

# 地域新聞からみる地域特有のメタデータ -大韓航空機撃墜事件を事例に-

石橋豊之 柊和佑 三河正彦 安藤友晴



# 自己紹介

- 石橋豊之
  - 1989/11/29生まれ
  - 経歴
    - 稚内高校→航空自衛隊→稚内北星学園大学→筑波大学大学院→稚内北星学園大学
    - 共同研究者は過程でお世話になっている方々
- 現在は助教
  - COC事業推進室長
    - PBL関連も担当
  - 図書館情報学系科目が本職



\* 2017稚内北星サンタラン

# 背景

- インターネットの発展
  - MLA連携とデジタルコンテンツのワンストップサービスの提供が求められている
- 他方で
  - 地方特有の情報の存在  
→ 地方独自の語彙や方言

# 地方の独自性

- 独自の要素
  - 語彙
  - 言い回し
  - 方言
  - 文脈
- この点を考慮する必要がある



# 地方の現状

- 人口減少
  - 過疎化
  - 超高齢化
- 稚内市の場合
  - 昭和50年の55,464 人がピーク
  - 現在: 34,801人 (2018/1)
    - 2000年段階: 43,774人

# 新聞

- 最新の情報をタイムリーに伝える媒体
  - 時事ネタが中心
- 地域新聞
  - その地域でのみ発行
  - 貴重な1次資料
    - その地域の将来的な研究をする上では非常に重要
    - デジタルアーカイブの必要性

# 本研究の目的

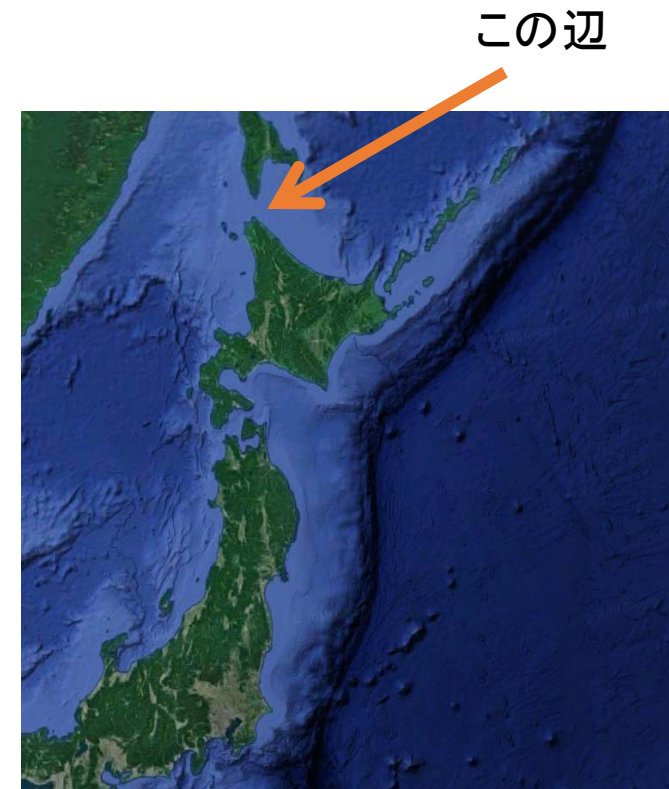
「地域の記憶」を記録するために、その周辺情報として地域住民、コミュニティ構成員の持つ情報をデジタルアーカイブに蓄積し、利活用できるメタデータを付与することを目的としている

- ・ そのうち本発表では  
本研究がメタデータとして残すべき地域特有の語彙をどのように抽出するか、  
その一手法を検討

# 主研究対象：北海道稚内市

- 周辺地域も含め過疎が進行
- 特徴
  - サハリンに近い
  - 自然が人を試してくる
    - 雪・風

試される大地  
北海道



出典: Google Map



# 宗谷地域研究所のプロジェクト

- 宗谷地域研究所[1]
  - 昨年発足
  - “宗谷本線、天北線のヒストリー & ストーリー” 可視化プロジェクトを展開
    - 宗谷地域の鉄路及びそれに関連する歴史等の可視化を目指している
      - 写真・映像等の資料を収集
      - オーラルヒストリー
- 将来的には、さらに広範囲にかけて地域の記録の保存・利用を目指している。

# 目指すべきゴール

- 「地域の記憶」のアーカイブ作業 = 特需
- ただ、地域住民は高齢化
  - 収集を急がないとダメ
- 五十年後に「使える」アーカイブが必要
  - 従来から図書館がやっていたこと

<b>C0</b> 総記	<b>C4</b> モデル
01 自動車	41 自動車モデル
02 事典	42 二輪車モデル
03 年鑑	43 その他のモデル
04 統計・白書	44 自動車カタログ
05 法規	45 二輪車カタログ
09 その他	49 その他
<b>C1</b> 伝記	<b>C5</b> モータースポーツ
11 経営者	51 モータースポーツ
12 技術者	52 レース・ラリー
19 その他	53 レース車
	59 その他
<b>C2</b> 社会	<b>C6</b> 技術
21 文化	61 自動車工学
22 環境	62 生産技術
23 施設	63 材料・燃料
24 安全	64 整備・修理
29 その他	65 技術論文
	66 関連発明
<b>C3</b> 産業	69 その他
31 自動車製造	
32 自動車部品	
33 自動車販売	
34 自動車運輸	
39 その他	

\* 豊田市立中央図書館  
の地域資料棚の分類

# 本研究の位置付け

- 五十年後になって語り部のいなくなる地域情報を残す
  - 地域研究、未来の地域活用、安全
- 地域住民自身がアップデート
  - 都市部ではわからない概念でも対応
    - 北海道では浜言葉・雪用語が多い
  - コスト減
    - 地域の市役所・図書館・大学を想定



