

科学研究費助成事業 研究成果報告書

平成 26 年 5 月 28 日現在

機関番号：10104

研究種目：基盤研究(B)

研究期間：2010～2013

課題番号：22300086

研究課題名(和文) 地方議会議録コーパスの構築とその学際的応用研究

研究課題名(英文) Construction of Japanese Regional Assembly Minutes Corpus and Its Interdisciplinary Research

研究代表者

木村 泰知 (KIMURA, Yasutomo)

小樽商科大学・商学部・准教授

研究者番号：50400073

交付決定額(研究期間全体)：(直接経費) 13,600,000円、(間接経費) 4,080,000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、研究者が利用可能な地方議会議録電子化コーパスを全国規模で構築しウェブ上で提供する。本研究の目的は以下の3点である。地方政治の研究者に向けた地方議会議録コーパスを構築しポータルサイト等を通して提供する。上記のコーパスを用いた政治学、社会言語学、情報工学の各分野の研究を行う。上記の研究成果を学際的に応用した政治情報システムの研究開発を行う。我々は、この地方議会議録コーパスを利用した政治学、社会情報学、情報工学の研究を行い、その成果を学際的に応用した政治情報システムの開発を行った。

研究成果の概要(英文)：Our goal is to construct a corpus of Japanese local assembly minutes, and provide available databases for each research toward both a research activity and an interdisciplinary collaboration. We carried out our tasks, which divided into three steps as follows: 1. We provide our assembly corpora with a lot of researchers through our web site (<http://local-politics.jp>). 2. We support researches, which include politics, social language and information engineering. 3. We construct a political information system, which enhances interdisciplinary collaboration. In this research, we achieved our goal, which supported corpus-based researches such as politics, social language and information engineering.

研究分野：総合領域

科研費の分科・細目：情報学・図書館情報学・人文社会情報学

キーワード：政治情報 会議録 地方議会

1. 研究開始当初の背景

日本政府が平成 11 年から進めてきた「平成の大合併」は、平成 17 年に施行された「合併特例新法」も影響し、平成 11 年 3 月末の時点で 3,232 存在した市町村の数が、平成 21 年 10 月時点で 1,772 に減少している(総務省の発表)。

この平成の大合併は、地方政治に関する研究に大きな影響を与えており、政治学では合併前後の違いに関する研究が数多く行われている。例えば、平野(2008)は平成の大合併前後に行われた市長選挙についての分析をしており、合併を行った市と行わなかった市の違いを当選者の属性から比較している。また、森脇(2008)は合併が地方議会や議員の活動に対して与えた影響を 856 議員にアンケート調査することで分析を行っている。

さらに、地方政治に関する研究は政治学以外の分野でも行われており、経済学の分野では川浦(2009)による「小規模自治体の多選首長は合併に消極的」という仮説を検証するために、北海道の地方議員、首長の情報を人手で調査している。また、我々も以前から、情報工学の分野において、一人ひとりの住民の興味や関心にマッチした政治情報を提供する住民本位型政治情報システムに関する研究開発(木村ら 2009)を行っており、北海道の 64 市町村を対象とした地方議会会議録のデータを収集、利用している。

こういった研究において、対象となるデータを独自に収集することは大きな負担であり、結果として小規模なデータに限定されてしまうといった研究遂行上の障害となることが多い。また、人文科学や社会科学の分野においてもコンピュータ上での処理が一般的になっているが、各研究者間で重複するデータの電子化作業などを個別に行っているといった非効率な状況も招いている。

- [1] 平野淳一, "「平成の大合併」と市長選挙", 日本選挙学会年報 選挙研究 第 24 巻第 1 号, pp32-39, 2008.
- [2] 森脇俊雅, "合併と地方議会活動: 議員アンケート調査の分析を中心にして", 日本選挙学会年報 選挙研究 第 23 巻, pp82-90, 2008.
- [3] 川浦昭彦, "Self-Serving Mayors and Local Government Consolidations in Hokkaido", 小樽商科大学・地域研究会 報告論文, 2009.
- [4] 木村泰知, 渋谷英潔, 高丸圭一, 地方議員と住民間の協働支援に向けたウェブの利用, 選挙研究第 25 巻第 1 号, pp.100-118, 2009.

