

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 22 日現在

機関番号：32207

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26370498

研究課題名(和文)学際的応用を考慮した地方議会議録コーパスの言語学的研究

研究課題名(英文)Linguistic Research on Corpus of the Local Assembly Minutes Considering Interdisciplinary Application

研究代表者

高丸 圭一 (Takamaru, Keiichi)

宇都宮共和大学・シティライフ学部・准教授

研究者番号：60383121

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：地方議会議録に出現するオノマトペ(擬音語・擬態語)の使用実態、語義と共起語の関係について分析を行った。また、多変量解析を用いてオノマトペ使用の地域差について分析を行った。地方議会議録にみられる気づかない方言、二人称代名詞の地域差等について分析を進めた。平成23～26年度の4年分の都道府県議会議録を対象として、TF・IDFを用いて特徴語を抽出し、発言内容の地域差、時間変化、個人差を分析した。

研究成果の概要(英文)：We analyzed local assembly minutes and revealed usage of onomatopoeia (mimetic words) and relation between co-occurrence words and meaning of onomatopoeia. In addition, regional differences of frequency of mimetic words were analyzed using multivariate analysis. Unnoticed dialects and regional differences of second person pronouns were also analyzed. Feature words of local assembly minutes were extracted using TF*IDF algorithm. The target period is FY 2011 - FY 2014. The regional differences, time changes, individual differences of the feature words were analyzed.

研究分野：社会言語学

キーワード：地方議会議録 オノマトペ 地域差 特徴語

1. 研究開始当初の背景

2010年現在、すべての都道府県議会および市区町村議会の73.4%が会議録をウェブに公開しており、地方自治体の取り組み(地方議会会議録)はインターネットを介して閲覧できるようになってきた。地方議会会議録は、都道府県または市区町村議会における議員や首長、行政職員などの発言の記録である。属性(年齢・性別・肩書きなど)が明らかで特定の自治体に居住する者の発言が、地域別・年度別に記録されたものであると捉えることができる。一つの地方議会の会議録を遡れば通事的な言語変遷を辿ることができ、全国の地方議会会議録を横断的に調査すれば、地域差などを分析することができるため、言語資源として注目に値する。また地方議会会議録は、言語研究のみならず、政治学、自然言語処理等の様々な研究分野で利用すべき資源であるが、自治体ごとに個別に提供されているため横断的な分析・調査は手つかずである。木村科研(No.22300086)では、地方自治体の議会事務局がウェブに公開している会議録を収集・整形し、関係データベースに登録して学際的に利用することを目指した研究が行われている。

2. 研究の目的

全国の地方議会会議録を収録したコーパスを利用して、地方議会発言の言語的特徴および言語的変異を分析し、コーパス言語学、社会言語学に寄与する知見を得ることを目的とする。さらに、その成果を学際的に応用可能な形でまとめることを目指す。地方議会会議録は重要な言語資源であるにもかかわらず、これに焦点を当てた研究は少ない。発言年度、地域、発言者の属性等が明らかで、日本全国を網羅している地方議会会議録を研究対象とすることが本研究の特色である。本研究では、公的な文書である会議録に、修正されずに記録される気づかない方言、方言以外の会議録の地域差(全文の差、議題の差、表現方法・記述方法の差)、テキストマイニング分析で得られる年度、発言者、地域による言語的特徴などの重要な知見を得ることが予想される。これらはコーパス言語学的な研究成果であると同時に、日本語学、社会言語学の成果としても価値を持つ。さらには、地域行政の動向に関係する諸分野で役立つという点でも意義がある。

3. 研究の方法

(1) データ検証・追加収集

木村科研が平成23年までに収集した地方議会会議録の提供を受けて研究を進める。この会議録コーパスは発言文字列に自治体名、開催年月日、発言者名などの情報がすでに付

与されており有用であるが、データ収集およびタグ付けをプログラムによって自動的に行う。本研究では正確な分析を行うために、データベースの各フィールドにデータが適切に格納されているかをまず人手で確認し、適宜修正を行う。

