

保守系・リベラル系議員の衆議院予算委員会質問における 言葉遣いの特徴分析

樋口心

s1511175@u.tsukuba.ac.jp

掛谷英紀

kake@iit.tsukuba.ac.jp

筑波大学

1. はじめに

政治や社会問題に関する言語資源の分析に計算機科学の手法を持ち込む研究は、欧米では数多くある[1,2]。たとえば、上院議員[3]や最高裁判事[4]のイデオロギー傾向の分析、法案成立の可否を予測する研究[5]などが行われている。

日本でも、政治に関する言語資源を機械学習することによって、議員や政党の特徴を分析する研究はいくつも行われている。畑中らは国会会議録を最大エントロピー法[6]で機械学習することで、政党の特徴を分析する手法を提案している[7]。橋本らは、政治家のレビューサイトにおけるレビューのテキストを最大エントロピー法で機械学習し、その結果をアマゾンのブックレビューとリンクさせ、支持する政党別に本の推薦を行うシステムを提案している[8]。また、中川らはランダムフォレストを用いた国会会議録のイデオロギー分析を行っている[9]。さらに、大南、掛谷らは短命に終わる議員・大臣の発言の特徴を分析している[10-12]。

党派による言葉遣いの違いについて、最近米国で興味深い研究成果が発表された。DupreeとFiskeは、過去25年間、計74回にのぼる大統領選挙の演説を調査し、民主党の候補の演説において、聴衆に黒人が多い場合と白人が多い場合では言葉遣いが異なるという結果を得た[13]。民主党の候補は、黒人が多い場合のみ、わざと平易な言葉を選び共感を呼び起こすという戦略が用いられていることが見出された。さら

に、同研究では一般の被験者に対して白人宛と黒人宛にメールの文面を作らせる実験も行われた。その結果、リベラル派であると自認する被験者について、やはり黒人相手の場合、文面に平易で温かみのある言葉を有意に多用する傾向があり、一方保守派を自認する被験者についてはメールの相手によって、使用する言葉に違いがないことが分かった。この結果は、普段差別反対を主張しているリベラル派の言葉遣いの中に、むしろ差別意識が見出されることを示している点で、注目に値する。

そこで、本研究は、日本においては保守派とリベラル派の間で、言葉遣いの違いにどのような違いが見られるかを調べることを目的とする。具体的には、国会会議録をもとに、リベラル系議員と保守系議員の間に国会質問における言葉遣いにどのような違いが見られるかを分析する。分析対象とする言語資源としては、最も多様な話題を含む衆議院予算委員会における国会質問を用いる。

2. 研究手法

本研究の分析手法について示す。まず、国会会議録検索システム[14]から入手した国会会議録の答弁の文書データを、形態素解析ツール MeCab[15]を用いて形態素ごとに分割する。次に、**読点の直前3語または4語を末尾表現と定義し、表出形をそのまま機械学習の素性として使用する。**末尾表現に着目したのは、

その年ごとの政治のテーマに依らないこと、言葉遣いとして無意識の違いが出やすいことが理由である。このように抽出した素性をもとに、学習データおよびテストデータを作成する。その後、学習データに対して、各カテゴリ間の特徴を機械学習し、テストデータを使用して得られたシステムの精度を算出する。精度を算出する際にはクロスバリデーションを用いる。以上の手順を図1に示す。

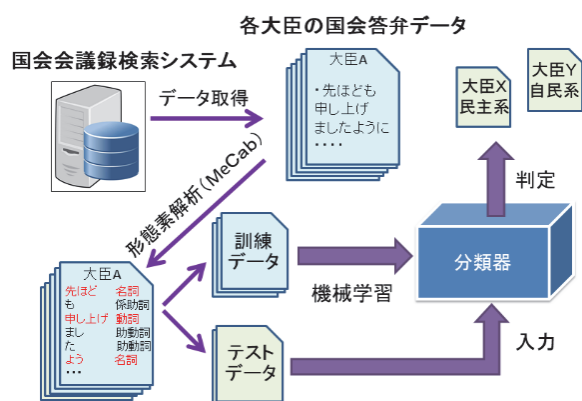


図1 システムの概要

本研究ではリベラル系議員と保守系議員を比較するために、両カテゴリ衆議院予算委員会の野党の発言を収集する。具体的には、保守系議員を、民主党政権時代の2009/9/16～2012/12/25の期間野党であった自民党と公明党の予算委員会所属の衆議院議員、リベラル系議員を、自民・公明連立政権時代の一部である2015/4/16～2018/7/31の期間野党であった民主党および分裂後の旧民主党系の予算委員会所属の衆議院議員と定義する。衆議院予算委員会を対象にした理由は、上述の通り政府の政策が横断的に話されていることによる。

上記の条件を満たす議員のリストを表1と表2に示す。今回の実験では両カテゴリの人数・データ量が不均衡であると、単語出現比率や機械学習の結果に悪影響を及ぼす。そこで、当該期間に予算委員会の委員であっても発言数が著しく多い、もしくは少ない、発言が無い議員は分析対象から除いている。

