

令和 2 年 5 月 31 日現在

機関番号：10104

研究種目：基盤研究(B) (一般)

研究期間：2016～2019

課題番号：16H02912

研究課題名(和文) 議論の背景・過程・結果を関連づける地方政治コーパスの構築とその学際的応用

研究課題名(英文) Construction of A Local Political Corpus that Links the Background, Process, and Results of Discussions and Its Interdisciplinary Application

研究代表者

木村 泰知 (Kimura, Yasutomo)

小樽商科大学・商学部・教授

研究者番号：50400073

交付決定額(研究期間全体)：(直接経費) 13,300,000円

研究成果の概要(和文)：本研究では、地方政治に関する研究の活性化・学際的応用を目指して、「議論の背景」「議論の過程」「議論の結果」を関連づけるコーパスを構築する。具体的には、新聞記事(議論の背景)、地方議会会議録(議論の過程)、条例(議論の結果)の3つの言語資源を地域、時間、課題の観点から関連づけることを目指す。本研究の目的は以下の3点である。A)3つの言語資源(地方議会会議録・条例・新聞記事)の収集・整理 B)地域・時間・課題などの観点から3つの言語資源を関連づけるコーパスの構築 C)上記の成果を用いた、政治学・経済学・言語学・情報工学における応用研究

研究成果の学術的意義や社会的意義

本課題では、地方政治に関する研究の活性化・学際的応用を目指して、政治学・経済学、社会言語学などの研究分野での利用を促進する。本研究では、地方議会会議録(議論の過程)と条例(議論の結果)および新聞記事(議論の背景)を関連づけるコーパスを構築するために、まず地方議会会議録に加え、条例、新聞記事を収集、整理する。その上で、議論の過程、結果、背景を関連づけるために、会議録に含まれる議論と条例、および、新聞記事を地域・時間・課題の3つの観点から結びつける。

研究成果の概要(英文)：In this research, a corpus that correlates "background of discussion", "process of discussion" and "result of discussion" is constructed with the aim of activating and interdisciplinary research on local politics. Specifically, we aim to link three language resources, newspaper articles (background of discussions), minutes of local assembly (process of discussions), and ordinances (results of discussions) from the viewpoint of region, time, and issues. The purpose of this research is the following three points. A) Collection and organization of three language resources (local assembly minutes, ordinances, newspaper articles) B) Build a corpus that links three language resources from the perspective of region, time, and issues C) Applied research in politics, economics, linguistics, and information engineering using the above results

研究分野：自然言語処理

キーワード：政治情報 地方議会 議会会議録 コーパス構築 議論の背景 議論の過程 議論の結果

## 1. 研究開始当初の背景

地方分権や地方創生などの議論が盛んになっており、地方政治が注目を集めている。このため、地方議会におけるすべての発言が書き記された地方議会会議録の分析も重要性を増している。幾つかの研究分野において、自治体がウェブ公開している地方議会会議録を対象とした調査・分析が行われている。このような研究では、自治体によって個別に公開された議会会議録を逐一収集しており、調査・分析を始めるまでの準備に時間と労力がかかるという問題があった。

このように地方議会会議録コーパスは、主として地域差のある大規模話しことばコーパスとしての特徴を活かした研究成果(自然言語処理や言語分析)をあげているが、地方議会会議録コーパスの重要な応用対象である地方政治に関する分析において、現在のところ研究がほとんど進んでいない。地方議会会議録コーパスを利用した研究をさらに学際的に推進するために、政治学、経済学諸分野の研究者に意見を求め、学問的要求を調査したところ、議論の過程としての議会会議録に加えて、議論の結果(制定された条例や、採決における投票結果)の情報があることが望ましいこと[政治経済学分野]や、その地方独自の課題や自治体の地域特性など、議論の背景となるものに関する情報が必要であること[政治学]が明らかになった。

## 2. 研究の目的

本研究では、地方政治に関する研究の活性化・学際的応用を目指して、「議論の背景」「議論の過程」「議論の結果」を関連づけるコーパスを構築する。具体的には、新聞記事(議論の背景)、地方議会会議録(議論の過程)、条例(議論の結果)の3つの言語資源を地域、時間、課題の観点から関連づけることを目指す。本研究の目的は以下の3点である。