(2) 地域差の分析

会議録は書き起こしの過程で整文が行われ、方言は共通語に修正される場合がある。しかし、議論の文脈上必要な方言や、「離合」や「～してみえる」の例のように気づかない方言はそのまま記載される。まず、検索・可視化システムを整備し、コーパスの全文検索によって、さまざまな表現の出現状況の確認し、分布の地図化を行う。

次に、議会発言に出現するオノマトペ(擬音語・擬態語)を対象として、出現頻度、係り先などの地域差について分析を進める。オノマトペは短いひらがなまたはカタカナの文字列であり、書きことばにおいても特殊拍(長音・促音・撥音)の多様な挿入が生じるため、自動抽出が容易ではない。このため、オノマトペ抽出手法について研究を進める。その上で、地方議会会議録に見られるオノマトペの出現頻度の地域差についての分析を行う。さらに、オノマトペの語義と共起語(係り先動詞)の関係についての分析を進める。

(3) 特徴語の分析

全国47都道府県議会の本会議、収集期間は平成23年4月の統一地方選挙から平成27年4月の統一地方選挙の前(平成27年3月)までの4年間を分析対象とする。統一地方選挙の時期以外に議員選挙を行う議会も存在するが、収集範囲をこの4年間に設定することにより、議員の出入りがもっとも少なくなるという利点がある。対象となる本議会会議録は、約400万文、1.5GBである。

TF・IDF法を用いて、都道府県別、発言時期別、議員別の特徴語を抽出し、それぞれの特徴比較を試みる。基本的には、ある文書において出現頻度(TF)が高い語がその文書における特徴語であると考えられる。ただし、その語が複数の文書で共通に用いられる語であるとき、語はその文書における特徴語といえないため、他の文書で共通に使われる語の順位が下がるような係数(IDF)をTFに掛け、単語を順位付けする手法である。都道府県別の特徴語抽出では、1つの議会における発言全体を1文書として、47文書(都道府県議会)の特徴語をそれぞれ求める。発言機関別の特徴語では、全議会における四半期(3カ月)の間の発言を1文書にまとめ、全16文書(4年×四半期)の特徴語を求める。議員別の特徴語は1議員の発言を1文書とし、議会ごとに発言議員の人数分の文書の特徴語を求め

る。

特徴語の抽出では、形態素解析を行い、各単語の頻度を求める。ただし、形態素解析では、例えば「離島航空路線」は「離島」「航空」「路線」の3語に分割され、「地域医療」は「地域」「医療」の2語に分割される。このまま「路線」や「地域」といった語の頻度を数えても、地域や期間の特徴を捉えることは難しいため、名詞連続(複合名詞)については、ひとまとまりの語として取り扱うことが適当である。そこで、本稿では中・長単位解析ツール Comainu を用い、長単位を基準として、品詞情報が「名詞-普通名詞-一般」のものを対象に、特徴語の抽出を行う。

4. 研究成果

(1) 地域差の分析

全国 403 自治体から収集した 2010 年度の地方議会会議録コーパス約 1000 万文(3 億語)を対象として、現代の話しことばにおけるオノマトペを分析した。形態素解析を用いて、186,416 例(異なり語数 982 語)のオノマトペ候補を抽出した。地方議会会議録コーパスを分析対象とした本研究の第一の知見は、やや改まった話し方をするであろう公的な場での話しことばにおいてもオノマトペが豊富に使用されていることである。特に、「しっかり」「どンドン」など、政策の推進を求めるときに使用されると考えられる表現や「はっきり」といった厳格な判断や言及を求める表現が高い頻度で出現した。さらに、「そろそろ」「ゆっくり」「じっくり」といった政策実行の時期や議論の進行に関連するとみられるオノマトペなどが出現することを確認した。また、オノマトペの出現確率は西日本においてやや高い傾向があり、統計的有意差が見られるものの、特定の地方で顕著に多く出現するという特徴は見られないことを明らかにした。

次に、手作業によって形態素解析における抽出誤りを除外した 155 語(12,512 例)のオノマトペを対象に出現傾向の地域差を分析した。全体的な傾向として、西日本においてオノマトペの出現確率が有意に高いことが明らかになった。また、地方別の出現頻度を対応分析した。第 1~3 次元において、それぞれ「九州沖縄」「近畿」「中国四国」に出現するオノマトペがそれぞれ分離して表れることを確認した。(図 1,2)九州沖縄地方では方言的語義の「ぴしゃっ」が多く出現することを確認した。漠然とした大まかなさまを表す「ばくっ」が近畿地方を中心として分布していることを確認した。また、中国四国地方で多く観察された「きらら」は山口県において固有表現(名称)に多用されていた。