本研究では言語データを収集するために、国会会議録検索システム検索用 API を利用して分析対象の議員の発言を収集する。上述の通り、民主党系(リベラル系)および自民党系(保守系)の議員の過去の衆議院予算委員会における発言から、両カテゴリの特徴分析を試みる。なお、国会会議録は図2のような構造を持つため、「○ 役職(氏名)」の部分を取り除き、それに続く部分をその人物の発言として抽出する。

以上の条件で抽出すると、取得発言データの合計は

民主党系大臣:1109333 字

自民党系大臣:1345163 字

となった。これらのデータをもとに学習データを作成し、機械学習を行う。

表1 当該期間に衆議院予算委員会の委員で、かつ発言を収集した議員のリスト

旧野党	伊東良孝	自民党	現野党	阿部知子	民進党
	遠山清彦	公明党		逢坂誠二	立憲党
	塩崎恭久	自民党		井出庸生	希望党
	下村博文	自民党		岡田克也	民主党
	加藤紘一	自民党		岡本充功	民進党
	橋慶一郎	自民党		階猛	民主党
	金子一義	自民党		岸本周平	民主党
	金田勝年	自民党		宮崎岳志	民主党
	高木陽介	公明党		古川元久	民主党
	佐田玄一郎	自民党		細野豪志	民主党
	山本幸三	自民党		山井和則	民主党
	小泉進次郎	自民党		篠原孝	民主党
	小池百合子	自民党		初鹿明博	民進党
	小野寺五典	自民党		小熊慎司	希望党
	小里泰弘	自民党		小山展弘	民進党
	菅義偉	自民党		松野頼久	民進党
	菅原一秀	自民党		西村智奈美	民進党
	赤澤亮正	自民党		青柳陽一郎	立憲党
	谷川弥一	自民党		石関貴史	民進党
	谷畑孝	自民党		前原誠司	民主党
	馳浩	自民党		大串博志	民主党
	町村信孝	自民党		長妻昭	民主党
	田村憲久	自民党		津村啓介	希望党
	東順治	公明党		辻元清美	民主党
	富田茂之	公明党		渡辺周	民進党
	武部勤	自民党		馬淵澄夫	民主党
	野田毅	自民党		本村賢太郎	民主党
	齋藤健	自民党		落合貴之	立憲党

表 2 当該期間に衆議院予算委員会の委員で、かつ発言を収集しなかった議員のリスト

旧野党	鳩山邦夫	自民党	現野党	稲富修二	希望党
	安倍晋三	自民党		岡本あき子	立憲党
	園田博之	自民党		山内康一	立憲党
	岸田文雄	自民党		大西健介	民進党
	山本拓	自民党		玉木雄一郎	民主党
	大口善徳	公明党		後藤祐一	民主党
	大野功統	自民党		緒方林太郎	民進党
	田野瀬良太郎	自民党		小川淳也	民主党
	石破茂	自民党		福島伸亨	民進党

○国務大臣(小泉純一郎君) 基本においては、障害者にとって住みやすい町をつついでいこう。そこで郵政省といたしまして、今国会にはテレビの字幕放送、解説放送そして車いす用公衆電話ボックスの設置等、そういう助成等を図るための法案を提出して、去る五月十八日に成立いたしました。
各省協力しながら、いわゆるノーマライゼーションといいますが、完全参加と平等という、障害者も健全者と同様にいろんな事業に参加していく、障害者に優しい町づくりに郵政省としても協力していきたい、そう思っております。

図 2 国会会議録の構造

3. 実験

収集した発言を表 1 に示されている議員毎にまとめ、どちらのカテゴリに分類されるか二値分類問題として機械学習を行う。2 章で説明したように今回は読点から末尾 4 語のみを末尾表現として素性に用いる。また、クロスバリデーションはリーブワンアウト法で行う。

上記のデータセットに基づき、最大エントロピー法による機械学習を行った結果を表 4 に示す。正解率が 71.4% となり、末尾表現だけによって自民党系野党、民主党系野党を言い当てることある程度可能である。このことは、両カテゴリの末尾表現の使用頻度に目立った違いがあることを示唆する。

今回の実験では両カテゴリについて大量の末尾表現を収集している。よって、単語の出現頻度も各カテゴリの特徴を得られる重要な情報源となる。そこで、両カテゴリの素性から、カテゴリによって出現頻度が大きく異なる素性を抽出する。特定の議員のみが極端に使用している末尾表現を各カテゴリの特徴語として抽出してしまわないように、発言を収集した議員 56 人のうち半数の 28 人以上が使用している末尾表現のみ抽出し

た。使用割合の偏りが大きい順に、自民党系議員に特徴的な素性を表 5、民主党系議員に特徴的な素性を表 6 に示す。

自民党系議員で特徴的な素性である「わけでありませす」は「そう思うわけでありませす」、「なんですよ」は「チャンスなんですよ」といった使われ方が主にされている。他の特徴語を見ても語尾が「ませす」・「ませました」で終わる末尾表現が多く、自分の主義・主張を述べるような発言が多いことがわかる。一方、民主党系議員の特徴的な素性である「ということませすね」は「趣旨でおっしやつたということませすね」、「わけませすね」は「おっしやつているわけませすね」といった使われ方が主にされている。同様にほかの特徴語をみても「ませすか」「ませすね」で終わる末尾表現が多く、相手の言質を取るような発言が多いことがわかる。