\* 2017年稚内北星

# ただし

- いわゆる方言データの充実が必要
  - 地方ごとに概念が異なるかも
  - 「辞書」に載っていない言葉も多い
  - 口語の扱いをどうするか
    - アイヌ語、ロシア語由来も
    - 地域住民が当たり前と感じる言葉は収集対象にすらならない
- そもそも、どんなデータになるかもわからない

# 五十年後に“使える”データ

- 生データ
- 生データの解釈に必要な語彙
  - 方言
    - 方言研究者と連携したい
  - 地域で呼ばれている地名
    - 地図には載っていない愛称・略称
  - 地域の常識・生活の知恵
    - 雪に関する知識・特有の産業に関する知識
- 五十年後の都市部の人間が解釈できる情報  
にしておく = 使える
  - 厳密に管理・整理されたデータにすることは考えない

# 本研究でやりたいこと

- プロジェクト
  - 地域情報資源を収集する
  - 資料に関連する地域住民の持つ情報を集める
  - 都市部の人間が、五十年後に理解するために必要な情報を理解するための情報を追加
  - データを追加しながら地域でアップデートする仕組みを組み込む



# 本研究でやりたいこと

- プロジェクト
  - 地域情報資源を収集する
  - 資料に関連する地域住民の持つ情報を集める
  - 都市部の人間が、五十年後に理解するために必要な情報を理解するための情報を追加
  - データを追加しながら地域でアップデートする仕組みを組み込む

# 記事の分析

- 対象
  - 日刊宗谷
    - 宗谷総合振興局内を対象とした地域新聞
    - 1948年刊行
  - 過去の記事をjpg形式で保存されている
    - テキスト化
    - KH Coder[2]を用いて分析

# 記事の分析

- 対象記事
  - 1983年に発生した大韓航空機撃墜事件
    - サハリン沖 = 稚内周辺の事件
  - 全国的にも報じられている
    - 全国紙でも大々的に取り上げている
  - 地域新聞、全国新聞（朝日新聞）との記事の比較が可能

# 対象範囲

- 事件が報じられた9月2日から9月9日
  - 3日～5日は対象の新聞社が社員研修旅行で臨時休刊
  - 全33記事
- 朝日新聞全国版の記事の分析も実施
  - 9月2日、6日、7日
  - 全46記事
  - 聞蔵を使用

# 比較方法

- 日刊宗谷の上位頻出語150
- 朝日新聞の上位頻出語150
- 両者を比較  
→重複していない単語の抽出

抽出語	出現回数
海域	39
本部	36
現場	18
対策	18
フェリー	16
海	13
漁業	13
設置	12
会社	11
漁船	11
警備	11
付近	11
北緯	11
海馬島	10



抽出語	出現回数
海域	39

記事全体からみた単語であることに注意



個々の記事のメタデータとして利用できるわけではない

警備	11
付近	11
北緯	11
海馬島	10

# メタデータスキーマ実装への課題

- メタデータ同士の結びつけ
- 海馬島で考えると
  - 宗谷だけでも3島存在
  - 「海馬島」という情報のみだとどの「海馬島」を連想するか分からない
  - 他の要素（地理・事件）を加えることで判定している
  - その要素を調査・分析する必要がある

# メタデータスキーマ実装への課題

## メタデータスキーマの検討

既存のものか**新規**に作成するか

既存: NewsML[3], ENMAP[4]

新規: Dublin Core, Schema.org[5]を利用して  
ローカライズさせていく

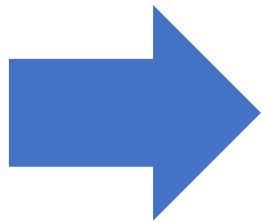
- 他の要素（地理・事件）を加えることで判定している
- その要素を調査・分析する必要がある

# メタデータスキーマ実装への課題

- 現在行なっていること
  - 新聞記事の要約
    - 稚内地域住民と旅行者で実施
      - それぞれ4名ずつに実施
    - 日刊宗谷の10記事を140字に要約
    - 理解時にひっかかる単語を調べさせる
  - 結果
    - 先ほどの14単語は「海馬島」を除くとひっかからない
    - ゆかりのない学生は漁業に関連するような単語にひっかかる傾向のみ

# 今後の展開

- (人的・経済的) コストの低下
- 地域独自の記事への応用
- 画像などの要素への対応



さらなる分析が必要  
そのためには、より多くのデータを調査

# プロジェクトへの応用

- 五十年後を待って進めるのは意味がない
- 地域情報をアップデートしながらより良い方法を模索
- 他地域への応用を検討しながら何と無くあちこちで使えるシステムを目指す



# 出典

- [1] 稚内北星学園大学. “稚内北星学園大学 宗谷地域研究所「“宗谷本線、天北線のヒストリー&ストーリー”可視化プロジェクト””. 稚内北星学園大学. [http://www.wakhok.ac.jp/tiiki\\_kenkyuu.html](http://www.wakhok.ac.jp/tiiki_kenkyuu.html) ,(accessed 2018-02-28).
- [2] 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. ナカニシヤ出版, 2014, 235p.
- [3] International Press Telecommunications Council. “NewsML-G2”. International Press Telecommunications Council. <https://iptc.org/standards/newsml-g2/>, (accessed 2018-02-28).
- [4] 時実象一. 欧州における新聞デジタル・アーカイブ Europeana Newspapers. 情報の科学と技術. 2017, 67(1), p.34-37.
- [5] 兼岩憲. セマンティックwebとリンクトデータ. コロナ社, 2017, 229p. 参照部分は、 p.138-139.