

地域新聞 DAの検索のためのメタデータ検討プロジェクト

Metadata Evaluation Project for Enhancing the Searchability of Regional Newspaper Digital Archive



石橋豊之 ISHIBASHI, Toyoyuki

中部大学人文学部コミュニケーション学科 助教

大阪府出身北海道育ち。筑波大学大学院図書館情報メディア研究科博士後期課程単位取得退学。修士（図書館情報学）。専門は図書館情報学。地域情報の組織化、図書館における社会教育等の研究を行っている。

ABSTRACT

The purpose of this project is to examine metadata for searching local newspapers during the digitization process. Given that local newspapers primarily target the residents of their respective areas, they are often of modest scale, making it challenging to establish paid databases akin to national or regional newspapers. Conversely, local newspapers serve as valuable resources for researching information about the region. This project focuses on the "Nikkan Soya," distributed exclusively within the Soya Subprefecture. This project employs the methodology previously utilized in the analysis of articles in the "Nikkan Soya". This report outlines the aforementioned methodology and discusses future directions.

1. プロジェクトの目的

本プロジェクトの目的はプロジェクト名からも分かるとおり、地域新聞をデジタルアーカイビングした際の検索に有効なメタデータを検討するものである。ここでいう地域新聞とは、“県レベル以下の特定の地域社会に力点を置いて記事を収集・編集し、特定地域の人々を主たる読者として販売している新聞”のことを指す[1]。このような地域新聞は、対象がその地域の住民に限定されるため比較的小規模のものが多く、全国紙やブロック紙のように有料データベースを提供することは難しいものである。他方で、その地域に関する情報を調査する手段として地域新聞は重要である。

このうち本プロジェクトでは宗谷総合振興局内でのみ流通している『日刊宗谷』を対象としている。他の地域新聞にも言えることだが、全国紙と異なりその地域の内容が記事の中心となるため、その地域特有の語彙が存在している。この点、筆者は以前にも『日刊宗谷』の記事と全国紙の記事を比較し、地域特有の語彙の分析を行なったことがある[2]。また、『日刊宗谷』の記事のみを対象とし分析も行なった[3]。本プロジェクトはこのうち後者の研究で用いた手法を用いている。

2. 宗谷総合振興局について

本プロジェクトの対象地域は北海道稚内市であり、宗谷総合振興局内に属している。北海道内でも最北端に位置していることから他の都市圏から離れているのが特徴である。加えて、人口減少が続いている地域でもある。2024年1月現在の宗谷総合振興局内の人口は56,466人となっている[4]。今後も振興局内では人口減少が見込まれており、その地域住民が主たる読者となる新聞にとってはマイナス要素となる。また、北海道で最大の発行部数を誇る北海道新聞ですら2023年9月で夕刊が休刊となっている実態もある[5]。当然、地域新聞がその地域で刊行を続けていくことが重要なのは間違いない。他方で、公立図書館といった地域情報を記録する機関に関してはデジタル化を含めたアーカイビングを積極的に行う必要があると考えている。実際に置戸町立図書館では、地元で発行される『置戸タイムス』をデジタル化し、一部ではあるものの置戸町デジタ

ル郷土資料館として公開している[6]。

本プロジェクトも最終的には公立図書館といった地域のアーカイブ機関での運用を想定している。

3. 研究手法の概要

前述した以前行った手法の概要を説明する。『日刊宗谷』の1960年、70年、80年、90年の1月1日～10日までの記事のテキスト化を実施した。この際にはOCRソフトの「読取革命」を用いてテキスト化を行ったが、使用した画像の解像度の荒いものや古い記事に関しては認識率が低く多くの記事は手動で修正を行なっている。また、明らかに誤植とわかるものについても修正している。

次に新聞に掲載されるものは記事だけではないため、記事・コラム・小説・社説・川柳・囲碁・将棋・天気予報・占い・その他（慶弔録等）に分類した。また、複数の記事に適用される見出しを特集としている。

このうち記事のみを対象とし、KH-Coder 3を用いて形態素解析を行った[7]。この際の単語を分類するために、『ニュース・シソーラス』の第3版（1990）及び第4版（2004）を参照している。このシソーラスでは7つの分類項目と42分野（1990は41分野）と分野に当てはまる「対象とする領域・範囲」という項目が存在する。この「領域・範囲」に当てはまる単語を『新聞切抜・写真分類表：昭和48年版』（1973）も参考にしながら抽出した。この結果の一部を表1に示す。