表 4 判定結果

	再現率	適合率	自民党 判定数	民主党 判定数
自民党	57.1%	80.0%	16	12
民主党	85.7%	66.7%	4	24

表 5 自民党系議員の特徴的な素性の一部抜粋

末尾表現	民主党系 発言回数	自民党系 発言回数	使用割合
わけでありませす	185	465	2.51
思っております	170	295	1.74
ありませんか	61	101	1.66
なんですよ	172	258	1.5
どうませすか	84	123	1.46
ないんですよ	168	245	1.46
でありませました	53	76	1.43
ていませすか	68	91	1.34
てきませました	65	77	1.18
てありませました	57	66	1.16

表 6 民主党系議員の特徴的な素性の一部抜粋

末尾表現	民主党系 発言回数	自民党系 発言回数	使用割合
んじゃないか	65	24	2.71
いるわけですね	101	48	2.1
ているのか	55	27	2.04
いるわけですよ	92	47	1.96
ということですね	67	37	1.81
いかがですか	205	117	1.75
思うんですよ	108	64	1.69
わけですよ	67	40	1.68
いかがでしょうか	153	95	1.61
いるんですね	183	116	1.58

こうした違いが生じる原因の一つとして、民主党系の議員は野党時代が長いこと、質問する立場での発言に慣れている一方、自民党系の議員は与党でいる期間が長いので、政府の責任追及型の質問に慣れていなかったことが考えられる。

4. まとめ

本研究では国会会議録をもとに、保守系とリベラル系の議員の言葉遣いの特徴分析を試みた。末尾表現を素性とした機械学習の結果、71.4%の正解率を得た。よって、両者には言葉遣いに一定の特徴的な違いがあると考えられる。それぞれでよく見られる素性としては、保守系(自民党系)の議員では自らの主義主張を述べる発言が多い一方、リベラル系(旧民主党系)議員には言質をとることを意図した発言が多いことが分かった。

参考文献

- [1] Daniel Hopkins and Gary King: “A Method of Automated Nonparametric Content Analysis for Social Science,” *American Journal of Political Science*, 54(1), pp. 229-247, 2010.
- [2] John Wilkerson and Andreu Casas: “Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges,” *Annual Review of Political Science*, 20, pp. 529-544, 2017.
- [3] Daniel Diermeier, Jean-François Godbout, Bei Yu and Stefan Kaufmann: “Language and Ideology in Congress,” *British Journal of Political Science*, 42(1), pp. 31-55, 2011.
- [4] Benjamin E. Lauderdale and Tom S. Clark: “Scaling Politically Meaningful Dimensions Using Texts and Votes,” *American Journal of Political Science*, 58,(3), pp. 754-771, 2014.
- [5] Tae Yano, Noah A. Smith, and John D. Wilkerson: “Textual predictors of bill survival in congressional committees,” *NAACL HLT '12 Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 793-802, 2012.
- [6] Adam L. Berger, Stephen A. Della Pietra, and Vincent J. Della Pietra: “A Maximum Entropy Approach to Natural Language Modeling. *Computational Linguistics*,” 22(1), pp. 39-71, 1996.
- [7] 畑中允宏, 村田真樹, 掛谷英紀: “新聞社説・国会議事録に基づく言論のイデオロギー別分類,” *言語処理学会第 15 回年次大会発表論文集*, pp.408-411, 2009.
- [8] 橋本悠, 掛谷英紀: “Web 上のレビュー記事のイデオロギー分析とその応用,” *言語処理学会第 16 回年次発表論文集*, pp.740-743, 2010.
- [9] 中川侑, 武田拓也, 吉元涼介, 芳鐘冬樹: “ランダムフォレストを用いた国会会議録のイデオロギー分析,” *FIT2016 (第 15 回情報科学技術フォーラム)講演論文集*, 4, pp. 277-278, 2016.
- [10] 大南勝, 掛谷英紀: “国会会議録に基づく短命大臣の特徴分析,” *言語処理学会第 23 回年次大会発表論文集*, pp.442-445, 2017.
- [11] 大南勝, 掛谷英紀: “国会会議録に基づく短命大臣の特徴分析 第 2 報,” *言語処理学会第 24 回年次大会発表論文集*, pp.1167-1170, 2018.
- [12] 掛谷英紀, 大南勝: “国会会議録に基づく「チルドレン議員」の特徴分析,” *言語処理学会第 24 回年次大会発表論文集*, pp.1163-1166, 2018.
- [13] Cydney Dupree, Susan T. Fiske (2018): *Self-Presentation in Interracial Settings: The Competence Downshift by White Liberals*, *PsyArXiv Preprints*.
- [14] 国会会議録検索システム <http://kokkai.ndl.go.jp/>
- [15] MeCab: Yet Another Part-of-Speech and Morphological Analyzer <http://mecab.sourceforge.net/>