震に関するツイートデータを抽出し、頻出単語の集計を行った。方法は、地震が発生した2018年9月6日3時7分以降ある2018年9月6日～2018年9月8日の3日間の6, 12, 18, 24時のそれぞれ10分間のツイートをTwitter [7]の検索機能によって検索ワードを「札幌」に設定して収集した。(例:「札幌 since:2018-09-06_06:00:00_JST until:2018-09-06_06:10:00_JST」)検索ワードは「札幌」としたのは震央である北海道胆振地方中東部に比較的近く、北海道の都市の中で最も人口が多いことから、地震による被害を様々な人々が受けたと考えたためである。引用ツイートやリツイート、自動生成ツイートとみられるものは除外した総ツイート数は4383だった。形態素解析器にはMeCab [8]を使用し、名詞に分類された単語を集計した。集計結果から作成した表においては数字と意味を持たないアルファベット、記号は除いた。

3.1.2 考察

検索ワードが「札幌」であったため、「札幌」と「北海道」が多数検出され、続いて「地震」「停電」「復旧」「電気」が上位を占めた。「地震」に続いて「停電」が続くことから北海道胆振東部地震のもたらした大規模停電の影響の大きさが伺える。そのほかにも「断水」や「充電」、「余震」も多数検出され、これらの単語が大きな問題、あるいは関心であったこ

とが明らかになった。「無事」「大丈夫」「情報」は無事をTwitter上で報告しているだけでなく、「実家」「家族」「友人」などと合わせて被害にあった方々を心配するツイートが多かった。「みたい」が非常に多かったのは、人々が正しい情報が分からない中で断定を回避する表現を使用したからだと考えられる。「道路」や「信号」は道路陥没や停電に伴う信号機の停止などを指していて、インフラに多大な影響を与えたことから多く検出された。

表1 形態素解析による頻出単語の集計結果上位60

| 順位 | 単語 | 回数 | 順位 | 単語 | 回数 | 順位 | 単語 | 回数 |
|----|-----|------|----|-----|-----|----|------|-----|
| 1 | 札幌 | 4526 | 21 | 大丈夫 | 267 | 41 | 時間 | 143 |
| 2 | 北海道 | 838 | 22 | 心配 | 253 | 42 | 中止 | 142 |
| 3 | 地震 | 776 | 23 | 中 | 253 | 43 | 何 | 139 |
| 4 | 停電 | 711 | 24 | 被害 | 228 | 44 | 予定 | 139 |
| 5 | 復旧 | 626 | 25 | 連絡 | 207 | 45 | 余震 | 137 |
| 6 | 市 | 538 | 26 | 店 | 185 | 46 | ニュース | 134 |
| 7 | 電気 | 526 | 27 | 今 | 180 | 47 | 北 | 130 |
| 8 | 日 | 480 | 28 | 月 | 178 | 48 | 地域 | 129 |
| 9 | 区 | 422 | 29 | 今日 | 175 | 49 | 充電 | 126 |
| 10 | こと | 385 | 30 | 震度 | 175 | 50 | 水 | 122 |
| 11 | よう | 371 | 31 | 私 | 167 | 51 | 旭川 | 122 |
| 12 | 人 | 333 | 32 | 所 | 166 | 52 | これ | 120 |
| 13 | 情報 | 306 | 33 | ところ | 166 | 53 | 確認 | 117 |
| 14 | 方 | 302 | 34 | 断水 | 161 | 54 | 事 | 116 |
| 15 | 市内 | 295 | 35 | 実家 | 155 | 55 | うち | 116 |
| 16 | 無事 | 292 | 36 | 大変 | 153 | 56 | 駅 | 115 |
| 17 | みたい | 290 | 37 | 明日 | 150 | 57 | 中央 | 112 |
| 18 | 時 | 290 | 38 | 気 | 149 | 58 | 日本 | 109 |
| 19 | そう | 285 | 39 | 営業 | 149 | 59 | — | 108 |
| 20 | さん | 272 | 40 | 状況 | 145 | 60 | 家 | 108 |

形態素解析による頻出単語の集計

3.2 共起ネットワーク

3.2.1 分析手法と結果

頻出単語の集計後、同一ツイート上にある単語同士の組み合わせをカウントし、共起関係を測った。共起関係とは、1つの文中にある単語が現れたとき、その文中に限られた単語が頻繁に出現する関係のことを示す。本論文 [9] では同一ツイート上にある単語同士を共起関係であると定義した。