- A) 3つの言語資源(地方議会会議録・条例・新聞記事)の収集・整理
- B) 地域・時間・課題などの観点から3つの言語資源を関連づけるコーパスの構築
- C) 上記の成果を用いた、政治学・経済学・言語学・情報工学における応用研究

## 3. 研究の方法

本研究では、地方議会会議録(議論の過程)と条例(議論の結果)および新聞記事(議論の背景)を関連づけるコーパスを構築するために、まず地方議会会議録に加え、条例、新聞記事を収集、整理する。その上で、議論の過程、結果、背景を関連づけるために、会議録に含まれる議論と条例、および、新聞記事を地域・時間・課題の3つの観点から結びつける。また、構築したコーパスを用いて、政治学・経済学・言語学・情報工学における応用研究の支援を行う。本研究期間において以下の3点を行う。

- A) 3つの言語資源(地方議会会議録・条例・新聞記事)の収集・整理  
地方議会会議録・条例は、自治体の公式ウェブサイトから収集し、新聞記事は、有料の新聞記事コーパスを購入する。特に、地方議会会議録・条例は、自治体によりウェブ上に掲載する形式が異なるため、異なる形式ごとのプログラムを作成する。
- B) 地域・時間・課題の観点から3つの言語資源を関連づけるコーパスの構築  
3つの言語資源を用いて、地域課題に関する議題の「背景」「過程」「結果」を「時間(発言日時、新聞の日付など)」「地域(自治体名・地域区分名など)」「課題(キーワード・固有名詞など)」の観点から結びつける。
- C) 上記の成果を用いた、政治学・経済学・言語学・情報工学における応用研究の支援  
本課題では、特に、政治学・経済学の研究分野での利用を促進する。  
政治学の分野では、研究分担者の河村准教授(東北大)が、地方政治を専門とする政治学者であり、地方議会会議録コーパスを利用する研究を進めている。応募者は、H27に採択された河村(東北大)が研究代表者となっている研究(研究課題番号:15H02790)の連携研究者になっており、政治学分野の支援を行う。  
また、経済学の分野では、研究分担者の川浦教授(同志社大)が、地方議会会議録を用いた研究を進めており、本課題で構築するコーパスを用いて研究を行う予定である。

## 4. 研究成果

本研究では、3つの言語資源(地方議会会議録・条例・新聞記事)に対して、対象期間を2011年から2015年まで、対象自治体を47都道府県とした。地方議会会議録では、都道府県議会会議録を「発言テーブル」と「発言者テーブル」で構成される関係データベースの形で整理した。発言テーブルは下記のように17のフィールドを定義した。

1. 識別子 (主キー, 発言文に固有に振られた ID)
2. 自治体名 (「北海道」「青森県」「宮城県」…)
3. 回 (定例会の回数)
4. 号 (各定例会の中の号数)
5. 年 (開催年 (和暦))
6. 月 (開催月)
7. 日 (開催日)
8. 開催期間
9. 表題 (例:「平成24年 2月 定例会 (第335回) -02月27日-04号」)
10. 発言者の役割 (「議長」「質問者」「答弁者」)
11. 発言者 ID (発言者テーブルを参照する外部キー)
12. 発言者名 (例:「畠山和純君」「畑正芳君」)
13. 発言者の役職 (例:「議長」「議会事務局議事課長」「知事」)
14. 発言文 (例:「次に, 冬の節電対策について伺います。」)
15. 発言以外の記録文 (例:「(拍手)」「[…君登壇]」「----- (区切り線)」)
16. 原本 URL
17. HTML ファイルのパス

また、発言者テーブルには、9つのフィールドを定義した。

1. 発言者 ID (主キー, 発言テーブルの11番目の項目と対応)
2. 都道府県名 (例:「北海道」「青森県」「宮城県」)
3. 自治体名 (例:「宮城県」)
4. 氏名 (例:「畠山 和純」)
5. フリガナ (例:「ハタケヤマ カズヨシ」)
6. 対象行政区 (例:「気仙沼市」)
7. 生年 (投票日の西暦から投票日時点の年齢を減じて算出)
8. 性別 (「男」or「女」)
9. 職名 (例:「議会事務局議事課長」「知事」「会計管理者」「教育長」)