このほか、方言語彙の出現分布(図 3)、二

人称の使用(図 4)、文末表現(図 5)などの変異を明らかにした。

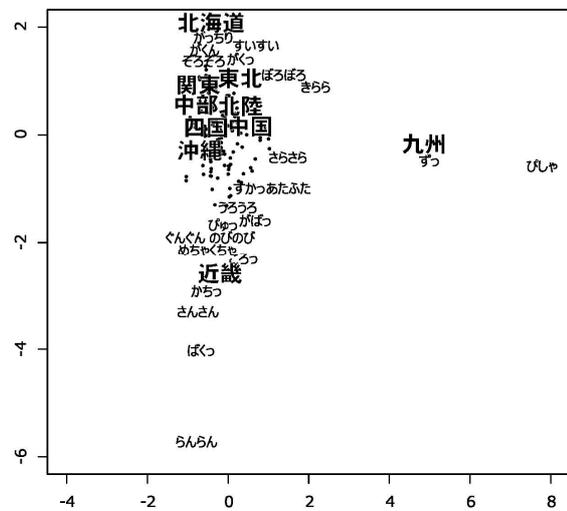


図 1 オノマトペ使用地域の対応分析
(横軸：第 1 次元，縦軸：第 2 次元)

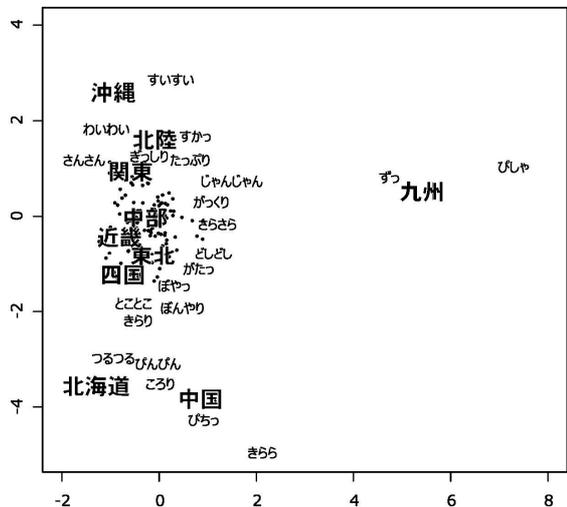


図 1 オノマトペ使用地域の対応分析
(横軸：第 1 次元，縦軸：第 3 次元)

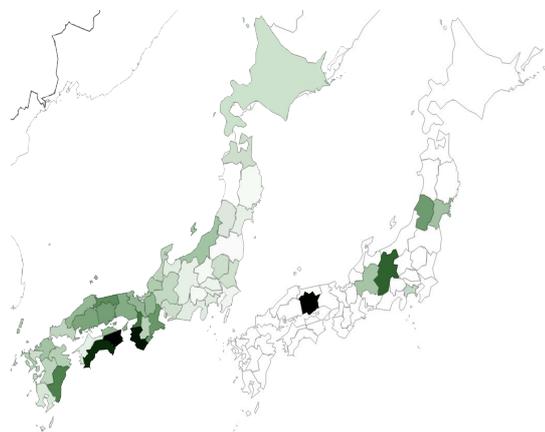


図 2 二人称の出現分布
(「あなた」(左)「おめえ」(右))