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、研究者が利用可能な地方政治電子化コーパスを全国規模で

構築しウェブ上で提供する。また、そのコーパスを利用した政治学、社会情報学、情報工学の研究を行い、その成果を学際的に応用した政治情報システムの開発を行っている。本研究の目的は以下の 3 点である。

地方政治の研究者に向けた電子化コーパスを構築しポータルサイト等を通して提供する。

上記のコーパスを用いた政治学、社会言語学、情報工学の各分野の研究を行う。

上記の研究成果を学際的に応用した政治情報システムの研究開発を行う。

3. 研究の方法

本研究では、地方政治を研究対象とする様々な分野の研究者が利用できる大規模な地方政治電子化コーパスの構築、および、その提供方法としてのポータルサイトとウェブ API の公開を行う。コーパスは、今までに行ってきた北海道の地方議会会議録データの自動収集や加工の技術を活用し、全国の市町村の議会会議録を対象とした大規模な構築を行う。

次に、構築されたコーパスの学術的貢献の一環として、コーパスを利用した、政治学、社会言語学、情報工学における以下の研究を行う。政治学では「有権者の関心と地方議会会議内容との相違の分析」、社会言語学では「発話スタイルの変遷と地域差の分析」、情報工学では「地方議会会議録を対象とした情報抽出」に関する研究を行う。

また、上記の研究で得られるであろう知見は、我々がこれまでに行ってきた住民本位型政治情報システムの研究開発においても役立つことが期待され、最終年度までに、これらの知見を学際的に応用した研究成果として全国の市町村を対象とした政治情報システムの研究開発を行う。

4. 研究成果

本研究では、研究の目的 に書いた通り、地方政治の研究者に向けた電子化コーパスの構築を行った。現時点では、全国の自治体のうち 422 市区町村の会議録を収集し、約 80GB のデータが存在する。

また、研究目的 の各分野の研究としては、情報工学をはじめ、社会言語学、政治学の各分野において研究成果をあげている。特に、社会言語学の分野では、会議録に含まれる地域特有の表現を容易に見つけることが可能となり、研究背景に記述した「会議録データを独自に収集する大きな負担」を支援することができた。

さらに、最終年度は、研究目的 の学際的応用研究に向けて、我々の構築した地方議会会議録コーパスを研究者に利用してもらうことを試みた。地方議会会議録コーパスを利用する研究では、社会言語学の観点から、地

方議会会議録に含まれるオノマトペに着目した地域特有の言語表現の解析を行い、政治学の観点から、被災地の地方議会と地方議員についての考察を行い、情報工学の観点から、地方議会に含まれる政治課題の自動抽出及び、地域政治情報コーパスへの賛否の積極性に関する注釈付けを行った。

5. 主な発表論文等

〔雑誌論文〕(計 5 件)

1. 筒井貴士, 我満拓弥, 大城卓, 菅原晃平, 永井隆広, 洪木英潔, 木村泰知, 森辰則. 地方議会会議録コーパスの構築および政治情報システム構築を目標としたアノテーションの一提案. 自然言語処理, Vol. 21, No. 2, pp. 125--156, (2014).
2. 高丸圭一, 地方議会では何が話題になっているのか 宇都宮市議会会議録のテキストマイニング, 都市経済研究年報, 13, pp.162-173, 宇都宮共和大学都市経済研究センター, 2013年11月
3. 木村泰知, 関根聡, 主辞に基づく政治問題抽出手法, 人工知能学会論文誌, Vol.28, No.4, pp.370-378, 2013.
4. 木村泰知, 洪木英潔, 高丸圭一, 乙武北斗, 小林哲郎, 森辰則, "地方議員マッチングシステムにおける能動的質問のための質問生成手法", 人工知能学会論文誌, Vol. 26, No. 5, pp.580-593, 2011.
5. 高丸圭一, 洪木英潔, 木村泰知, "全国の市町村議会会議録のウェブ公開とデータ提供の状況", 都市経済研究年報, 第11号, pp.47-72, 宇都宮共和大学都市経済研究センター, 2011年11月