47 都道府県の異なる形式で公開されている会議録は、このデータベースに格納することで、自治体、年度、議員名などの観点から、検索できる全文検索に加えて、クロス表やマップ検索などの機能も追加した (<http://local-politics.jp/47pref>)。



本研究では、基盤となる会議録データベースを軸として、社会言語学、政治学、経済学の観点から応用研究を進めた。特に、政治・経済分野における応用研究としては、「選挙」に焦点を当て、「選挙結果と発言量」や「選挙結果と議員の発言内容」についての研究を進めた。

他にも、異なる言語資源として、条例および新聞についても考慮しながら、会議録と Wikipedia のタイトルを結びつける研究を進めた。QA Lab PoliInfo2 (<https://poliinfo2.net/>) の Entity Linking タスクでは、設議会会議録から条例を見つけるデータセットの構築を行った。下記のサイトからは、データセットのダウンロードも可能になっている。

<https://github.com/kmr-y/NTCIR14-QALab-PoliInfo-FormalRunDataset>  
<https://poliinfo.github.io/>

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件／うち国際共著 0件／うちオープンアクセス 0件）

1. 著者名 Yasutomo Kimura, Minoru Sasaki	4. 巻 7
2. 論文標題 Stance Classification Using Political Parties in Tokyo Metropolitan Assembly Minutes	5. 発行年 2018年
3. 雑誌名 The Seventh International Conference on Data Analytics	6. 最初と最後の頁 46-49
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yasutomo Kimura, Yuzu Uchida, Keiichi Takamaru	4. 巻 11
2. 論文標題 Speaker Identification for Japanese Prefectural Assembly Minutes	5. 発行年 2018年
3. 雑誌名 Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA)	6. 最初と最後の頁 1-5
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Ototake Hokuto, Sakaji Hiroki, Takamaru Keiichi, Kobayashi Akio, Uchida Yuzu, Kimura Yasutomo	4. 巻 14
2. 論文標題 Web-based system for Japanese local political documents	5. 発行年 2018年
3. 雑誌名 International Journal of Web Information Systems	6. 最初と最後の頁 357 ~ 371
掲載論文のDOI（デジタルオブジェクト識別子） 10.1108/IJWIS-12-2017-0085	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計26件（うち招待講演 1件／うち国際学会 1件）

1. 発表者名 洪水英潔, 木村泰知, 乙武北斗, 内田ゆず, 高丸圭一, 阪本浩太郎, 石下円香, 三田村照子, 森辰則, 神門典子
2. 発表標題 NTCIR-14 QA Lab-PoliInfo Formal RunにおけるSummarization Taskの結果と評価
3. 学会等名 言語処理学会第25回年次大会(NLP2019)
4. 発表年 2019年

1. 発表者名 木村泰知, 渋谷英潔, 森辰則, 神門典子
2. 発表標題 英国地方議会議録コーパスの構築に向けた分析
3. 学会等名 言語処理学会第25回年次大会(NLP2019)
4. 発表年 2019年

1. 発表者名 木村泰知, 佐々木稔
2. 発表標題 東京都議会の会派を用いたStance classificationの試み
3. 学会等名 言語処理学会第25回年次大会(NLP2019)
4. 発表年 2019年

1. 発表者名 佐々木稔, 木村泰知
2. 発表標題 Gaussian LDAを用いた地方議会議録のトピック分析
3. 学会等名 言語処理学会第25回年次大会(NLP2019)
4. 発表年 2019年

1. 発表者名 木村泰知, 渋谷英潔, 乙武北斗, 内田ゆず, 高丸圭一, 阪本浩太郎, 石下円香, 三田村照子, 神門典子
2. 発表標題 NTCIR-14 QA Lab-PoliInfoの Formal Run Dataset の構築
3. 学会等名 言語処理学会第25回年次大会(NLP2019)
4. 発表年 2019年

1. 発表者名 松森拓真, 谷内健太, 木村泰知
2. 発表標題 Wikipediaのカテゴリ情報を用いた地方議会会議録における発言に対する情報理解支援
3. 学会等名 言語処理学会第25回年次大会(NLP2019)
4. 発表年 2019年