例:「今は停電と水は断水まではいってないけど濁っ

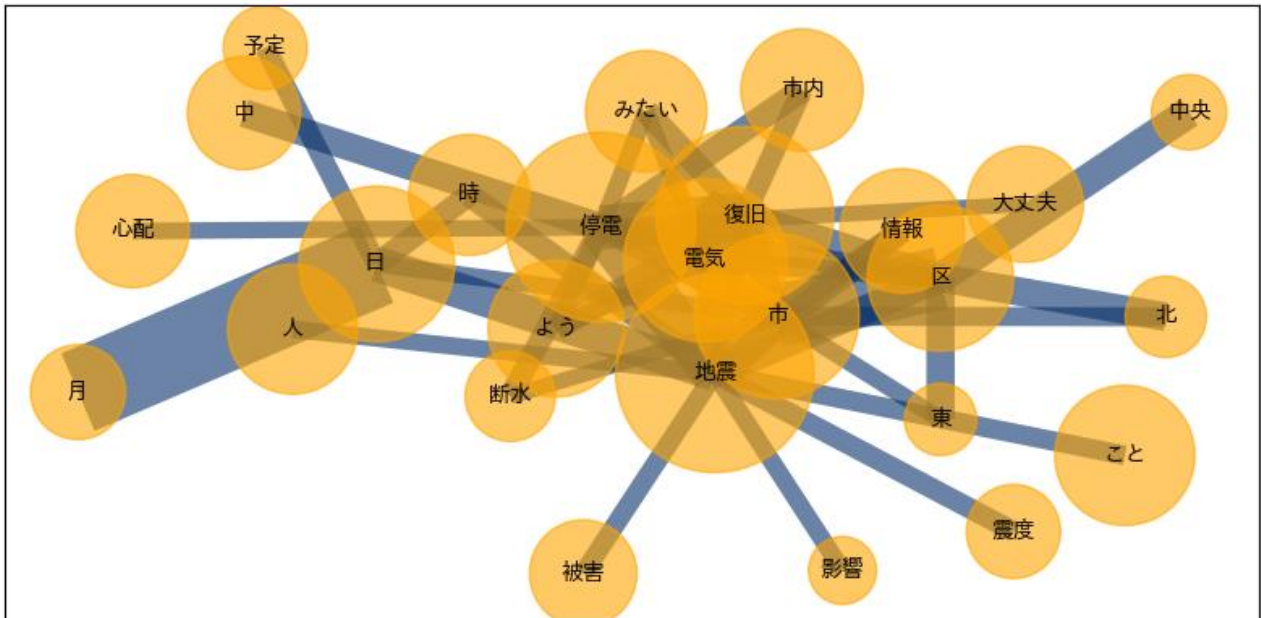


図 4 共起ネットワーク

てて不便です。」

「今」「停電」「水」「断水」「不便」の5つの名詞が検出され、(今, 停電)、(今, 水)、(今, 断水)、(今, 不便)、(停電, 水)、(停電, 断水)、(停電, 不便)、(水, 断水)、(水, 不便)、(断水, 不便)の10種類の組み合わせが共起関係としてそれぞれカウントされる。各ツイートを一つの集合として捉え、単語の出現順序は考慮しない。そのため共起関係は無向グラフで描かれる。表1から検索ワードである1位の「札幌」とそれに深く関係する2位の「北海道」がその他の単語と共起関係をもつのは自明であり、この2単語は除外した。これを総ツイート4383に対して行いカウントされた結果が図4である。図に表示されている単語は単独で90回以上集計され、なおかつ70回以上他の単語と共起関係を示し、かつ、数字と意味を持たないアルファベット、記号を除いたものに限っている。黄色い円の広さはその単語の頻出度合いを表し、青い線は太いほど、その単語同士が共起関係にあることを示している。位置は意味を持たない。

3.2.2 考察

図4から上記の共起関係を持っていることが明らかになった。特に「停電」、「電気」、「地震」について以下で示す。右側の単語は「停電」、「電気」、「地震」とそれぞれ共起関係を持つ単語を示し、その順序はカウントされた回数が多い順、つまり共起関係が強い順で並べた。

停電 ↔ 地震、中、復旧、断水、市内、みたい、影響、大丈夫、心配

電気 ↔ 復旧、停電、区、よう、地震、市、みたい、こと、水道、水、市内

地震 ↔ 停電、市、日、区、情報、時、震度、

こと、被害、影響、人、電気、復旧、東

が結びついていて、「停電」、「電気」、「地震」に対し、ツイートした人がどのようなことに関心を抱いていたのかが図4では視覚的に明らかになった。北海道胆振東部地震は北海道全域にわたる大規模停電が起こったため、その復旧がどれほどかかるかに非常に関心があったようだ。また停電に伴う断水の被害も受けていて、それらが頻出の単語だった。「市」、「日」、「時」、「区」はどの地域がどの程度の震度の地震が起こり、いつ停電しているのかというような情報が発信されていたからである。

3.3 停電戸数と共起関係強度の時系列変化

3.3.1 分析手法と結果

図4の共起ネットワークにおいて、強い共起関係を持っていた(停電, 復旧)とそれと近い意味で同じく強い共起関係を持っていた(電気, 復旧)の時系列による共起関係の増減を調査することによって、時系列で解消していった停電戸数とこれらの2つの共起関係の関連性を調査した。方法は、2018年9月6日6時~2018年9月8日6時までの期間で6, 12, 18, 24時でそれぞれ共起関係を測り、それを集合の