〔学会発表〕(計 15 件)

1. 高丸圭一, 内田ゆず, 乙武北斗, 木村泰知, 地方議会会議録におけるオノマトペの出現傾向に関する基礎的検討-少数の自治体に高頻度で出現するオノマトペについて-, 言語処理学会第20回年次大会, P5-16, 2014.
2. Fumitoshi Ashihara, Yasutomo Kimura and Kenji Araki, Clause Classification-based Retrieval of Demand from Local Assemblies Minutes, Sep2-9, Pacling2013.
3. 高丸圭一, 地方議会会議録コーパスにおける出現確率の相関を用いた文末表現の地域差の分析, 社会言語科学会第33回研究大会, pp.174-177, 2014.

4. 河村和徳, 被災地から考える 2013 年参議院選挙 - ネット選挙解禁を視野に, 『VOTERS』14号, 6-7頁, 2013.
5. 河村和徳, 被災地から考える地方議会と地方議員像, 『ガバナンス』, 34-36頁, 2014.
6. 湯淺壘道, 指定管理者制度と個人情報保護, 情報処理学会電子化知的財産・社会基盤研究会, 2013.
7. 高丸圭一, 乙武北斗, 洪木英潔, 木村泰知, 森辰則, 形態素 N-gram を用いた地方議会会議録コーパスの地域変異検出の試み 文末表現を例に 言語処理学会第19回年次大会発表論文集, pp.737-740, 2013.
8. 乙武北斗, 高丸圭一, 洪木英潔, 木村泰知, 森辰則, 地方議会会議録コーパスの学際的応用を目的とした n-gram データの構築およびウェブUIの試作, 言語処理学会第19回年次大会発表論文集, pp.733-736, 2013.
9. 筒井貴士, 我満拓弥, 洪木英潔, 木村泰知, 森辰則, 地域政治情報コーパスへの賛否の積極性に関する注釈作業, 言語処理学会第19回年次大会発表論文集, pp. 874-877, 2013.
10. 葦原史敏, 木村泰知, 荒木健治, 節の分類情報を用いた地方議会会議録における要求・要望表現抽出, NLC 研究会, 第2回 テキストマイニング・シンポジウム, pp.1-6, 2012.
11. 木村泰知, 洪木英潔, 高丸圭一, 乙武北斗, 森辰則, 地方議会会議録コーパスの構築とその利用, JSAI2012 人工知能学会全国大会第26回, 3B3-NFC-4-3, 2012.
12. 木村泰知, 関根聡, 主辞に着目した政治問題の定義と注釈付け, 言語処理学会第18回年次大会論文集, F1-6, 2012.
13. 菅原晃平, 大城卓, 齋藤誠, 永井隆広, 洪木英潔, 木村泰知, 森辰則, 地方議会会議録コーパスの拡充における問題点の分析と対処, 言語処理学会第18回年次大会論文集, P1-15, 2012.
14. 葦原史敏, 木村泰知, 荒木健治, 地方議会会議録における要求・要望表現抽出の提案, 言語処理学会第18回年次大会論文集, P1-27, 2012.
15. 大城卓, 渡邊裕斗, 洪木英潔, 木村泰知, 森辰則, 地方政治情報システムのための地方議会会議録への注釈付けタグセットの提案, 言語処理学会第18回年次大会論文集, P3-9, 2012.

〔その他〕

ホームページ等
<http://local-politics.jp/>

6. 研究組織

(1)研究代表者

木村泰知 (KIMURA Yasutomo)
小樽商科大学・商学部・准教授
研究者番号：50400073

(2)研究分担者

乙武北斗 (OTOTAKE Hokuto)
福岡大学・工学部・助教
研究者番号：20580179

河村和徳 (KAWAMURA Kazunori)
東北大学・情報科学研究科・准教授
研究者番号：60306868

高丸圭一 (TAKAMARU Keiichi)
宇都宮共和大学・准教授
研究者番号：60383121

森辰則 (MORI Tatsunori)
横浜国立大学・環境情報研究科・教授
研究者番号：70212264

湯淺壘道 (YUASA Harumichi)
情報セキュリティ大学院大学・教授
研究者番号：60389400