1. 発表者名 乙武北斗, 高丸圭一, 内田ゆず, 木村泰知
2. 発表標題 述語項構造に基づく地方議会会議録からの意見抽出手法の検討
3. 学会等名 第34回ファジィシステムシンポジウム講演論文集
4. 発表年 2018年

1. 発表者名 木村泰知
2. 発表標題 政治情報における自然言語処理の課題
3. 学会等名 日本知能情報ファジィ学会誌
4. 発表年 2018年

1. 発表者名 木村 泰知, 渋谷 英潔, 阪本 浩太郎, 石下 円香, 三田村 照子, 神門 典子
2. 発表標題 政治情報の信憑性判断における情報アクセス技術の役割 - NTCIR-14 QA Lab-PoliInfoタスクの提案 -
3. 学会等名 情報処理学会 第236回自然言語処理研究会
4. 発表年 2018年

1. 発表者名 乙武 北斗、高丸 圭一、内田 ゆず、木村 泰知
2. 発表標題 一般公開版「都道府県議会議録検索システム」の概要
3. 学会等名 人工知能学会全国大会（第32回）
4. 発表年 2018年

1. 発表者名 内田 ゆず、高丸 圭一、乙武 北斗、木村 泰知
2. 発表標題 都道府県議会議録コーパスを用いた議員の議会活動の可視化に向けて
3. 学会等名 人工知能学会全国大会（第32回）
4. 発表年 2018年

1. 発表者名 木村泰知，関根聡，乾健太郎
2. 発表標題 地方議会議録の要約に向けて
3. 学会等名 言語処理学会第24回年次大会(NLP2018)
4. 発表年 2018年

1. 発表者名 木村泰知，戸嶋咲穂，洪木英潔
2. 発表標題 新聞記事における政治家の発言の引用記述と議会議録との対応関係の調査 フェイクニュース検出に向けて
3. 学会等名 言語処理学会第24回年次大会(NLP2018)
4. 発表年 2018年

1. 発表者名 松森拓真, 木村泰知, 坂地泰紀
2. 発表標題 地方議会議録における発言文の推定
3. 学会等名 言語処理学会第24回年次大会(NLP2018)
4. 発表年 2018年

1. 発表者名 井原 大将, 内田 ゆず, 高丸 圭一, 木村 泰知, 江崎 浩
2. 発表標題 全地方議会議録の横断検索に向けたデータ収集とデータ構造の検討
3. 学会等名 第233回自然言語処理研究会
4. 発表年 2017年

1. 発表者名 木村泰知, 内田ゆず, 高丸圭一
2. 発表標題 都道府県議会議録のパネルデータ作成に向けた発言者情報の付与
3. 学会等名 第33回ファジィシステムシンポジウム講演論文集
4. 発表年 2017年

1. 発表者名 田中琢真, 坂地泰紀, 小林暁雄, 木村泰知, 増山繁,
2. 発表標題 地方議会の議案収集に向けた議案一覧抽出の試み
3. 学会等名 電子情報通信学会技術研究報告
4. 発表年 2017年



1. 発表者名 田中琢真, 小林暁雄, 坂地泰紀, 内田ゆず, 乙武北斗, 高丸圭一, 木村泰知, 増山繁
2. 発表標題 都道府県議会会議録を対象とした議題・議案表現の自動抽出に向けた検討
3. 学会等名 人工知能学会全国大会
4. 発表年 2017年

1. 発表者名 木村泰知, 小林暁雄, 坂地泰紀, 内田ゆず, 高丸圭一, 乙武北斗, 吉田光男, 荒木健治,
2. 発表標題 議論の背景・過程・結果を関連づける地方政治コーパスの構築の試み
3. 学会等名 人工知能学会全国大会
4. 発表年 2017年

1. 発表者名 坂地泰紀, 酒井浩之, 小林暁雄, 内田ゆず, 乙武北斗, 高丸圭一, 木村泰知
2. 発表標題 都道府県議会会議録からの意見や意志を表す発言の抽出
3. 学会等名 言語処理学会第23回年次大会(NLP2017)
4. 発表年 2017年