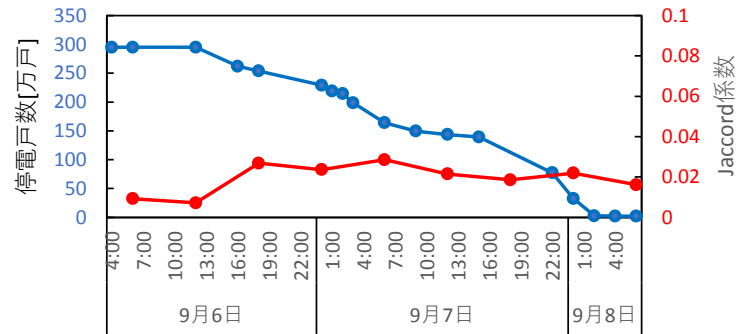


図5 停電戸数と共起関係の時系列変化

類似度を測る Jaccard 係数を用いて値を算出した。

$$\text{Jaccard 係数} = \frac{A \cap B}{A \cup B} \quad (1)$$

A : 「停電」「電気」のいずれかを要素に持つ共起の発生回数

B : 「復旧」を要素に持つ共起の発生回数として値を算出

3.3.2 考察

図5は青色のグラフが北海道全域の停電戸数を表すのに対して、赤色のグラフは(停電, 復旧)と(電気, 復旧)の共起関係を合計することで算出した Jaccard 係数を示している。停電が解消し始めてから初めて計測された9月6日18時以降に(停電, 復旧)と(電気, 復旧)の Jaccard 係数は大幅に増加し、その後もツイートされ続けていることから停電からの復旧をツイート上で報告する人が多かったのだと思われる。このことから Twitter が停電復旧時、重要な情報発信・収集源であることが言え、現実状況の変化によって共起関係の強度が変化することを示した。

4 まとめ

本研究では、北海道胆振東部地震における大規模停電の影響を評価するため、停電情報の可視化および形態素解析を用いたツイート分析を行った。

停電情報の可視化では、北海道電力が公開している停電情報に基づいてデータを収集し、GISを用いて停電地域の遷移状況の可視化を行った。これにより停電戸数および停電地域の遷移状況を明らかにすることができた。

形態素解析を用いたツイート分析では、(1) 頻出単語 (2) 共起ネットワーク (3) 停電戸数と共起関係強度について3つの分析を行った。頻出単語の分析では、MeCabを用いて形態素解析を行い名詞に分類された単語の集計を行った。集計結果では、「地震」

「停電」「復旧」「電気」が上位を占めており、北海道胆振東部地震による停電の影響の大きさが伺えた。そのほかにも「断水」「道路」「信号」等が検出され、インフラに対する影響も見られた。共起ネットワークの分析では、同一ツイート上にある単語同士の組み合わせをカウントすることで共起関係を測った。その結果、特に「地震」との共起関係が最も強い単語は「停電」であることが示され、地震による停電の影響が甚大であることが明らかになった。停電戸数と共起関係度の分析では、Jaccard 係数を用いて停電の復旧過程がどのようにツイートで反映しているのかを分析した。その結果、9月6日18時頃に Jaccard 係数は大幅に増加し、その後もツイートされ続けていることから停電からの復旧をツイート上で報告する人が多かったのだと考えられる。このことから Twitter が停電復旧時、重要な情報発信・収集源であることが言え、現実状況の変化によって共起関係の強度が変化することを示した。

本研究は、大規模停電による横断的な影響をツイート分析することで網羅的に調査することを目的としていた。しかし、ツイート内容には偏りがあるため網羅的に分析するのは難しく、今後この点を検討する必要がある。

参考文献

- [1] 2018年北海道胆振東部地震・大阪府北部の地震被害調査報告書；土木学会，2019
- [2] 西村泰紀，梶谷義雄，多々納裕一；大規模災害による宿泊業への影響評価—市町村宿泊旅行統計とアンケート調査に基づく東日本大震災のケーススタディ；土木学会論文集 D3（土木計画学），Vol.69, No.5, 2013, pp.I_217-I_227
- [3] 庄司学，高橋大；Twitterを利用した地震災害時における停電情報の時空間分布のセンシング；平成27年度電気学会電子・情報・システム部門大会
- [4] 湊淳，鈴木慎吾，村上雄太郎，伊多波正徳，小澤哲；形態素解析を用いた事故・リスクに関する記述の分析；日本感性工学論文誌，Vol.13, No.2, 2014, pp.341-346
- [5] Twitter；北海道電力株式会社；https://twitter.com/Official_HEPCO
- [6] 国土交通省；国土数値情報；<http://nlftp.mlit.go.jp/ksj/index.html>（2019/6/24 確認）
- [7] Twitter；<https://twitter.com/>
- [8] MeCab：Yet Another Part-of-Speech and Morphological Analyzer；<https://taku910.github.io/mecab/>（2019/10/12 確認）
- [9] Z. S. Harris；Co-Occurrence and Transformation in Linguistic Structure；Vol.33, No.3, Part 1, pp.283-340
- [10] 吉川武時，長尾真編；言語の機械処理；三省堂，1984