表 1 日刊宗谷における分類・分野・対象とする領域範囲

分類	分野	対象とする領域・範囲の一部
政治	行政	管内、支庁（あるいは宗谷総合振興局）、浜森辰雄、市役所
	政治	各政党（自民党、社会党、共産党等）
経済	経済	商工会議所、景気、所得、倒産
	水産	漁業、漁師、組合、水産、スケソ（スケトウダラ）、漁獲、底引船（底曳船）、不漁、漁場
	農林	組合、農業、酪農、開拓、生乳、林業、肉牛
	運輸	空港、自動車、宗谷線、天北線、東日本海フェリー
	労働	機動職業訓練、税務署
社会	社会	観光、消防団、福祉、町内会、人口、ロータリークラブ
	健康・病気	健康、病院、医療、保健、保健所
	行事	正月、成人式、交礼会、新春座談会、初日の出
事件	災害	交通事故、交通安全、火災、吹雪、除雪、海難、野犬
	司法	道警、派出所
文化	自然・生物	海岸、宗谷岬、流水
	スポーツ	バレーボール、柔道、スケート、スキー
国際	国際	サハリン（あるいは樺太）、ソ連

この際には時期が限定的であったためあてはまらない分類・分野が存在していた。また地域新聞であるが故に軍事・外交などそもそも取り扱いの少ないと思われる項目も存在している。領域・範囲に関しては『ニュース・シソーラス』でも用いられている用語も存在し、これらは地域特有ではなく一般的な語彙と言えるだろう。他方で、『ニュース・シソーラス』の分野とは異なる分

野にした方が記事の内容上妥当なものも存在した。例えば、消防団は分野としては災害に位置付けられているが、登梯式のように災害と関係のない記事も多いため、社会に位置付けた。

以上のように単語を分類したが、記事のさらなるテキスト化を行う必要があったため、大学院生に協力を仰ぎテキスト化を依頼した。以前はOCRをかけていたが、前述した通り精度が低かったため、今回は手動で取り組んでもらっている。対象は、1990年1月の1186記事であり、このうち343記事は以前に行っているため、843記事が新規となっている。しかし、本稿執筆段階で終了していないため、終了次第分析を行い、今後学会等で報告していく予定である。

4. 他の地域情報との組み合わせについて

他の地域情報との組み合わせとしてはまず地理情報が考えられる。GIS (Geographical Information System: 地理情報システム) を用いて記事や新聞に掲載される広告の位置を可視化することは以前から行ってきた。例えば下記の図1は、QGISというフリーソフトウェアのGISを用いて日刊宗谷に掲載された広告元の所在地をプロットしたものである。図2は広告の内容(書かれているもの)を表示したもののだが、現状はあまり意味をなしていない。このようなプロットをする際の課題としては、広告に住所が記載されていなかったり住所が過去の地名のものであり現在では使われていないケースが発生することである。この場合、前者は『稚内商工名鑑』といった別の資料を用いて調査し、後者は以前行った地名の変遷を組織化した際の情報を用いて解決した[8]。なお、この際には、現在の地名を取得した後「CSVアドレスマッチングサービス」を用いて緯度・経度を取得してプロットをしている[9]。



図 1 広告元を QGIS にプロット

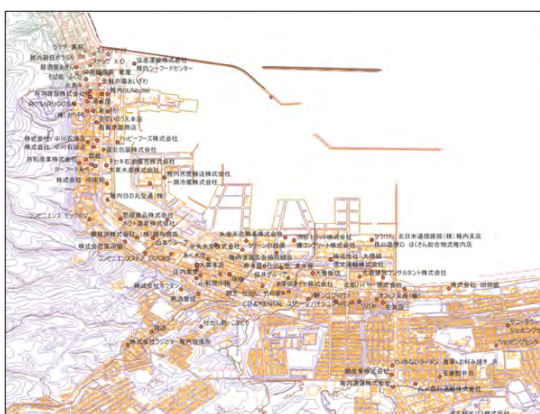


図 2 広告内容を QGIS にプロット

こうした可視化を行うことにより、当時の稚内市の企業の変遷等を視覚的に示すことができる(例えば、昭和55年の広告のほとんどは現在の中央周辺であったのに対し平成2年では他の地域のもが増加していることがわかる)。また、記事に関しても〇〇で事件があったと書かれていてもその〇〇がどこかわからなければ情報の大きな欠落となる。もちろん現在の記事であれば大きな問題はないものの過去の記事の場合、現在では使われていない地名(公的な地名や通称など)が用いられているケースも当然ある。位置情報を付与し可視化させることはこうしたケースにも対応できる。