1. 発表者名 木村泰知, 小林暁雄, 坂地泰紀, 内田ゆず, 高丸圭一, 乙武北斗, 吉田光男, 川浦昭彦
2. 発表標題 地方政治コーパス構築における従来の成果と現在の課題 - 政治・経済分野の応用研究に向けたパネルデータの構築-
3. 学会等名 言語処理学会第23回年次大会(NLP2017)
4. 発表年 2017年

1. 発表者名 田中琢真, 小林暁雄, 坂地泰紀, 内田ゆず, 乙武北斗, 高丸圭一, 木村泰知
2. 発表標題 都道府県議会議録を用いた地方政治コーパスの構築の試み
3. 学会等名 情報処理北海道シンポジウム 2016
4. 発表年 2016年

1. 発表者名 章原史敏, 木村泰知, 荒木健治
2. 発表標題 地方議会における自治体の政策・取組を引用する議員の発言の抽出
3. 学会等名 第9回テキストマイニング・シンポジウム, 電子情報通信学会技術研究報告
4. 発表年 2016年

1. 発表者名 田中 琢真, 小林 暁雄, 坂地 泰紀, 内田 ゆず, 乙武 北斗, 高丸 圭一, 木村 泰知
2. 発表標題 地方政治コーパス構築に向けた都道府県議会議録からの発言データの抽出
3. 学会等名 第32回ファジシステムシンポジウム
4. 発表年 2016年

1. 発表者名 Yasutomo Kimura, Keiichi Takamaru, Takuma Tanaka, Akio Kobayashi, Hiroki Sakaji, Yuzu Uchida, Hokuto Ototake, Shigeru Masuyama
2. 発表標題 Creating Japanese Political Corpus from Local Assembly Minutes of 47 Prefectures
3. 学会等名 Coling 2016 workshop, The 12th Workshop on Asian Language Resources (国際学会)
4. 発表年 2016年

1. 発表者名 木村泰知
2. 発表標題 議論の背景・過程・結果を関連づける地方政治コーパスの構築に向けて
3. 学会等名 電子情報通信学会「言語理解とコミュニケーション(NLC)」「思考と言語(TL)」合同研究会(招待講演)
4. 発表年 2016年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

地方議会会議録プロジェクト <a href="http://local-politics.jp/">http://local-politics.jp/</a> NTCIR-14 QA Lab PoliInfo <a href="https://poliinfo.github.io/">https://poliinfo.github.io/</a> NTCIR-14 QA Lab PoliInfoのデータセット公開 <a href="https://github.com/kmr-y/NTCIR14-QALab-PoliInfo-FormalRunDataset">https://github.com/kmr-y/NTCIR14-QALab-PoliInfo-FormalRunDataset</a>
---

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	川浦 昭彦 (Kawaura Akihiko) (10271610)	同志社大学・政策学部・教授  (34310)	
研究分担者	乙武 北斗 (Ototake Hokuto) (20580179)	福岡大学・工学部・助教  (37111)	
研究分担者	荒木 健治 (Araki Kenji) (50202742)	北海道大学・情報科学研究科・教授  (10101)	

## 6. 研究組織（つづき）

	氏名 (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	河村 和徳 (Kawamura Kazunori) (60306868)	東北大学・情報科学研究科・准教授  (11301)	
研究分担者	高丸 圭一 (Takamaru Keiichi) (60383121)	宇都宮共和大学・シティライフ学部・教授  (32207)	
研究分担者	湯浅 壘道 (Yuasa Harumichi) (60389400)	情報セキュリティ大学院大学・その他の研究科・教授  (32721)	
研究分担者	吉田 光男 (Yoshida Mitsuo) (60734978)	豊橋技術科学大学・工学（系）研究科（研究院）・助教  (13904)	
研究分担者	坂地 泰紀 (Sakaji Hiroki) (70722809)	東京大学・大学院工学系研究科（工学部）・特任講師  (12601)	
研究分担者	内田 ゆず (Uchida Yuzu) (80583575)	北海学園大学・工学部・准教授  (30107)	
研究分担者	渋木 英潔 (Shibuki Hideyuki) (30423982)	国立情報学研究所・情報社会相関研究系・特任研究員  (62615)	
研究分担者	小林 暁雄 (Kobayashi Akio) (20710130)	豊橋技術科学大学・工学（系）研究科（研究院）・助教  (13904)	