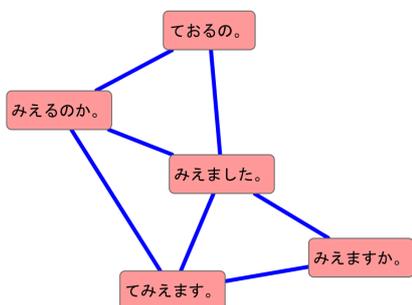


図3 出現地域が類似する文末表現のネットワーク表現

(2) 特徴語の分析

都道府県間の比較では、各地の政治課題が多く見られた。特に、医療施設（例えば「ふくしま国際医療科学センター」「さいたま赤十字病院」「東千葉メディカルセンター」）は新設や移転にあたりそれぞれの地域で活発な議論があり、上位に挙がった。このほか、茨城県の「森林湖沼環境税」（期間延長）、東京都の「二〇二〇年大会」（オリンピック）、長野県の「浅川ダム」「F-V断層」（地震への対応）、鳥取県の「環境大学」（公立化）なども分析対象期間に議論になった政治課題であるといえる。

また、地域の特産物・ブランド・キャラクタ等の名称、例えば「鳴門わかめ」「すだち」「あか牛」「とちまるショップ」が多く見られた。結果の一例を表1に示す。

表1 都道府県別の特徴語の一例

	特徴語		
	第1位	第2位	第3位
宮城県	仙台牛	貞山運河	水産業復興特区
茨城県	森林湖沼環境税	いばらきづくり	つくば国際戦略総合特区
千葉県	東千葉メディカルセンター	北千葉道路	千葉ニュータウン
東京都	都有地	都内中小企業	二〇二〇年大会
石川県	小松基地	のと里山海道	IRいしかわ鉄道
鳥取県	子育て王国	環境大学	まんが博
大分県	安心・活力・発展プラン	大分県づくり	豊後牛

発言期間別の特徴語の比較から政治課題変遷を読み取ることができることを示した。例えば、宮城県の平成23年度第1~2四半期に、東日本大震災関連の語が上位に多く見られる様子が観察された。議員別の特徴語の比較からは各議員の政治的関心分野を読み取ることができることを示した。例えば、栃木県議会W議員の特徴語は「モニタリングボス

ト」「干しシイタケ」「二重被害」「キノコ生産者」「クマザサ」であり、議員が放射能汚染の被害について集中的に発言していることが読み取れる。

有権者（一般市民）が議会会議録の全て読むことは労力がかかるため、現実的ではないが、このように特徴語を抽出することで、それぞれ議員の議会における発言傾向を簡単につかむことが可能となる。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕(計6件)

高丸圭一, 内田ゆず, 木村泰知「地方政治コーパスにおける都道府県議会会議録パネルデータの基礎分析」『宇都宮共和大学シティライフ学論叢』, 第18号, pp.136-155 (査読無)

Yasutomo Kimura, Keiichi Takamaru, Takuma Tanaka, Akio Kobayashi, Hiroki Sakaji, Yuzu Uchida, Hokuto Ototake and Shigeru Masuyama「Creating Japanese Political Corpus from Local Assembly Minutes of 47 Prefectures」『Proceedings of the 12th Workshop on Asian Language Resources』 ISBN 978-4-87974-722-8 (査読有)

(<https://aclweb.org/anthology/W/W16/W16-5410.pdf>)

井上史雄「日本語 2人称代名詞の地方議会会議録における出現」『明海日本語』 Vol.21, pp.1-16 (査読有)

(<http://www.urayasu.meikai.ac.jp/japanese/meikainihongo/21/meikainihongo21.pdf>)

高丸圭一, 内田ゆず, 乙武北斗, 木村泰知「係り先動詞に着目したオノマトペの語義分類に関する検討」『知能と情報』, Vol.28, No.4, pp.693-699 (査読有)

(DOI: 10.3156/jsoft.28.693)

二階堂整, 川瀬卓, 高丸圭一, 田附敏尚, 松田謙次郎「地方議会会議録による方言研究 セミフォーマルと気づかない方言」方言の研究, 1, pp.299-324 (査読有)

高丸圭一, 内田ゆず, 乙武北斗, 木村泰知「地方議会会議録コーパスにおけるオノマトペ 出現傾向と語義の分析」人工知能学会論文誌, Vol.30, No.1, SP2-K, pp.306-318 (査読有)

(DOI:10.1527/tjsai.30.306)

〔学会発表〕(計12件)

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「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」言語

処理学会第23回年次大会 C1-5 (2017-03)
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「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」言語処理学会第23回年次大会,
D4-3 (2017-03)