このほか、広告そのものも現状画質の問題が存在するものの、記事とは別途データベース化しある種のデジタルミュージアムのような形式にしたいとも考えている。

次に、公立図書館におけるレファレンス資料の組み合わせについても考えている。現在は『世

界・日本・北海道宗谷地方年代史メモノート』という稚内市史第1巻編纂に携わった橋村氏が作成した資料を用いる予定である。こちらは、10年ほど前に市立図書館に寄贈されたものであり、各年代の宗谷地方に関する時事ネタ等をまとめたものである。その内容の精緻さからレファレンスサービス等でも活用されている。このような二次資料と一次資料にもなりうる地域新聞を紐付けることは図書館におけるレファレンスサービスを行う上では有効であると考ええる。ただし、こちらのノートは全12巻であるものの日刊宗谷刊行以降のものは9巻以降であるため、研究の対象も9巻以降となる。

最後に本プロジェクトを通じて地域情報の有用性についても示していければと考えている。というのも現在稚内市では稚内市史第3巻編纂に向けて動いている[10]。このような編纂の際には多くの資料や情報が収集されることになる。実際に前述した橋村氏が編纂の際に用いた膨大な資料が最近になり稚内市立図書館に寄贈された。この資料の中には市役所にも残っていないような地図など現在となつては貴重なものも多々ある。本プロジェクトでも今後こうした資料を活用していく予定であり、このような過去の資料を活用し有効性を示すことは、今後新たに市史編纂に際して収集される資料や情報のデジタル化や利活用にもつながるだろう。幸い第3巻の刊行は稚内市制施行80年目の節目となる2028年を予定しているため時間的な余裕はある。今年度は様々な事情で当初予定よりもプロジェクトを進めることができなかったものの、本格的な市史編纂が始まる前までにはさらなる成果を出したいと考えている。

【参考文献】

- [1] 吉岡雅光. 地域新聞の地域求心力. 立正大学人文科学研究所年報. 1997, 別冊 (11), pp.15-29. 参照部分は、p.16.
- [2] 石橋豊之, 柊和佑, 三河正彦, 安藤友晴. 域新聞からみる地域特有のメタデータ～大韓航空機墜落事件を事例に～. デジタルアーカイブ研究報告. 2020, 2, pp.8-14.
- [3] 石橋豊之, 柊和佑, 三河正彦, 安藤友晴. 地域新聞のデジタルアーカイブへの利活用のための地域特有の語彙の抽出と可視化に関する研究: 日刊宗谷を事例として. デジタルアーカイブ学会誌. 2022, 6 (s3) , s202-s205.
- [4] “宗谷管内の住基ネットにおける人口【参考値】”. 宗谷総合振興局. <https://www.souya.pref.hokkaido.lg.jp/ts/tss/souyatoukei18jyuuki.html> ,(accessed 2024-02-20).
- [5] “夕刊を休刊します。ご愛読ありがとうございました*朝刊と道新デジタルのご購読をお願いいたします”. 北海道新聞社. <https://kk.hokkaido-np.co.jp/post-1885/> ,(accessed 2024-02-20).
- [6] 今西輝代教. “置戸町立図書館の資料とデジタルアーカイブ”. 地域資料サービスの展開. 蛭田廣一編. 日本図書館協会, 2021, p.14-38, (JLA図書館実践シリーズ, 45).
- [7] 樋口耕一. 社会調査のための計量テキスト分析: 内容分析の継承と発展を目指して. 第2版, ナカニシヤ出版, 2020, 264p.
- [8] 石橋豊之, 柊和佑, 三河正彦, 安藤友晴. 地域包括的支援サービスを志向した意思伝達のための語彙収集に関する研究. デジタルアーカイブ学会誌. 2019, 3 (2), p. 203-206.
- [9] “CSVアドレスマッチングサービス”. 東京大学空間情報科学研究センター. <https://geocode.csis.u-tokyo.ac.jp/cgi/geocode.cgi?action=start> ,(accessed 2024-02-20).
- [10] 市史第3巻発刊へ: 教育、観光での活用視野に. 日刊宗谷. 2023-08-31.