乙武北斗, 内田ゆず, 高丸圭一, 木村泰知
「オノマトペが修飾する名詞の抽出と分
析 「-の」「-(と)した」を対象として」
第32回ファジィシステムシンポジウム,
WD3-2 (2016-08)

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「地方政治
コーパス構築に向けた都道府県議会会議
録からの発言データの抽出」第32回ファ
ジィシステムシンポジウム, WD3-4
(2016-08)

内田ゆず, 高丸圭一, 乙武北斗, 木村泰知
「BCCWJにおけるオノマトペの語義とコ
ロケーションの分析」JSAI2016 第30回人
工知能学会全国大会,
1K4-0S-06a-2(2016-06)

乙武北斗, 内田ゆず, 高丸圭一, 木村泰
知「表層格に着目したオノマトペ共起語の
抽出と分析」NLP2016, 言語処理学会第22
回年次大会, P7-7, pp.195-198 (2016-03)
高丸圭一「地方議会会議録に見られるオノ
マトペ」日本語学会2015年度秋季大会ワ
ークショップ「地方議会会議録から見える
日本語のバリエーション」

高丸圭一, 内田ゆず, 乙武北斗, 木村泰知
「現代オノマトペ・コロケーション・デー
タベース構築のための係り先動詞の分析」
第31回ファジィシステムシンポジウム,
pp.415-420 (2015-09)

内田ゆず, 高丸圭一, 乙武北斗, 木村泰知
「BCCWJ コアデータにおけるオノマトペ
出現実態の分析」JSAI2015 第29回人工
知能学会全国大会, 3G4-0S-05b-5
(2015-05).

二階堂整, 川瀬卓, 高丸圭一, 田附敏尚,
松田謙次郎「地方議会会議録による方言研
究の可能性」日本方言研究会第99回研究
発表会発表原稿集, pp.57-64 (2014-10)

高丸圭一, 内田ゆず, 乙武北斗, 木村泰知
「地方議会会議録コーパスを用いたオノ
マトペの分析」第6回コーパス日本語学ワ
ークショップ予稿集, pp.83-92 (2014-09)

木村泰知, 洪木英潔, 内田ゆず, 乙武北斗,
高丸圭一, 森辰則「地方議会会議録におけ
るオノマトペの自動抽出手法の提案」第
30回ファジィシステムシンポジウム,
pp.638-641 (2014-09)

〔その他〕

ホームページ

・高丸研究室

<http://www.takamaruzeni.com/>

(研究代表者による研究業績、および、本研
究成果の概要を公開している。)

・地方議会会議録コーパスプロジェクト

<http://local-politics.jp/>

(研究グループで取り組んでいる地方議会
会議録分析の概要、および、本研究で行っ
た地方議会会議録の特徴語分析の結果を
公開している。)

・オノマトペを分析する研究グループの情報
発信サイト

<http://ono-collo.com/>

(研究グループで取り組んでいるオノマト
ペ研究の概要、および、本研究で行った会
議録に含まれるオノマトペの概要を公開
している。)

新聞記事

日本経済新聞「現代言葉考」(井上史雄)
の以下の記事に本研究の成果が用いられて
いる。

・言葉の乱れ政治家にも一因(平成28年8月
28日)

・会議録に残る議員の雑言(平成28年9月4日)

・地方議員の言い方、関西が丁寧(平成28年
9月25日)

・秋の空は女心か 男心か(平成28年10月9日)

・地方議会、「不祥事」表現にも差(平成28
年10月23日)

・西日本、方言の使い分け発達(平成28年10
月30日)

6. 研究組織

(1) 研究代表者

高丸 圭一 (TAKAMARU Keiichi)

宇都宮共和大学・シティライフ学部・准教授
研究者番号: 60383121

(2) 研究分担者

木村 泰知 (KIMURA Yasutomo)

小樽商科大学・商学部・准教授

研究者番号: 50400073

乙武 北斗 (OTOTAKE Hokuto)

福岡大学・工学部・助教

研究者番号: 20580179

井上 史雄 (INOUE Fumio)

東京外国語大学・その他の部局等・名誉教授

研究者番号: 40011332

内田 ゆず (UCHIDA Yuzu)

北海学園大学・工学部・准教授

研究者番号: 80583575